

High dielectric constant gate oxides for metal oxide Si transistors

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

2006 Rep. Prog. Phys. 69 327

(<http://iopscience.iop.org/0034-4885/69/2/R02>)

[View the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 130.75.19.25

The article was downloaded on 23/06/2011 at 15:14

Please note that [terms and conditions apply](#).

High dielectric constant gate oxides for metal oxide Si transistors

John Robertson

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

Received 6 October 2005

Published 14 December 2005

Online at stacks.iop.org/RoPP/69/327

Abstract

The scaling of complementary metal oxide semiconductor transistors has led to the silicon dioxide layer, used as a gate dielectric, being 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. These oxides had not been extensively studied like SiO_2 , and they were found to have inferior properties compared with SiO_2 , such as a tendency to crystallize and a high density of electronic defects. Intensive research was needed to develop these oxides as high quality electronic materials. This review covers both scientific and technological issues—the choice of oxides, their deposition, their structural and metallurgical behaviour, atomic diffusion, interface structure and reactions, their electronic structure, bonding, band offsets, electronic defects, charge trapping and conduction mechanisms, mobility degradation and flat band voltage shifts. The oxygen vacancy is the dominant electron trap. It is turning out that the oxides must be implemented in conjunction with metal gate electrodes, the development of which is further behind. Issues about work function control in metal gate electrodes are discussed.

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

Contents

	Page
1. Introduction	329
1.1. Scaling and gate capacitance	329
1.2. Equivalent oxide thickness and equivalent capacitance thickness	331
2. Choice of High K oxide	333
2.1. K value	333
2.2. Thermodynamic stability	334
2.3. Crystalline or amorphous oxides	335
2.4. Interface quality	336
2.5. Band offset	336
2.6. Defects	338
3. Materials chemistry of high K oxides	338
3.1. Deposition	338
3.2. Alloy crystallization	342
3.3. Atomic diffusion	343
3.4. The interfacial layer	344
4. Bonding and electronic structure	346
4.1. Nature of bonding in high K oxides	346
4.2. Dielectric constants	351
4.3. Band offsets	352
4.4. Bonding at ZrO_2 and HfO_2 –Si Interfaces	355
4.5. Oxides on Ge	361
4.6. Other binary oxides	361
4.7. Perovskites	363
4.8. LaAlO_3	366
5. Electronic structure of defects	368
6. Electrical quality	370
6.1. Charge trapping and conduction processes	371
6.2. Mobility degradation	375
6.3. V_T stability	379
7. Work functions and metal gates	380
7.1. Introduction	380
7.2. Effective work functions	380
7.3. Metal gate systems	384
7.4. Dipole layers versus simple metals; fully-silicided silicon (FUSI)	386
7.5. Interaction of poly-Si with HfO_2	387
8. Summary	390
Acknowledgments	390
References	390

1. Introduction

1.1. Scaling and gate capacitance

The complementary metal oxide semiconductor (CMOS) field effect transistor (FET) made from silicon is the most important electronic device. This has arisen because of its low power consumption and because of its performance improvement over forty years according to Moore's Law of scaling. This law notes that the number of devices on an integrated circuit increases exponentially, doubling over a 2–3 year period. The minimum feature size in a transistor decreases exponentially each year. The semiconductor Roadmap defines how each design parameter will scale in future years to continue this trend, as shown in table 1 and figure 1.

The scaling cannot go on forever, and the limits to Moore's law are often said to be lithography and the need for very short wavelengths of light to pattern the minimum feature size. It turns out that materials are now a key constraint. First, the maximum current density in interconnects between transistors recently led to copper replacing aluminium as the conductor. Then, RC time delays led to the SiO₂ inter-metal dielectric being replaced by materials of lower dielectric constant such as SiO₂:F or porous SiOCH alloys. But the most serious problem in logic circuits is now the FET 'gate stack', which is the gate electrode and the dielectric layer between the gate and the silicon channel.

The thickness of the SiO₂ layer presently used as the gate dielectric is now so thin (under 1.4 nm) that the gate leakage current due to direct tunnelling of electrons through the SiO₂ becomes too high, exceeding 1 A cm⁻² at 1 V (figure 2), so that power dissipation increases to unacceptable values [1–7]. In addition it becomes increasingly difficult to make and measure accurately such thin films. Finally, the reliability of SiO₂ films against electrical breakdown declines in thin films. These reasons lead to a desire to replace SiO₂ 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

$$C = \frac{\varepsilon_0 K A}{t}, \quad (1)$$

where ε_0 is the permittivity of free space, K is the relative permittivity, A is the area and t is the oxide thickness. Hence, the solution to the tunnelling problem is to replace SiO₂ with a physically thicker layer of new material of higher dielectric constant (permittivity) K (figure 3). This will keep the same capacitance but decrease the tunnelling current. These new gate oxides are called 'high K oxides'.

For the electrical design of a device the precise material does not matter, 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

$$t_{\text{ox}} = \text{EOT} = \left(\frac{3.9}{K} \right) t_{\text{hi}K}. \quad (2)$$

Here 3.9 is the static dielectric constant of SiO₂. The objective is to develop high K oxides which allow scaling to ever lower values of EOT.

The gate leakage problem has been apparent since the late 1990s [8], but then the criteria for the choice of oxide were not known. In about 2001, the choice of oxide narrowed to HfO₂, but the problems of making HfO₂ into a successful electronic material appeared very difficult. It was debated whether high K oxides would be used, but instead the device engineers might use novel device designs to circumvent the problem. However, the increasing

Table 1. Summary of 2003 Roadmap. Node, gate length, EOT of high power (CPU) and low standby power devices (mobile), gate oxide material and gate electrode material. Earliest introduction of high K would be late in 45 nm node.

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} high power	1.5	1.3	1.1	0.9	0.8	0.6	0.5	0.5
T_{ox} low 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. TaC _x , TaSiN _x			

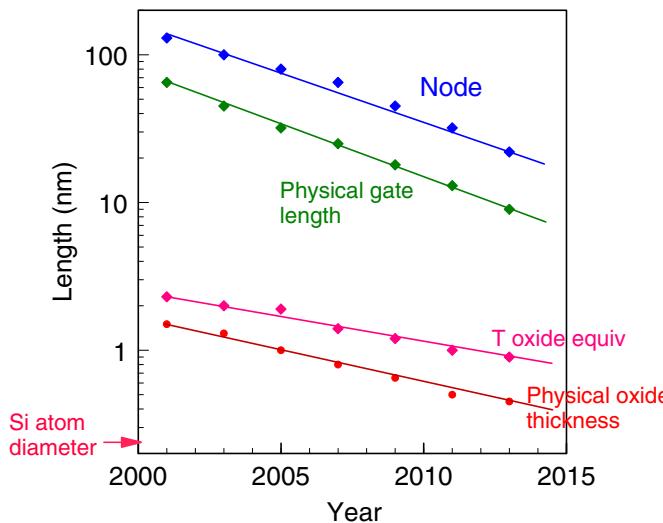


Figure 1. The scaling of feature size, gate length and oxide thickness according to the 2003 semiconductor Roadmap.

importance of the low-power sector of electronics, in mobile phones, lap-tops, etc meant that the problem must be confronted [1]. Low standby power CMOS requires a leakage current of below $1.5 \times 10^{-2} \text{ A cm}^{-2}$ rather than just 1 A cm^{-2} . There have been many difficulties in manufacturing high K oxide layers of sufficient quality but these have gradually been overcome. Announcements [9] indicated that enough of the problems are now solved such that high K oxides could be implemented in the 45 nm node, but recent announcements are more pessimistic.

Four key problems have been identified for the successful introduction of high K oxide [10]:

- (1) the ability to continue scaling to lower EOTs,
- (2) the instabilities caused by the high defect densities,
- (3) the loss of carrier mobility in the Si channel when using high K oxides,
- (4) the shifts of the gate voltage threshold and the need for metal gates.

Thus, this paper reviews the choice of oxides, their deposition, thermal stability, stability in device structures, electronic structure, interface properties, band offsets, electronic defects

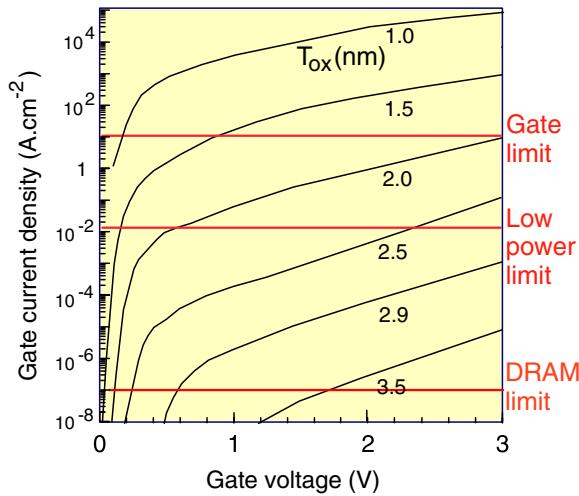


Figure 2. Leakage current versus voltage for various thickness of SiO_2 layers, from Lo *et al* [8].

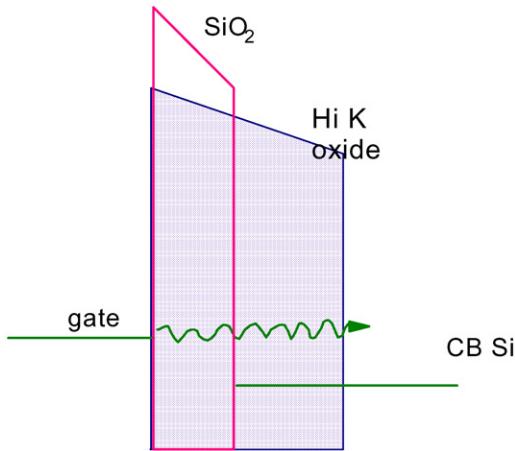


Figure 3. Schematic of direct tunnelling through a SiO_2 layer and the more difficult tunnelling through a thicker layer of high K oxide.

and carrier mobilities to understand what we have achieved so far, and how to solve these four problems. Problem (4) remains unresolved.

1.2. Equivalent oxide thickness and equivalent capacitance thickness

In CMOS FETs, the gate capacitance is 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], (figure 4). These three capacitances add as

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

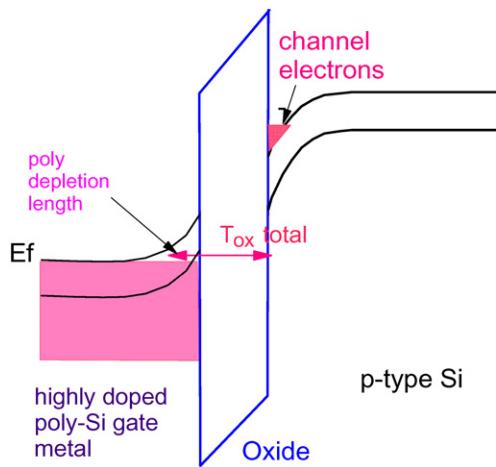


Figure 4. The three contributions to the capacitance of the gate/electrode stack: channel, dielectric and gate depletion.

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 the gate stack as

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

The channel capacitance C_{Si} arises due to quantum delocalization; the two-dimensional electron gas in the Si channel cannot lie infinitely close to the channel surface but must delocalize a few angstroms into the Si. 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 (poly-Si) because of its compatibility with SiO_2 . 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 angstroms, whereas a good metal has a higher carrier density and a depletion depth of only 0.5 Å. This depletion effect can be removed by replacing the poly-Si with a normal metal, which has a higher carrier density. Typical gate metals are TaC_x , TiN , TaSiN or Ru .

The gate metal is chosen primarily for its work function (WF) and its thermal robustness. The WF of the gate electrode determines the gate threshold voltage needed to turn the device into inversion. There are three choices [1] as illustrated in figure 5. In a CMOS circuit 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 WF should correspond to the mid gap energy of Si, about 4.6 eV. This is the simplest choice but also the worst in terms of device properties. The harder choice is to use a different metal for NMOS and PMOS gates. NMOS requires a gate metal with a WF close to the Si conduction band (CB) energy, 4.0 eV below the vacuum level. Such metals are quite reactive. On the other hand, PMOS requires a metal with WF close to the Si valence band or 5.1 eV. This metal would be very noble like Pt, but such metals are difficult to etch. Simple elemental metals turn out to be problematic, because they tend to react with SiO_2 or HfO_2 . Instead, we must use one of the high stability ‘diffusion barrier’ metals such as TiN. But these do not have a large range of WFs. Thus, ‘metal gates’ is a separate and difficult topic, which turns out to require considerable development, as discussed in section 7. A final point is that it is non realized that some high K oxides such as HfO_2 react with poly-Si, so that not only is it desirable to replace poly-Si with a metal gate, it is necessary.

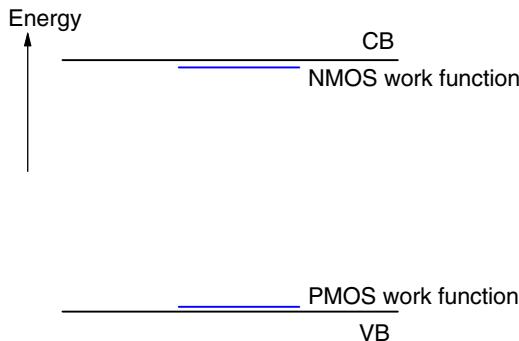


Figure 5. Desired work function of metal gates for NMOS and PMOS.

2. Choice of High K oxide

Silicon dioxide is the key reason that microelectronics technology uses Si and not some other semiconductor. As a semiconductor, Si has an average performance, but in most 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, etc) 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 nanometre scale. Its only problem is that it is possible to tunnel across it when very thin. Hence, we must lose 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. Its K value must be high enough to 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 to processing to 1000°C for 5 s (in present process flows).
4. It must act as an insulator, by having band offsets with Si of over 1 eV to minimize carrier injection into its bands.
5. It must form a good electrical interface with Si.
6. It must have few bulk electrically active defects.

Interestingly, once SiO_2 is replaced as the dielectric, the key advantage of Si is lost, so other semiconductors could be considered for MOSFETs. Indeed, this is happening; first Si–Ge layers, then perhaps even III–Vs.

2.1. K value

The first of the above requirements is that the oxide's K value should be over 12, preferably 25–30. There is a trade off with the band offset condition, which requires a reasonably large band gap. Table 2 and figure 6 show that the K of candidate oxides tends to vary inversely with the band gap, so we must accept a relatively low K value [11]. There are numerous oxides with extremely large K s, such as SrTiO_3 , which are candidates for dielectrics in DRAM capacitors [12], but these have a too low band gap. In fact, a very large K is undesirable

Table 2. Static dielectric constant (K), experimental band gap and (consensus) CB 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 ₃ N ₄	7	5.3	2.4
Al ₂ O ₃	9	8.8	2.8 (not ALD)
Ta ₂ O 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 ₂ O ₃	30	6	2.3
Y ₂ O ₃	15	6	2.3
a-LaAlO ₃	30	5.6	1.8

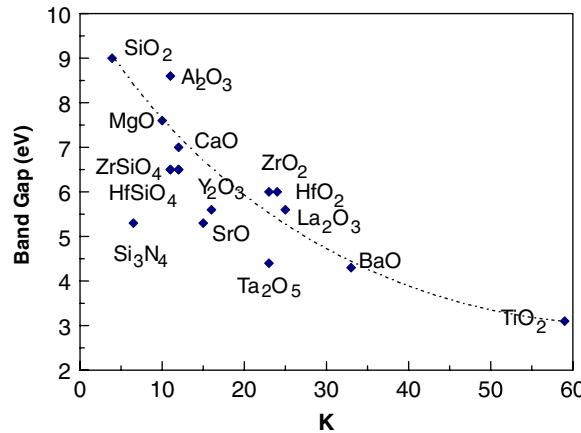


Figure 6. Static dielectric constant versus band gap for candidate gate oxides, after Robertson [11].

in CMOS design because they cause undesirable large fringing fields at the source and drain electrodes [13].

2.2. Thermodynamic stability

The second requirement arises from the condition that the oxide must not react with Si to form either SiO₂ or a silicide according to the unbalanced reactions



This is because the resulting SiO₂ layer would increase the EOT and negate the effect of using the new oxide. In addition, any silicide formed by (6) is metallic and would short out the field effect.

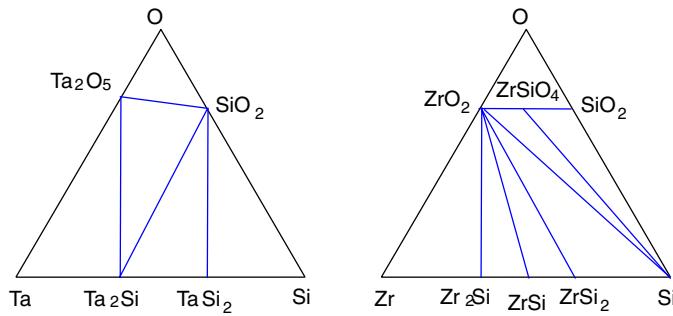


Figure 7. Comparison of ternary phase diagrams of metastable Ta–Si–O and stable Zr–Si–O systems.

This condition requires that the oxide have a higher heat of formation than SiO_2 . Hubbard and Schlom [14, 15] found that this restricts us to very few oxides, 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 . The group II oxides such as SrO are not favoured as they react with water. Hence this leaves us Al_2O_3 , ZrO_2 , HfO_2 , Y_2O_3 , La_2O_3 , Sc_2O_3 and some 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 two very similar elements. However, the thermodynamic data of some oxides was slightly inaccurate. It was subsequently found that ZrO_2 is slightly reactive with Si [15, 16] and can 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 and Lu_2O_3 also have a lower K than La_2O_3 . The other lanthanides Pr_2O_3 , Gd_2O_3 , etc are comparable to La [17–21].

One way to represent the stability of an oxide in contact with Si is on a ternary phase diagram with tie lines [1]. Figure 7 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 line but not directly. On the other hand, ZrO_2 and ZrSiO_4 and any composition in $(\text{ZrO}_2)_{1-x}(\text{SiO}_2)_x$ are connected by tie-lines and are in equilibrium in contact.

2.3. Crystalline or amorphous oxides

The third condition is kinetic stability, that the oxide must withstand the processing, a rapid thermal anneal for 5 s at 1000 °C. We must choose to use a crystalline or an amorphous oxide. If an amorphous oxide is desired, this is a strenuous condition in that most high K oxides are not good glass formers, unlike SiO_2 . In particular, HfO_2 and ZrO_2 crystallize at much lower temperatures and, if used, would be used as nano-crystalline phases. Al_2O_3 is a reasonably good glass former and is the next best in this respect. Ta_2O_5 is a glass former but was eliminated as reactive. All the other oxides crystallize well below 1000 °C.

The crystallization 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 [22, 23]. This then retains the stability against crystallization to close to 1000 °C. However, silicates have significantly lower

K values. If K value were the only condition, aluminates would be preferred to silicates, because of their higher K . The K value roughly follows a linear rule of mixtures with composition. The addition of nitrogen is found to be beneficial to Hf silicate, further raising its crystallization temperature, so that Hf silicates can just pass this criterion [24].

The arguments against a crystalline oxide are that its grain boundaries will cause higher leakage currents and possible diffusion paths for dopants. On the other hand silicates have a much lower K than the simple oxides, so they cannot ‘scale’ to as low EOTs. In fact, Lee *et al* [25] and Kim *et al* [26] find that the leakage currents of amorphous and nanocrystalline HfO_2 are similar.

The fact that crystallized HfO_2 does not cause higher leakage current has convinced many companies (e.g. Intel and Freescale) to favour binary oxides because of their higher K , whereas Texas Instruments [24] favours silicates because of their easier integration and lower trap densities.

2.4. Interface quality

The oxide is in direct contact with the Si channel. The carriers in the channel flow 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. Oxide grain boundaries could introduce extra defects. Therefore, there are two ways of ensuring 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 could configure its interface bonding to minimize 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 are isotropic, so that fluctuations in polarization from differently oriented oxide grains will not scatter carriers. Finally, amorphous phases have no grain boundaries.

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

2.5. 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 [11, 27], as shown schematically in figure 8. SiO_2 has a wide gap of 9 eV, so it has large barriers for both electrons and holes. However, in oxides with a narrower band gap like SrTiO_3 (3.3 eV), their bands must be aligned almost symmetrically with respect to those of Si for both barriers to be over 1 eV. In practice, the CB offset is 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 [11]. It is interesting that these are the *same* oxides that pass the thermal stability criterion. This is because a high heat of formation correlates with a wide band gap in ionic compounds.

Al titanate was recently proposed as a possible gate oxide, because of its low atomic diffusion rates [28]. However, further study of its electrical properties are needed, as it could have a rather high leakage, based on its estimated band gap and band offset. Similarly, tantalates remain unlikely candidates.

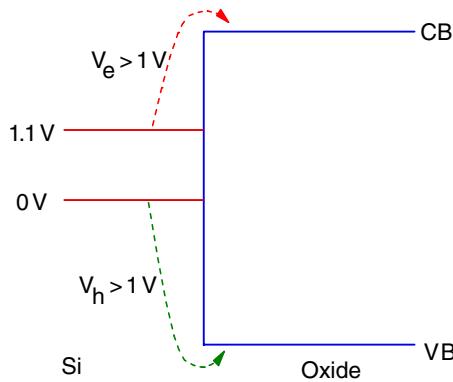


Figure 8. Schematic of band offsets determining carrier injection in oxide band states.

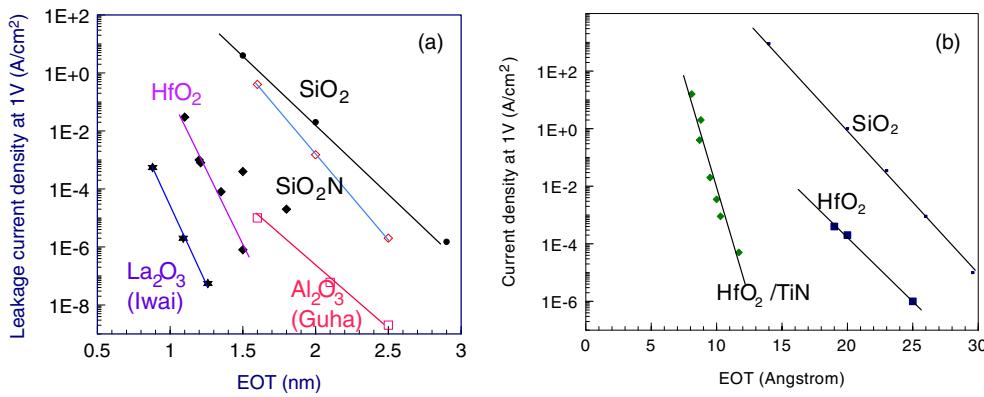


Figure 9. (a) Leakage current density versus EOT for various high K oxides for HfO_2 [29], ZrO_2 [30], Al_2O_3 [29, 31] and La_2O_3 [19]. (b) Leakage current density versus EOT for HfO_2 with poly-Si gates and TiN gates, after [32, 33].

The leakage current for various high K oxides is plotted as a function of EOT in figure 9. Figure 9(a) shows data for HfO_2 from Gusev *et al* [29], for ZrO_2 from Gusev *et al* [30], for Al_2O_3 of Guha *et al* [29, 31] and for La_2O_3 from Iwai *et al* [19]. Figure 9(b) compares data for HfO_2 films with poly-Si electrodes and HfO_2 with TiN electrodes, from Tsai *et al* [32] and the IMEC group [33].

Device scaling will require the use of ever smaller EOTs. Which oxides are the most suitable for such end of Roadmap devices? Yeo *et al* [34] defined a theoretical figure of merit, k , for direct tunnelling, based on the barrier height ϕ , tunnelling mass m^* and K

$$J = J_0 \exp(-2k \cdot t),$$

$$k = (2m^*\phi)^{1/2} \left(\frac{K}{3.9} \right),$$

where t is the EOT. Lanthanides have the lowest leakage in figure 9(a) and have the highest figure of merit because they have the highest CB offset, as shown in section 4.3. Hf alloys are presently preferred because La oxides are hygroscopic. Eventually La_2O_3 or La compounds

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

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

such as LaAlO_3 may be used, according to the 2003 Roadmap (table 1). EOTs beyond 0.8 nm require the minimization of any interfacial SiO_2 (see section 2.4) and may eventually favour epitaxial oxides. But there are presently huge problems to this option.

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 re-bond 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 [35]. 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. Table 3 summarizes the advantages and disadvantages of various deposition methods. 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 deposits in line of sight, so it does not give good coverage.

A method for producing high purity thin oxides is to deposit metal by electron beam evaporation or sputtering, which is highly controllable to small thickness, and then oxidize the deposited metal by ozone or UV assisted oxidation. The advantage is that this produces less damage than oxide sputtering. Ideally, the oxygen partial pressure should be controlled to oxidize just the metal, not the Si.

The preferred industrial scale methods are chemical vapour deposition (CVD) and atomic layer deposition (ALD). CVD uses a volatile metal compound as a precursor which is introduced into the chamber and oxidized during deposition onto the substrate. The CVD

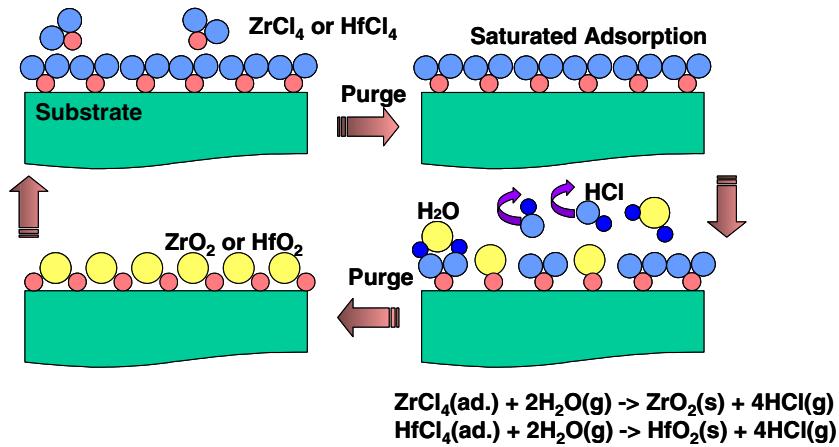
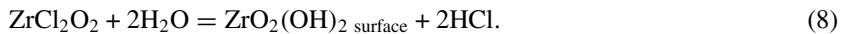
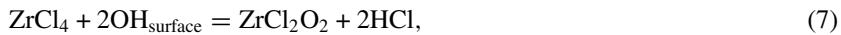


Figure 10. Schematic of the cyclic process of atomic layer deposition. Thanks to McIntyre.

precursors can be metal chlorides such as ZrCl_4 and HfCl_4 or metal organics such as tetra-butoxyl Zr, in which case it is called metal organo CVD (MOCVD). 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.

ALD is a method of cyclic deposition and oxidation [36]. It was originally developed to produce conformal, pin-hole free insulating films for electro-luminescent displays. As shown schematically in figure 10, the surface is exposed to the precursor such as ZrO_2 which is absorbed as a saturating monolayer (ML). The excess precursor is then purged from the chamber by an Ar or nitrogen pulse. A pulse of oxidant such as H_2O , H_2O_2 or ozone is then introduced which must then fully oxidize the adsorbed layer to the oxide and a volatile by-product such as HCl . The excess oxidant is then purged by another Ar pulse, and the cycle is repeated.

The effective chemical reactions are



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 oxidizes 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, have high purity, be 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 their chlorides. However, these have low volatility and tended to clog. A wide range of new precursors is being developed [36,37].

ALD was developed to produce highly conformal, pinhole-free insulating films (figure 11). The advantage of ALD is that it can grow the thinnest films of any method 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 the 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

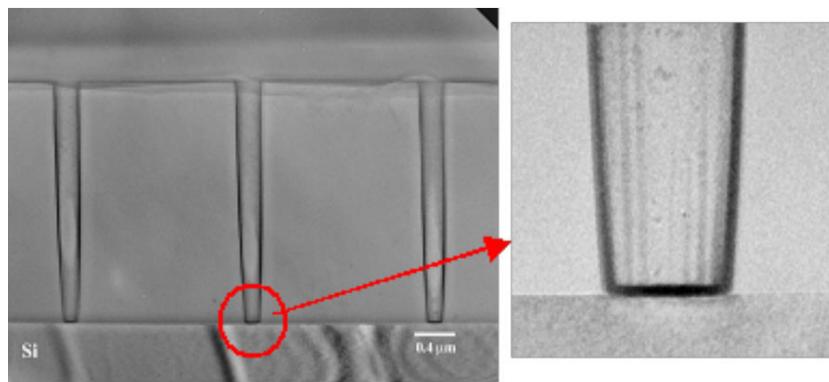


Figure 11. Scanning electron microscope image of trench structure showing excellent coverage by ALD HfO_2 . Thanks to McIntyre.

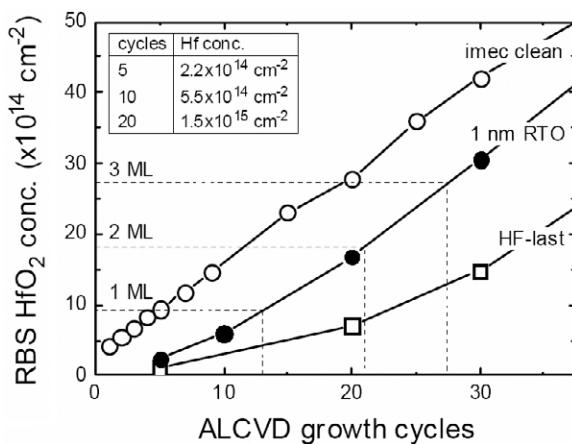


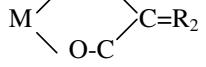
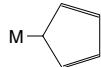
Figure 12. Film thickness versus number of ALD cycles, for different Si surface preparations, showing the nucleation delay on HF-last Si.

excellent method for producing Al_2O_3 , using trimethyl-aluminium as a precursor [36]. 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 at below one ML because of steric hindrance. This is not a significant disadvantage, it just takes more cycles to grow a certain thickness.

The biggest problem with ALD has been nucleation. 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 surfaces but had a slow initial growth rate [38, 39] (figure 12). This meant that oxide films even 3 ML thick were not fully ‘closed’ [38]. It was found that nucleation occurred much more readily on a slightly pre-oxidized Si surface [40]. Thus, ALD is usually carried out on a ‘chemical oxide’ (SiO_2) surface formed by an ozone or oxidative cleaning of Si. This limits the minimum EOT presently achievable by ALD. It was shown that the nucleation problems on Si-H terminated surfaces

Table 4. Precursor molecules for ALD

Type	Formula	Comments
Chlorides	MCl ₄	Non-volatile, corrosive byproduct
Alkyls	M-R	Al, Zn OK, but few others
Alkoxides	M-OR	Limited stability, better for MOCVD than ALD
Diketonates		Low reactivity, high melting point
Alkyl imide	M = N-R	Thermal stability
Alkyl amides	M - NR ₂	Good thermal stability
Alkyl amido	M - (N = RR') _n	Good thermal stability
Pentadienyl		Low reactivity or low volatility
Nitrates	M (NO ₃) _n	

for Al₂O₃ depended on the ALD precursor not the oxidant, so that the choice of precursor is the key to overcoming this problem [40].

The various types of ALD precursor are listed in table 4. The precursors are designed to have the desired volatility, reactivity and stability [36, 37, 41], as noted in table 4. Except for halides, the molecule is formed by adding organic ligands. The ligands create the volatility. Volatility is aided by using monomer ligands, which do not pack so easily in the solid state. The metal valence is satisfied by polar bonds from N or O atoms. These bonds define the stability and reactivity.

The first precursors were the metal halides. However, these are being superceded by other molecules because of their low volatility and corrosive byproducts (HCl etc). The alkoxides and diketonates are commercially available oxygen containing precursors, of reasonable reactivity. The nitrogen based ligands are becoming more favoured. It turns out that amido compounds have the desired reactivity to give ALD on oxide surfacers and also to nucleate on Si-H terminated surfaces. Beta-diketonates are useful for lanthanide oxides. ALD has now been shown to be able to produce metal oxide films, metal nitrides (for electrodes) and even metals [41], after reduction from the metal.

Chabal *et al* [42] followed the progress of an HfO₂ film being grown at 200C by ALD by FTIR, as in figure 13. The film was grown from tetrakis ethyl, methyl amino hafnium (TEMAH) or Hf(N(CH₃)(C₂H₅))₄ at 100C with water as oxidant on an H terminated Si surface. First, the infra-red spectrum showed that there is an abrupt interface during growth, with Hf-O-Si bonds gradually replacing Si-H bonds. The HfO₂ growth can be seen from the Hf-O modes. The continuing presence of Si-H even after 3 ML of oxide has grown shows that there remains island growth with discontinuous nucleation. A separate broadened and red-shifted Si-H mode arises as the Si-H group becomes surrounded by HfO₂. The reaction of the precursor, not the oxidant is seen to be crucial.

The film was then subjected to thermal annealing [42]. The film possessed O-H modes from the oxidant (actually O-D modes, as the oxidant is D₂O to allow it to be tracked). The post-deposition anneal is used to density the film and remove contaminants such as organics, water, etc. It is found that at 400–500C, the OH modes disappear and Si-O modes appear. An interfacial SiO₂ layer 0.9 nm thick has grown. It turns out that ALD leaves an oxygen-rich film. When the oxidant is water, the film contains –OH groups. Annealing would try to evolve

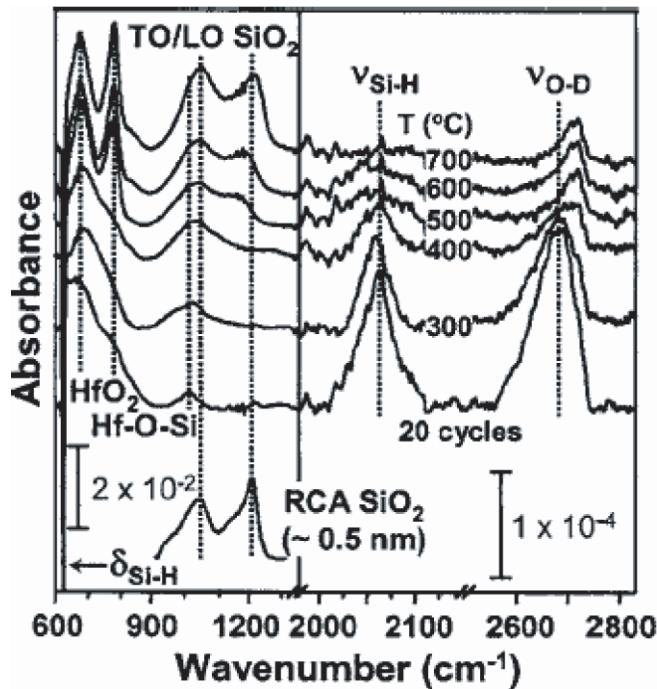


Figure 13. Annealing of the infra red spectrum of a HfO₂ film grown by ALD with water as oxidant, showing the evolution of hydrogen [42].

these as H₂O molecules. However, as the film densifies, H₂O is too large to diffuse through. It is only possible for H to diffuse either as H⁺ ions or H₂ molecules, desorbing as H₂. Further annealing to higher temperatures expels the excess oxygen, and this creates the interfacial SiO₂ layer by oxidizing the underlying Si. There is no interfacial layer during growth itself.

3.2. Alloy crystallization

Silicate and aluminite alloys of Zr, Hf and La oxides are often used instead of the pure metal oxides to have a higher resistance to crystallization [22–24, 43]. Zr silicate has been the most widely studied in this regard. Crystallization directly to the crystalline silicate ZrSiO₄ is inhibited by kinetics. Instead, Maria *et al* [44] showed that crystallization occurred by a phase separation of the silicate into ZrO₂ and SiO₂ phases, followed by a crystallization of the ZrO₂ component. This can be seen for HfO₂–SiO₂ alloys in the high-resolution transmission electron microscope images in figure 14 for two different compositions by Stemmer *et al* [45].

The phase diagram of the ZrO₂–SiO₂ system is known reasonably well [45–47] (figure 15(a)). That of HfO₂–SiO₂ is less well known but should be similar. The key factor is that ZrO₂ and SiO₂ liquids are immiscible over a range of composition. This is attributed to the high ionic charge of Zr. This ‘miscibility gap’ can be extrapolated to lower temperatures to define a solid phase miscibility gap. It also defines a spinodal region in which the alloy can spontaneous phase separate to lower its free energy [46]. The glass transition temperature is also marked in figure 15(a); it falls in ZrO₂ rich silicates. Thus, Zr silicates crystallize by two mechanisms. For 20–60 mol% Zr, silicates crystallize by spinodal decomposition

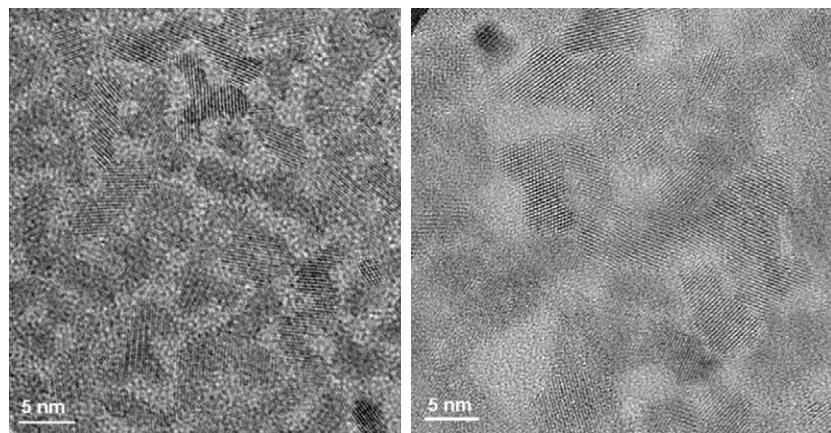


Figure 14. Plan view TEM image of crystallization in HfO₂/SiO₂ alloy system (a) 40% HfO₂ and (b) 80% HfO₂ [45]. Thanks to Stemmer.

followed by crystallization. This leads to small grain sizes. Silicates with over 60% Zr crystallize by the kinetically limited nucleation and growth of crystalline ZrO₂. This was confirmed by extensive transmission electron microscopy (TEM) and x-ray scattering studies on Hf silicates by Stemmer *et al* [47]. The La silicate phase diagram in figure 15(b) [44] is qualitatively similar to that of ZrSiO₄ except that the two-phase region is further towards SiO₂.

In contrast, the phase diagrams of aluminates such as ZrO₂–Al₂O₃ show no miscibility gap [48] (figure 15(c)), so they are more resistant to crystallization [49]. However, it turns out that aluminates have higher densities of electronic defects, so that silicates are often preferred to aluminates as gate oxides. The MIRAI group have introduced Hf, Al oxide laminates, that is unmixed aluminates [50].

Despite the use of silicates, they still cannot fully achieve the 1000 °C stability requirement. The final improvement in performance comes with adding some nitrogen [24, 51]. The N reduces the diffusion coefficient of oxygen in the alloys, and this reduces the crystallization rate enough such that the silicate can withstand 1000 °C.

3.3. Atomic diffusion

We noted that a gate oxide must withstand processing to temperatures of the 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 to diffuse through it. All these aspects require the gate oxide to have low atomic diffusion coefficients. Interestingly, the proposed oxides HfO₂ and ZrO₂ belong to the class of fast oxygen ion conductors like CeO₂, of interest in solid oxide fuel cells. Clearly, for gate oxide use, diffusion must be low.

There have been extensive measurements of the atomic diffusion rates of Hf, O, B and P in HfO₂ and Hf silicate measured after implantation by secondary ion mass spectroscopy (SIMS) and nuclear reaction profiling [52–56]. This is reviewed by Baumvol *et al* [54]. The mixing of oxide and Si layers has also been studied by medium energy ion scattering (MEIS) which measures the element profile.

A great advantage of alloying with SiO₂ is that the Si sites in silicates are covalently bonded to oxygen. This greatly lowers the oxygen diffusion rate. The basic silicate is found

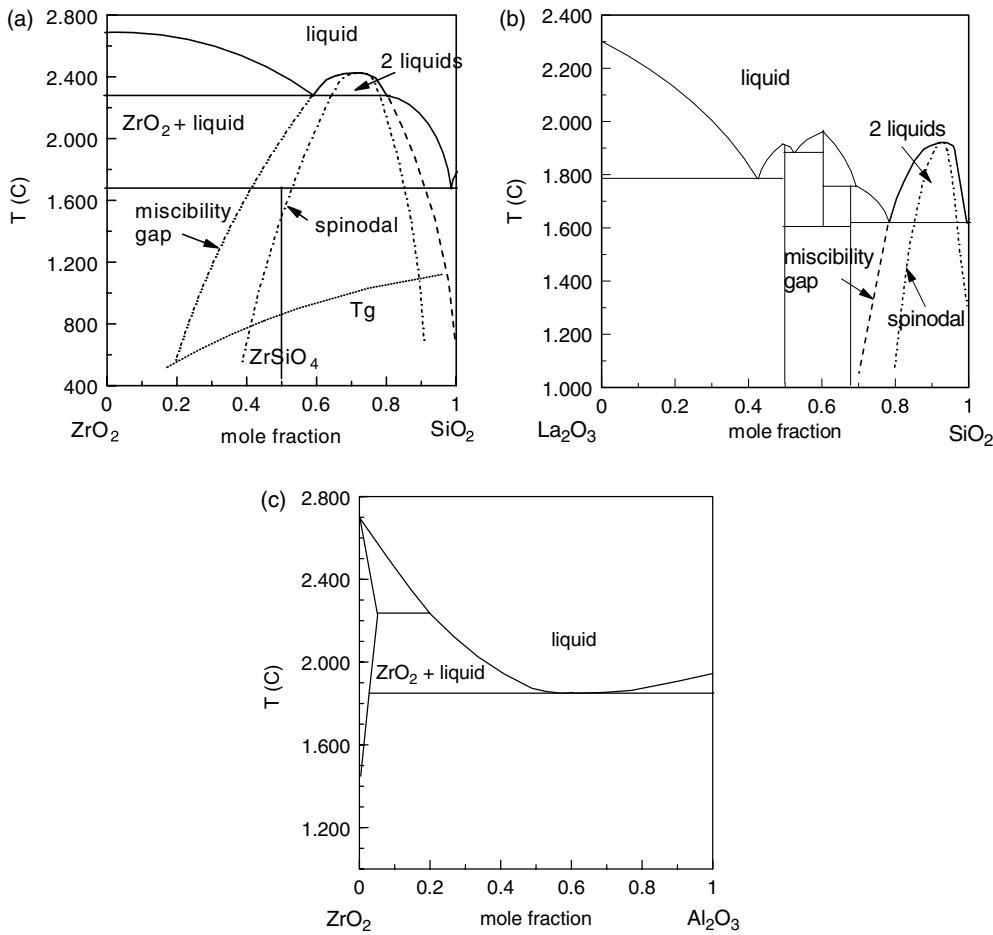


Figure 15. (a) Phase diagram of $\text{ZrO}_2/\text{SiO}_2$ showing miscibility gap. After Kim and McIntyre [46]. (b) Phase diagram of $\text{La}_2\text{O}_3/\text{SiO}_2$ with miscibility gap. After Maria *et al* [44]. (c) Phase diagram of $\text{ZrO}_2/\text{Al}_2\text{O}_3$. After Zhao *et al* [48].

to perform adequately in most respects. However, nitrogen addition lowers diffusion rates further, further raising the crystallization temperature [24].

Another key role of the oxide is to block dopant diffusion from any poly-Si gate electrode [57]. 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 could be a short circuit diffusion path, so here N acts to block diffusion by stopping crystallization and the formation of any grain boundaries [25].

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 [58]. There are advantages and disadvantages to this interfacial layer, 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

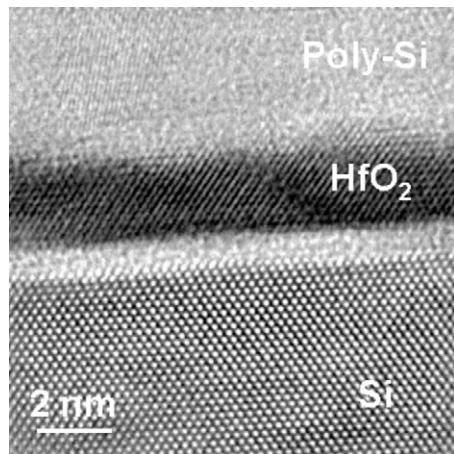


Figure 16. HRTEM cross section showing interfacial layer of SiO₂ below the HfO₂ layer. Thanks to S Stemmer.

formula

$$1/C = 1/C_1 + 1/C_2, \quad (9)$$

which becomes

$$\text{EOT} = t_{\text{SiO}_2} + \text{EOT}_{\text{hi}K}. \quad (10)$$

Thus, an extra SiO₂ layer is undesirable as it adds to the overall EOT. In fact, the K of SiO₂ (3.9) is so small that a SiO₂ layer can rapidly use up the EOT allocation. It is a severe limitation to scaling.

The SiO₂ layer does not arise from the direct reaction of HfO₂ with Si. It arises from the diffusion of O through the HfO₂ layer to oxidize the Si underneath. The SiO₂ layer usually grows during the post-deposition annealing stage, and not during growth. Narayanan *et al* [59] proved this for the case of YO. This can be avoided by adding silicate or N to the HfO₂ layer to reduce diffusion rates. However, scaling requirements will reduce the ability to use silicates in the future because they lower K . It also arises for ALD oxide, because the oxide is initially oxygen-rich with OH groups, and it loses the excess O to oxidize the underlying Si [42].

The second reason an SiO₂ layer exists is that it was intentionally put there, for beneficial reasons. Firstly, a ‘chemical oxide’ acts as a nucleation layer for ALD growth of HfO₂ [39,40].

The SiO₂ layer is also introduced because it improves the overall electrical quality of the Si–oxide interface (see later). The Si–SiO₂ interface is well understood and can be processed to be of high quality. In principle, it can be made with a very low defect concentration, by annealing. A SiO₂ layer will also space the Si channel from the high K oxide, which can lessen a reduction in carrier mobility due remote scattering (see section 6).

A disadvantage of an interfacial oxide is that it may not have the same quality as SiO₂ produced by thermal oxidation of Si [60–65]. It may be defective. Copel *et al* [62] have used a number of techniques such as MEIS to study the profile and composition of interfacial oxides under HfO₂. They found that they are SiO₂ despite sometimes appearing to have higher K values than thermal oxide. Z-contrast electron energy loss spectroscopy (EELS) found a similar result [63–65]. This is because it is sub-stoichiometric. Given its thickness, it is also likely to contain some Hf ions. Bersuker *et al* [65] have emphasized that defects in this layer contribute to degraded channel mobilities.

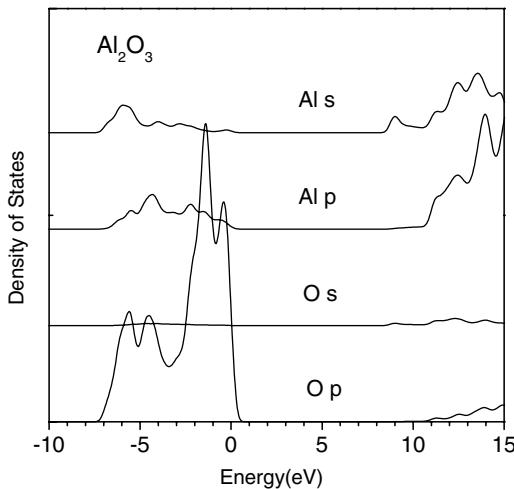


Figure 17. Density of states of Al₂O₃ in corundum structure. Note O 2p-like valence band and 8.8 eV band gap. Gap corrected to experimental value.

It is an advantage to be able to control the thickness of the interfacial SiO₂ layer and if necessary remove it entirely. This can be done in two ways. Firstly, Si and SiO₂ 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 [66]. 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 SiO₂ to displace Si [67, 68].

4. Bonding and electronic structure

4.1. Nature of bonding in high *K* oxides

The oxides of interest except for Al₂O₃ are transition metal oxides. Figure 17 shows the density of states (DOS) of Al₂O₃. The top of the valence band lies at 0 eV and the band gap lies from 0 to 8.8 eV. The bonding in Al₂O₃ is more ionic than in SiO₂ and its atoms have ionic coordinations. However, its electronic DOS does resemble that of SiO₂. Its valence band consists mainly of O p states and a CB of mainly Al s,p states.

A more typical case is the transition metal oxide, ZrO₂. ZrO₂ films are amorphous at lower temperatures but crystallize relatively easily. ZrO₂ is stable in the monoclinic structure at room temperature, it transforms to the tetragonal structure above 1170°C and it can be stabilized in the cubic fluorite structure by the addition of Y [69]. HfO₂ is similar. In cubic and tetragonal ZrO₂, Zr has 8 oxygen neighbours and each oxygen has four Zr neighbours, while in monoclinic ZrO₂ each Zr atom has 7 oxygen neighbours. Tetragonal ZrO₂ is derived from cubic ZrO₂ by displacing oxygens along the *z* axis towards 4 of the Zrs.

Figure 18 shows the bands and DOS of cubic ZrO₂. It has an indirect gap of 5.8 eV, the experimental value [69]. (In all plots in this section, the band gap has been corrected for the band gap error in local density formalism to the experimental value.) Recent calculations find that the tetragonal phases have the widest gaps (table 5) [70, 71]. The valence band is 6 eV wide, and it has a maximum at X formed from O p states. The CB minimum in the cubic phase 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 t₂ states 5 eV higher (at Γ). The partial DOS shows considerable

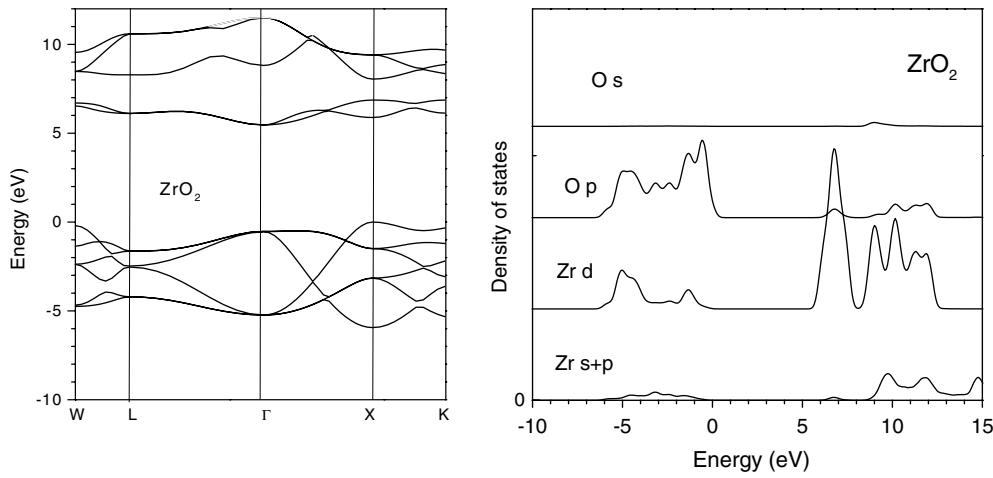


Figure 18. Bands and density of states of cubic ZrO_2 . Band gap corrected to experimental value.

Table 5. 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

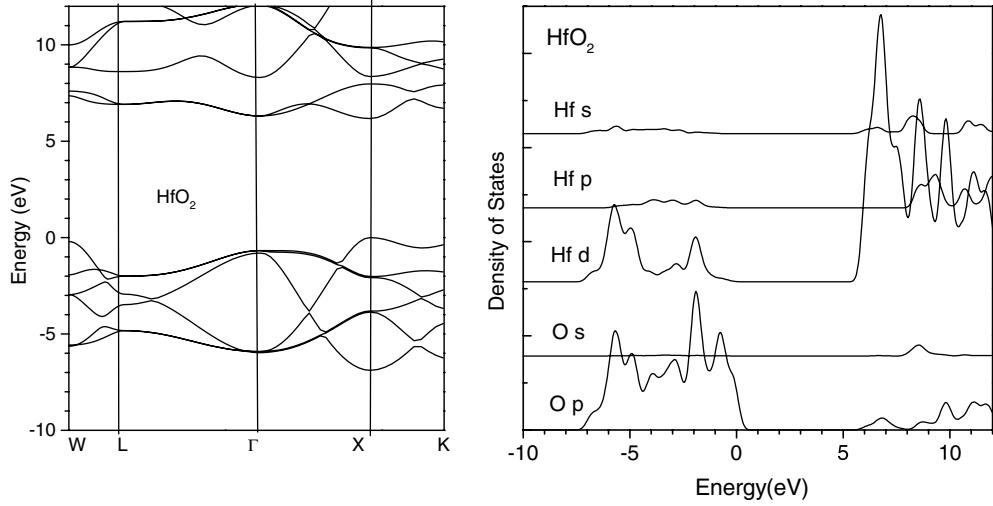


Figure 19. Bands and density of states of cubic HfO_2 . Gap corrected to experimental value.

charge transfer, with the valence band being strongly O p states, and CB on Zr d states, with 30% admixture [72]. The valence DOS of the monoclinic and tetragonal phases are similar. In the CB, the lower symmetry phases tend to wash out the crystal field splitting somewhat. The CB DOS tends to show 5 peaks in all phases.

The band structure and DOS of HfO_2 is very similar to that of ZrO_2 except that the crystal splitting of the Hf 5d states in the CB is larger than in ZrO_2 (figure 19).

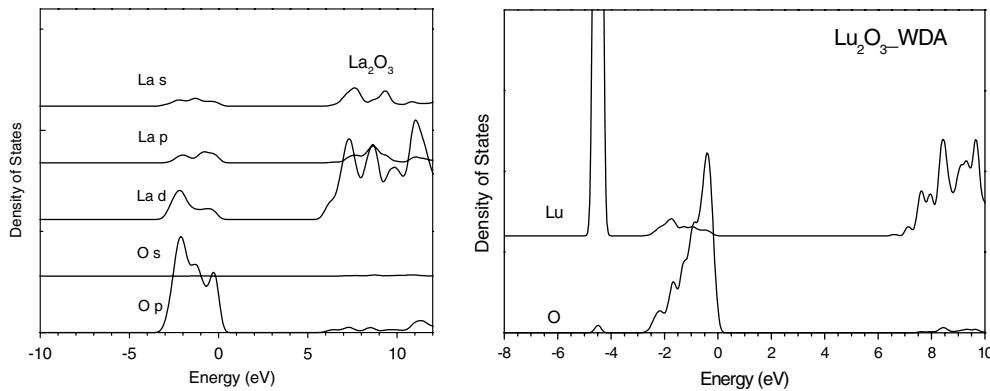


Figure 20. Density of states of La_2O_3 and Lu_2O_3 . Gap corrected to experimental value.

Crystalline La_2O_3 has the La_2O_3 structure in which La is seven-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 localized on O p states and the CB in on La d with some La s,p states starting at 8 eV [72]. 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 . Figure 20 also shows the DOS of Lu_2O_3 , another lanthanide of interest.

TiO_2 is another group IV oxide which has various structures including rutile. Its gap is much narrower, about 3.3 eV.

Of the group IIIA metal oxides, Y_2O_3 has the cubic bixbyite (defect spinel) structure. This has a large 80 atom unit cell in which there are two types of Y sites, both seven-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 [72]. The valence band is again only 3 eV wide. The partial DOS shows the valence band is largely O p states. The CB minimum has mixed Y d,s character.

In these cases, the CB minimum at Γ lies at the metal d energy while the top of the valence band are non-bonding O 2p states, lying at the O 2p energy. Thus the band gap is a simple gap between O 2p valence states and metal d states. The band gap is proportional to the metal atomic d orbital energy, as noted by Lucovsky *et al* [73].

ZrSiO_4 is typical of the transition metal silicates. Crystalline ZrSiO_4 has the body-centred tetragonal structure. The Zr and Si atoms are organized in chains. Each Zr atom has eight O neighbours. Each Si has four O neighbours in a tetrahedral arrangement. These coordinations may carry over to the amorphous phases and amorphous alloys. Its partial DOS is shown in figure 21. The band gap is about 6.5 eV [72]. The valence band is about 7 eV wide [74]. The CBs form two blocks. The lower CB is due to Zr d states and lies between 6.5 and 8 eV, and a second CB due to Si–O antibonding states lie above 9 eV.

Recent EXAFS spectra suggest that the coordinations in Zr silicate alloys do not follow the zircon structure but are locally more like in pure ZrO_2 or pure SiO_2 [75]. This implies local phase separation.

It is an important general rule that the CB 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 minimum of the silicates has a 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 [76].

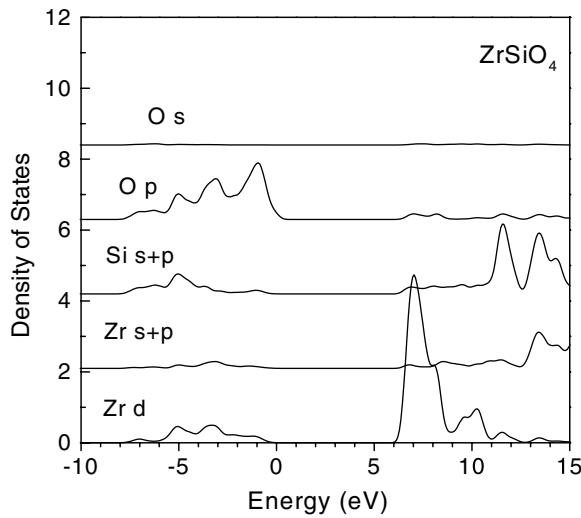


Figure 21. Density of states of crystalline ZrSiO_4 . Gap corrected to experimental value.

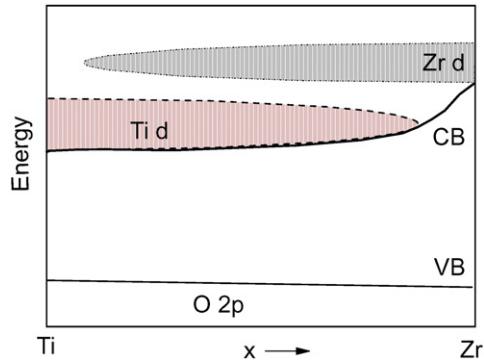


Figure 22. Schematic band diagram for $(\text{Zr}, \text{Ti})\text{O}_2$ showing sharp increase in gap near ZrO_2 .

The silicates illustrate a key property of disorder in these transition metal oxides. The bands of semiconductor alloys like GaAs–AlAs follow the ‘band’ picture, in which, to first approximation, the band gap interpolates linearly between the end members, with perhaps a small bowing. In contrast, transition metal oxides follow the atomic limit, where the DOS of the alloy is the sum of the DOSs of the components. Thus, the band gap is that of the narrower gap oxide (A) and then suddenly shoots up to the larger gap near pure B (see figure 22). This occurs in $\text{Zr}_x\text{Si}_{1-x}\text{O}_2$. It also occurs in, for example, $\text{Zr}_{1-x}\text{Ti}_x\text{O}_2$. Thus, any Ti will lower the gap of the alloys across almost the whole composition range like an extreme form of bowing. In the Ti-poor end, the Ti d states would be localized gap states in ZrO_2 gap.

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 23 shows the partial DOS of SrTiO_3 . The band gap is direct and 3.3 eV wide. The lowest CBs are $\text{Ti } d_{xy} t_2$ states followed by the $\text{Ti } d_{z^2}$ 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

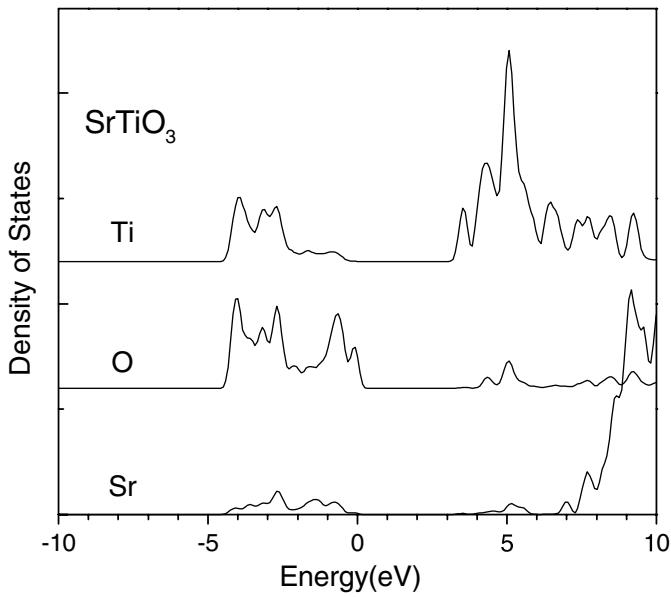


Figure 23. Density of states of cubic SrTiO_3 . Gap corrected to experimental value.

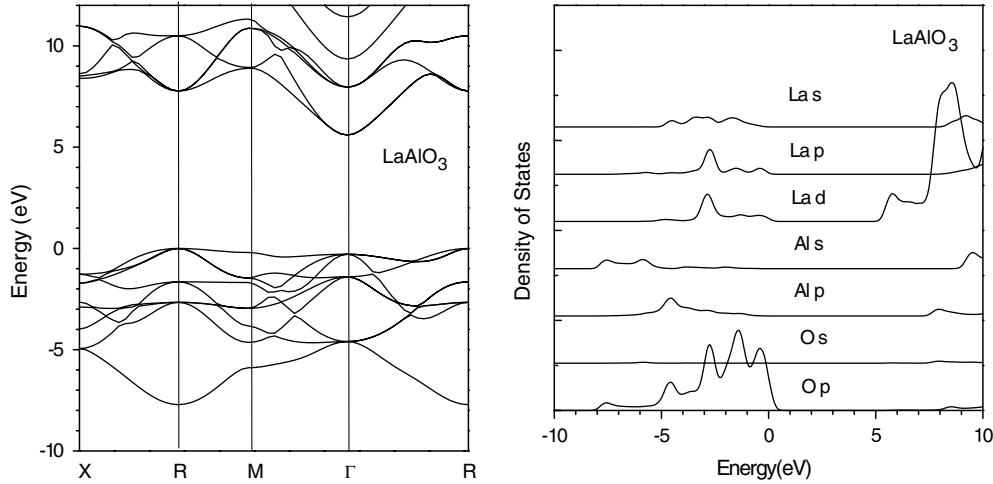


Figure 24. Bands and density of states of cubic LaAlO_3 . Gap corrected to experimental value.

from the band gap, and the ion can be considered to be essentially fully ionized and passive. On the other hand, the Ti–O bond is polar but only about 60% ionic.

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 24. The band gap is taken as 5.6 eV from recent ellipsometry work [77].

These various band structures have been calculated using the local density approximation (LDA) in the generalized gradient approximation (GGA). It is well known that LDA

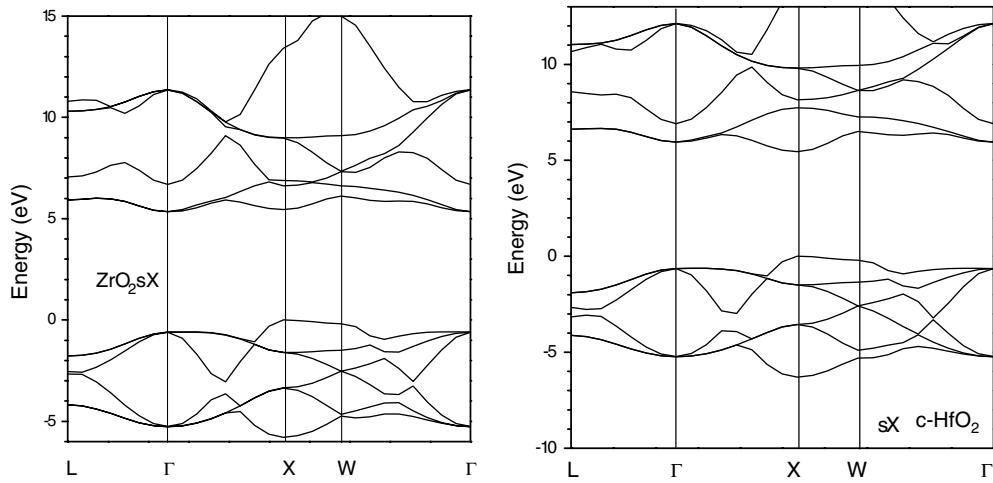


Figure 25. Bands of cubic ZrO₂ and HfO₂ calculated by sX.

underestimates the band gap of semiconductors and insulators. This has been corrected in the diagrams shown by the so-called scissors operator—an empirical upward rigid shift of all CBs to fit the experimental gap. It would be useful if *ab initio* methods gave the correct band gap. The GW method is one of these, and GW bands of ZrO₂ and HfO₂ are known [70, 78]. However, it is computationally expensive. A number of less expensive methods are known. We have used two of these, the screened exchange (sX) and the weighted density approximation (WDA) to provide the band structures of these oxides. These two methods are explained in detail elsewhere [72, 79–81]. These parameter-free methods are very valuable for calculations on the defect states, where empirical corrections have less basis. The bands of cubic ZrO₂ and HfO₂ using sX [72] are shown in figure 25. Table 5 gives the band gaps for the various phases in these approximations.

4.2. Dielectric constants

The static dielectric constant 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 equals the square of the refractive index, $n - \kappa_e = \varepsilon_\infty = n^2 \cdot \varepsilon_\infty$ values are typically 4–5 for the wide gap oxides of interest. Thus they are *not* the main source of the high K . The large static dielectric constant arises from the lattice contribution

$$\kappa - n^2 = \sum \frac{Ne^2 Z_T^{*2}}{m\omega_{TO}^2}. \quad (11)$$

Here, N is the number of ions per unit volume, e is the electronic charge, Z_T^* is the ion's transverse effective charge, m is the reduced ion mass, ω_{TO} is the frequency of the transverse optical phonon and the sum is over the ions. 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 have been calculated in the local density formalism of the various phases of HfO₂ and ZrO₂ [82, 83]. This is a good means to understand the differences and the anisotropies. Rignanese *et al* [83] found that the tetragonal phase has the largest and most anisotropic K but not by as much as found earlier by Vanderbilt and Zhao [82].

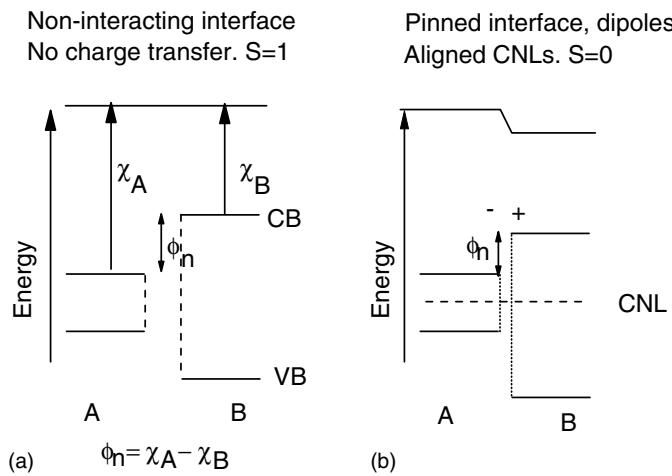


Figure 26. Schematic of how charge transfer at semiconductor interface controls its band line up. (a) No charge transfer and (b) charge transfer.

These calculations have been extended to lanthanides such as Lu_2O_3 and La compounds [84–86].

4.3. Band offsets

The band offset between the oxide and Si defines the barrier for injection of electrons or holes into the oxide bands. The electron barrier or CB offset tends to be the smaller of the two. The CB offset is one of the key criteria in the selection of a gate oxide. It must be over 1 eV to give an adequately low leakage current [11, 27]. The band offset for most high K oxides has been measured by methods such as photoemission.

This band line up at an interface is controlled by a dipole formed by charge transfer across the bonds at the interface. For two non-interacting surfaces, there is no charge transfer, no dipole and the CB line up is given by the difference between the electron affinities (the energy of the CB edge below the vacuum level)(figure 26). This is known as the Schottky limit. If the surfaces interact, charge transfer occurs across the interface and the resulting interface dipole modifies by this offset. The charge transfer acts to align a reference 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 reference energy level below the two CBs and is independent of the vacuum levels. Most high K oxides are intermediate between the two limits.

The band line up at an interface is controlled by a dipole formed by charge transfer across the interface bonds [87–89]. The dipole consists of two components, a component intrinsic to the bulk oxide and a component which depends on the specific interface bonding [89–91]. The intrinsic component is of interest because the specific bonding at the interface is usually not known. The intrinsic component is often the main component. However, the interface specific component can be important. In that case, there is no unique offset value for a given oxide on Si. This can be useful as it allows us to vary offsets by varying the interface chemistry.

Let us consider the intrinsic component of the offset. A particular model is that of metal induced gap states (MIGS)[92–95]. 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 for the neutral surface. 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 CB above the metal Fermi level, is given by

$$\phi_n = S(\Phi_M - \Phi_S) + (\Phi_S - \chi_S), \quad (12)$$

where Φ_M is the metal WF, Φ_S is the CNL 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 [96]

$$S = \frac{1}{1 + (e^2 N \delta / \varepsilon_{\infty})}, \quad (13)$$

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, can contribute to S [90]. Nevertheless the MIGS model often gives reasonably good predictions.

The model is extended to the band offsets between semiconductors. Charge transfer tends to align the CNL of the bulk oxide with the CNL of the bulk Si. The CB offset is given by [11]

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

Here, χ_a is the EA of the oxide, χ_b is the EA of the semiconductor and $\Phi_{S,a}$ and $\Phi_{S,b}$ are the CNLs 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 CB edge. S is a constant, the Schottky barrier pinning factor, which is found by Mönch [93] to vary empirically with the electronic component of the dielectric constant of the wider gap material (the oxide) as

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

This CNL model is a zeroth-order but fully determined model of band offsets, in which the CNLs are determined by the bulk electronic structure of oxide and of Si. The local bonding at the interface does not enter in this model. (The CNL acts like the mean electronegativity of the semiconductor. The charge transfer can be considered to arise from a difference of the two bulk electronegativities across the interface.)

The predicted CB offsets in this model [11, 72] are given in table 6 and figure 27 for the various oxides. Table 5 compares these with the experimental values measured by photoemission, internal photoemission or barrier tunnelling [97–113]. 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 CB or the Si CB to the oxide valence band, depending on the Si doping and of the polarity 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 [98, 103]. $SrTiO_3$ indeed has a small CB offset [97]. There is now recent data by Hattori *et al* [111] for La_2O_3 which agrees well with the prediction of 2.3 eV. La_2O_3 and $LaAlO_3$ have particularly large CB offsets [111, 110] which means they could be the second generation high K oxides with lowest leakage. The experimental and calculated values for Lu_2O_3 and Gd_2O_3 are also in reasonable agreement [113]. There is good agreement for Al_2O_3 made by oxidation [99, 100]. The largest exception is the internal photoemission of Afanasev *et al* [101] for Al_2O_3 . This is because

Table 6. Comparison of the calculated CB offset (by LDA method) and experimental values for various gate oxides, by various authors.

	Calculated (eV)	Experiment (eV)	References
SiO ₂		3.1	Alay [109]
Ta ₂ O 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	Afanasev [83]
Al ₂ O ₃	2.4	2.8	Ludeke [81]
		2.2 ^a	Afanasev [83]
a-LaAlO ₃	1.0	1.8	Edge [87]
La ₂ O ₃	2.3	2.3	Hattori [88]
Y ₂ O ₃	2.3	1.6	Miyazaki [89]

^a ALD.

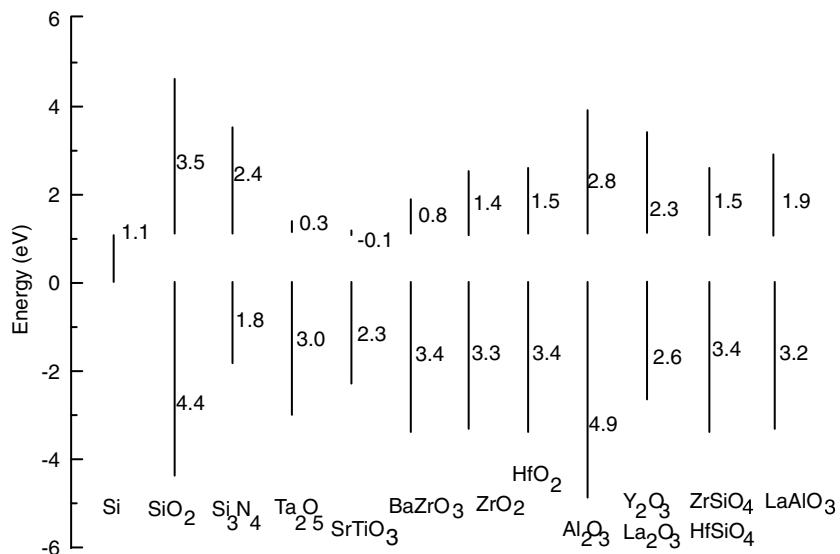


Figure 27. Predicted barrier heights for a range of high K gate oxides, after [11].

these authors used Al₂O₃ films grown by ALD whose band gap is much less (6.8 eV) than that of the pure bulk oxide (8.8 eV).

It is seen that only Zr, Hf, Al, Y and La 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–V.

Lucovsky *et al* [73, 104] 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 CB of the oxide is pure metal d, and so its energy tends to follow the band offset.

The fact that the MIGS model is insufficient is seen by the case of the (111)Si : NiSi₂ interface, where the Schottky barrier height has two values depending on whether the local

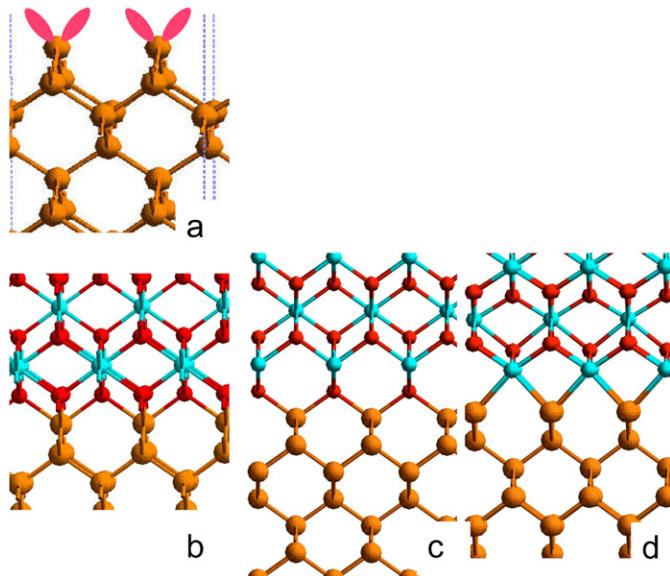


Figure 28. Bonding at (100)Si : ZrO₂ interfaces. (a) Si(100) surface, (b) OOZr, (c) OZr and (d) Zr terminated interfaces.

interface bonding is a type A or type B interface [90]. There must be an explicit component depending on interface bonding which allows band offset engineering. We shall return to this in section 4.4 and later.

4.4. Bonding at ZrO₂ and HfO₂-Si Interfaces

The simple MIGs model of the oxide interface is surprisingly successful. Nevertheless, we need a more detailed description of the Si-oxide interface. It is important to know the detailed bonding at the Si-oxide interfaces for three reasons. Firstly, the band offset does depend on the interface bonding. Secondly, imperfect interfaces will have defects which can create states in the gap which trap charge. Thirdly, the control of WFs in the case of metal gates requires this understanding of interface chemistry.

It is useful to consider epitaxial oxide systems in order to understand the bonding principles in more detail [114–126]. We choose the Si : ZrO₂ system because it is a reasonably well lattice-matched interface and ZrO₂ has (when Y doped) the high symmetry cubic lattice. The lattice constants of Si and ZrO₂ are 5.43 Å and 5.07 Å, respectively. This allows ZrO₂ to be grown epitaxially on the Si(100) cube face [127, 128], with the ZrO₂ cube face lying directly on top of the Si cube face. This is expressed as ZrO₂(100)//Si(100), and with the [001] directions of Si and oxide parallel, that is ZrO₂[001]/Si[001]. The ZrO₂ : Si system is representative of HfO₂ and also of other cubic oxide systems such as the bixbyite series of Y-rich oxides (Y,La)₂O₃ [129–131]. The lattices of Si and fluorite structures are closely related so that high symmetry interface structures are possible [132–134].

The ideal (100) surfaces of ZrO₂ are polar, being oxygen terminated. However, we can form bulk non-polar units of ZrO₂ as O²⁻Zr⁴⁺O²⁻ (or ‘OZrO’) units by assigning Os alternately up or down to give non-polar faces [122].

Now consider the ideal Si(100) face (figure 28(a)). Here, each Si atom has 2 dangling bonds (DBs). These will create states in the gap. If we place a non-polar OZrO unit on this

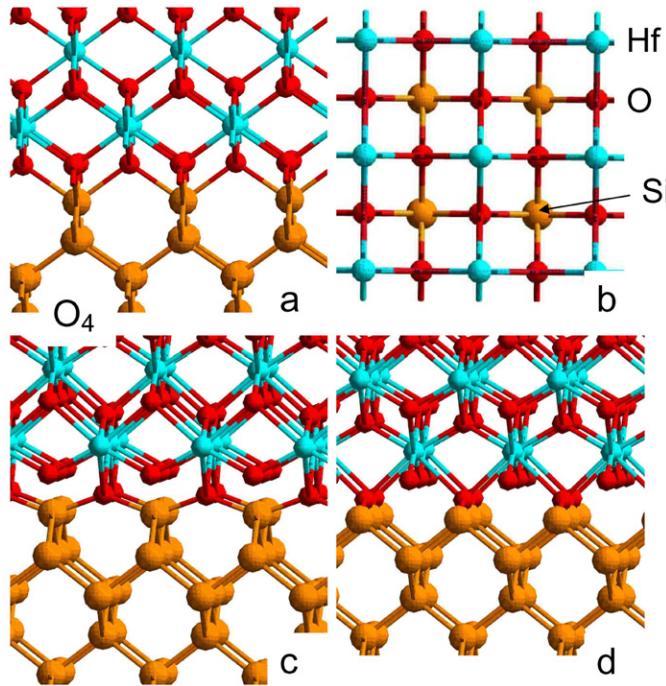


Figure 29. Various calculated interface configurations of (100)Si:ZrO₂ interface. (a) Ideal O₄ interface, (b) top view along [001], (c), (d) relaxed O₄ interface viewed from the [110] and [1̄10] directions.

(100)Si, there will be no reaction and this will still leave Si DB states in the Si gap. These gap states will create a metallic interface and prevent the field effect action needed for an FET. Therefore it is not a useful interface.

If instead we put a polar OOZrO unit on the Si(100), the first O forms two Si–O bonds with each silicon. This O, being divalent, saturates the two Si DBs to form a Si–O–Si bridge. Then, the non-polar OZrO unit is added on top. The whole ZrO₂ lattice can be built up on top of this interface by adding further non-polar OZrO layers. This interface will have no gap states and is what we want.

This also works with a ZrO terminating unit. In this case, the ZrO is formally Zr²⁺O²⁻ and the Zr has two unsatisfied valences. These can make two polar Zr–Si bonds to the Si DBs. This also gives an insulating interface with all valences satisfied. The two examples show that epitaxial growth of ZrO₂ on (100)Si is possible, with valence satisfaction and insulating interfaces, provided that the polar faces of ZrO₂ are used. The three cases of OO, O and Zr termination are shown in figures 28(b)–(d).

We have carried out detailed total energy calculations of various atomic models of (100) interfaces to test these ideas using the GGA of the LDA [119–122]. Some of the interfaces are shown in figures 28–30. Figure 28(c) shows the ideal Si:OZrO interface, which has only one layer of four-fold coordinated oxygen sites at the interface. We find this interface to be metallic, as expected from the above discussion. Figure 29(a) shows the ideal Si:OOZrO interface, with a double oxygen layer at the interface. Here the interfacial oxygens are six-fold coordinated initially, bonded to two Sis and four Zrs. It is found that the interfacial oxygens relax to form the structure shown from two directions in figures 29(c) and (d). Those oxygens

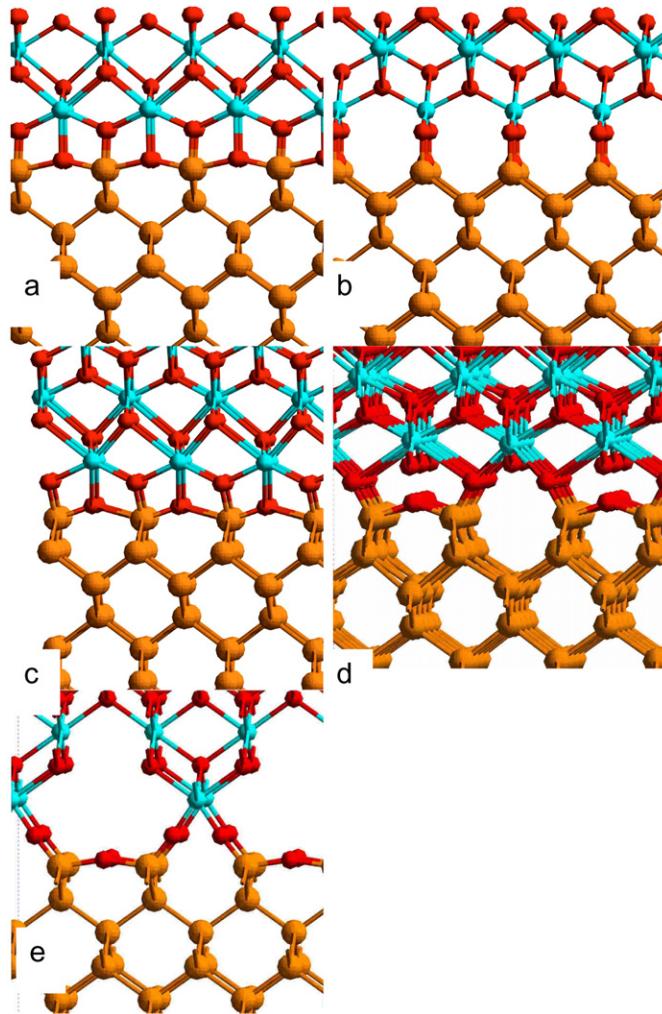


Figure 30. Calculated configurations of O_3 interface from the (a) $[110]$ and (b) $[1\bar{1}0]$ directions, the (c) O_{3T} , (d) O_{3B} and (e) the O_{2A} interfaces.

lying in the Si–O–Si bridges relax towards the silicon layer. The other two oxygens relax up towards the ZrO_2 layer. This replicates our simple discussion above. This interface is denoted the O_4 .

Another interface can be constructed with the oxygens being initially three-fold coordinated to one Si atom and two Zr atoms. This is denoted the O_3 interface. The oxygen bonding is then similar to $ZrSiO_4$. This interface structure relaxes to the configuration shown in figures 30(a) and (b). Here, half of the oxygens are bonded to two Sis and one Zr, and the other half are bonded to two Zrs and one Si. The top layer Sis are each five-fold coordinated. This interface is also insulating.

A third O-terminated interface O_{3T} with three-fold coordinated oxygens is possible as shown in figure 30(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 O₃ sites are no longer planar and this allows it to gain stability.

A fourth O-terminated structure O_{3B} is shown in figure 30(d). Here, one DB of each Si is used in a lateral Si–O–Si bridge [110]. This leaves one DB to bond to the ZrO_2 layer. However, this needs an extra half ML 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* [125]. 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_{2A} interface (figure 30(e)). On the interface Sis, one of the two Si DBs forms a Si–O–Si bridge with its neighbouring Si. This also occurs at the $(100)Si : SiO_2$ interface. The other Si DB then forms a Si–O–Zr bridge to the first Zr layer. This Si–O–Zr bridge is a covalent unit. Above this Zr, the rest of the ZrO_2 bonding is ionic, as in normal bulk ZrO_2 . This interface has a 2×1 symmetry. A centred 2×2 cell is also possible. The interesting thing here is that this interface could be formed by ALD deposition, according to molecular dynamics simulations [125]. The precursor $ZrCl_4$ is a covalently bonded molecule, and ALD is carried out on a partly pre-oxidized 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 ML.

Overall, these interfaces have the same number of oxygen atoms at the interface. The O_{3T} interface is found to be the most stable structure and then O_3 . The O_4 interface is marginally less stable than O_3 . Extensive testing finds that the O_{2A} is as stable as the O_3 interface [122]. 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 [128] measured the interface configuration at $(100)Si : ZrO_2$ by high-resolution TEM. 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 six-fold coordinated Zr_6 , as in figures 31(a) and (b). This structure relaxes so that the terminal Zr –Si bond lengthens. Figures 31(c) and (d) shows another interface in which Zr is ten-fold coordinated, with the Zr bonded to four oxygens, four Sis in the top layer and to two more Sis in the layer under that. This bonding is similar to that in $ZrSi_2$. Our calculation finds that the Zr_{10} is the more stable of these two Zr -terminated interfaces by 0.3 eV per interface Si (figure 32).

The calculations find that the three interfaces, O_4 , O_3 , O_{3T} and O_{3B} and Zr_6 are insulating [119]. 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. Dong *et al* [126] recently carried out an extensive study with similar results. Chang *et al* [123, 124] calculated the surface electronic structure of some $Si : ZrO_2$ interface configurations, some of which were metallic. Similarly, Fiorentini and Gulleri [114] calculated the stabilities of some $Si : HfO_2$ interfaces.

The band offsets have been derived from the calculations of the various interface structures of $Si : ZrO_2$ from the calculated alignment of the DOS of Si and oxide (figure 33). This gives the offset of the valence bands. The offset of the CBs is not given well by the LDA calculations, as LDA underestimates the band gap. The VB offset is calculated and the CB offset is then found by adding the experimental band gaps to the VB offset. A second method to calculate the band offsets uses the internal electrostatic potential method. In this case, the energy of the VB edge above the average electrostatic potential is found for the bulk oxide and bulk Si (figure 34). Then the difference in electrostatic potential across the interface is found for a supercell.

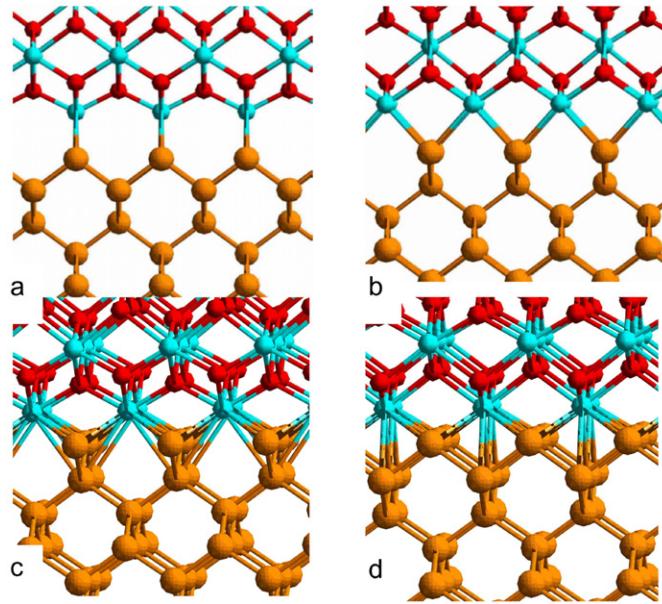


Figure 31. Calculated metal-terminated configurations of Zr_6 and Zr_{10} interfaces. (a), (b) Zr_6 interface viewed from $[110]$ and $[1\bar{1}0]$ directions, respectively. (c), (d) Zr_{10} interface viewed from $[110]$ and $[1\bar{1}0]$ directions.

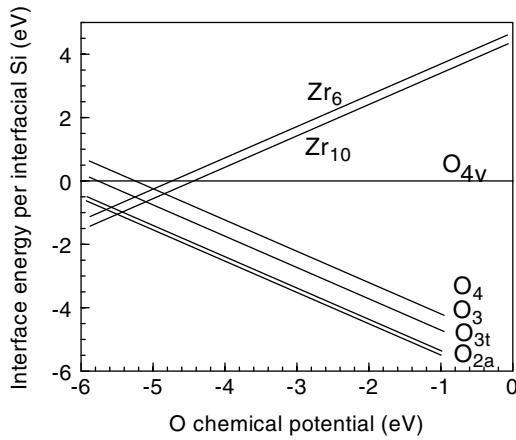


Figure 32. Interface energies of various interface $(100)\text{Si}:\text{ZrO}_2$ configurations versus oxygen chemical potential. See Dong *et al* [126] for details.

It is found that the VB offset is fairly 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. In contrast, the VB offset for Zr-terminated interfaces has a large variation of 0.9 eV between Zr_{10} and Zr_6 . The VB offset is much larger for Zr_6 . This is due to a large change in the interface dipole [122]. It can be related to the Zr–Si distance projected along the surface normal. This distance is short for Zr_{10} , so the dipole with Zr^+ is small, so the VB offset is less for this case. Summarizing, the interface-specific dipole is quite small for the O-terminated interfaces, but there is a large dipole change for the Zr-terminated interfaces. An even larger change of offsets was found by

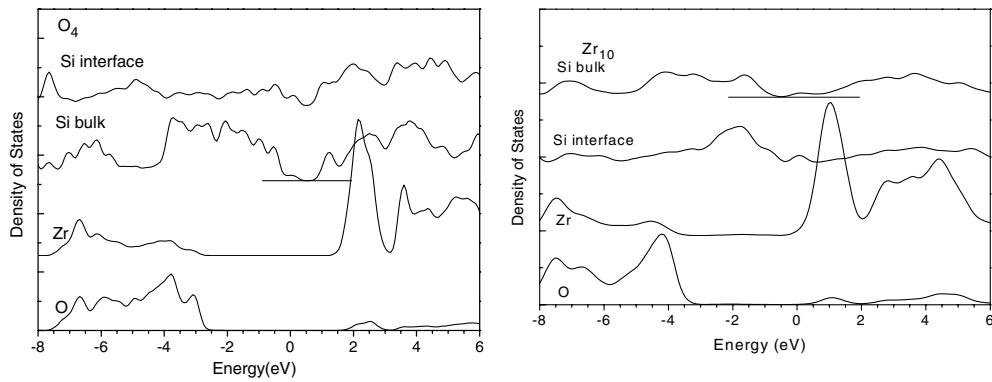


Figure 33. Local DOS of (100)Si:ZrO₂ for the O₄ and Zr₁₀ interfaces.

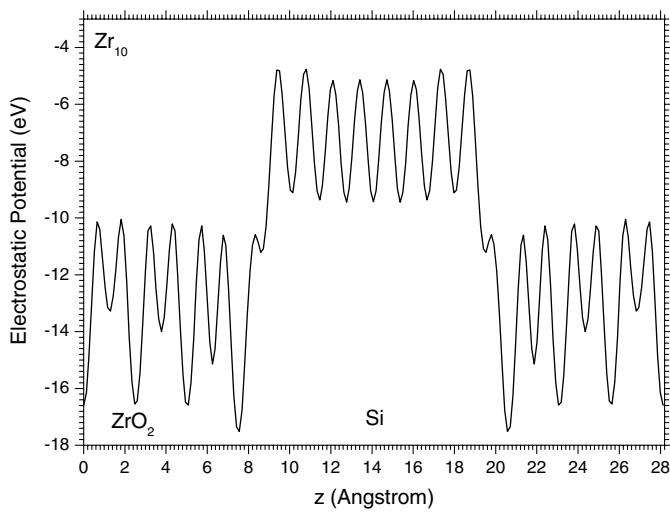


Figure 34. Average electrostatic potential along a supercell containing Zr₁₀ interfaces. Potential averaged normal to z axis.

Dong *et al* [126]. The possibility to change the offset is seen experimentally on Ni on ZrO₂ interfaces [135] and is consistent with the calculations of Klinik *et al* [136]. The variation in band offset [134] can be even larger at interfaces of more ionic compounds such as Si:CaF₂. These results emphasize how much the interface dipole can depend on termination and go beyond the MIGS model [91]. Their importance lies in their implication for metal gates. A similar situation occurs for the Si:SrTiO₃ interfaces grown by McKee *et al* [155], see later.

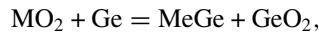
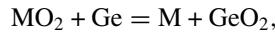
The relative constancy of the band offset for O-terminated ZrO₂ interfaces is valuable technologically. It means that the band offset of a ZrO₂ gate oxide does not depend on the surface orientation. It is therefore relatively constant for the polycrystalline or amorphous oxide interfaces. This is very convenient, as it means there will be a larger process window for oxide formation. It is also similar to the established case of Si:SiO₂ where the band offset is constant between Si faces [137]. On the other hand, the band offsets at the two Zr-terminated interfaces differ.

The interface of ALD HfO_2 on Si(100) has been modelled by various groups [125, 138, 139]. ALD nucleates better on slightly oxidized surfaces, so the starting situation is modelled by the hydroxyl-terminated 2×1 reconstructed Si(100) surface. Each surface Si has one OH group, and is bonded to the adjacent via an Si–O–Si bridge. The first ALD reaction occurs by the reaction of the precursor such as HfCl_4 with the –OH groups to form HfCl_3O^- groups. These groups are then oxidized to form Si–O–Hf bridges. This first layer is a covalent bridge because steric hindrance around the large precursor molecules prevents the higher density ionic structure forming. The subsequent layers do grow into bulk HfO_2 whose bonding is ionic. The interface has a 2×1 symmetry. The ALD growth of Al_2O_3 has also been simulated [140, 141] and follows the same pattern.

4.5. Oxides on Ge

ALD is now the standard way to grow high K oxides on Si. The basics of ALD are described in section 3.1. The difficulty of nucleating HfO_2 on H-terminated Si (HF last Si) led to the process of growing HfO_2 on O-terminated Si. This ‘chemical oxide’ is created by either cleaning Si by an oxidative treatment (ozone or H_2O_2 clean) or growing an initial very thin SiO_2 layer. This obviously creates a Si– HfO_2 interface with an interfacial SiO_2 layer. The HfO_2 will grow amorphous and convert to nanocrystalline during heat treatment. There is unlikely to be an epitaxial relationship.

Recently, Ge has started being used as a substrate because of its higher carrier mobility than Si. Si–Ge is also of interest as their strained layers also show higher mobility than Si. ZrO_2 and HfO_2 have been grown on Ge [142–145]. GeO_2 is proportionately less stable than SiO_2 . The equivalent reactions to (5) and (6),



are more endothermic. Interfacial GeO_2 forms less readily. Consequently, when ZrO_2 is grown by ALD on Ge, the absence of an interfacial layer allows some direct contact between ZrO_2 and Ge, and epitaxial regions occur [143, 144]. The large lattice-mismatch means that epitaxy is not good with frequent dislocations, but nevertheless it occurs.

The problems with growth on Ge are that the diffusion rates of most species are much faster in Ge than in Si. Hf diffuses into the Ge substrate. Processing temperatures should be lowered to near 400 °C [142], but this is not always possible. There is an effort to develop diffusion barriers such as N or P to work with Ge-oxide interfaces [145]. The interface reactions have been traced by MEIS and back-etched photoemission experiments [145, 147]. The band offsets for Hf and Zr oxides on Ge are quite similar to those on Si, the CB offset being higher by 0.1 eV [146, 147].

4.6. Other binary oxides

Many of the binary oxides ZrO_2 , HfO_2 , CeO_2 , Y_2O_3 and Gd_2O_3 have lattice constants which are close to Si. This raises the possibility of lattice-matched growth, and perhaps eventually forming perfect Si-oxide interfaces by molecular beam epitaxy (MBE). CeO_2 is cubic with a 0.37% mismatch to Si. Y_2O_3 has the cubic bixbyite structure with a lattice constant of 10.61 Å, which has a 2.5% mismatched to twice the Si unit cell of 5.43 Å. Y_2O_3 can be matched to Si to within 0.18% by alloying with La_2O_3 with the larger radius La ion. It would be very

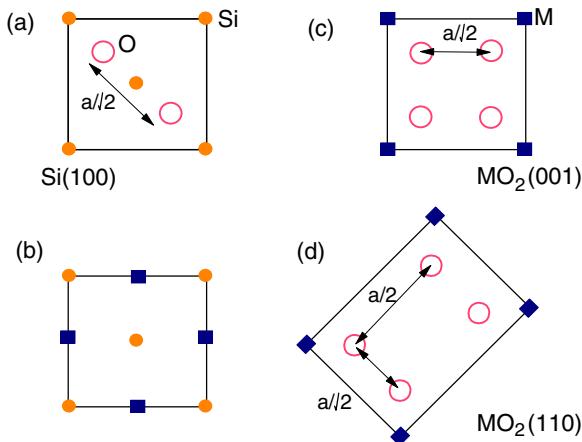


Figure 35. Schematic of bonding on Si(100) for (a) initial oxygen layer, (b) initial Zr layer. (a) favours (110) growth, (b) favours (100) growth.

useful if these oxides grew in the cube-on-cube (100)/(100) orientation on Si. In practice, this rarely occurs. Many MBE studies have been carried out. The oxides will grow readily as (111)/(111) on Si [129–131, 148–151]. They will also grow on Si(100) as oxide(110)/Si(100). The latter situation is particularly bad because this forms two domains of oppositely oriented oxide with [100]Si // oxide[110] and [1̄10]. Why?

The reason is thought to be as follows. Figure 35 shows a top view of the Si(001) surface. The two Si DBs per site point along [110] directions. Incident oxygen atoms will bond to these to form Si–O–Si bridges oriented along [110] because oxygen is divalent. This gives an O-terminated square surface cell with oxygens separated by $a/\sqrt{2}$, where a is the Si lattice constant [130]. Compare this with the O-terminated cubic fluorite lattice of say CeO₂, and the oxygens form a square cell separated by $a/2$. This does not match to the oxygen spacing on O-terminated Si.

On the other hand, make a (110) surface of the fluorite. The surface oxygens are separated by $a/2$ in [110] and by $a/\sqrt{2}$ along [1̄10]. Thus, this lattice matches to the O-terminated Si at least along one direction. In the other direction, there is a fair match if the average is taken over 2 cells of oxide per 3 cells of Si. Thus, if Si is O-terminated, there is better overall matching for (100)Si//(110)oxide, and this accounts for the observation.

The way to solve this problem is to use metal-first growth, as found by Osten *et al* [152] for Gd₂O₃ on Si. The stages growth are monitored by taking XPS of the core levels of Si and oxide for the growth of the first few oxide MLs. The 2p core level of a Si atom in SiO_x shows a 1.2 eV shift per O bond neighbour. This is due to the charge transfer to O⁻. The Si shows a negative shift when bonded to a metal like Zr or Gd. Growth is controlled so that the first ML forms Si–metal bonds. Then the O₂ partial pressure is controlled so that the oxide grows normally, forming M–O bonds. O diffusion through these oxides is easy, so that eventually O can diffuse to the interface and oxidize the initial Si–M bonds to Si–O–M bonds. This creates the eventual Si–O terminated interface but frozen into the configuration determined by the first Si–M growth. It is possible that this occurred in the method of Wang and Ong [128] after desorption of SiO.

This is consistent with molecular dynamics simulations of Zr on Si(100). Blochl *et al* [153] founds that for a half ML of Zr, the Zr adopts the hollow site between four Sis. This is

the same as Zr at our Zr_{10} interface. The Zr lies in the same symmetry at Zr_{10} as it does at the O_4 interface. Thus, the Zr hollow site will initiate the Zr_{10} interface, and this will subsequently oxide to give the O_4 interface of oxide(100)//Si(100).

4.7. Perovskites

$SrTiO_3$ has been widely studied as a substrate for high T_C oxide superconductors. Its alloy $Ba_xSr_{1-x}TiO_3$ (BST) has been widely studied as a high dielectric constant dielectric and a possible replacement for SiON in DRAM capacitors [12]. Thus $SrTiO_3$ was initially of interest as a possible gate dielectric. This interest was lessened when it became clear that $SrTiO_3$ was reactive with Si and that it had a small CB offset. Nevertheless, $SrTiO_3$ has a very good lattice match to Si, and its growth on Si gives important information [154–158].

$SrTiO_3$ (STO) is cubic and has a lattice constant of 3.91 Å. Si has a lattice constant of 5.43 Å. The (001) face of the $SrTiO_3$ lattice can be matched to the (001) face of Si if the $SrTiO_3$ lattice is rotated by 45° so that the [100] $SrTiO_3$ direction lies parallel to [110]Si direction, that is, Si(001) // $SrTiO_3$ (001) and Si[100] // $SrTiO_3$ [110].

One solution to the lack of stability of STO in contact with Si was proposed by McKee *et al* [154, 155]. They grew a sequence of mutually stable compounds, $SrSi_2$ on Si, SrO on $SrSi_2$, and then grew STO on SrO . This allows the growth of bulk STO on Si. The growth process was evidenced by lattice images from high-resolution TEM and by elemental mapping across the interface by EELS. They also grew $BaTiO_3$ on Ge [155]. FETs have been made from these Si : STO structures [159].

In a second solution, Droopad *et al* [156–158] achieved growth of STO on (001)Si with a silicate type interface. FETs were made from the Si : STO structures [160]. They had quite good characteristics and leakage. Subsequent TEM, however, showed evidence of growth of SiO_2 interface layers, presumably during post-deposition annealing, as expected from the reactivity of STO and Si.

To understand the behaviour in Si : STO systems, we first consider the bonding at high symmetry interfaces. The lattice of bulk $SrTiO_3$ consists of alternate layers of SrO and TiO_2 stacked along [001]. The layers are neutral in terms of formal charges. Thus, the simple SrO - and TiO_2 -like faces are non-polar and have a closed electronic shell without states deep in the band gap. It is an advantage that we get non-polar faces without further manipulation.

The ideal Si(100) surface was shown before. It has two Si DBs (DBs) per surface atom. The real surface undergoes a dimer reconstruction, in which pairs of Sis move towards each other and form a lateral Si–Si bond, figure 36(a). This removes one DB per Si. The dimer reconstruction of the surface leaves one DB per surface Si. The DBs are half-filled orbitals with one electron per state, and they would form a state in the Si band gap and give a metallic interface. It is important for an electrically useful interface that all interface states are removed from the Si gap.

If we try to build $SrTiO_3$ by placing either of the non-polar SrO or TiO_2 faces onto the dimerized (100) Si surface, we do not get an insulating interface. The oxide side of the interface is a closed shell, but the Si DBs remain on the Si side and give states in the Si gap.

The way to remove these DB states is to first passivate the Si surface with a half ML of Sr [117, 118, 121], as in figure 36(b). The Sr is divalent and electropositive. It can transfer its two valence electrons, one to each of the Si DBs, to make Si anions. This fills the DB states. More importantly, if the interaction is right, the Sr s orbital will repel the Si DB states into the Si valence band, so that the states no longer lie in the Si band gap, as in figure 2. Then the Si surface is fully passivated. This is the Sr-terminated interface. It is then possible to build up

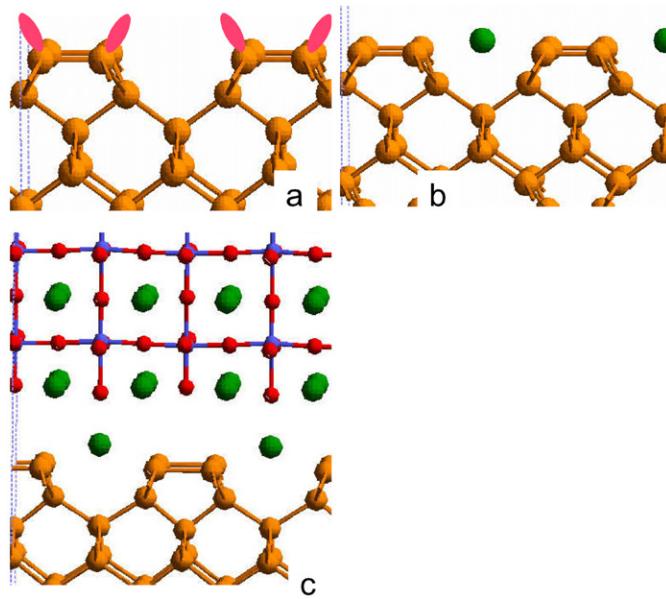


Figure 36. Bonding at a Sr-terminated (100)Si:SrTiO₃ interface. (a) Dimerized (100)Si, (b) 0.5 ML of Sr and (c) added SrTiO₃.

a SrTiO₃ crystal on top of this surface by adding alternately SrO and TiO₂ planes, as shown in figure 36(c). This interface can be called Sr-terminated STO. The bonding at the interfacial layer can be written as $(\equiv \text{Si}^-)_2 \text{Sr}^{2+}$, where the dashes denote covalent bonds. This is the same as in Zintl compounds [118].

A second interface, oxygen-terminated Si:STO can be constructed. We start again with the 2×1 dimerized Si(100) surface as shown in figure 37(a). We now bond one oxygen atom to each surface Si to form a neutral, non-bridging oxygen radical, as in figure 37(b). This still has a half-filled O DB state. This will lead to a metallic interface. So again, we can add 0.5 ML of Sr to this surface to passivate this surface, as shown in figure 37(c). The Sr transfers its 2 electrons, one to each of the oxygens, to fill their DB states. This gives closed shell non-bridging oxygen ions. This is also a passivated surface with no states in the Si band gap. The O DB states are well below the Si valence band maximum. We can then build up the rest of the SrTiO₃ layer on top of this interface, with the SrO or TiO₂ layer first, as shown in figure 37(d). The bonding at the interface can be written as $(\equiv \text{SiO}^-)_2 \text{Sr}^{2+}$, which is similar to a Sr metasilicate.

The O-terminated interface is reached at a higher O partial pressure than the Sr-terminated interface. At a still higher O partial pressure, the oxygen begins to oxidize the Si surface layer, converting the Si–Si dimers into Si–O–Si bridges, see figure 37(e). This interface is also a closed shell without states in the Si band gap. In both figures 1 and 2, the surface cell has 2×1 symmetry. It gives an easier picture than the centred 2×2 cells used earlier.

The electronic structure of these various interfaces were calculated by Forst *et al* [118], Demkov *et al* [116] and Peacock and Robertson [117, 121]. The insulating character was confirmed. A very interesting result found was that the O-terminated interface had a larger CB offset than the Si-terminated interface. The CB offset was estimated as 1.4 eV, over the minimum 1 eV required.

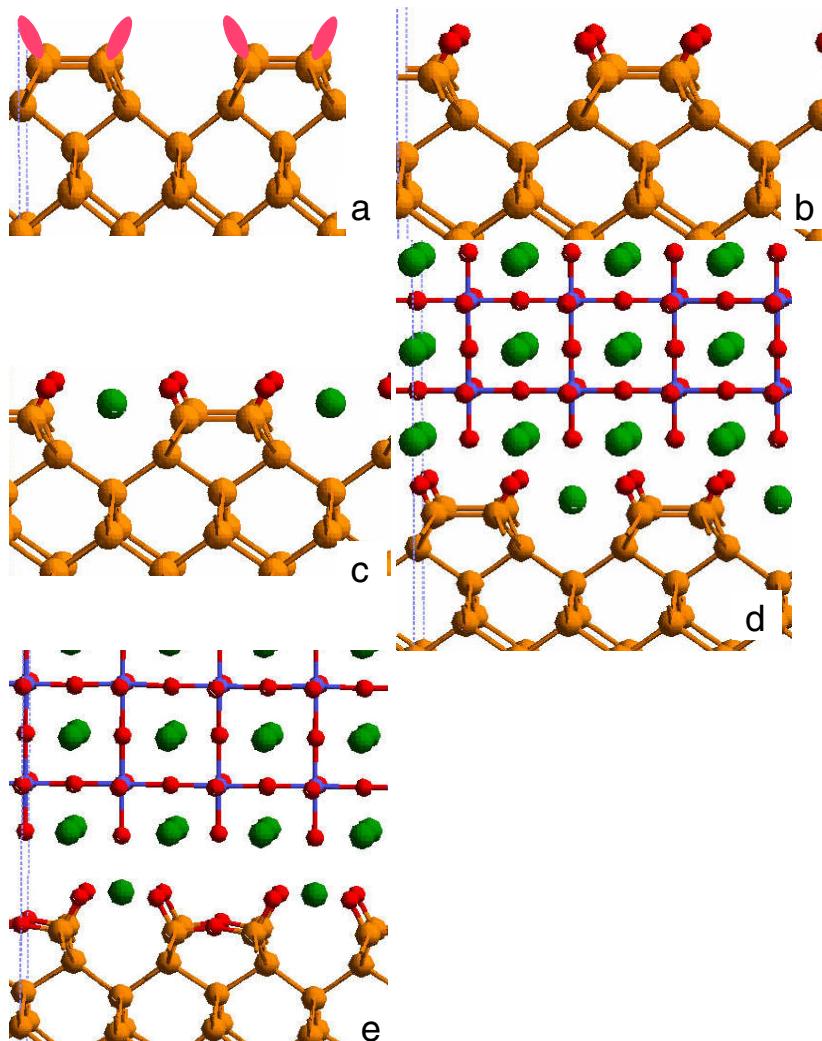


Figure 37. Bonding at a O-terminated (100)Si : SrTiO₃ interface. (a) Dimerized (100)Si, (b) added non-bridging oxygens, (c) 0.5 ML of Sr, (d) added SrTiO₃ and (e) with Si–O–Si bridges on the Si side.

The different offsets can be understood in terms of interface dipoles. At the Si-terminated interface, the charge transfer to create the Si⁺ and Sr²⁺ sites occurs *parallel* to the interface. There is no net charge transfer from the neutral SrO layer to the Si, so there is *no interface dipole*.

At the O-terminated interface, with its smaller VB offset of 1.1 eV, there is again charge transfer *parallel* to the interface to create the Sr²⁺ and O[−] sites. There is now a new dipole due to the polar Si–O bonds lying *normal* to the interface. This interface dipole reduces the VB offset. Thus, the interface stoichiometry from O-poor to O-rich modifies the dipole and band offset. Interface III has a similar band offset to interface II. This is because the Si–O–Si bridges lie laterally and give little extra dipole normal to the interface. The offset of interface IV with a TiO₂ layer next to the interface has a similar band offset, again because SrO and TiO₂ layers are neutral and have no charge transfer normal to the interface.

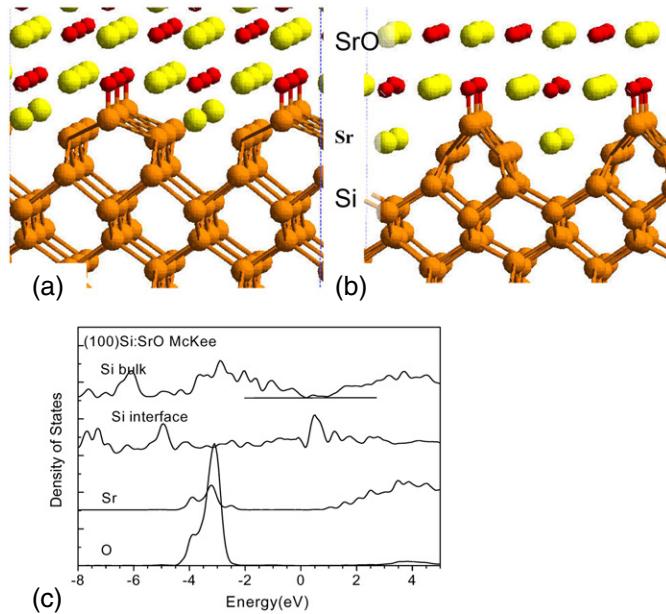


Figure 38. Simplified (100)Si:SrO interface according to the model of McKee *et al* [154]. (a) Unrelaxed, (b) relaxed and (c) calculated DOS.

We now return to the lower symmetry interface structure of McKee *et al* [154] derived from TEM and EELS (figure 38). It has a larger 4×2 surface unit cell than those in figure 1. Half the Sis in the top layer are missing, and 0.25 ML of Sr is placed in the resulting channels. The upper Sis are displaced to lie above four Si atoms in the next layer down, forming a ‘four-leg stool’. Above this lies a full SrO layer, with Os lying immediately over the Sis. Due to its larger surface unit cell, we carried out calculations on a Si:SrO analogue of this interface without TiO₂ layers, as in figure 3(a). It still captures the essential physics.

We find that the relaxed interface in figure 38(b) has a strong relaxation of the Si in the next to top layer. The DOS for this interface is shown in figure 38(c). We find that it has gap states, which essentially meet in mid-gap. These gap states arise from the non-tetrahedral bonding of the top layer and next-layer silicons. These gap states would cause the capacitance–voltage (CV) plot of this interface to stretch out, so the interface would have a poor electrical performance as an FET. However, an interface with very good CV properties has been shown [159], which cannot have states in the gap. The original interface structure was confirmed by TEM [154]. We must therefore conclude that the interface used in the CV measurements differs in some respect from that used in the TEM. For example, the annealing to make the FET could have modified the interface.

4.8. LaAlO₃

LaAlO₃ is another perovskite like SrTiO₃. It has a small 1.3% lattice mismatch to Si. It is stable next to Si as La and Al oxides are both stable next to Si [14, 161]. It has low oxygen diffusion coefficients, so that an interfacial SiO₂ layer does not grow, making it a desirable gate oxide if epitaxial oxides were ever used [162]. It has a 5.6 eV gap [77]. It has a relatively

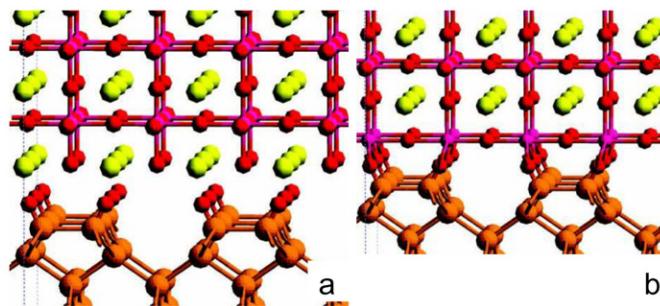


Figure 39. (100)Si : LaAlO₃ interfaces. (a) O-terminated interface with LaO last layer, (b) with AlO₂ last layer.

large CB offset of 1.8 eV [110, 163]. Experimentally, it has proved impossible so far to grow epitaxial LaAlO₃ layers on Si [110], except with a SrTiO₃ intermediate layer. The LaAlO₃ layers are amorphous. Recently, it was possible to grow Si on (100)LaAlO₃, which means that the Si : LaAlO₃ interface can be studied [164].

The difference in behaviour of LaAlO₃ to SrTiO₃ arises from the metal valence. La and Al are both 3-valent, so that La and Al both could attempt to occupy the same lattice site, which would tend to make it amorphous.

The 3–3 metal valence means that the component LaO and AlO₂ layers on the bulk cell have charges +1 and –1. This has a large effect on the possible surface and interface structures. Neutrality requires each +1 charged LaO layer to be compensated by –1/2 charge layer on either side of it. And similarly a –1 charged AlO₂ layer would be compensated by +1/2 layers on either side.

Theoretically, there are a number of high symmetry interface structures [121], La-terminated and AlO terminated. A LaO-terminated layer having +1e per O will want to donate 1/2e to the Si side. It can do this if Si presents a face with non-bridging oxygens, as in figure 39(a). The NBOs each have charge –1e. We call this type I LaO termination. But its overall charge is –1/2e per O.

Alternatively, an LaO will donate 1/2e to Si DBs, to make negative Si DBs, as at the Sr-terminated STO interface. The Si–O contact is not a good bond because both Si and O sites are negatively charged. Thus, this interface is rather unstable.

Thirdly, an AlO₂ layer has a formal charge of –1. It will donate +1/2e to the Si layer. This Si⁺ can then accept a dative bond from one oxygen in the AlO₂ layer, as shown in figure 39(b). This oxygen is now lower than the other oxygens. All these interfaces are calculated to have no states in the Si gap.

The need to compensate the net formal charge of the LaO and AlO₂ layers creates problems in forming high symmetry interfaces. Proper compensation occurs if the final layer has a net charge of ±1/2. This occurs by allowing vacancies. Oxygen vacancies have a formal charge of +2, so that a cell with 1 in 4 vacancies would be needed. Alternatively, use La vacancies. Assume that the Si layer is terminated by non-bridging oxygens of formal charge –1 and that these oxygens lie in the LaO layer. Each O must give –1/2e to the adjacent AlO₂ layer, is –1e due to being non-bridging, or a net –3/2e. Thus 3 La ions per 6 Os gives neutrality. Experimentally, Klenov *et al* [164] found that the Si : LaAlO₃ interface had a 3 × 1 reconstruction, with 2 out of 3 columns appearing to have heavy La ions attached. This is consistent with a 3 × 2 cell, if the 2-fold repeat is disordered.

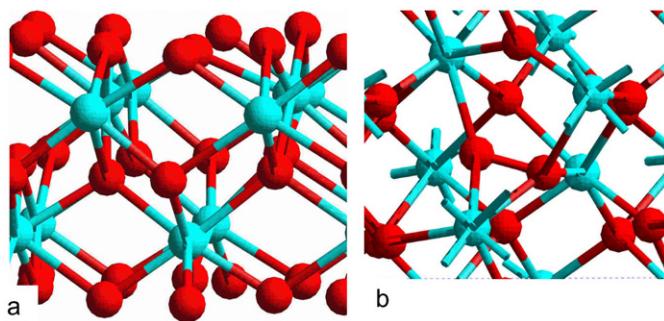


Figure 40. The relaxed structure of (a) neutral oxygen vacancy and (b) the neutral oxygen interstitial in ZrO_2 .

Recently, Forst *et al* [165] calculated the stability of various interfaces, and found that LaO terminated interfaces were more stable. A version of 3×1 interface was found to give no gap states.

5. Electronic structure of defects

One problem with high K oxides is that they contain a much higher defect concentration than SiO_2 . SiO_2 has such a low defect concentration for two reasons. First, its large heat of formation (large Si–O bond strength) means that off-stoichiometry defects such as O vacancies are costly and so they are rare. The second is that SiO_2 has a low coordination [166]. Its covalent bonding means that the main defects are DBs, and the low coordination allows the SiO_2 network to relax to remove any DBs by re-bonding the network. This occurs particularly for defects at the Si : SiO_2 interface.

The high K oxides differ from SiO_2 in that their bonding is ionic and they have a higher coordination number. The more ionic bonding and higher coordination numbers mean that the high K oxides are poorer glass formers [35, 166]. The effect of poor glass-forming ability and high coordination is that the oxides have larger non-equilibrium defect concentrations. The oxides still have high heats of formation, so the equilibrium defect concentrations should be low. However, the non-equilibrium defect concentration is high because the oxide network is less able to relax to rebond and remove defects.

The oxygen vacancy and oxygen interstitial (figure 40) are the two most likely intrinsic defects in ZrO_2 and HfO_2 in terms of their formation energies [167, 168]. Defects at the metal site would cost more energy.

First, let us consider the oxygen vacancy in ZrO_2 and HfO_2 . Recall that the valence band of ZrO_2 consists mainly of O p states and the CB consists mainly of Zr d states. The CB is split into e and t_2 states by the crystal field. This is the simple model of ZrO_2 as O^{2-} and Zr^{4+} ions.

The structure and electronic structure of the oxygen vacancy and oxygen interstitial in ZrO_2 and HfO_2 have been calculated by various groups [70, 167–176] including Foster *et al* [167, 168] and Xiong *et al* [174, 175]. Many used LDA methods. The LDA produces good total energies and structures, but it is well known that LDA under-estimates band gaps, by 60% in the case of ZrO_2 . It can also give the wrong localization of gap states even when filled. Thus it is necessary to correct for the band gap under-estimate by LDA, and this often done by the scissors operator (moving the gap to fit the experimental value) and moving the defect level in various ways. Foster *et al* [167] widened the gap but left the vacancy level

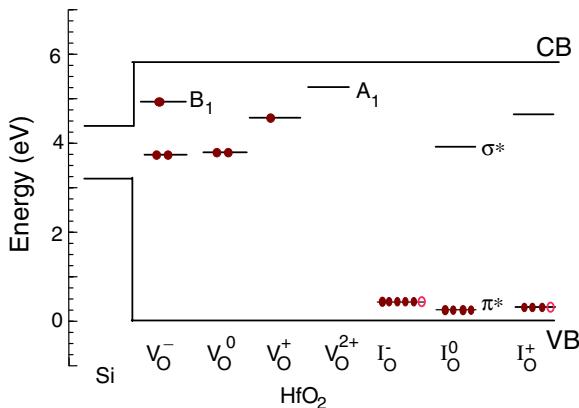


Figure 41. Molecular orbital diagram of relaxed oxygen vacancy in HfO_2 in various charge states.

at the same energy above the VB edge. 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 lying below the Si VB edge. Other groups moved the vacancy level up with the CB [169], so that it lies high in the gap. Perhaps a more natural adjustment is neither of these, but to move the vacancy level in proportion to its LDA energy in the gap.

The alternative is to use methods which require no empirical correction [70, 79–81]. Kralik *et al* [70] calculated the energy level of the ideal neutral O vacancy by the GW approximation, which is generally regarded as the most accurate but the 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 *et al* [175] used the sX method [79] to calculate the defect excitation energies for ZrO_2 and HfO_2 . A supercell of 48 atoms was used. The geometry is relaxed in GGA and the energy levels calculated in sX.

The O vacancy creates a singly degenerate gap state of A_1 symmetry. It is occupied by 2 electrons for the neutral vacancy. The state is strongly localized on the d orbitals of the adjacent Hf ions. In c- HfO_2 , the state of the relaxed neutral vacancy is calculated to lie at 3.8 eV above the oxide VB, figure 41. The ionic positions do not relax much from their ideal positions in the neutral vacancy, the Hf–Hf distance is 3.59 Å compared with 3.54 Å in bulk HfO_2 .

At the positive vacancy, V^+ , the A_1 defect state is now singly occupied. Its net positive charge causes the adjacent Hf^{4+} ions to relax outwards from the vacancy. The Hf–Hf spacing becomes 3.74 Å. This relaxation causes the defect state to move upwards from 3.2 to 4.7 eV (figure 41). At V^{2+} , the A_1 state is now empty. The greater positive charge causes further outward relaxation of the Hfs; so the Hf–Hf separation is now 3.90 Å. The gap state now lies higher at 5.2 eV due to the relaxation. A similar behaviour is found for the vacancy in the other poly-types.

The vacancy can also trap one or two electrons. A trapped electron causes the adjacent Hf ions to distort asymmetrically, pulling down an extra, singly-degenerate state of B_1 symmetry out of the CB. It is singly occupied for V^- and doubly occupied for V^{2-} . The A_1 state is full in both cases. The complete spectrum of the vacancy levels is summarized in figure 41.

The strong effect of lattice relaxation on the energy levels of the vacancy is due to the ionic bonding and the strong localization of the defect wavefunction on the Hf ions adjacent to

the vacancy. The wavefunction extends to the neighbouring ions. It is not as localized as the vacancy wavefunction in a very ionic solid like MgO, where it is localized within the vacancy itself [177].

The present vacancy level energies are much more consistent with experiment than the previous results, as discussed in section 6.1. Takeuchi *et al* [178] used spectroscopic ellipsometry on HfO₂ films oxidized to different levels to identify an absorption band at 4.5 eV. They attribute this to transitions from the HfO₂ valence band to the oxygen vacancy, and so place the V_O level at 4.5 eV in the gap. Kerber and Cartier [179] noted that this trap must lie just above the Si conduction edge. Mitard *et al* [180] placed trapping levels at \sim 1.5 and 0.8 eV below the oxide CB.

The other main defect is the oxygen interstitial (figure 40(b)). It can have a number of charge states, see figure 42. Their energy levels were calculated using the weighted density approximation (WDA) [81, 175]. The simplest is the closed shell species I²⁻ equivalent to an extra O²⁻ ion. In this state, the interstitial is well separated from the other oxygen anions and it adds filled O 2p states just to the valence band. At the I⁻ or O⁻, removing 1 electron leaves a hole at the VB edge. This ion moves slightly closer to another O²⁻ without actually forming a direct O–O bond; the O–O distance is 2.0 Å [168]. The neutral I⁰ or O⁰ interstitial has two holes in the O 2p levels. The two holes allow this interstitial to form a true O–O bond, giving the dumb-bell shaped superoxy anion O₂²⁻. The O–O bond length is 1.49 Å. This O–O 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 lying at -3.0 eV in the valence band and at $+0.3$ eV just above the VB edge (figure 43).

The σ^* state of I⁰ could trap an electron, in which case this would 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. This O⁺ ion forms a shorter O–O bond with a O²⁻ ion of length 1.39 Å, giving a dumbbell O₂⁻ ion. This gives rise to an empty σ^* state in the upper gap at 4.5 eV, (figure 43). It also has a filled π state and a half-filled π^* state. The Fermi level lies in the π^* state which lies just above the valence band top. This O₂⁻ ion is called the superoxyl 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 (ESR) in HfO₂ thin films [181].

The oxygen vacancy and interstitial have a similar behaviour in the other binary oxides. Figure 44 shows the O vacancy levels for La₂O₃. The La ions next to the vacancy also relax outwards for the positively charged vacancy. The A₁ vacancy state rises upwards in response. The overall levels lie slightly deeper in the gap than in HfO₂. The vacancy also supports negatively charged states, in the B₁ state. Overall, the vacancy levels lie slightly higher above the Si CB edge than in HfO₂ because of the larger CB offset of La₂O₃.

The O interstitial is similar in the silicates as in HfO₂. The vacancy levels in HfSiO₄ and ZrSiO₄ behave slightly differently than in the binary oxides [176] because now the vacancy site can have both Hf and Si neighbours.

6. Electrical quality

We have so far described the production, characterization 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 better than SiO₂. There are three aspects to this: charge trapping, mobility and gate threshold shifts.

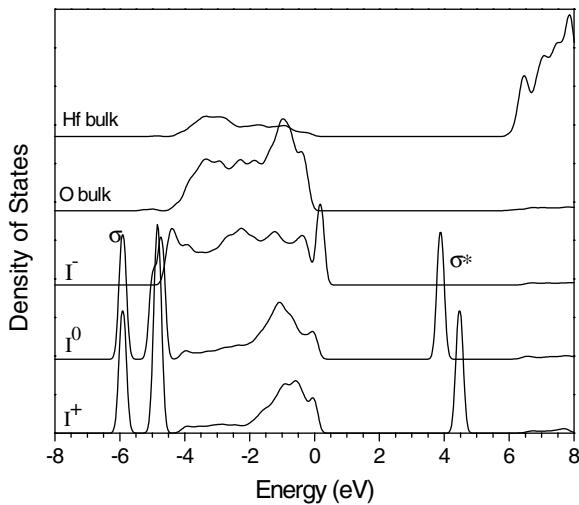


Figure 42. Calculated local DOS of the O interstitial in ZrO₂, in various charge states.

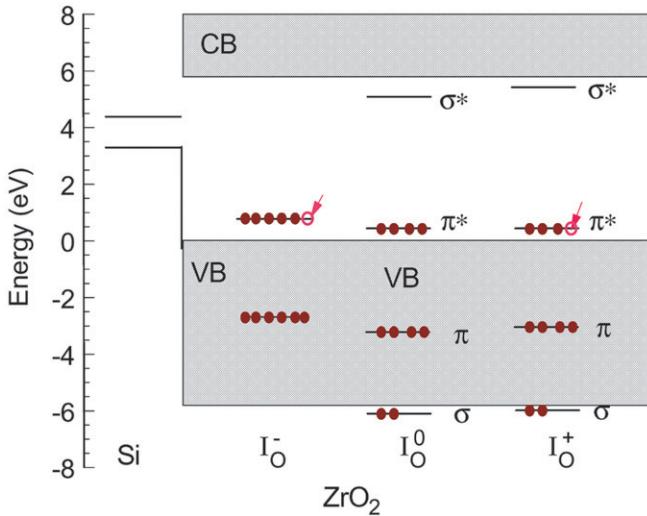


Figure 43. Molecular orbital diagram of the O interstitial in ZrO₂, showing energy levels and electron occupancies.

6.1. Charge trapping and conduction processes

We have noted that high K oxides possess a larger bulk density of defects and trapped charge than SiO₂. Charge trapping leads to instability in the flat band voltage and gate threshold voltage. It is seen as hysteresis on a drive current versus gate voltage plot. The effect can be demonstrated by charge pumping experiments. It is notable that HfSiO_x gate oxides have less hysteresis than HfO₂ 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. A clearer understanding of its origin would also help.

The origin of this trapped charge is becoming clearer. The first source is intrinsic defects in the oxide and interface traps. Zafar *et al* [182, 183] showed that trapping in HfO₂ and

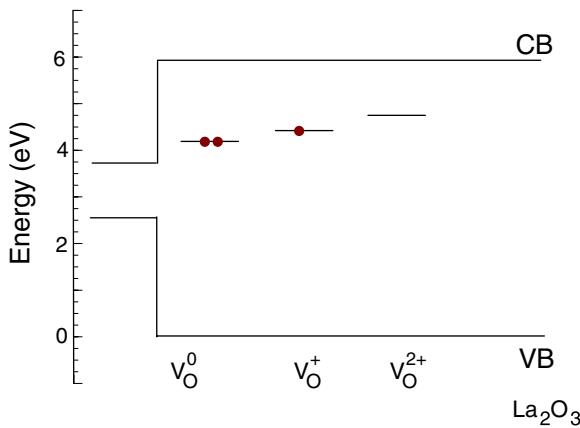


Figure 44. Molecular orbital diagram of relaxed oxygen vacancy in La_2O_3 in various charge states.

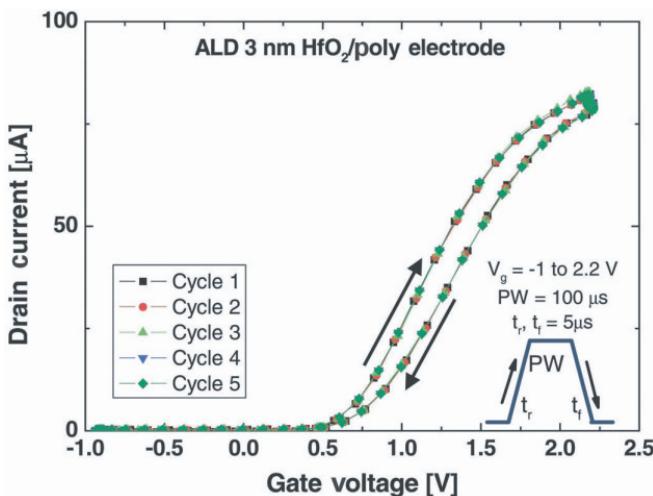


Figure 45. Electron trapping in HfO_2 gate oxide layer. The hysteresis between the up and down ramps shows the presence of sizable trapping. The identical curves for up and down show that no new defects are created [188].

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.

The chemical nature of the traps can be detected in their paramagnetic configuration by ESR. So far, most of the defects found by ESR have been those related to the Si DB at the interface, called the P_b centre [184–186]. Recently, Lenahan and Conley [186] identified two paramagnetic defects by ESR in bulk HfO_2 produced by ALD and subjected to corona discharging: the Hf^{3+} ion (an electron trapped at Hf^{4+} or V^+) and the superoxy radical (or oxygen interstitial). These are the same centres which were previously identified in ZrO_2 powder used in catalysis [187].

Figure 45 shows the effect of transient charge trapping data in the gate oxide on device characteristics, from Bersuker *et al* [188]. The gate voltage is 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* [179] interprets this as fast trapping and detrapping of electrons in the oxide. Similar results are found by Shanware *et al* [189]. Pantisano *et al* [190] showed that these are electron traps in the bulk HfO₂, by varying the thickness of a SiO₂ interfacial layer.

Takeuchi *et al* [178] recently used spectroscopic ellipsometry on HfO₂ films oxidized 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. Kerber *et al* [179] noted that the instability data were consistent with an electron trap level lying just above the Si CB edge. This is consistent with our calculated levels for V⁻.

Electron trapping rates have been analysed [180, 182, 191]. The most complete interpretation of electron detrapping over a wide temperature range was given by Mitard *et al* [180]. They noted trapping rates spanning six decades. They gave levels at 0.8 and 1.5 eV below the HfO₂ CB edge, or more accurately at 0.7 and 1.3 eV above the midgap of the underlying Si. Overall, the electron trapping data are very consistent with the dominant trapping level being the oxygen vacancy. The trap level is consistent with the calculated levels of the negative oxygen vacancy V⁻ in figure 32. The energy spectra of the traps can also be derived by inelastic tunnelling spectroscopy [192], charge pumping and CV methods [193]. CV tends to favour interface traps.

The oxygen interstitial configuration was shown in figure 40(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 and 0.5 eV with respect to the HfO₂ 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 charged 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 HfO₂ films by Lenahan and Conley [186].

The trapped charge can be reduced by annealing. This can be carried out in forming gas (N₂/H₂ mixture) or other nitrogen containing gases such as ammonia. The objective is to reduce the hysteresis in figure 45 to 7 mV. This is only so far possible in silicates. Annealing is also 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.

A possible interpretation of how nitrogen passivates O vacancies was given by Umezawa *et al* [194]. They noted that two nitrogens substituting for oxygen next to a neutral oxygen vacancy, VN₂, has no state in the gap. This has been confirmed by Xiong and Robertson [195] using sX calculations. This occurs because each N would create a hole in the VB, so the two electrons of V⁰ fall into the holes, giving the V an effective local positive charge. The Hf ions adjacent to the vacancy relax outwards, as in the simple V²⁺, but here the relaxation repels the A₁ level completely out of the gap. The configuration can also be viewed as a closed shell system, so it should not give rise to a gap state. The absence of gap states accounts for a drop in leakage currents due to tunnelling [194]. Gavartin *et al* [196] have studied various other configurations of N in HfO₂. These still give gap states, so they are not passivated.

Recently, implantation of fluorine ions has been found to be beneficial [197, 198], probably because F is useful at passivating oxygen vacancies. This supports the idea that oxygen vacancies are the problem.

The deposition of a poly-Si or metal gate creates a reducing ambient, which will create O vacancies in the gate dielectric. In HfO₂, these can diffuse into the bulk. The fact that trapping is a strong function of the gate electrode material also supports the idea that O vacancies are the cause of trapping [199].

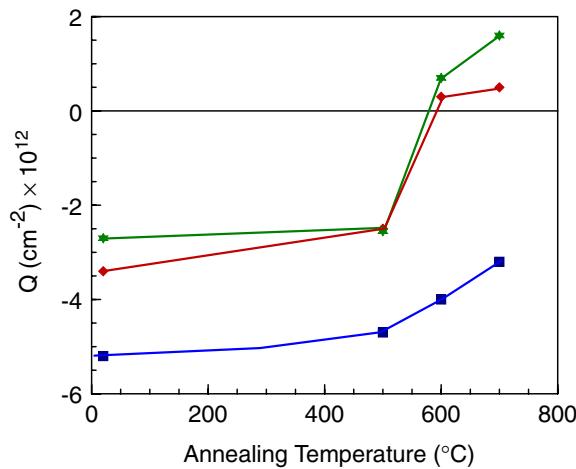


Figure 46. Variation of trapped charge with annealing temperature, after Houssa *et al* [200].

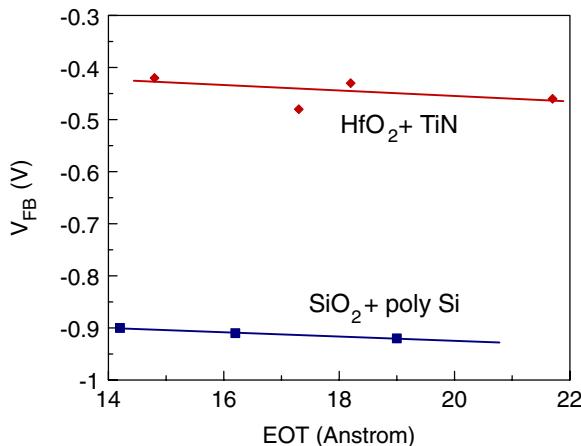


Figure 47. Low bulk fixed charge as revealed by CV plot for HfO₂ gate oxide, after Datta *et al* [202].

Figure 46 shows the variation of trapped charge and interface state density in ALD ZrO₂ with annealing temperature [200]. It is interesting that the trapped charge changes sign at 500 °C when annealed. Houssa *et al* [200] speculates that the positive charge can be due to protons in the oxide trapped on O²⁻ ions (that is as OH⁻ ions). Carter *et al* [201] noted that a higher post deposition anneal (PDA) at 700 °C rather than that at 400 °C used for SiO₂ was more beneficial for passivating traps. Figure 47 shows that a fixed charge of only 10¹¹ cm⁻² has been achieved with HfO₂ gate oxide by annealing by Datta *et al* [202].

The conduction processes in insulating thin films can be classified as limited by bulk or interface. Bulk processes include hopping through traps (Poole–Frenkel) and space charge limited current (SCLC). Interface limited processes are Schottky emission, Fowler Nordheim tunnelling, trap-assisted tunnelling (TAT) and direct tunnelling.

For HfO₂ and ZrO₂, Poole–Frenkel hopping must be dominant for thicker films, over 4 nm. For thinner films, Houssa *et al* [203] attributes conduction to TAT. Tunnelling should

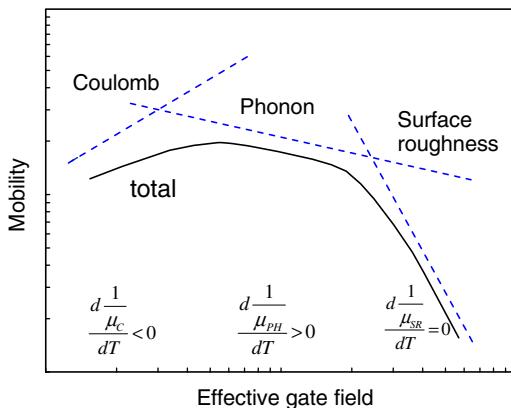


Figure 48. Schematic carrier mobility versus vertical field in FETs in the universal mobility model, showing the mechanisms which limit the mobility and their temperature dependences.

be essentially independent of temperature, hopping T-dependent. The technologically relevant films of HfO₂ tend to be only 2–3 nm thick. In this range, TAT occurs for $V < 0.5$ V [25, 26], and direct tunnelling occurs for higher voltages. In all cases, conduction is by electrons because the CB barrier is lower. The relevant traps are oxygen vacancy levels. In La₂O₃, with its larger CB offset, SCLC is also found [204].

In the tunnelling regimes, Zafar *et al* [205] note that the barrier height can be extracted from the temperature dependence of the conductivity. It should be possible to check the consistency of barrier heights derived from this method and internal photoemission for such films.

6.2. Mobility degradation

The objective of device scaling is to create smaller, faster devices. Speed follows the 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. 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 Sun, Plummer [206] and others. The most recent version is by Takagi *et al* [207] 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 scattering processes add up to a total scattering rate ν ,

$$\nu = \nu_1 + \nu_2 + \nu_3,$$

so the processes, limiting mobility add according to Matthiessen’s rule,

$$\frac{1}{\mu} = \frac{1}{\mu_C} + \frac{1}{\mu_{PH}} + \frac{1}{\mu_{SR}}, \quad (16)$$

where C = Coulombic scattering, PH = phonon scattering and SR = surface roughness. The mobility is limited by different mechanisms at different gate fields, as each obeys a different power law with field, see figure 48. At low fields, mobility is limited by Coulombic scattering by trapped charges in the oxide and/or channel and/or the gate electrode interface, at moderate field it is limited by phonon scattering and at high fields by scattering by surface roughness. The

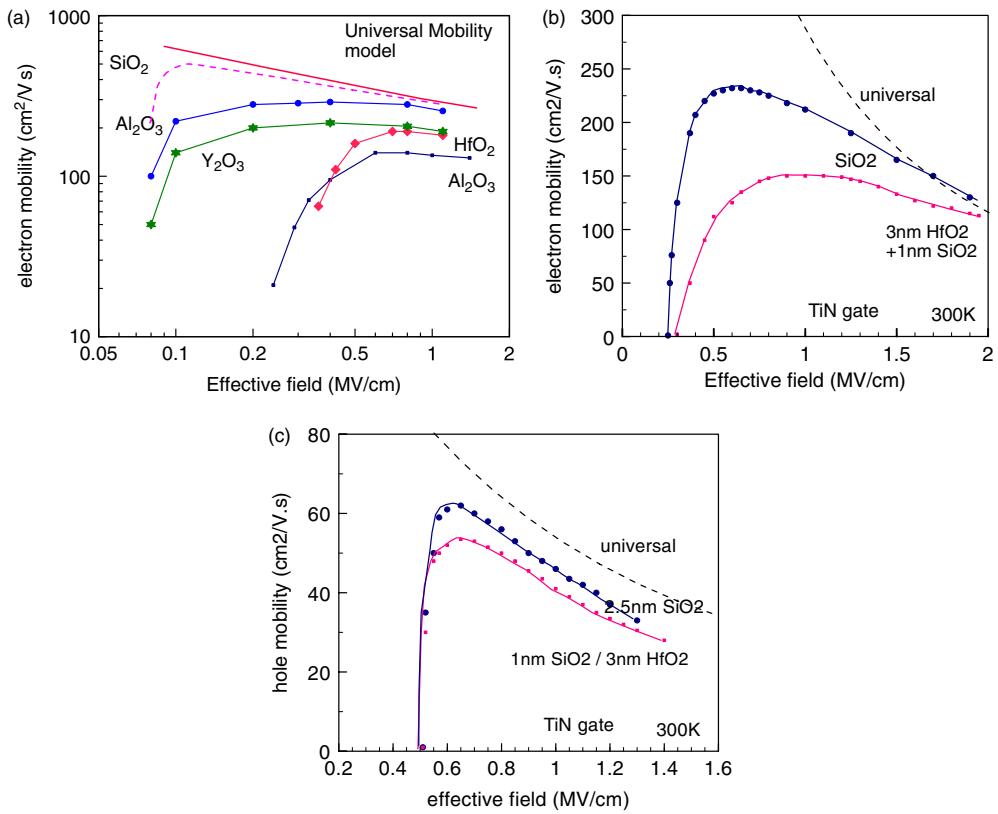


Figure 49. Electron mobility of Si, for various gate oxides, after Gusev *et al* [28]. (b), (c) electron and hole mobility of FETs on 3 nm HfO₂ gate oxide on 1 nm SiO₂, with TiN gate electrodes, showing the stronger degradation of electron mobility.

different mechanisms also show different temperature dependences, with Coulomb scattering and roughness scattering being T independent.

CMOS devices with a SiO₂ gate oxide have a mobility close to the universal limit. In that case, the mobility is limited mainly by roughness of the Si:SiO₂ interface. The mobilities in devices with high K gate oxides presently lie well below the universal curve [6, 29, 32, 39, 208–217]. This is particularly true of NMOS devices. The reduction in mobility for PMOS devices is fractionally less. Figure 49 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 strongly debated, between two likely mechanisms. The first possibility is remote phonon scattering (RPS) by low energy polar phonon modes in the oxide, as noted by Fischetti *et al* [212]. The second possibility is remote Coulomb scattering (RCS) by large amounts of trapped charge [10, 214–216]. RCS is clearly important, as some high K oxides such as Al₂O₃ show mobility degradation [29], have much more trapped charge than SiO₂ and *no soft modes*.

Saito *et al* [214] introduced a general model including the above effects. In their variant, scattering arose mostly from charge defects in the oxide and from fluctuations in the dielectric constant from anisotropic oxide crystallites.

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

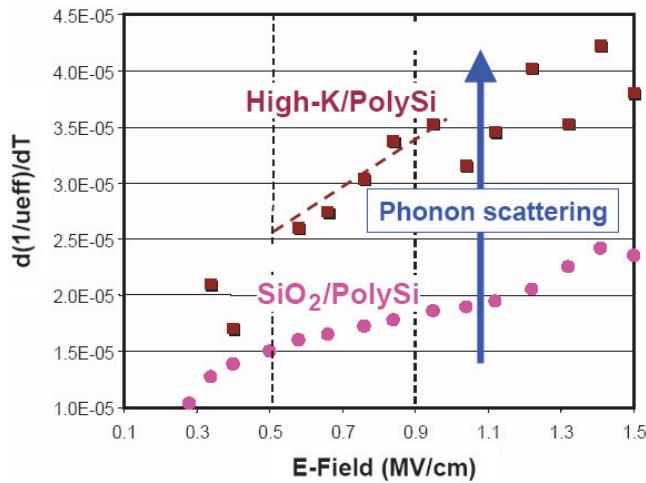


Figure 50. Measured T dependence of mobility for NMOS, after Chau *et al* [9,213].

by Ma *et al* [218]. It can be excluded by direct measurements of Hall effect mobility, which showed a direct reduction [219].

Fischetti *et al* [212] noted that the high K of most oxides results from their low-lying polar vibration modes, see section 4.2. The oxides are incipient ferroelectrics, and these soft polar modes would drive the ferroelectric instability if their frequency fell to zero. These polar modes have a strong coupling to scatter carriers in the Si channel—hence ‘remote scattering’. On the other hand, in SiO₂ such polar modes have a much higher frequency and do not have a large coupling. Fischetti *et al* [212] modelled the effect for various oxides and SiO₂. It was found to be pronounced in ZrO₂ and HfO₂. The effect is smaller in ZrSiO₄ or HfSiO₄ which are covalently bonded without the soft modes. It is also a small effect in Al₂O₃ which has no soft modes.

The importance of RPS is that it is intrinsic, and so it will be a *fundamental* limit to mobility in small EOT devices. On the other hand, RCS is an extrinsic effect, which one can hope to reduce by processing and reducing trap densities. RCS would be preferred technologically.

The two mechanisms can be distinguished by their temperature, field and 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 can increase at higher temperatures (see figure 48). Ren *et al* [220] and Chau and Datta [9,202,213] 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 50. Thus, the RPS mechanism is important. Ren *et al* [220] used HfO₂ gate oxide, as did McIntyre *et al* [198]. Ren’s analysis is more complex in that they distinguish scattering by phonons in the oxide and in the Si. Chau *et al* [9,213] suggest that RPS would be screened out by metal gate electrodes, such as TiN.

The second method is to plot the mobility against oxide thickness, as in the work of Murto *et al* [10], Ragnarsson *et al* [39], Casse *et al* [216], and Kirsch *et al* [217]. The mobility is most reduced in thin high K oxide [208], see figure 52 later.

The case with a variable thickness ALD-grown SiO₂ layer below a fixed thickness (30 Å) ALD HfO₂ layer gives much insight [216]. The mobility is expressed as an ‘additional mobility’

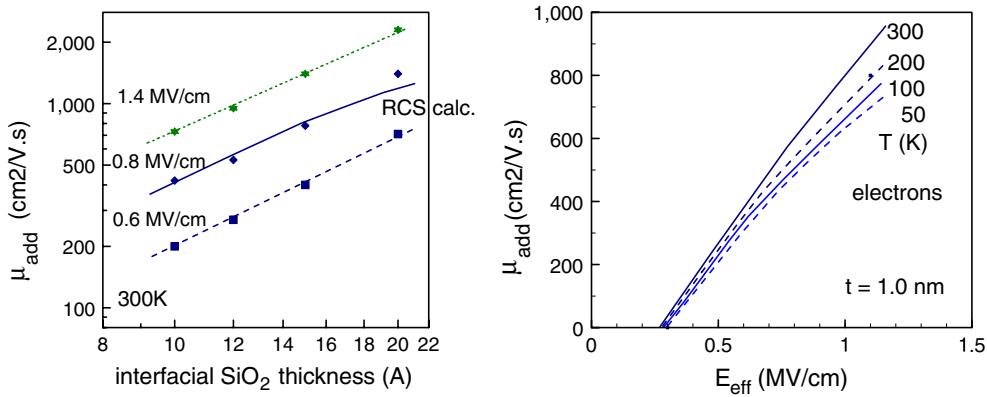


Figure 51. (a) ‘Additional’ mobility due to HfO₂ overlayer versus SiO₂ sublayer thickness, showing exponential dependence expected of RCS. (b) Additional mobility versus gate field, at low fields and different temperatures, after Casse *et al* [216].

compared with the reference mobility for a gate stack with only 25 Å SiO₂ gate oxide,

$$\frac{1}{\mu_{\text{add}}} = \frac{1}{\mu_{\text{HfO}_2}} - \frac{1}{\mu_{\text{ref}}}.$$

Chasse found that the reduction decreased exponentially with SiO₂ layer thickness, varying with SiO₂ thickness, t , as

$$\Delta \left(\frac{1}{\mu_{\text{add}}} \right) = \frac{1}{\mu_0} \exp(-2\beta \cdot t),$$

where μ_{add} is the additional mobility, μ_0 is a constant, t is the SiO₂ layer thickness and β equals the thermal wavevector, $\beta = (2mkT)^{1/2}/\hbar$. k is Boltzman’s constant. According to theory, both RPS and RCS from the HfO₂ layer will give the *same* dependence on t , and this is indeed found. This shows that the mobility reduction is due to remote scattering from an effect in the HfO₂ layer [216].

Generally, RCS is dominant at lower fields and phonon scattering at moderate fields [208]. The excess scattering was found to vary linearly with gate field at low gate fields for varying SiO₂ thickness (figure 51). This linear dependence is as expected for RCS. This is a strong evidence for a contribution from RCS. The temperature dependence of the excess could also be derived. The total mobility plot was then re-assembled from the Mattheisen’s rule, and it was found to describe well the experimental data and its T dependence [216]. Thus, although the mobility is reduced at both low and medium fields, the main cause is RCS.

It is interesting that the roughness scattering at high gate fields is not truly a fundamental parameter; it must be a material specific property of the Si:SiO₂ interface [221] which takes on a universal behaviour in well-made samples. Second, the reason that there is a larger reduction in mobility for electrons than holes is that the absolute mobility value of electrons is higher, so that an increased RCS has a proportionally greater effect for electrons.

A value of the charge causing the excess RCS can be extracted. It is found to be over 10¹³ cm⁻², which is larger than the trapped charge measured from CV plots. It is difficult to account for this. One possibility is that the charge is in the form of dipoles [21, 68].

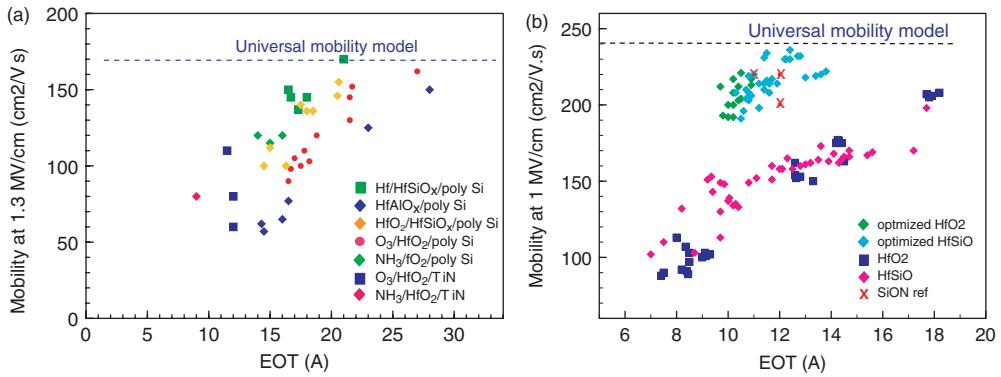


Figure 52. Mobility versus EOT for NMOS, showing how the mobility is reduced below the universal value for thinner oxide layers, after Kirsch *et al* [217].

These observations have allowed groups to develop an engineering solution to the problem of mobility reduction. There should be a moderately thick high quality SiO₂ layer next to the channel and a HfO₂ layer of top. The total EOT should be that required by device design. Figure 52(a) shows the mobility trend versus EOT, for early samples [10] and figure 52(b) for later samples [217]. The early samples had a mobility decreasing strongly at low EOT. The SiO_x interfacial layer is believed to be oxygen deficient from EELS data. The more recent samples with slightly thicker, more stoichiometric SiO₂ layer show a more moderate reduction, especially for the upper samples in figure 52(b). The mobility is now close to the SiON reference case for the optimized samples, shown in green. The latter case is the ‘engineering solution’. The thickness and stoichiometry of SiO₂ interfacial layer clearly requires careful control. It is interesting in the later samples, for both the optimized and unoptimized cases, that HfO₂ and HfSiO_x follow the same trend. Thus, the absence of soft phonons in HfSiO₄ has no effect—again pointing to defect scattering.

6.3. V_T stability

The third major problem for high K oxides is the shift of flat band voltages V_{FB} . Shifts in V_{FB} give shifts in the threshold voltage V_T of the transistor. V_T differs from V_{FB} by the voltage to invert the Si, $2\phi_F$ [222]. V_{FB} is derived from the CV curve of a CMOS capacitor. High K oxides can have a large trapped charge density and a large interface charge density. We again assume that the gate oxide consists of a SiO₂ layer lying under a high K layer. V_{FB} can be found from Poisson’s equation, in terms of the EOT value, t ,

$$V_{FB} = \Phi_{ms} + \frac{Q_{it}}{\varepsilon_0 K_{SiO_2}} t + \frac{Q_b K_{hiK}}{2\varepsilon_0 K_{SiO_2}} t^2. \quad (17)$$

Here, Φ_{ms} is the difference in WFs of the Si and the gate electrode, Q_{it} is the interface fixed charge at the high K –SiO₂ interface, Q_b is the bulk fixed charge density in the high K layer, K_{SiO_2} is the dielectric constant of SiO₂ (3.9) and K_{hiK} is the dielectric constant of the high K layer.

Q_b and Q_{it} values can be derived by fitting V_{FB} values for a thickness series, if we assume that Q_b and Q_{it} are material constants, independent of the high K layer thickness. One should always remember that this is not necessarily true. Extrapolating V_{FB} to zero t gives Φ_{ms} . V_{FB}

is usually small for SiO_2 gate oxides, but larger and unstable values can occur for high K oxides. This is a problem, as the FETs must operate with supply voltages under 1 V in future.

There are 3 sources of V_T shifts: first Q_b and Q_{it} in the oxide [217], second an interface charge at the gate electrode interface and finally Fermi level pinning at the gate electrode interface [223]. There are two components to Q_b and Q_{it} , a steady charge and a transient charge [179]. In devices with SiO_2 gate oxide, only Q_{it} varies, whereas with high K oxides, trapping can vary Q_b and Q_{it} . Fast electron trapping and detrapping is an important part of V_T shifts.

Ultimately, extra defects are created and actions such as positive bias temperature instability (PBTI), negative bias temperature instability (NBTI) and various types of breakdown can occur [183, 224]. These are beyond the scope of this pages.

7. Work functions and metal gates

7.1. Introduction

The purpose of the gate electrode in CMOS is to shift the surface Fermi level E_F of the Si channel to the appropriate band edge, to invert it [222]. An NMOS FET consists of a p-doped Si channel. Its gate electrode of low WF (~ 4.05 eV) will move E_F at the channel surface to its CB, inverting the channel (figure 5). A PMOS device has a n-doped Si channel and a gate electrode with WF 5.15 eV which inverts the channel by shifting its E_F into its valence band. This requires a change in WF of 1.1 eV, the Si band gap.

In present CMOS, the gate electrodes are not real metals but polycrystalline Si doped highly n-type or p-type, respectively, for NMOS and PMOS. Their WFs are 4.05 eV and 5.15 eV, respectively, just as those required. Poly-Si has the advantage that it is refractory, easily deposited, and compatible with SiO_2 and the process flows. However, doped poly-Si has limited carrier density, and so it contributes a depletion length of order 2 Å to the equivalent capacitance thickness (ECT) of the gate stack. It is proposed to replace poly-Si with real metals of the appropriate WF to lower the ECT. Real metals have higher electron densities, and so their depletion lengths are under 0.5 Å.

At a practical level, poly-Si has been found to be fundamentally incompatible with ZrO_2 or HfO_2 oxide and will require the replacement of poly-Si gate with metal gates. It was found that the reducing ambient during the CVD deposition poly-Si from silane causes a gross reduction of the ZrO_2 or HfO_2 , leading to silicide formation, leakage paths, Hf–Si bonds and nuclei for the large grain poly-Si growth [30, 225, 226]. This has led to the realization that high K gate oxides and metal gates must be introduced simultaneously, rather than in two separate steps.

The gate metals to be used must be ‘band edge metals’, with WFs equal to the band edge energies of Si, 4.05 and 5.15 eV. More midgap values would increase the sub-threshold slope S' of the FET turn-on characteristic, $S' = \partial \ln I_{SD} / \partial V_{GS}$, which is undesirable. The problems involved in using real metals are finding a metal with the correct WF, ensuring the thermal stability of that metal in contact with the oxide, whether SiO_2 or a high K oxide, and generally ensuring a process compatibility. Generally, PMOS metals with large WF will be too noble and difficult to etch, while NMOS metals with small WF will be too reactive.

7.2. Effective work functions

Let us first consider the WF problem. True WFs are referenced to the vacuum level. In the Schottky limit and in the absence of fixed charges, the flat band voltage of a MOS capacitor is

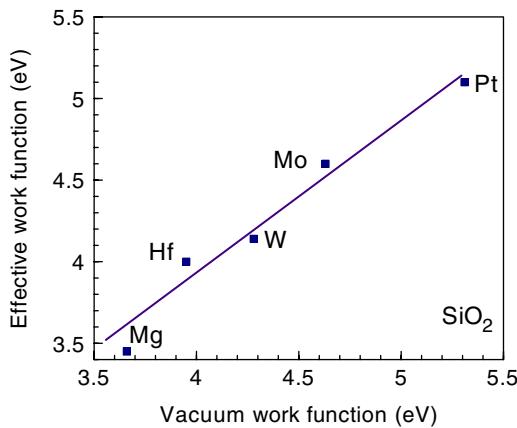


Figure 53. Electron barrier height of metals on SiO_2 , versus metal WF.

given by [222]

$$V_{\text{FB}} = \Phi_M - \Phi_S.$$

Inverting this equation, an effective WF of the gate metal, $\Phi_{M,\text{eff}}$, can be derived from the measured flat band voltage of the CV plot of the MOS capacitor,

$$\Phi_{M,\text{eff}} = V_{\text{FB}} + \Phi_S,$$

by referencing to the WF Φ_S of the Si substrate, 4.05 V or 5.15 V for a n-type or p-type Si, respectively.

For metal electrodes on an arbitrary oxide, we can define a pinning factor as the change of flat band voltage divided by the change in the metal's vacuum WF,

$$S = d\phi_{\text{FB}}/d\Phi_M.$$

Linearizing this model leads to another definition of 'effective WF', $\Phi_{m,\text{eff}}$, as

$$\Phi_{M,\text{eff}} = \Phi_{\text{CNL},d} + S(\Phi_{M,\text{vac}} - \Phi_{\text{CNL},d}).$$

Here, $\Phi_{m,\text{vac}}$ is the metal's vacuum or true WF and $\Phi_{\text{CNL},d}$ is the CNL energy of the oxide, measured from the vacuum level [227].

Yeo *et al* [227] extracted the effective WFs of various metals on SiO_2 from their barrier heights (internal photoemission), CV plots or tunnelling, as shown in figure 53. The slope of the plot is $S = 0.86$. This is close enough to 1.

Yeo *et al* [227] then analysed the data for metals on HfO_2 and ZrO_2 layers using the data then available. They extracted a slope of $S = 0.53$ and $S = 0.41$, respectively, (close to the theoretical value).

They then argued that the effective WF of metals on HfO_2 must be able to shift over the range $\Delta\Phi_{m,\text{eff}} = 1.1$ eV, the Si gap, requiring the *vacuum* WF to shift over the range

$$\Delta\Phi_{M,\text{vac}} = 1.1/S,$$

which is 2.2 V for HfO_2 . With this condition, there are *no* metals possible to perform as band edge metals. This would be a serious fundamental limit to high K oxides.

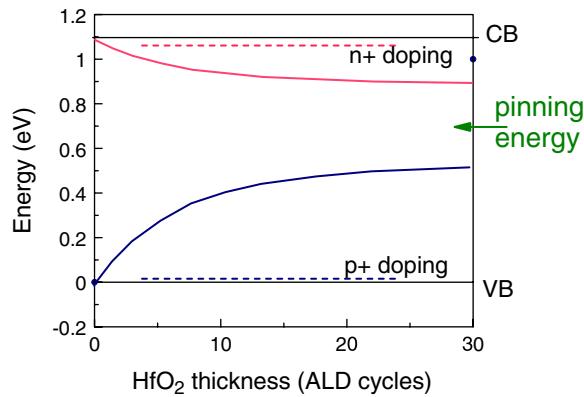


Figure 54. Schematic of flat band voltage shifts versus HfO_2 layer thickness on SiO_2 on Si, from n-type and p-type poly-Si gate electrodes, after Hobbs *et al* [230].

The next worrying result was that Hobbs *et al* [223, 228–230] analysed the evolution of flat band voltages of doped poly-Si gates on a gate oxide consisting of a SiO_2 layer plus a variable thickness layer of HfO_2 on top. They found that the poly-Si was not able to shift the E_F of the Si channel from its CB to VB. Instead, E_F appeared to be slightly ‘pinned’ towards a voltage in the upper Si gap, see figure 54.

This led to the realization that high K gate oxides were intrinsically incompatible with poly-Si gates, and that the introduction of high K gate oxides must be accompanied with the introduction of metal gates. Two hopefully separate and significant process changes would need to occur simultaneously. But there is no consensus on which gate metals to use. Further, the results of Yeo *et al* [227] suggest that there can be no metal which could work for HfO_2 . This was a serious situation and requires detailed analysis of the data and the ideas.

First we check the data of metals of different WFs onto HfO_2 on Si. The barrier height of the metals to the HfO_2 valence band edge has been measured by photoemission, by tunnelling or by internal photoemission, or the band alignment can be deduced from CV measurements. The data indicate that the *barrier heights* change with metal by much less than the change in the WF.

Sayan *et al* [231] measured the VB offset by photoemission for Hf and Pt on HfO_2 , as shown in figure 55(a). Si is also included after allowing for its band gap. This gives $S \sim 0.5$. Afanasev and Stesmans [232] measured the Schottky barrier height of Al, Ni and Au on HfO_2 by internal photoemission (figure 55(b)) and found a similar S value. However, the barrier values are different to Sayan’s, as these are also included in figure 55(a). The barrier heights for ZrO_2 are shown in figure 56, and these also give a value of $S \sim 0.5$. Zafar *et al* [234] and Afanasev *et al* [233] have also measured barrier heights recently.

Schaeffer *et al* [235] derived the V_{FB} of metals on HfO_2/Si MOS capacitors by CV measurements on thickness series (figure 57). V_{FB} is extracted assuming any fixed charge is constant, equation (17). They found that V_{FB} changed by less than 50% of the change in vacuum WF. An extreme case is LaB_6 which has a very low vacuum WF of 2.6 eV. Schaeffer *et al* [235] found a pinning factor S closer to 0.2 than 1. Thus their data showed an even weaker dependence. Note, each of the metals was subjected to annealing as in front-end processing, whereas Sayan’s and Afanasev’s data are not. This can lead to a reaction of the metal and oxide. Schaeffer *et al* [236] and Cartier [253] also found that the WF of some simple metals

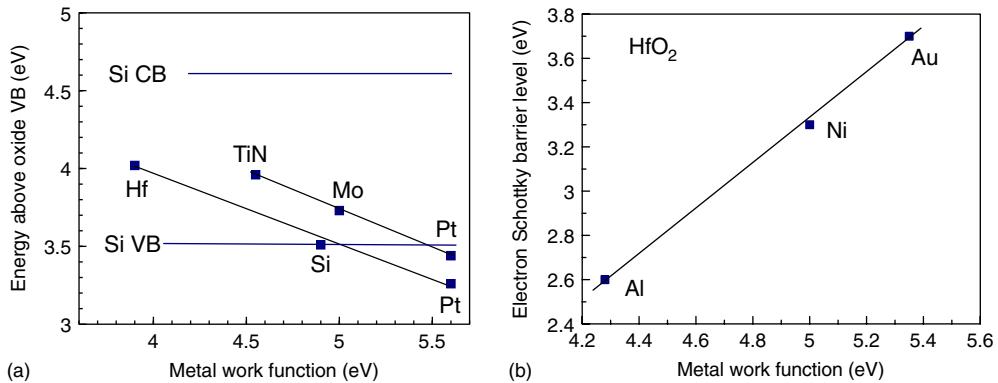


Figure 55. (a) VB offset of Pt and Hf layers on HfO₂ films, as measured by photoemission [232] versus vacuum WF. (b) CB barrier heights for metals on HfO₂ measured by internal photoemission [233] versus vacuum WF.

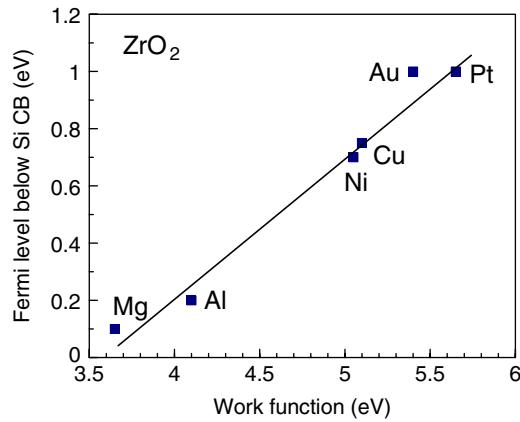


Figure 56. CB barrier heights on ZrO₂ measured by internal photoemission [102] versus vacuum WF.

such as Pt or Re could be switched between values due to the effect of hydrogen or oxygen vacancies.

Recently, Mahji *et al* [237] recognized that it was unrealistic to extract V_{FB} based on assuming a constant fixed charge in HfO₂ layers (figure 58). Instead, they used a ‘wedding cake’ of a SiO₂ layer etched back to different thicknesses, with a constant thickness HfO₂ layer deposited on top. Then, V_{FB} is extrapolated versus SiO₂ thickness, whose fixed charge should be constant. They found a much wider range of effective WF values (figure 58). Again, the metals had been subjected to a 1000 °C high temperature anneal, as in CMOS, and they primarily tested extreme refractory metal systems such as nitrides and silicides. For proprietary reasons, most of the metals are confidential. Nevertheless, they found a much wider WF range than Schaeffer *et al* [235]. The key observation is the similarity of the slope for HfO₂ and SiO₂. This indicates that there is little Fermi level pinning on HfO₂ in this data set.

The answer to the problem posed by Yeo *et al* [227] is as follows. They determined the *barrier height* of various metals on HfO₂, which does vary at a lower rate than the vacuum

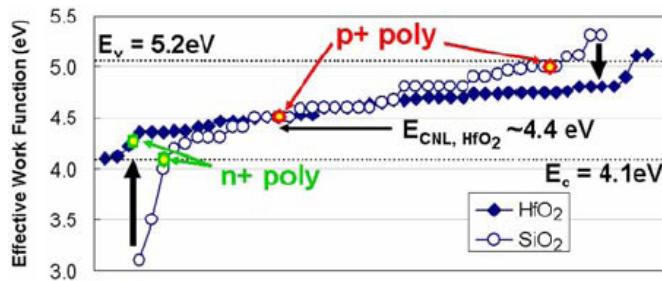


Figure 57. Effective WFs of various metals on SiO_2 and HfO_2 , derived from CV measurements on thickness series. After Schaeffer *et al* [235].

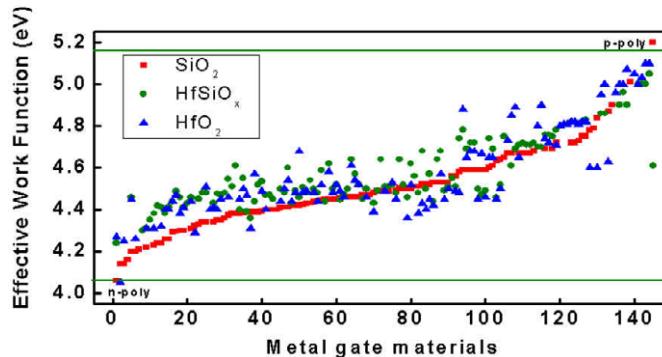


Figure 58. Effective WFs of various metals on SiO_2 , HfO_2 and HfSiO_x , derived from CV measurements on thickness series ‘wedding cake’ oxide. After Mahji *et al* [237].

WF. This is correctly expressed by the S factor, with $S \approx 0.5$. This occurs because a dipole layer is created at the metal-oxide interface.

However, the CMOS capacitor consists of *two* metal-oxide interfaces. At the flat band voltage, the potential at each side is the same. If the dielectric inside is SiO_2 , there are no dipole layers. If the dielectric is HfO_2 , there is an equal and opposite dipole layer at each interface. Thus, the potential felt at a Si channel on the Si side should be that applied by a gate metal. It will vary over the full range of the real WF. The potential inside the HfO_2 will vary by less, as seen by the barrier heights. But that is *internal* to the HfO_2 .

This effect is seen in experiment. Ren and Kwong [238] and Yang *et al* [239] have measured CV shifts of Ni, Ta, Ti and Pt on HfO_2 , the latter with a replacement gate process, and derived the effective WFs, see figure 59. The total shift is 1.15 V for Ti to Pt, and the slope is $S = 0.91$, similar to that for SiO_2 .

Thus, it should be possible to find two band edge metals suitable for NMOS and PMOS. Indeed a number of candidate NMOS and PMOS metals have recently appeared. For NMOS, TaC_x (4.2 eV) and TaTb_xN (4.23 eV) appear promising [240, 241], while for PMOS Ru (5.03 eV) and (Ta,Al)N with WF of 5.0 eV are suitable [242, 243].

7.3. Metal gate systems

A notable factor for metal gates is that most metals are not refractory, unlike poly-Si. They tend to react with the oxide, even with SiO_2 . The net effect of this is to shift the effective WFs towards midgap [237, 244]. A simple demonstration of this effect is shown in figure 60.

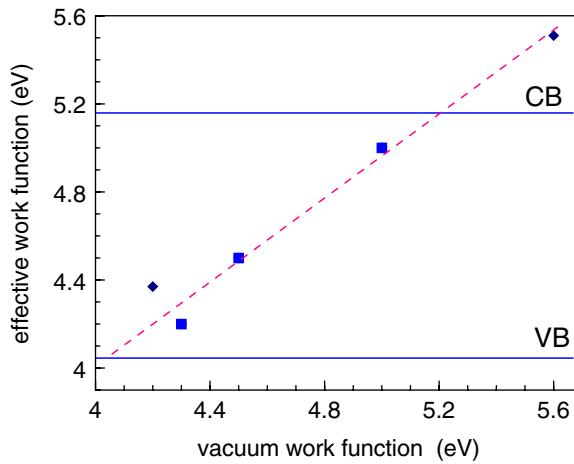


Figure 59. Effective WF of various metals on HfO_2 versus vacuum WF, from CV. Data from [239, 240].

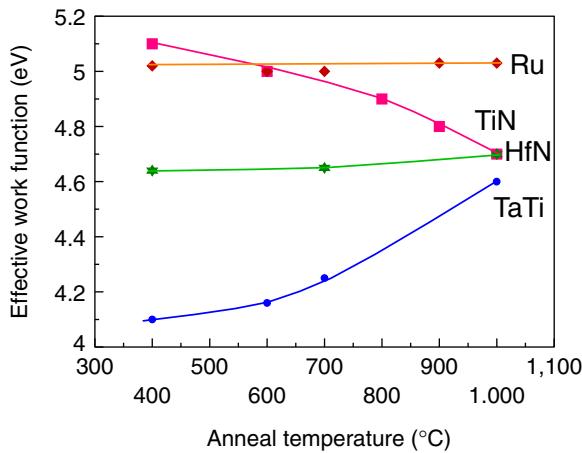


Figure 60. Effective WF of various metals, as a function of annealing temperature. For Ru, TaTi, HfN and ALD TiN after Yu [244]. TiN data is not necessarily representative.

The search for suitable gate metal systems has been based on various strategies. The first has been simple metals or metals compounds. The work functions of various elemental metals is tabulated in table 7 for convenience [245]. Of the simple metals, few show a stable WF after high temperature annealing. Ru is one of the few [242].

This has led to an intensive study of the WFs of compounds with high thermal stability such as nitrides, carbides and silicides of transition metals—high temperature diffusion barrier materials. Unfortunately, there are no vacuum WF data on these materials to use as a guide. It is likely that the WFs vary over a narrower range than of the pure metals, due to band filling. TiN (4.55 eV), TaN, HfN are all stable enough, but most of these are midgap WFs. TaSiN is one system with very high stability. It is midgap but its WF can be reduced at higher Si contents [246].

Table 7. Work functions (after Michaelson [245]) and atomic Mulliken electronegativities of various metals. The work functions include both a bulk electronegativity and surface dipole component.

Element	Work function (eV)	Mulliken electronegativity of atom (eV)
Si	4.85	4.77
Ge	5.0	4.6
Mg	3.66	3.75
Al	4.28	3.23
Sc	3.5	3.34
Ti	4.33	3.45
V	4.3	3.6
Cr	4.5	3.72
Mn	4.1	3.72
Fe	4.5	4.06
Co	5.0	4.3
Ni	5.15	4.4
Cu	4.65	4.48
Zn	4.33	4.45
Ga	4.2	3.2
Y	3.1	3.19
Zr	4.05	3.64
Nb	4.3	4.0
Mo	4.6	3.9
Ru	4.71	4.5
Rh	4.98	4.30
Pd	5.12	4.45
Ag	4.26	4.44
La	3.5	3.1
Hf	3.9	3.8
Ta	4.25	4.11
W	4.55	4.40
Re	4.96	4.02
Os	4.83	4.9
Ir	5.27	5.4
Pt	5.65	5.6
Au	5.1	5.77

The second strategy is to check metal alloy systems, such as Ta–Mo, Ru–Y, where the WF is tuned over a wide range by the composition [247, 248]. The WF of the N, Si alloys can also be tuned by composition. For example (Ti, Al)N can vary from 4.36 to 5.1 eV [249].

7.4. Dipole layers versus simple metals; fully-silicided silicon (FUSI)

The vacuum WF is defined as the energy required to take an electron out of the solid to outside its surface, nearby. This often depends on the face of the metal. The WF has two components, an internal potential and the surface dipole. The internal potential cannot vary with face, for a conservative field. But the dipole can depend on surface, and this causes the surface dependence. The effective WF of a gate metal is a quantity of the solid, no external surfaces are involved or should surface dipoles.

An interesting idea to create systems with the desired effective WF is to use interface dipoles to vary WFs. The voltage difference across a dipole is

$$V = \frac{eN \cdot Q \cdot d}{\epsilon_0},$$

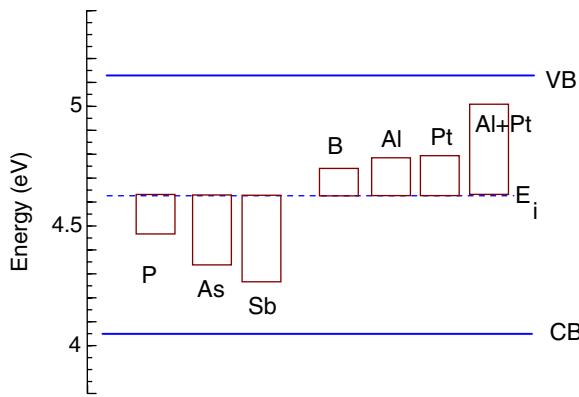


Figure 61. Flat voltage shifts due to doped FUSI gate metals on SiO_2 after Cabral *et al* [251].

where N is the areal density of dipoles, Q is their effective charge and d is the dipole length. These dipoles create the change in Schottky barrier heights for the different terminations of HfO_2 interfaces on Si, in section 4.3.

The same effect can be used to modify metal—oxide interfaces. Possibilities are layers of two different metals, control of termination of metal nitrides or varying the fraction of oxygen in the termination of the oxide. The effect can be large if N is all the interface atoms. Intel’s interesting ‘metal A’ and ‘metal B’ [9] could be based on this approach.

An example of WF control by dipole layers is FUSI gates [250, 251]. In a FUSI gate, a poly-Si gate is converted into a silicide such as NiSix by reaction with the metal Ni. Silicides form by diffusion in a low temperature reaction. NiSix is useful because it occurs by the Ni diffusing through the Si, as the opposite of Si out-diffusion would leave voids at the bottom. The WF of FUSI electrodes can be varied by pre-doping the poly-Si by the group III or V dopants. An example is seen in figure 61 [251] for doped silicides on SiO_2 . Clearly, you cannot electronically dope a metal, as it has a large DOS at the Fermi level. The effect happens because the dopants are insoluble in the silicide, so the advancing silicide front expels the dopants ahead of it, in a ‘snow plough’ effect. This creates a dipole layer, which modifies the effective WF in the same polarity as substitutional doping would. However, whereas doping of Si pins the Fermi level at WF of the dopant levels, dipole layers give rise to a completely variable WF shift, with no pinning. This can be a disadvantage. The microscopic mechanism of snow-ploughing has been proved by ion scattering and photoemission study [252]. Considerable work must be carried out to understand the efficiency of the WF change as a function of the oxide, on SiO_2 versus HfO_2 [250, 251].

A disadvantage of FUSI is that it is a gate last process, which interrupts the whole process integration, for example, of using stressors to engineering band structures of Si–Ge channels to increase mobilities.

7.5. Interaction of poly-Si with HfO_2

The interaction of poly-Si gates with HfO_2 causes the effective WF of the Si to become relatively pinned. 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* [223, 228–230]. They indicated that the problem arises from an interaction between the HfO_2 and the poly-Si gate material. In principle, the

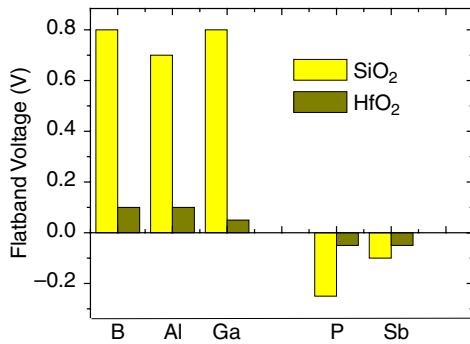


Figure 62. Flat band voltage shifts for versus poly-Si dopant for poly-Si electrodes on HfO_2 . Showing that the shifts are mainly a function of dopant polarity, after Cartier [253].

data could be accounted for by fixed charges, dopant diffusion or interface traps. However, the range of tests [228–230] suggests that the origin is the interaction of the gate and the HfO_2 gate oxide. Figure 62 shows that the large V_T shifts with p-type poly-Si are not simply due to boron penetration, they are the same for all p-type dopants B, Al and Ga [253].

To an extent, the observed pinning behaviour is expected from the MIGS model of Schottky barriers, as the 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 54 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 [229]. 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 CNLs 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. The associated band bending behind the pinning model is supported by direct photoemission and barrier height measurements [254–256].

A possible explanation was provided by Hobbs *et al* [223, 230]. 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 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. Thus, Hobbs *et al* [223] and also Chau [9] 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* [230]. These calculations were extended to much more interface configurations by Xiong *et al* [257].

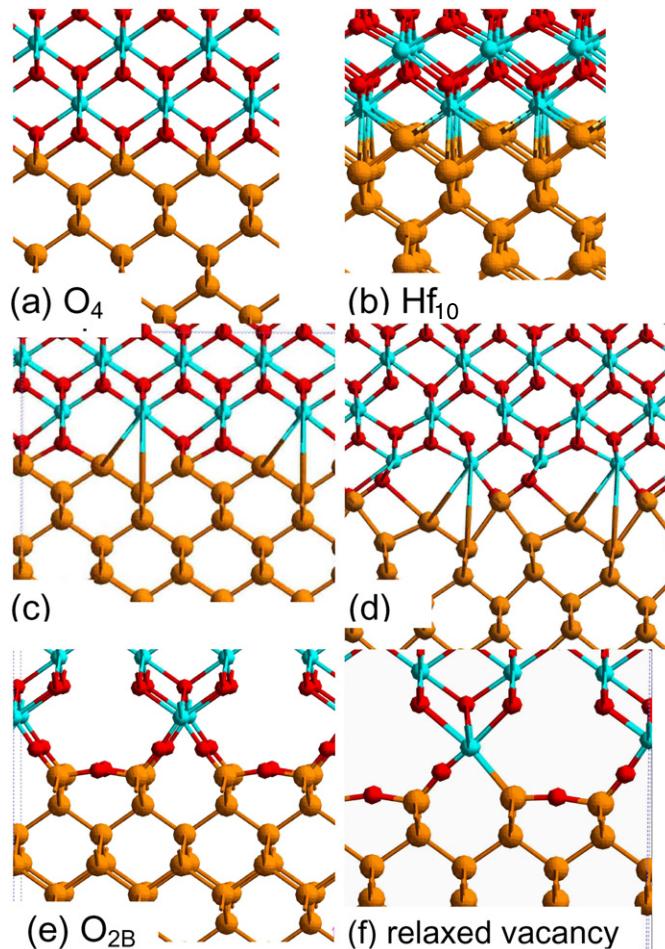


Figure 63. (a) Ideal O₄ interface, (b) ideal Hf₁₀ interface, (c) relaxed O vacancy at O₄ interface, (d) relaxed O vacancy at the O₃ interface, (e) ideal O_{2B} interface and (f) relaxed O vacancy at the O_{2B} interface.

Figure 63 compares model [100]HfO₂ : Si interfaces without and with Hf–Si bonds. We note that the most symmetric O₄ interface could be continuously transformed into the Hf₁₀ interface by the removable of interface O atoms. The O₄ interface when relaxed has 2 Si–O bonds, the Hf₁₀ 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 minimize its total energy. The local DOS was calculated, and it was found that an interface state causes 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.

Other interface configurations were tried. Figure 63 shows the 2 × 1 symmetry 2-fold coordinated O-terminated interface studied by Fonseca [125], 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 V_T

shifts which appear when poly-Si gates are used with HfO₂ gate oxide. The specific interface configuration is not restrictive.

Hobbs *et al* [230] also found that poly-Si on Al₂O₃ 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 Al₂O₃ has negative fixed charge. The new model attributes this effect to interaction at the gate interface, not to fixed charge. Al₂O₃ appears to behave differently because an O interface vacancy does not rebond to form Al–Si bonds but leaves a Si DB. The Si DB state lies in the lower gap, about 0.2 eV above the VB. This occurs as follows. In Al₂O₃, Al exists in both 6-fold and 4-fold sites, with formal charge of +3 and –1 at each. These balance if the ratio is right. At the interface, Al only exists at 4-fold sites in the oxide. This requires negative charge, which comes from the gate. So there is a pinning effect. The second model of Fermi level pinning at poly-Si : HfO₂ interfaces was provided by Shirashi *et al* [258]. They proposed that p-type metals induce the formation of O vacancies in the oxide, and that charge transfer from vacancy to the poly-Si generates a dipole layer at the interface which tends to pin E_F higher in the gap. This model is effective for p-metals.

8. Summary

This paper has reviewed the materials chemistry, bonding and electrical behaviour of oxides needed to replace SiO₂ 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. HfO₂ and Hf silicate have emerged as the preferred oxides. The necessary deposition and processing to produce working devices have been achieved. However, the oxides need to be optimized 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 problematic as they give rise to defect levels close to the Si CB. Flat band voltage instability is due to electron trapping at the O vacancy. Mobility degradation is largely due to remote charge scattering by charged defects in the oxide. DC flat band voltage shifts tend to be due to interaction and reaction of the gate electrode and the high K oxide. This is acute for poly-si gates. It appears that high K oxides will only be implemented together with metal gates. Metal gates are an equally severe problem, which need much more study.

Acknowledgments

The author would like to thank P W Peacock and K Xiong for many calculations, P McIntyre and S Stemmer for illustrations and M Copel, E Gusev, V Afanasev, J Schaeffer, E Cartier, P Mahji, G Bersuker, G Lucovsky, G Reimbold and L Pantisano for preprints and discussions.

References

- [1] Wilk G, Wallace R M, Anthony J M, 2001 *J. Appl. Phys.* **89** 5243
- [2] Wallace R M and Wilk G D 2003 *Crit. Rev. Solid State Mater. Sci.* **28** 231
- [3] Wallace R M and Wilk G (ed) 2002 *MRS Bull.* **27**

- [4] Robertson J 2004 *Eur. Phys. J. Appl. Phys.* **28** 265
- [5] Huff H and Gilmer D (ed) 2004 *High K gate Dielectrics* (Berlin: Springer)
- [6] Houssa M (ed) 2003 *High Dielectric Constant Materials: VLSI MOSFET Applications* (London: IOP)
- [7] Demkov A A and Navrotsky A (ed) 2005 *Materials Fundamentals of Gate Oxides* (Dordrecht: Springer)
- [8] Lo S H, Buchanan D A, Taur Y and Wang W 1997 *IEEE Electron Device Lett.* **18** 209
- [9] Chau R 2003 *Int. Workshop on Gate Insulator* (Tokyo 2003) <http://www.intel.com/research/silicon/micron.htm#high>
- [10] Murto R W, Gardner M I, Brown G A, Zeitoff P M and Huff H R 2003 *Solid State Technol.* **46** 43
- [11] Robertson J 2000 *J. Vac. Sci. Technol. B* **18** 1785
- [12] Kingon A I, Kingon A I, Maria J P and Streiffer S K 2000 *Nature* **406** 1032
- [13] Plummer J D and Griffin P B 2001 *Proc. IEEE* **89** 240
- [14] Hubbard H J and Schlom D G 1996 *J. Mater. Res.* **11** 2757
- [15] Schlom D G and Haeni J H 2002 *MRS Bull.* **27** 198
- [16] Copel M, Gribelyuk M and Gusev E 2000 *Appl. Phys. Lett.* **76** 436
- [17] Gutowski M, Jaffe J E, Liu C L, Stoker M, Hegde R I, Rai R S and Tobin P J 2002 *Appl. Phys. Lett.* **80** 1897
- [18] Wu Y H, Yang M Y, Chin A, Chen W J and Kwei C M 2000 *IEEE Electron Device Lett.* **21** 341
- [19] Iwai H *et al* 2002 *Tech. Digest IEDM* (IEEE)
- [20] Kwo J *et al* 2000 *Appl. Phys. Lett.* **77** 130
- [21] Fissel A, Osten H J and Bugiel E 2003 *J. Vac. Sci. Technol. B* **21** 1765
- [22] Ohmi S, Takeda M, Ishiwara H and Iwai H 2004 *J. Electrochem. Soc.* **151** G279
- [23] Wilk G D and Wallace R M 1999 *Appl. Phys. Lett.* **74** 2854
Wilk G D, R Wallace and Anthony J M 2000 *J. Appl. Phys.* **87** 484
- [24] Visokay M R, Chambers J J, Rotondaro A L P, Shanware A and Colombo L 2002 *Appl. Phys. Lett.* **80** 3183
- [25] Lee B H, Kang L, Nieh R, Qi W J and Lee J C 2000 *Appl. Phys. Lett.* **76** 1926
- [26] Kim H S, Marshall A, McIntyre P C and Saraswat K C 2004 *Appl. Phys. Lett.* **84** 2064
- [27] Robertson J and Chen C W 1999 *Appl. Phys. Lett.* **74** 1168
- [28] Auciello O, Fan W, Kabius B, Saha S, Carlisle J A, Chang R P H, Lopez C, Baragiola R A 2005 *Appl. Phys. Lett.* **86** 042904
- [29] Gusev E P *et al* 2001 *Tech. Digest—Int. Electron Devices Meeting* p 455
- [30] Gusev E P, Cartier E, Buchanan D A, Gribelyuk M and Copel M 2001 *Microelectron. Eng.* **59** 341
- [31] Guha S, Cartier E, Bojarczuk N A, Bruley J, Gignac L and Karasinski J 2001 *Appl. Phys. Lett.* **90** 512
- [32] Tsai W *et al* 2003 (IMEC) *Tech. Digest IEDM* paper 13.2
- [33] DeGendt S 2004 *IEDM: Int. Electron Devices Meeting* (San Francisco) (Short course)
- [34] Yeo Y C, King T J and Hu C 2002 *Appl. Phys. Lett.* **81** 2091
- [35] Robertson J 2005 *Solid State Electron.* **49** 283
- [36] Ritala M, Kukli K, Rahtu A, Raisanen P I and Leskela M 2000 *Science* **288** 319
Ritala M 2004 *High K Gate Dielectrics* ed M Houssa (Bristol: IOP)
- [37] Jones A C and Chalker P R 2003 *J. Phys. D: Appl. Phys.* **36** R80
- [38] Green M L *et al* 2002 *J. Appl. Phys.* **92** 7168
- [39] Ragnarsson L A, Pantisano L, Kaushik V, Saito S I, Shimamoto Y, DeGendt S and Heyns M 2003 *Tech. Digest IEDM* p 87
- [40] Frank M M, Chabal Y J, Green M L, Delabie A, Brijs B, Wilk G D, Ho M Y and Baumvol I J R 2003 *Appl. Phys. Lett.* **83** 740
Frank M M, Chabal Y J and Wilk G D 2003 *Appl. Phys. Lett.* **82** 4758
- [41] Hausmann DM, Kim E, Becker J and Gordon R G 2002 *Chem. Mat.* **14** 4350
Hausmann DM and Gordon R G 2003 *J. Crystal Growth* **249** 251
Becker J S, Kim E and Gordon R G 2004 *Chem. Mat.* **16** 3497
Lim B S, Rahtu A and Gordon R G 2003 *Nature Mat.* **2** 749
- [42] Ho M T, Brewer R T, Wiehlunski L S, Chabal Y J, Moumen N and Boleslawski M 2005 *Appl. Phys. Lett.* **87** 133103
- [43] Rayner G B, Kang D and Lucovsky G 2003 *J. Vac. Sci. Technol. B* **21** 1783
- [44] Maria J P, Wicaksana D, Kingon A I, Busch B, Schulte H, Garfunkel E and Gustafsson T 2001 *J. Appl. Phys.* **90** 3476
- [45] Stemmer S, Li Y, Foran B, Lysaght P S, Streiffer S K, Fuoss P and Seifert S 2003 *Appl. Phys. Lett.* **83** 3141
- [46] Kim H and McIntyre P C 2002 *J. Appl. Phys.* **92** 5094
- [47] Stemmer S, Chen Z, Levi C G, Lysaght P S, Foran B, Gisby J A and Taylor J R 2003 *Japan. J. Appl. Phys.* **42** 3593

- [48] Zhao C *et al* 2002 *J. Non-Cryst. Solids* **303** 144
- [49] Ho M Y, Gong H, Wilk G D, Busch B W and Green M L 2002 *Appl. Phys. Lett.* **81** 4218
- [50] Toriumi A *et al* 2005 *Microelectron. Eng.* **80** 190
- [51] Lee J C *et al* 2003 *Tech Digest IEDM* p 4.4
- [52] Morais J *et al* 2002 *Appl. Phys. Lett.* **81** 2995
Morais J *et al* 2001 *Appl. Phys. Lett.* **79** 4192
- [53] Quevedo-Lopez M A *et al* 2002 *Appl. Phys. Lett.* **81** 1074
Quevedo-Lopez M A *et al* 2003 *Appl. Phys. Lett.* **82** 4669
- [54] de Almeida R M C and Baumvol I J R 2003 *Surf. Sci. Rep.* **49** 1
- [55] Busch B W, Pluchery O, Chabal Y J, Muller D A, Opila R, D A Muller, Kwo J R and Garfunkel E 2002 *Mater. Res. Soc. Bull.* (March) p 206
- [56] Ferrari S and Scarel G 2004 *J. Appl. Phys.* **96** 144
- [57] Green M L, Gusev E P, Degraeve R and Garfunkel E L 2001 *J. Appl. Phys.* **90** 2057
- [58] Stemmer S 2004 *J. Vac. Sci. Technol. B* **22** 791
- [59] Narayanan V, Guha S, Copel M, Bojarczuk N A, Flaitz P L and Bribelyuk M 2002 *Appl. Phys. Lett.* **81** 4183
- [60] Copel M and Reuter M C 2003 *Appl. Phys. Lett.* **83** 3398
- [61] Gribelyuk M A *et al* 2002 *J. Appl. Phys.* **92** 1232
- [62] Copel M, Reuter M C and Jamison P 2004 *Appl. Phys. Lett.* **85** 458
- [63] Wilk G D and Muller D A 2003 *Appl. Phys. Lett.* **83** 3984
- [64] Baik H S, Kim M, Park G S, Song S A, Varela M, Franceschetti A, Pantelides S T and Pennycock S J 2004 *Appl. Phys. Lett.* **85** 672
- [65] Bersuker G *et al* 2004 *Japan. J Appl. Phys.* **43** 7899
- [66] Copel M 2003 *Appl. Phys. Lett.* **82** 1580
- [67] Watanabe H, Saitoh M, Ikarashi N and Tatsumi T 2004 *Appl. Phys. Lett.* **85** 449
- [68] Copel M, Cartier E, Narayanan V, Reuter M C, Guha S and Bojarczuk N 2002 *Appl. Phys. Lett.* **81** 4227
- [69] French R H, Glass S J, Ohuchi F S, Xu Y N and Ching W Y 1994 *Phys. Rev. B* **49** 5133
- [70] Kralik B, Chang E K and Louie S G 1998 *Phys. Rev. B* **57** 7027
- [71] Robertson J, Xiong K and Clark S J 2005 *Thin Solid Films* **496** 1
- [72] Peacock P W and Robertson J 2002 *J. Appl. Phys.* **92** 4712
- [73] Lucovsky G, Zhang Y, Whitten J L, Schlot D G and Freeouf J L 2004 *Microelectron. Eng.* **72** 288
- [74] Rignanese G M, Gonze X and Pasquarello A 2001 *Phys. Rev. B* **63** 104305
- [75] Morais J, Miotti L, Bastos K P, Teixeira S R, Baumvol I J R, Rotondaro A L P, Chambers J J, Visokay M R, Colombo L and MartinsAlves M C 2005 *Appl. Phys. Lett.* **86** 212906
- [76] Kato H, Nango T, Miyagawa T, Katagiri T, Seol K S and Ohki Y 2002 *J. Appl. Phys.* **92** 1106
- [77] Lim S G, Kriventsov S, Jackson T N, Haeni J H, Schlot D G, Balbashov A M, Uecker R, Reiche P, Freeouf J L and Lucovsky G 2002 *J. Appl. Phys.* **91** 4500
- [78] Drabowski J, Zavadinsky V and Fleszar A 2001 *Microelectron. Reliab.* **41** 1093
- [79] Bylander B M and Kleinman L 1990 *Phys. Rev. B* **41** 7868
- [80] Asahi R, Mannstadt W and Freeman A J 1999 *Phys. Rev. B* **59** 007486
- [81] Rushton P P, Tozer D J and Clark S J 2002 *Phys. Rev. B* **65** 235203
- [82] Zhao X and Vanderbilt D 2002 *Phys. Rev. B* **65** 233106
- [83] Rignanese G M, Gionze X, Jun G, Cho K J and Pasquarello A 2004 *Phys. Rev. B* **69** 184301
- [84] Shevlin S A, Curioni A and Andreoni W 2005 *Phys. Rev. Lett.* **94** 146401
- [85] Bonera E, Scarel G, Fanciulli M, Delugas P and Fiorentini V 2005 *Phys. Rev. Lett.* **94** 027602
- [86] Delugas P, Fiorentini V and Filippetti A 2005 *Phys. Rev. B* **71** 134302
- [87] Baldereschi A, Baroni A and Resta R 1988 *Phys. Rev. Lett.* **61** 734
- [88] van de Walle C G 1989 *Phys. Rev. B* **39** 1871
- [89] Franciosi A and van de Walle C G 1996 *Surf. Sci. Rep.* **25** 1
- [90] Tung R T 2000 *Phys. Rev. Lett.* **84** 6078
- [91] Tung R T 2001 *Phys. Rev. B* **64** 205310
- [92] Mönch W 1987 *Phys. Rev. Lett.* **58** 1260
- [93] Mönch W 1994 *Surf. Sci.* **300** 928
- [94] Tejedor C, Flores F and Louis E 1977 *J. Phys. C: Solid State Phys.* **10** 2163
- [95] Tersoff J 1984 *Phys. Rev. Lett.* **52** 465
- [96] Cowley A W and Sze S M 1965 *J. Appl. Phys.* **36** 3212
- [97] Chambers S A, Liang Y, Yu Z, Dropad R, Ramdani J and Eisenbeiser K 2000 *Appl. Phys. Lett.* **77** 1662
- [98] Miyazaki S 2001 *J. Vac. Sci. Technol. B* **19** 2212

- [99] Maria D J 1974 *J. Appl. Phys.* **45** 5454
- [100] Ludeke R, Cuberes M T and Cartier E 2000 *Appl. Phys. Lett.* **76** 2886
- [101] Afanasev V V, Houssa M, Stesmans A and Heyns M M 2001 *Appl. Phys. Lett.* **78** 3073
- [102] Afanasev V V, Houssa M, Stesmans A and Heyns M M 2002 *J. Appl. Phys.* **91** 3079
Afanasev V V 2002 private communication
- [103] Sayan S, Garfunkel E and Suzer S 2002 *Appl. Phys. Lett.* **80** 2135
- [104] Rayner G B, Kang D, Zhang Y and Lucovsky G 2002 *J. Vac. Sci. Technol. B* **20** 1748
- [105] Oshima M, Toyoda S, Okumura T, Okabayashi J and Kumigashira H 2003 *Appl. Phys. Lett.* **83** 2172
- [106] Li Q, Wang S J, Li K B, Huan A C H, Chai J W, Pan J S and Ong C K 2004 *Appl. Phys. Lett.* **85** 6155
- [107] Reynault O, Barrett N T, Samour D and Quiais-Marthon S 2004 *Surf. Sci.* **566** 526
- [108] Afanasev V V, Stesmans A, Chen F, Shi X and Cambell S A 2002 *Appl. Phys. Lett.* **81** 1053
- [109] Fulton C C and Nemanich R J 2004 *J. Appl. Phys.* **96** 2665
- [110] Edge L F, Schlom D G, Chambers S A, Cicerella E, Freeouf J L, Hollander B and Schubert J 2004 *Appl. Phys. Lett.* **84** 726
- [111] Hattori T *et al* 2004 *Microelectron. Eng.* **72** 283
- [112] Ohta A, Yamaoka M and Miyazaki S 2004 *Microelectron. Eng.* **72** 154
- [113] Seguini G, Bonera E, Spiga S and Scarel G 2004 *Appl. Phys. Lett.* **85** 5316
- [114] Fiorentini V and Gulleri G 2002 *Phys. Rev. Lett.* **89** 266101
- [115] Fissel A, Dabrowski J and Osten H J 2002 *J. Appl. Phys.* **91** 8986
- [116] Zhang X, Demkov A A, Li H, Hu X, We H and Kulik J 2003 *Phys. Rev. B* **68** 125323
- [117] Peacock P W and Robertson J 2003 *Appl. Phys. Lett.* **83** 5497
- [118] Forst C J, Ashman C, Schwarz K and Blochl P E 2004 *Nature* **427** 56
- [119] Peacock P W and Robertson J 2004 *Phys. Rev. Lett.* **92** 057601
- [120] Robertson J and Peacock P W 2004 *Microelectron. Eng.* **72** 112
- [121] Robertson J and Peacock P W 2004 *Phys. Status Solidi B* **241** 2236
- [122] Peacock P W, Xiong K, Tse K Y and Robertson J 2005 *Phys. Rev. B* submitted
- [123] Puthenkovicham R, Carter E A and Chang J P 2004 *Phys. Rev. B* **69** 155329
- [124] Puthenkovicham R, Carter E A and Chang J P 2004 *J. Appl. Phys.* **96** 2701
- [125] Fonseca L R C, Demkov A A and Knizhnik A 2003 *Phys. Status Solidi B* **239** 48
- [126] Dong Y F, Feng Y P, Wang S J and Huan A C H 2005 *Phys. Rev. B* **72** 045327
- [127] Wang S J *et al* 2001 *Appl. Phys. Lett.* **78** 1604
- [128] Wang S J and Ong C K 2002 *Appl. Phys. Lett.* **80** 2541
- [129] Guha S, Bojarczuk N A and Narayanan V 2002 *Appl. Phys. Lett.* **80** 766
- [130] Narayanan V, Guha S, Bojarczuk N A and Ross F M 2003 *J. Appl. Phys.* **93** 251
- [131] Apostolopoulos G, Vellianitis G, Dimoulas A, Alexe M, Scholz R, Fancuilli M, Dekadjevi D T and Wiemer C 2002 *Appl. Phys. Lett.* **81** 3549
- [132] Cherns D, Anstis G R, Hutchison J L and Spence J C H 1982 *Phil. Mag. A* **46** 849
- [133] Hamann D R 1988 *Phys. Rev. Lett.* **60** 313
- [134] Satpathy S and Martin R M 1989 *Phys. Rev. B* **39** 8494
- [135] Dong Y F, Wang S J, Chai J W, Feng Y P and Huan A C H 2005 *Appl. Phys. Lett.* **86** 132103
- [136] Klinik A A, Iskandaraova I M, Bagaturyants A A, Potapkin B V and Fonseca L R C 2005 *J. Appl. Phys.* **97** 064911
- [137] Alay J L and Hirose M 1997 *J. Appl. Phys.* **81** 1606
Keister J W *et al* 1999 *J. Vac. Sci. Technol. B* **17** 1831
- [138] Deminsky M *et al* 2004 *Surf. Sci.* **549** 67
- [139] Garavin J, Fonseca L, Bersuker G and Shluger A L 2005 *Microelectron. Eng.* **80** 412
- [140] Elliott S D and Greer J C 2004 *J. Mater. Chem.* **14** 3246
- [141] Puurunen R L 2005 *J. Appl. Phys.* **97** 121301
- [142] Chui C O, Ramanathan S, Triplett B B, McIntyre P C and Saraswat K C 2002 *IEEE Electron Device Lett.* **23** 473
- [143] Kim H, Chiu C O, Saraswat K C and McIntyre P C 2003 *Appl. Phys. Lett.* **83** 2647
- [144] Gusev E P, Shang H, Copel M, Gribelyuk M, Kozlowski P and Zabel T 2004 *Appl. Phys. Lett.* **85** 2334
- [145] Kim H, McIntyre P C, Chui C O, Saraswat K C and Cho M H 2004 *Appl. Phys. Lett.* **85** 2902
- [146] Chui C O, Lee D I, Singh A A, Pianetta P A and Saraswat K C 2005 *J. Appl. Phys.* **97** 113 518
- [147] Wang S J, Huan A C H, Foo Y L, Chai J W, Pan J S, Li Q, Dong Y F, Feng Y P and Ong C K 2004 *Appl. Phys. Lett.* **85** 4418
- [148] Ami T, Ishida Y, Nagasawa N, Machida A and Suzuki M 2001 *Appl. Phys. Lett.* **78** 1361

- [149] Nagata H, Tsukahara T, Gonda S, Yoshimoto M and Koinuma H 1991 *Japan. J. Appl. Phys.* **30** L1136
- [150] Kwo J, Hong M, Kortan A R, Queeney K L, Chabal Y L, Opila R L, Muller D A, Chu S N G, Sapjeta B J and Lay T S 2001 *J. Appl. Phys.* **89** 3920
Dimoulas A *et al* 2001 *J. Appl. Phys.* **90** 4224
Dimoulas A *et al* 2002 *J. Appl. Phys.* **92** 426
Seo J W, Fompeyrine J, Guiller A, Norga G, Marchiori C, Siegwart H and Loucquet J P 2003 *Appl. Phys. Lett.* **83** 5211
- [151] Fissel A, Osten H J and Bugiel E 2003 *J. Vac. Sci. Technol. B* **21** 1765
- [152] Osten H J *et al* 2005 unpublished
- [153] Forst C J, Blochl P E and Schwarz K 2003 *Comput. Mater. Sci.* **27** 670
- [154] McKee R A, Walker F J and Chisholm M F 1998 *Phys. Rev. Lett.* **81** 3014
- [155] McKee R A, Walker F J and Chisholm M F 2001 *Science* **293** 468
McKee R A, Walker F J, Nardelli M B, Shelton W A and Stocks G M 2003 *Science* **300** 1726
- [156] Droopad R, Yu Z Y and Ramdani J 2001 *J. Crystal Growth* **227** 936
- [157] Hu X *et al* 2003 *Appl. Phys. Lett.* **82** 203
Li H *et al* 2003 *J. Appl. Phys.* **93** 4521
- [158] Yang G Y, Finder J M, Wang J, Wang Z L, Yu Z, Ramdani J, Droopad R, Eisenberger K W and Ramesh R 2002 *J. Mater. Res.* **17** 204
- [159] Jeon S, Walker F J, Billman C A, McKee R A and Hwang H 2002 *Tech. Digest IEDM* p 955
- [160] Eisenbeiser K *et al* 2000 *Appl. Phys. Lett.* **76** 1324
- [161] Sivasubramani P, Kim M J, Gnade B E, Wallace R M, Edge L F, Schlom D G, Craft H S and Maria J P 2005 *Appl. Phys. Lett.* **86** 201901
- [162] Edge L F *et al* 2004 *Appl. Phys. Lett.* **84** 4629
- [163] Afanasev V V, Stesmans A, Zhao C, Caymax M, Heeg T, Schubert J, Jia Y, Schlom D G and Lucovsky G 2004 *Appl. Phys. Lett.* **85** 5917
- [164] Klenov D O, Schlom D G, Li H and Stemmer S 2005 *Japan. J. Appl. Phys.* **44** L617
- [165] Forst C J, Schwarz K and Blochl P E 2005 *Phys. Rev. Lett.* **95** 137602
- [166] Lucovsky G 2001 *J. Vac. Sci. Technol. A* **19** 1553
- [167] Foster A S, Sulimov V B, Gejo F L, Shluger A L and Nieminen R N 2001 *Phys. Rev. B* **64** 224108
- [168] Foster A S, Gejo F L, Shluger A L and Nieminen R N 2002 *Phys. Rev. B* **65** 174117
- [169] Torii K, Shirashi K, Miyazaki S, Yamabe K, Boero M, Chikyow T, Yamada K, Kitajima H and Arikado T 2004 *Tech. Digest IEDM* p 129
- [170] Shen C, Li M F, Wang X P, Yu H Y, Feng Y P, Lim A T L, Yeo Y C, Chan D S H and Kwong D L 2004 *Tech. Digest IEDM* p 733
- [171] Pemmaraju C D and Sanvito S 2005 *Phys. Rev. Lett.* **94** 217205
- [172] Kang J, Lee E C, Chang K J and Jin Y G 2004 *Appl. Phys. Lett.* **84** 3894
- [173] Robertson J, Xiong K and Falabretti B 2005 *IEEE Trans. Dev. Mater. Reliab.* **5** 84
- [174] Xiong K and Robertson J 2005 *Microelectron. Eng.* **80** 408
- [175] Xiong K, Robertson J and Clark S J 2005 *Appl. Phys. Lett.*
- [176] Prunade J M and Artacho E 2005 *Phys. Rev. B* **71** 094113
- [177] Klein B M, Pickett W E, Boyer L L and Zeller R 1987 *Phys. Rev. B* **35** 5802
- [178] Takeuchi H, Ha D and King T J 2004 *J. Vac. Sci. Technol. A* **22** 1337
- [179] Kerber A, Cartier E 2003 *IEEE Electron Device Lett.* **24** 87
- [180] Mitard J, Leroux C and Reimbold G 2005 *Proc. NATO workshop (St Petersburg)* (Berlin: Springer)
- [181] Kang A Y, Lenahan P M and Conley J F 2003 *Appl. Phys. Lett.* **83** 3407
- [182] Zafar S, Callegari A, Gusev E and Fischetti M V 2003 *J. Appl. Phys.* **93** 9298
- [183] Zafar S, Gusev E P and Cartier E 2005 *IEEE Trans. Dev. Mater. Reliab.* **5** 45
- [184] Stesmans A and Afanasev V V 2003 *Appl. Phys. Lett.* **82** 4074
- [185] Cantin J L and von Bardeleben H J 2002 *J. Non-Cryst. Solids* **303** 175
- [186] Lenahan P M and Conley J F 2005 *IEEE Trans. Dev. Mater. Reliab.* **5** 90
- [187] Matta J *et al* 1999 *Phys. Chem. Chem. Phys.* **1** 4975
- [188] Bersuker G, Zeitzoff P, Brown G, Huff H R 2004 *Mater. Today* **7** 26
- [189] A Shanware *et al* 2003 *Tech. Digest IEDM* paper 38-6
- [190] Pantisano L, Cartier E, Kerber A, Degraeve R, Lorenzini M, Rosmeulen M, Groesenken G and Maes H E 2003 *Tech. Digest VLSI* p 12A3
- [191] Gusev E P and D'Emic C P 2003 *Appl. Phys. Lett.* **83** 5223
- [192] He W and Ma T P 2003 *Appl. Phys. Lett.* **83** 5461
- [193] Fedorenko Y G *et al* 2004 *Appl. Phys. Lett.* **84** 4771

- [194] Umezawa N, Shiraishi K, Ohno T, Watanabe H, Chikyow T, Torii K, Yamabe K, Yamada K, Kitajima H and Arikado T 2005 *Appl. Phys. Lett.* **86** 143507
- [195] Xiong K and Robertson J 2006 *J. Appl. Phys.* submitted
- [196] Gavartin J L, Shluger A L, Foster A S and Bersuker G I 2005 *J. Appl. Phys.* **97** 053704
- [197] Tseng H H *et al* 2005 *IEDM (Washington, DC)*
- [198] McIntyre P *et al* 2005 *IEDM (Washington, DC)*
- [199] Gusev E P *et al* 2004 *Tech. Digest IEDM* p 729
- [200] Houssa M, Afanasev V V, Stesmans A and Heyns M M 2000 *Appl. Phys. Lett.* **77** 1885
- [201] Carter R J, Cartier E, Kerber A, Pantisano L, Schram T, DeGendt S and Heyns M 2003 *Appl. Phys. Lett.* **83** 5331
- [202] Datta S *et al* 2003 *IEDM (Washington, DC)* paper 28.8
- [203] Houssa M, Touminen M, Naili M, Afanasev V, Stesmans A, Haukka S and Heyns M M 2000 *J. Appl. Phys.* **87** 8615
- [204] Iwai H 2004 private communication
- [205] Zafar S, Cabral C, Amos R and Callegari A 2002 *Appl. Phys. Lett.* **80** 4858
- [206] Sun S C and Plummer J D 1980 *IEEE Trans. Electron. Device* **27** 1497
- [207] Takagi S I, Toriumi A, Iwase M and Tango H 1994 *IEEE Trans. Electron. Device* **41** 2357
- [208] Ragnarsson L A, Guha S, Copel M, Cartier E, Bojarczuk N A and Karasinski J 2001 *Appl. Phys. Lett.* **78** 4169
- [209] Guha S, Gusev E P, Okorn-Schmidt H, Ragnarsson L A and Bojarczuk N A 2002 *Appl. Phys. Lett.* **81** 2956
- [210] Hiratani M, Saito S, Shimamoto Y and Torii K 2002 *Japan. J. Appl. Phys.* **41** 4521
- [211] Onishi K, Krishnan S A and Lee J C 2003 *IEEE Trans. Electron. Device* **50** 384
- [212] Fischetti M V, Neumayer D A and Cartier E A 2001 *J. Appl. Phys.* **90** 4587
- [213] Chau R, Datta S, Doczy M, Doyle B, Kavalieros J and Metz M 2004 *IEEE Electron. Device Lett.* **25** 408
- [214] Saito S, Hisamoto D, Kimura S and Hiratani M 2003 *Tech. Digest IEDM* 33.3
- [215] Torii K *et al* 2003 *Microelectron. Eng.* **65** 447
- [216] Casse M *et al* 2005 *IEE Trans. Electron. Device* submitted
- [217] Kirsch P D *et al* 2005 *Sematech Workshop (Austin, TX, September 2005)*
- [218] Zhu W, Han J P and Ma T P 2004 *IEEE Trans. Electron. Device* **51** 98
- [219] Ragnarsson L A, Bojarczuk N A, Karaninski J and Guha S 2003 *IEEE Electron. Device Lett.* **24** 689
- [220] Ren Z, Fischetti M V, Gusev E P, Cartier E A and Chudzik M 2003 *Tech. Digest IEDM* paper 33.2
- [221] Lucovsky G 2004 private communication
- [222] Streetman B G and Banerjee S 2000 *Solid State Electronic Devices* (Englewood Cliffs, NJ: Prentice Hall) p 270
- [223] Hobbs C *et al* 2003 *VLSI Symp. (Kyoto)* p 9
- [224] Ribes G *et al* 2005 *IEEE Trans. Dev. Mater. Reliab.* **5** 5
- [225] Perkins C M, Triplett B B, McIntyre P C, Saraswat K C and Sher E 2002 *Appl. Phys. Lett.* **81** 1417
- [226] Gilmer D C *et al* 2003 *Microelectron. Eng.* **69** 138
- [227] Yeo Y C, King T J and Hu C 2002 *J. Appl. Phys.* **92** 7266
- [228] Samavedam S B *et al* 2003 *Tech. Digest IEDM* paper 13.1
- [229] Hobbs C *et al* 2004 *IEEE Trans. Electron. Devices* **51** 971
- [230] Hobbs C *et al* 2004 *IEEE Trans. Electron. Devices* **51** 978
- [231] Sayan S, Garfunkel E and Robertson J 2003 *Proc Electrochem. Soc.*
- [232] Afanasev V and Stesmans A 2002 *Appl. Phys. Lett.* **80** 1261
- [233] Afanasev V V, Stesmans A, Pantisano L and Schram T 2005 *Appl. Phys. Lett.* **86** 232902
- [234] Zafar S *et al* 2005 *Tech. Digest VLSI* paper 4A2
- [235] Schaeffer J K *et al* 2004 *Tech. Digest IEDM* p 287
- [236] Schaeffer J K, Fonseca L R C, Samavedam S B, Liang Y, Tobin P J and White B E 2004 *Appl. Phys. Lett.* **85** 1826
- [237] Majhi P *et al* 2005 *Sematech Workshop (Austin, TX, September 2005)*
- [238] Ren C, Kwong D L 2004 *IEEE Electron. Device Lett.* **25** 580
- [239] Yang H *et al* 2005 *Appl. Phys. Lett.* **86** 092107
- [240] Tseng H H *et al* 2004 *Tech. Digest IEDM* p 821
- [241] C Ren *et al* 2005 *IEEE Electron. Device Lett.* **26** 75
- [242] Suh Y S, Lazar H, Chen B, Lee J H and Misra V 2005 *J. Electrchem. Soc.* **152** F138
- [243] Alshareef H N *et al* 2005 *Appl. Phys. Lett.* submitted
- [244] Yu H Y, Kwong D L 2004 *IEEE Electron. Device Lett.* **25** 337
- [245] Michaelson H B 1977 *J. Appl. Phys.* **48** 4729
- [246] Sun Y S, Heuss G and Misra V 2004 *J. Vac. Sci. Technol. B* **22** 175
- [247] Chen B, Suh Y, Lee J, Misra V and Cabral C 2005 *Appl. Phys. Lett.* **86** 053502
- [248] Lee C K, Kim J Y, Hong S N and Misra V 2005 *J. Mater. Sci.* **40** 2693

- [249] Cha T H *et al* 2002 *Appl. Phys. Lett.* **81** 4192
- [250] Gusev E P *et al* 2004 *Tech. Digest IEDM* p 79 (FUSI)
- [251] Cabral C *et al* 2004 VLSI (Kyoto) *Tech. Digest* p 184
- [252] Copel M, Pezzi R P and Cabral C 2005 *Appl. Phys. Lett.* **86** 251904 FUSI
- [253] Cartier E *et al* 2005 *VLSI Symp. (IEEE, Kyoto, 2005)*
- [254] Ludeke R, Narayanan V, Gusev E P, Cartier E and Chey S J 2005 *Appl. Phys. Lett.* **86** 122901
- [255] Lim D, Haight R, Copel M and Cartier E 2005 *Appl. Phys. Lett.* **87** 072902
- [256] Afanasev V V, Stesmans A, Pantisano L and Chen P J 2005 *Appl. Phys. Lett.* **86** 072107
- [257] Xiong K, Peacock P W and Robertson J 2005 *Appl. Phys. Lett.* **86** 012904
- [258] Shirasihi K *et al* 2004 *Jpn. J. Appl. Phys.* **43** L1413 (*Tech Digest VLSI (Kyoto, 2004)* p 108)