Theoretical Evaluation of Zirconia and Hafnia as Gate Oxides for Si Microelectronics

Vincenzo Fiorentini and Gianluca Gulleri

INFM and Dipartimento di Fisica, Università di Cagliari, Cittadella Universitaria, I-09042 Monserrato (CA), Italy
(Received 7 August 2002; published 9 December 2002)

Parameters determining the performance of the crystalline oxides zirconia (ZrO_2) and hafnia (HfO_2) as gate insulators in nanometric Si electronics are estimated via *ab initio* calculations of the energetics, dielectric properties, and band alignment of bulk and thin-film oxides on Si (001). With their large dielectric constants, stable and low-formation-energy interfaces, large valence offsets, and reasonable (though not optimal) conduction offsets (electron injection barriers), zirconia and hafnia appear to have considerable potential as gate oxides for Si electronics.

The performance needs of modern information technology are forcing Si-based ultra-large-scale-integrated (ULSI) devices into the domain of nanometric dimensions. This downscaling implies, among others, the effective continuing reduction of the physical thickness of insulating gate-oxide layers in CMOS (complementary metal-oxide-semiconductor) devices. Amorphous SiO_2, the natural oxide of Si technology, is now nearing its fundamental size limits, with physical thicknesses currently down to 2 unit cells [1]. This leads to uncomfortably large (> 1 A/cm^2) leakage currents and increased failure probabilities. The main reason for the strong reduction of gate-oxide thickness in device downscaling is the need for increasing capacitances in the CMOS conducting channel. In a CMOS, the gate-oxide layer dominates the series capacitance of the channel. An increase in capacitance can be obtained by reducing the dielectric thickness d/ε of the oxide layer, having physical thickness d and relative dielectric constant ε. Because of its small dielectric constant, SiO_2 as a gate oxide has emerged as one of the key bottlenecks in device downscaling [1,2].

It thus appears that, if Moore's law [3] on ULSI circuit component density—and, hence, circuit performance—is to remain valid in the next decade, a replacement will have to be found for silica as a gate insulator. The basic selection criteria are (i) larger dielectric constant ("high-κ"), (ii) interface band offsets to Si as large as or comparable to those of silica (especially the electron injection barrier), (iii) epitaxy on Si energetically not too costly, and (iv) thermodynamical stability in contact with Si. Much research is currently ongoing on a number of candidate oxides [2]. Low-leakage research transistor structures have been realized among others using hafnia (HfO_2) and zirconia (ZrO_2) [2,4], both amorphous and crystalline. However, their interface and dielectric properties are virtually unknown. In this work, we address the expected performance, in terms of the above criteria, of the latter compounds through first-principles density-functional calculations of the structure, energetics, thermodynamical stability, dielectric constants, and band

offsets of crystalline hafnia and zirconia thin films epitaxially grown on Si(001).

Our density-functional theory calculations in the generalized gradient approximation (GGA) [5] use the VASP code [6] and the ultrasoft [7] pseudopotentials provided therewith. Semicore states are treated as core for Hf and Zr; test calculations done including the semicore as valence using the all-electron projector augmented waves (PAW) [8] method as implemented in VASP [6] confirmed the pseudopotential results. Bulk optimizations were done in a 12-atom (conventional face-centered cubic or tetragonal) cell, while the interfaces are simulated by (001)-oriented oxide/Si superlattices contained in tetragonal cells of $c(2 \times 2)$ basal section, and in-plane lattice constant $a_{Si} = 5.461$ Å, our theoretical value for bulk Si. Interface supercells contain around 50 atoms depending on the local interface structure, with 9 layers (18 atoms) for the Si region and typically 11 layers (e.g., 24 oxygen and 10 Zr atoms) for the oxide region. The plane-wave basis cutoff is 350 eV; for the k-space summation, we use $4 \times 4 \times 4$ meshes for the bulk and $4 \times 4 \times 1$ meshes for the z-elongated interface supercells.

Bulk and Si-epitaxial structure.—Bulk hafnia and zirconia were studied in the fluorite, monoclinic, and Si-epitaxial structures. The lattice parameters for ZrO_2 are $a = 5.10$ Å for fluorite and $(a, b, c) = (5.186, 5.255, 5.351)$ Å, off-normal angle $\theta = 8.83°$ for monoclinic. For HfO_2, $a = 5.06$ Å for fluorite and $(a, b, c) = (5.108, 5.175, 5.280)$ Å, off-normal angle $\theta = 8.80°$ for monoclinic. The latter phase is favored over fluorite by 0.115 eV/formula unit for ZrO_2 and by 0.248 eV/formula unit for HfO_2. The results agree with experiment and with recent calculations [9–11]. The formation enthalpies ΔH_{ox} are -11.52 eV and -10.74 for hafnia and zirconia, respectively (close to experiment, as usual using GGA) compared to -8.30 eV for silica: therefore, both oxides are stable in contact with Si with respect to the decomposition into silica and metal. The same holds for the epitaxial phase discussed next, whose excess energy is only about 0.2 eV/formula above the monoclinic.

The tetragonal Si-epitaxial crystalline phase of each oxide was obtained imposing the in-plane lattice constant of Si and adjusting the axial ratio and internal coordinates in the 12-atom conventional cells. The axial ratios c/a_{Si} are 0.92 for ZrO_2 and 0.90 for HfO_2. We verified by variable-cell damped dynamics [6] that this tetragonal bulk is stable against monoclinic distortions. The Si-epitaxial configuration, depicted in Fig. 1 for ZrO_2, may be viewed as a z-stacking of cation-anion bilayers alternatingly oriented at 90° to each other, in which (a) metal cations are disposed in dimerized (110)-like rows (cation-cation distances within the rows 3.4 and 4.2 Å compared to 3.86 Å ideally) and (b) oxygen quadruplets, originally square in fluorite, elongate to rhomboids along the (110) rows bending slightly sideways. The cation (anion) coordination decreases from 8 to 6 (from 4 to 3), in partial analogy to the monoclinic structure [10].

The elastic energy $E_{elastic}^{epi}$ of the Si-epi distorted bulk is 0.23 eV/formula or 5.87 meV/Å^3 for ZrO_2 and 0.16 eV/formula or 4.37 meV/Å^3 for HfO_2 with respect to monoclinic bulk (i.e., both are slightly favored energetically over fluorite, whose occurrence is barred anyway by symmetry). While substantial, these energies are comparable to those of order ~4 meV/Å^3 involved (for much smaller strains) in nitride semiconductor epitaxy [12]. As we now discuss, the knowledge of the volume-specific epitaxial strain energy enables us to extract an area-specific interface energy, as well as to estimate the critical pseudomorphic growth thickness.

Interface energetics and offsets.—Assuming a $c(2 \times 2)$ basal section, we investigated for both materials several
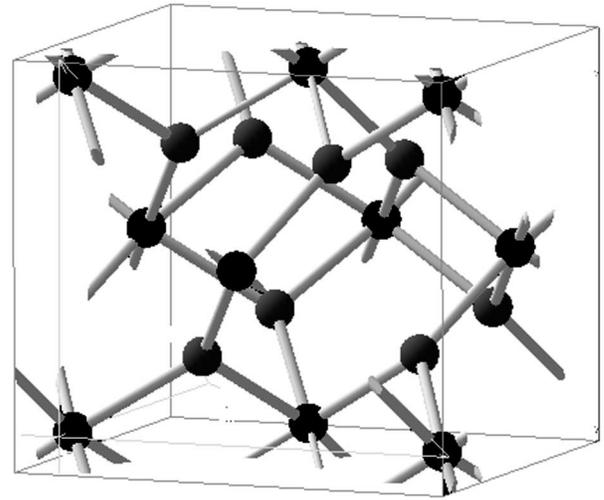

FIG. 1. Si-epitaxial ZrO_2. Grey (black) atoms: O (Zr).

local structures and terminations of oxide/Si (001) interfaces, e.g., Si/O, Si/metal, Si/metal-bilayer, mixed Si-metal layer/O, mixed Si-metal layer/O with 50% vacancies. The starting configuration of the oxide portion of the interface superlattices is assembled using the optimized Si-epi structure. The supercell length and atomic positions are then reoptimized: the axial ratio remains unchanged, and relaxations occur only in the first two interface-neighboring layers. The interface energy can be expressed as the difference of the energies E_{SL} of the interface cell and E_{bulk} of the corresponding bulk components, as

$$E_{form} = \frac{1}{2A}[E_{SL} - E_{bulk}] = \frac{1}{2A}[(2A\delta + n_{Si}V_{Si}E_{Si} + n_{ox}V_{ox}E_{ox}) - (n_{Si}V_{Si}E_{Si} + n_{ox}V'_{ox}E'_{ox})] = \delta,$$

with n the number of bulk units, V, V' and E, E' the corresponding volumes and energies per unit volume, A the basal superlattice area. The formation energy per unit area, δ, can be extracted unambiguously if the oxide bulk energy is calculated in the same strain state as in the superlattice (Si remains unstrained), as in that case $V_{ox}E_{ox} = V'_{ox}E'_{ox}$ and all volume-dependent terms drop off. Any other choice of the bulk energies inserts a volume dependence in the interface energy [12].

The interface cell may be stoichiometric, metal- or oxygen-deficient depending on its local structure. Its formation energy will, therefore, depend on growth conditions, metal-rich ones favoring oxygen deficit and O-rich favoring oxygen excess. Theoretically, this is described by fixing the chemical potentials of the constituents. Here, only one potential, e.g., oxygen's, is independent: $\mu_O = \mu_{O_2}/2$ means O-rich conditions and $\mu_O = \mu_{O_2}/2 + \Delta H_{ox}/2$ metal-rich ones.

The formation energies of the various interfaces are listed in Table I. The standard Si-O interface is favored in O-rich growth conditions. In metal-rich conditions, the preferred structure is the mixed Si-metal to 50% vacant

oxygen layer interface depicted in Fig. 2, which remarkably is the same as was recently obtained [13] in all-electron *ab initio* molecular dynamics simulations of metal deposition on, and oxidation of, Si (001). Notably, the two favored interfaces have large *negative* formation energies (referred, we remind, to the prestrained bulk). This energetic gain in interface formation will be counterbalanced by the excess energy of the film's upper surface and by the buildup of epitaxial elastic energy in the growing layer. An estimate of the critical thickness t_c for pseudomorphic growth over an area A then results from

$$AE_{form} + At_c E_{elastic}^{epi} + AE_{surf} = 0,$$

using which we predict that crystalline zirconia and hafnia thin films should grow pseudomorphically on Si (001): indeed, using our calculated values for, e.g., zirconia, and the GGA surface-energy estimate for the tetragonal phase $E_{surf} \simeq 0.05$ eV/Å^2 [14], we obtain $t_c \sim$ 18 and 27 Å for metal- and oxygen-rich conditions.

The interface band offsets are evaluated for each interface using the standard "bulk-plus-lineup" procedure

TABLE I.   Formation energies (eV/Å^2) of, and valence and conduction band offsets (eV) at, different Si (001)/oxide interfaces. The assumed growth conditions are indicated. The best offset/energetics combinations are displayed in underlined bold for metal-rich conditions and bold for oxygen rich conditions. All GW corrections are included.

| Material | $\rightarrow$ | HfO$_2$ | | | ZrO$_2$ | | |
Interface $\downarrow$	Growth	VBO	CBO	E_{form}	VBO	CBO	E_{form}
Si/O	O-rich	4.14	0.47	−0.16	4.08	**0.72**	**−0.21**
Si-M/O	stoich	4.40	0.19	0.17	4.18	0.62	0.12
Si/M	M-rich	3.96	0.65	0.12	4.72	0.08	0.07
Si-M/O vac	M-rich				3.91	<u>**0.89**</u>	<u>**−0.15**</u>
Si/O vac	stoich	4.62	−0.01	0.22	3.70	1.10	0.13

[15], expressing the valence offset (VBO) as the sum of the interface potential lineup and the valence-band-top differences of the separately considered bulks. The conduction band offsets (CBOs), hence the electron injection barriers, are estimated as CBO = $E_{\text{gap}}^{\text{oxide}} - E_{\text{gap}}^{\text{Si}} -$ VBO. The gap of Si is taken to be 1.1 eV; for both oxides, we use our GGA gaps corrected with the perturbative many-body (GW) eigenvalue corrections of Ref. [16] for ZrO$_2$, namely, 5.9 and 5.7 eV for zirconia and hafnia, respectively. These values are close to experiment for hafnia and near the bottom of the (large) experimental range for zirconia. We neglect spin-orbit corrections, which should be well below ±0.1 eV as the valence states are oxygen-derived. We do include, instead, the quasiparticle corrections to the bulk valence-band edges at the GW level: this is essential since these corrections are of order ~1 eV in oxides compared to typical ~0.1 eV in semiconductors. We apply to the VBOs an overall correction of −1.08 eV, resulting from the −0.15 eV correction [17] for Si and the −1.23 eV correction [16] for ZrO$_2$. Using the latter for both oxides introduces some uncertainty in the HfO$_2$ results, but unfortunately no GW data are currently available for hafnia.

In Table I we report the predicted VBOs and CBOs. Qualitatively, VBOs cluster around 4 eV, with appreciable structure dependence, and CBOs are in the range 0 to 1 eV. Interestingly, for zirconia the energywise most favorable structures have some of the largest conduction offsets.
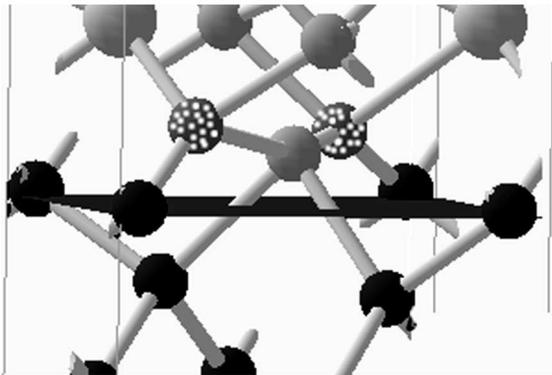
The high-end CBOs, ~1 eV, are smaller than, but comparable to, the 1.4–1.5 eV estimates by Robertson [18], who used a simple charge-neutrality-level model at the empirical tight-binding level.

Dielectric constants.—The lattice contribution to the dielectric tensor has been calculated for both oxides in the fluorite, monoclinic, and Si-epitaxial structures. We used a standard formalism to evaluate the zero-frequency dielectric constant [10] via the frequencies of zone-center IR-active modes and the transverse dynamical charges. The vibrational modes are calculated diagonalizing the zone-center dynamical matrix $-\partial F_i^\alpha / \partial u_j^\beta$, obtained differentiating by centered finite-differences (with displacements of 0.1 Å) the Hellmann-Feynman force component α on atom i with respect to the displacement of atom j along direction β. The dynamical charges are obtained by differentiation of the Berry-phase [19] polarization with respect to atomic displacements (of typically 0.05 Å).

Since the epi-oxides were optimized without constraints, they have no symmetry of practical use. The calculation of the dynamical charge tensor for all atoms in the epitaxial and monoclinic structures is thus rather demanding and currently in progress. In Table II we give estimates of the diagonal elements of the lattice dielectric tensor obtained using the dynamical charge tensor of the fluorite phase, which is diagonal and isotropic, and calculated to be $Z_{\text{Hf}}^* = 5.20$ and $Z_{\text{O}}^* = -2.60$ for HfO$_2$, $Z_{\text{Zr}}^* = 5.50$ and $Z_{\text{O}}^* = -2.75$ for ZrO$_2$. Of course, smaller

FIG. 2.   The [mixed metal-Si]/[O 50% vacant] interface (black: Si; grey: metal; white-dotted: O).

TABLE II.   Lattice dielectric tensor for fluorite, monoclinic, and Si-epitaxial XO$_2$ (the small off-diagonal elements for the monoclinic are not displayed for clarity), calculated using the dynamical charge tensor of fluorite. $\varepsilon_{\text{lat}}^{\text{ave}}$ is the orientational average measured by series capacitance in polycrystalline layers and obtained as $3/\varepsilon^{\text{ave}} = 1/\varepsilon^{xx} + 1/\varepsilon^{yy} + 1/\varepsilon^{zz}$.

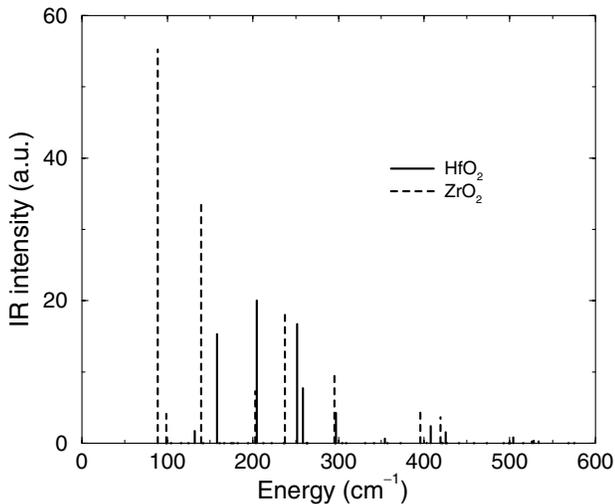|	$\varepsilon_{\text{lat}}^{xx}$	$\varepsilon_{\text{lat}}^{yy}$	$\varepsilon_{\text{lat}}^{zz}$	$\varepsilon_{\text{lat}}^{\text{ave}}$
HfO$_2$ fluorite	27.8	27.8	27.8	27.8
HfO$_2$ monoclinic	17.5	15.7	12.4	14.9
HfO$_2$ Si-epi	27.6	18.6	24.5	22.9
ZrO$_2$ fluorite	30.5	30.5	30.5	30.5
ZrO$_2$ monoclinic	24.7	18.3	14.6	18.4
ZrO$_2$ Si-epi	22.5	71.5	44.9	37.0

FIG. 3. Orientationally averaged IR intensity spectrum (mode dielectric constants) of Si-epitaxial HfO_2 (solid line) and ZrO_2 (dashed line).

dynamical charges such as found in monoclinic phases [10,11] will decrease the dielectric constant, especially the zz component. Using the monoclinic cation charge tensors of Refs. [10,11] and imposing the Friedel sum rule to obtain an average anion charge tensor, we estimated ε_{zz} to be 9.9 and 11.9 in monoclinic HfO_2 and ZrO_2, respectively, in fair agreement with previous results. Along with our fluorite values, also in good agreement with previous calculations, this gives us confidence on the reliability of our procedure.

With reference to Table II, for hafnia we find a reduction in dielectric constant compared to fluorite both in the Si-epi and monoclinic phases, though the latter is rather more dramatic, with a more than twofold decrease, in agreement with previous calculations [11]. For zirconia, we also find a similar, approximately twofold reduction of the monoclinic dielectric tensor compared to fluorite; notably, though, a drastic enhancement is found in the Si-epitaxial phase. This results from the large IR intensity of modes at about 90 to 140 cm^{-1}, as can be seen in Fig. 3, which shows the mode dielectric constants [10] for both materials in the Si-epi phase. The two lower-energy modes for zirconia (dashed lines) contribute mostly to the yy component, the third to the zz component. The pronounced softness of Si-epi zirconia is due to the backfolding of zone-border (X-point) modes.

We carefully checked against artifacts by accurately reoptimizing structures and repeating phonon calculations for different displacements. We are confident in our procedure also in view of the results for the other phases. The single zone-center IR-active mode of fluorite is $\omega = 230$ cm^{-1} for HfO_2 and $\omega = 258$ cm^{-1} for ZrO_2; for the latter, this agrees with recent predictions [9,10]; for the former, the frequency is 20% lower than in Ref. [11]. We obtained the same results (within 0.5% for the lattice constant and 2% for the frequency) with the all-electron PAW method with valence semicore [6,8].

The vibrational spectra of the epi and monoclinic phases will be reported elsewhere, but we note in passing that the results for the monoclinic are close to previous reports [11].

In conclusion, zirconia and hafnia as gate oxides appear to be interesting from the dielectric and epitaxy-energetic standpoints. The electron injection barriers of less than 1 eV are disappointing, as they are much smaller than at silica/Si interfaces. While insufficient for hot electrons, the barrier should be still acceptable for standard two-dimensional inversion layers, whose energy levels are ~100 meV above the interface triangular-well bottom [20], and also in view of the reduction of drive voltages to well below 1 eV in nanometric devices [2].

[1] H. R. Huff, J. Electrochem. Soc. **149**, S35 (2002); S. Thompson, P. Packan, and M. Bohr, Intel Techn. J. **Q398**, 1 (2001).
[2] D. A. Buchanan, IBM J. Res. Dev. **43**, 245 (1999); G. D. Wilk, R. M. Wallace, and J. M. Anthony, J. Appl. Phys. **89**, 5243 (2001); MRS Bull. **27**, No. 3 (2002); IBM J. Res. Dev. **46**, Nos. 2/3 (2002).
[3] G. E. Moore, Electronics **83**, 114 (1965); Proc. SPIE Int. Soc. Opt. Eng. **2438**, 2–17 (1995).
[4] Y. Kim *et al.*, Tech. Dig. Int. Electron Devices Meet., 20.2.1 (2001).
[5] J. P. Perdew, in *Electronic Structure of Solids*, edited by P. Ziesche and H. Eschrig (Akademie-Verlag, Berlin, 1991), p. 11.
[6] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996); Phys. Rev. B **54**, 11 169 (1996); G. Kresse and D. Joubert, *ibid.* **59**, 1758 (1999).
[7] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).
[8] P. E. Blöchl, Phys. Rev. B **50**, 17 953 (1994).
[9] G.-M. Rignanese, F. Detraux, X. Gonze, and A. Pasquarello, Phys. Rev. B **64**, 134301 (2001).
[10] X. Zhao and D. Vanderbilt, Phys. Rev. B **65**, 075105 (2002).
[11] X. Zhao and D. Vanderbilt, Phys. Rev. B **65**, 233106 (2002).
[12] F. Bernardini and V. Fiorentini, Phys. Rev. B **57**, R9427 (1998).
[13] P. Blöchl and C. Först (to be published).
[14] A. Christensen and E. A. Carter, Phys. Rev. B **62**, 16 968 (2000).
[15] M. Peressi, N. Binggeli, and A. Baldereschi, J. Phys. D **31**, 1273 (1998).
[16] B. Kralik, E. K. Chang, and S. G. Louie, Phys. Rev. B **57**, 7027 (1998).
[17] X. Zhu and S. G. Louie, Phys. Rev. B **43**, 14 142 (1991).
[18] J. Robertson, J. Vac. Sci. Technol. B **18**, 1785 (2000).
[19] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1992); VASP implementation by Martijn Marsman.
[20] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).