Energetics of silicon suboxides

D. R. Hamann
Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974
(Received 4 October 1999)

Substoichiometric oxides of Si are modeled by a simple and consistent set of structures that allow the energy penalty associated with the substoichiometry to be calculated. The energies are evaluated by generalized-gradient density-functional theory, and trends in the electronic structure are described.

The interface between crystalline Si and its stable oxide SiO_2 is of great technological importance. The fact that this interface can be formed with an extremely low areal density of defects is a key enabler of today's integrated-circuit electronics. As the lateral dimensions of the individual transistors forming these circuits continue to shrink, the thickness of the SiO_2 layer insulating the gate electrode from the conducting channel must also shrink to achieve the desired performance. Gate oxide layers as thin as 25 Å are manufactured today, and recent experiments suggest that acceptable insulating properties can be expected down to 12 Å.[1] At these dimensions, the interface is only a few molecular layers thick, and smoothness on an atomic scale is required. No crystalline phases of SiO_2 are sufficiently lattice matched to grow pseudomorphically on Si, and the oxide layers are found to be amorphous. The flexibility of the SiO_2 network, which consists of corner-sharing SiO_4 tetrahedra with a soft bond-bending force constant at the shared O vertices, is generally credited with permitting the formation of an interface that is defect-free in the sense of having all atoms fully coordinated.

Given its importance and its complexity, the Si-SiO_2 interface has been the subject of many experimental and theoretical studies. A key observation, based on x-ray photoemission spectroscopy, is that the interface appears to contain all possible ionic states of Si.[2,3] The Si $2p$ core spectra generally show five lines, two representing the bulk materials Si and SiO_2 whose relative intensities depend on oxide thickness and electron escape depth. In addition, ''suboxide'' lines associated with the interface are seen corresponding to Si^{1+}, Si^{2+}, and Si^{3+}. Concentrations are roughly equal at (100) interfaces, while Si^{1+} and Si^{3+} predominate at (111).[2,3] Theoretical studies of the interface have, from computational necessity, been restricted to periodic structures with a rather small repeat distance. They generally describe ''maximally abrupt'' interfaces in which the only suboxide species present is Si^{2+},[4,5] although plausible models specifically designed to agree with the photoemission concentrations have been proposed for (100).[6] The actual interface configuration is undoubtedly controlled by a complex combination of equilibrium energetics and growth kinetics, and simulating its formation in any moderately realistic way with first-principles calculations is a distant goal awaiting vast further increases in computing power. Force-field models have a better chance of simulating realistic structures and processes,[5] but have not been developed to explicitly deal with suboxide configurations.

The purpose of the present investigation is to provide both qualitative and quantitative information on the manner in which incomplete O coordination influences the energies of Si—O and Si—Si bonds. To accomplish this, I have constructed a series of simple and highly artificial model suboxide structures. In each of these, all bonds of the same type are symmetry equivalent and are as equivalent as possible to the bonds in reference configurations of pure Si and pure SiO_2. The structures all have complete bonding networks in the sense that all Si are fourfold and all O twofold coordinated, which is the situation believed to exist at the interface (except for a very low density of bonding defects). The zero-order picture of chemical bonding is that A—B bonds have a well-defined bond energy. The present approach calculates the first-order correction to this picture due to the interactions of neighboring bonds. Second-order effects of more distant neighbors, of course, will set a limit on the accuracy with which these results can be transferred to other structural arrangements.

The model structures were all developed with the idea of using the structure mistakenly assigned to β-cristobalite in early work[7] as a reference. This structure is essentially a diamond Si lattice with an O placed midway between each Si pair, with cubic space group symmetry $Fd3m$. The lattice constant must be increased far beyond that of real β-cristobalite to accommodate normal-length (1.62 Å) Si—O bonds with the symmetry-imposed 180° Si—O—Si bond angle. This model is shown in Fig. 1(a). The Si^{3+} suboxide model is formed by removing the O from all [111]-directed bonds and bringing the neighboring Si's to a normal (2.35 Å) Si—Si bond length. The resulting structure has the trigonal space group $R3m$, and as can be seen in Fig. 1(b) consists of layers of buckled honeycomb Si—O—Si rings joined in their layer-normal direction by Si—Si bonds.

The Si^{2+} model is constructed by removing the O's from [111] and $[\bar{1}\bar{1}1]$ bonds. Seen in Fig. 1(c), the resulting orthorhombic structure consists of zigzag Si—O—Si chains in the **a** lattice direction and zigzag Si—Si—Si chains in the **b** direction, and has space group *Imma*. Finally, the Si^{1+} model is the ''inverse'' of the Si^{3+} model, with the O's retained in the [111] bonds and removed elsewhere. Layers of buckled honeycomb Si rings are now joined by Si—O—Si links, Fig. 1(d). The space group is the same, $R3m$. The neutral reference, Si^0, is of course just ordinary bulk Si. The symmetries of these structures enforce the 180° Si—O—Si bond angle for all of them. This incurs some energy penalty

 ©2000 The American Physical Society

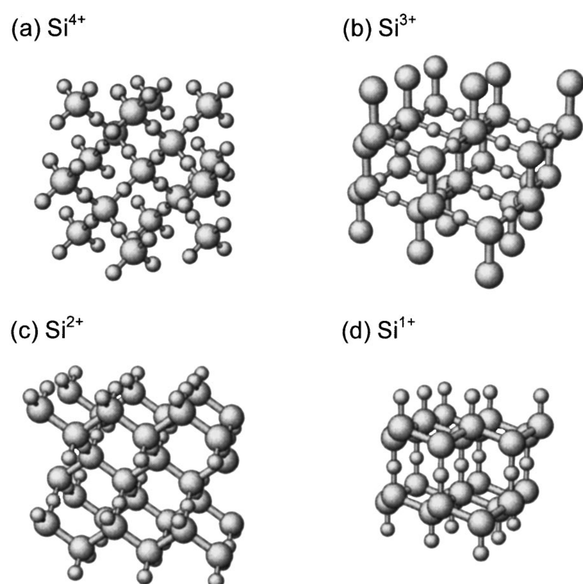

FIG. 1. Structural models.

compared to the 140° to 155° angles found in tetrahedrally coordinated Si polymorphs, but it is consistent throughout and is consistently subtracted out of the energy comparisons I will make. Exact tetrahedral symmetry about each Si is not enforced for the suboxide structures, but departures from the ideal 109.5° bond angles are found to be small in the fully relaxed structures. All models have two Si atoms per primitive unit cell.

To obtain the small expected energy differences between the references and the suboxide models with useful accuracy, well-converged density-functional calculations were performed. These employed adaptive curvilinear coordinates,[8] norm-conserving pseudopotentials[9] with nonlinear core corrections,[10] and the Perdew-Burke-Ernzerhof generalized-gradient density functional.[11] The author recently showed that the generalized-gradient approximation (GGA) gave excellent agreement with experiment in describing the high-pressure SiO_2 phase transition from α-quartz with fourfold-coordinated Si to stishovite with sixfold-coordinated Si, while the local-density approximation was qualitatively incorrect.[12] It therefore seemed imperative to use the GGA in treating the decreasing O coordination of Si in theses suboxide models.

The calculations were carried out with an average basis cutoff of 35 Ry, boosted near the O's to an effective cutoff of 120 Ry by the adaptive coordinate transformation,[8] well beyond the 90 Ry previously shown to give excellent convergence in similar calculations.[12] Optimization of the geometries involved just the cubic lattice parameter a for the reference Si^0 and Si^{4+} models, the lattice parameters a and c and one internal coordinate for the trigonal Si^{1+} and Si^{3+} models, and a, b, c, and one internal coordinate for orthorhombic Si^{2+}. The internal coordinates were relaxed via Car-Parrinello molecular dynamics,[13] while the lattice parameters were determined by fitting energy samples evenly spaced in volume, c/a, and a/b as appropriate with polynomial or Murnaghan[14] functions. Geometric optimization lowered energies by at most a few tenths of an eV compared to the ideal geometries with standard bond lengths and strictly tetrahedral angles. A Brillouin-zone sample of eight evenly spaced points[15] (full zone) was used to optimize the geometries, and the sample sizes were then increased to give convergence to ~0.01 eV. The Si^{1+} and Si^{2+} structures, despite their complete bonding networks and insulating or semiconducting ''parent'' compounds, turned out to be slightly semimetallic, with a few tenths of an eV overlap of the nominal valence and conduction bands, and required quite large **k** samples for the desired degree of convergence. The overlap is likely a consequence of the general tendency of density-functional calculations to underestimate band gaps.

TABLE I. Suboxide energy penalties ΔE, bond lengths, and total energies.

Species	ΔE (eV)	$R_{Si—Si}$ (Å)	$R_{Si—O}$ (Å)	E_{tot} (hartree)
Si^{4+}	0.0		1.617	−72.6983
Si^{3+}	0.24	2.338	1.633	−56.4915
Si^{2+}	0.51	2.383	1.651	−40.2831
Si^{1+}	0.47	2.416	1.665	−24.0972
Si^0	0.0	2.367		−7.9427

The results are summarized in Table I. Appropriate fractions of the end-member total energies representing the zero-order ''constant bond energy'' (CBE) model were subtracted from the three suboxide total energies, giving energy differences ΔE per Si. All the suboxide energies are positive, as might be expected on the basis that Si oxidation proceeds to completion in the bulk. It was not possible to anticipate either the magnitude of the suboxide energies or the trend among the species.

The models were designed so that the lengths of the Si—O and Si—Si bonds could vary completely independently in the course of the geometry optimization. The bond lengths shown in Table I show significant shifts from the reference models. The variation of the Si—O bond length from 1.665 Å in Si^{1+} to 1.633 Å in Si^{3+} is in the expected direction, toward the standard Si^{4+} value of 1.617 Å. The Si—Si trend is counterintuitive, however, showing the greatest increase from the reference length at Si^{1+} and a length in the Si^{3+} model that is slightly shorter than that calculated for Si^0. Evidence for a trend in this direction has, however, been found on (111) surfaces.[16]

As noted above, the energy decreases due to geometric relaxation were small on the scale of the ΔE's, so the question of understanding ΔE trends is largely decoupled from details of the geometry. Calculations of the energies of these structures with a widely used ionic pair-potential model[17] gave the minimum ΔE penalty for Si^{1+} (all these ΔE's were unphysically large, however). The fact that the Si^{3+} appears to be the most stable of the three species suggests that covalency and quantum resonance among the Si—O bonds play important roles. Figure 2 shows the densities of states of the models. Si^0 is aligned with the top of the valence band at energy 0, Si^{1+} and Si^{2+} are aligned with their Fermi levels at 0, and Si^{3+} and Si^{4+} are shifted so that the centers of gravity of their lowest O 2s bands are aligned with that of Si^{2+}. This latter alignment is physically justified by the fact that the O coordination and nominal O oxidation state is the same in all the structures, so the ''semicore'' O 2s provides a consistent reference energy with which to line up the undetermined

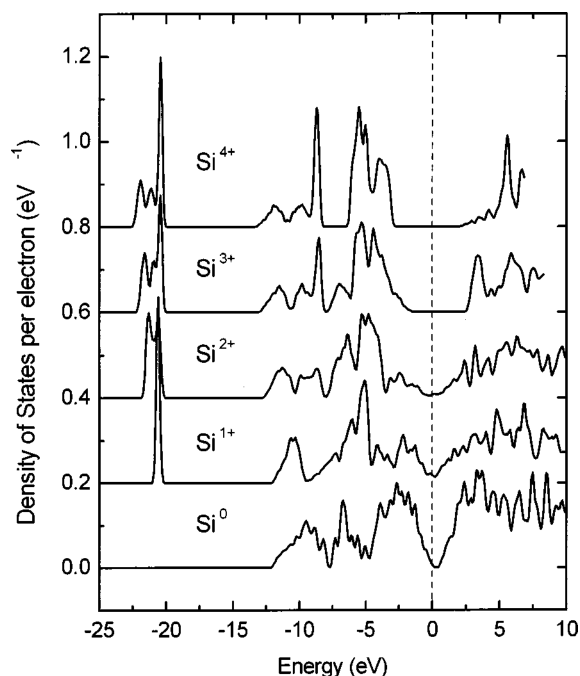

FIG. 2. Densities of states normalized by electron number and smoothed with an 0.3-eV FWHM Gaussian. Energy scales are aligned as described in text. Zeros are offset by 0.2 units for clarity.

average electrostatic potentials of the different models. The trend towards a larger average band gap with increasing valence is obvious and consistent with the trend of the CBE reference energies. As a leading correction, resonance among the much stronger Si—O bonds would be expected to yield more binding energy than among weaker Si—Si bonds. The resonance broadening of O-related features is most apparent in the $2s$ band, since it is obscured by the overlap with the Si-based density of states (DOS) contributions in the vicinity of the O $2p$ bonding and lone-pair bands. This appears to be the most plausible explanation of the observed ΔE trends. Note that the DOS plots in Fig. 2 are smoothed using an 0.3-eV full width at half maximum Gaussian, which obscures the 0.66-eV density-functional gap of Si^0 and exaggerates the very small Fermi-level DOS of Si^{1+} and Si^{2+}.

The energy scale found here for the ''suboxide penalty,'' 0.25–0.5 eV, is of the same order as the energy spread produced by different topologies (hence different Si—O—Si bond angle strain) among SiO_2 polymorphs.[18] It is also readily accessible through thermal energy at the temperatures of 800–1000 °C typically used in gate oxide growth. While a complex tradeoff among numbers of suboxide atoms, suboxide species, bond-length and bond-angle strain relief, and growth-kinetic constraints must determine the actual Si-SiO_2 interface structure, the roughly comparable ΔE's are consistent with the concentrations inferred from photoemission.[2,3] Note, however, that under the constraints of the constant O number, rearranging interface bonding topology to convert two Si^{2+} to one Si^{1+} and one Si^{3+} reduces the penalty by ~0.3 eV. Since a maximally smooth interface would consist predominantly of Si^{2+}, this is consistent with the roughening recently observed at the top (polysilicon) interface of a gate structure after annealing at 1050 °C.[1] Finally, I observe that the narrowed gap observed in the interface region by atomic-resolution electron-energy-loss spectroscopy[1] is consistent with the electronic structure trends of my highly oversimplified set of models, suggesting that despite their simplicity they bear relevance to the suboxide region of real interfaces.

Note added in proof. Recent calculations of more realistic interface structures show Si–O bond length trends in agreement with Table I.[19]

The author thanks D. A. Muller, M. S. Hybertsen, and P. H. Citrin for stimulating discussions.

[1] D. A. Muller *et al.*, Nature (London) **399**, 758 (1999).
[2] F. J. Himpsel *et al.*, Phys. Rev. B **38**, 6084 (1988); Z. H. Lu *et al.*, Appl. Phys. Lett. **63**, 2941 (1993); M. M. Banaszak Holl *et al.*, *ibid.* **65**, 1097 (1994); K. Raghavachari and J. Eng, Phys. Rev. Lett. **84**, 935 (2000).
[3] K. Ohishi and T. Hattori, J. Appl. Phys. **33**, L675 (1994).
[4] C. Kaneta *et al.*, Microelectron. Eng. **48**, 117 (1999).
[5] Y. Tu and J. Tersoff, cond-mat/9903424 (unpublished).
[6] A. Pasquarello, M. Hybertsen, and R. Car, Appl. Phys. Lett. **68**, 625 (1996); Phys. Rev. B **53**, 10 942 (1996).
[7] R. W. G. Wyckoff, *Crystal Structures*, 2nd ed. (Krieger, Malabar, FL, 1982), Vol. 1, p. 318.
[8] F. Gygi, Europhys. Lett. **19**, 617 (1992); Phys. Rev. B **48**, 11 692 (1993); **51**, 11 190 (1995); F. Gygi and G. Galli, *ibid.* **52**, R2229 (1995).
[9] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979); D. R. Hamann, Phys. Rev. B **40**, 2980 (1989).
[10] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
[11] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[12] D. R. Hamann, Phys. Rev. Lett. **76**, 660 (1996).
[13] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[14] F. D. Murnaghan, Proc. Natl. Acad. Sci. USA **30**, 244 (1944).
[15] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[16] M. T. Sieger *et al.*, Phys. Rev. Lett. **77**, 2758 (1996).
[17] S. Tsuneyuki *et al.*, Phys. Rev. Lett. **61**, 869 (1988).
[18] N. R. Keskar and J. R. Chelikowsky, Phys. Rev. B **46**, 1 (1992).
[19] K.-O. Ng and O. Vanderbilt, Phys. Rev. B **59**, 10 132 (1999).