PHYSICAL REVIEW B VOLUME 21, NUMBER 12 15 JUNE 1980

Theory of the electronic structure of the Si-SiO_2 interface

R. B. Laughlin and J. D. Joannopoulos
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

D. J. Chadi
Xerox Palo Alto Research Center, Palo Alto, California 94304

(Received 31 October 1979)

A theory of the Si-SiO_2 interface based on recent experimental findings for silicon surfaces and their oxidation is presented. It is proposed that a simple local-orbital picture can simultaneously describe silicon, its oxidation, and the Si-SiO_2 interface and that two dimensionality is not essential to physically meaningful calculations of interface local densities of states. Calculations are performed which show that interface states do not arise simply from the presence of a boundary. It is argued that band tailing at the interface, like that in amorphous silicon, is due primarily to strain rather than to charged centers, and that dangling bonds at the interface should give rise to an inhomogeneously broadened discrete level at midgap.

I. INTRODUCTION

Great activity in surface physics in recent years has produced a number of results for silicon surfaces and their oxidation which are potentially very useful in understanding the valence and conduction states of the Si-SiO_2 interface. Firstly, photoemission,[1-3] electron-energy-loss,[4] spin-resonance,[5] and infrared-absorption[6] measurements have provided a detailed picture of the electronic structure of the silicon (111) surface in agreement with self-consistent pseudopotential calculations.[7,8] An essential feature of this picture is a band of surface states[7,8] in the lower gap region formed from dangling bonds. This half-filled band is thought[9] to be responsible, via a Peierls instability, for the 2×1 reconstruction. Surface atoms relax[5] alternately in and out, the out atoms forming a filled band at the valence-band maximum, the down atoms forming an empty band 0.4 eV higher.[8,9] The filled band is observed in photoemission,[1,3] the empty one via infrared absorption,[6] and both are reproduced in a self-consistent calculation[8] when reasonable relaxations are introduced. Secondly, the results of the self-consistent calculations are extremely close to those of *non*-self-consistent empirical tight-binding calculations,[10] even for the 2×1 reconstruction, in which a substantial fraction of an electron[8] is purportedly transferred from in atoms to out. Thirdly, oxidation removes[1] dangling-bond surface states without introducing spins[5] and without increasing the density of states at the Fermi level.[11]

Unlike the silicon surface, the Si-SiO_2 interface cannot be approached easily with pseudopotentials. The problem is technical rather than physical. Pseudopotential calculations are not practical without periodicity in two dimensions, and it has thus far proved difficult to impose periodicity on Si-SiO_2 without invoking exotic bonding,[12] an extremely unphysical approximation in light of the probable tendency of bonds to saturate, if necessary, at the expense of periodicity. For all its structural complexity, however, the interface appears experimentally to have a simple[11] valence electronic structure. Features induced in the silicon density of states by chemisorption of less than a monolayer of oxygen lie within 2 eV of the bulk bands of the oxide.[11] The transition from oxidized silicon to SiO_2 appears smooth[11] in photoemission. The absence of any explicit manifestations of structural complexity in the density of states is a well-understood[13] occurrence in bonded-disordered materials. Features in the density of states are attributable[13] either to coordination or to topology. Those due to topology are simply absent in the glass. The existing photoemission data[11] suggest that, like the reconstructed 2×1 silicon surface, the interface may be described with *non*-self-consistent tight binding. This possibility, which is rather surprising in light of the amount of charge transfer in this problem and the localized[14] nature of the nonbonding oxygen p orbitals on an oxidized surface, has a number of important consequences. The alignment of the valence-band edges of the two materials, for example, can be predicted from the ionization potentials of free atoms. Within the context of a tight-binding description, a number of options[13] exists for modeling structurally complex (disordered) systems, such as the interface. One of these is the cluster-Bethe-lattice method.[15] This involves constructing a molecule locally representative of the actual system and then approximating the effect of embedding this molecule in a solid by attaching Bethe lattices[16] to all the dangling bonds at the periphery. The resulting cluster-Bethe-lattice system has Green's-function matrix elements at its center

very similar to those of the actual system[13] as modeled by a random network, but is much easier to solve.

In this paper, we use the empirical tight-binding and cluster-Bethe-lattice methods to establish a theoretical connection between the surface and the interface-valence and conduction states. The plan of the paper is as follows. In Sec. II, we explain the Bethe lattice, motivate its use in studying the interface, and outline its solution. In Sec. III, we discuss the problem of interpolating the tight-binding Hamiltonian between covalent silicon and ionic SiO_2. In Sec. IV, we present numerical results for a number of model interfaces having neither distorted nor broken bonds. We attempt to answer a number of elementary questions: What local densities of states near the interface would one expect for different structures? Can localized states arise simply from the presence of a boundary? Are there any signatures of interface bonding in the valence state which might be observed in photoemission? In Sec. V, we discuss the effects of distorted or disrupted bonding at the interface. In Sec. VI, we summarize our results.

II. THE BETHE-LATTICE INTERFACE

A Bethe lattice is a mathematical idealization of a material based on a tight-binding description of the crystalline bulk. An atom in the Bethe lattice has the same complement of orbitals, the same nearest-neighbor coordination, and the same interactions with these neighbors as an atom in the crystal, but the logical continuation of this rule to second neighbors and beyond produces a chain of nearest-neighbor interactions which never closes on itself. The Bethe lattice has three distinct advantages over a crystal as a structural model for a disordered material: (i) it is not periodic, (ii) it is not pointlessly complicated, and (iii) structure in the density of states of the crystal destroyed when it is made amorphous is correctly[17] absent in the Bethe-lattice density of states.

In Fig. 1, we compare the densities of states of the silicon and SiO_2 Bethe lattices with x-ray photoemission spectra (XPS) of *crystalline* silicon[17] and *amorphous* SiO_2.[18] But for the peak at −6 eV in the silicon XPS, which is absent in amorphous silicon, all the major features in the experimental densities of states appear in the theoretical curves. These include two bands in SiO_2—one centered at −7 eV, one centered at −12 eV, both about 5 eV wide, the lower one possessing at peak near −11 eV; one band in silicon, about 10 eV wide, with a peak near −11 eV and one band in silicon, about 10 eV wide, with a peak near −3 eV. The gaps of both Bethe lattices are slightly larger than the measured gaps due to the inability[19] of the Bethe lattice to correctly describe the long-wavelength states at the band edges. The solutions of these Bethe lattices and the agreement of their valence and conduction states with experiment and with other calculations have been discussed previously.[19,20] The tight-binding parameters used in the calculations are reproduced in Table I.

Breaking one of the bonds in the silicon Bethe lattice produces a localized state at midgap analogous to the midgap surface band[7-10] on unrelaxed silicon (111), as well as resonances analogous to the more deeply bound,[10] bands of surface states and resonances in silicon (111). In this sense the Bethe lattice with a dangling bond is a physically meaningful model of the surface. The ability of the Bethe lattice to reproduce the major features in the surface density of states stems from the high degree of localization[8] of the wave functions on individual surface sites. Dangling bonds, for example, isolated from one another on the (111) surface, interact weakly to form a narrow band.[7-10] Two dimensionality is responsible for the width of the band, but not its center of mass or its orbital character. As bond formation is analogous[20] in silicon and SiO_2, it is reasonable from a one-electron point of view to assume that a Bethe lattice also models SiO_2 surfaces reliably. Experimentally, it is not yet clear whether this is the case. In all likelihood the strong excitonic effects[8] in SiO_2 seriously complicate the one-electron picture at the surface. These should not, however, affect an interface with silicon since proximity to a screening medium seems to severely reduce these effects in chemisorption (see Sec. III). Therefore, although the two-dimensional structure of the interface is not known, local densities of states should be calculable using two Bethe lattices bonded together as a structural model.

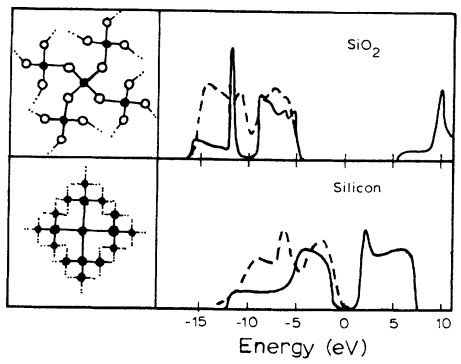

FIG. 1. Schematic drawings and densities of states of the SiO_2 and silicon Bethe lattices. Dashed curves are x-ray-photoemission spectra of amorphous SiO_2 (Ref. 18) and crystalline silicon.[17] The tight-binding parameters used in the calculations are listed in Table I.

TABLE I. Tight-binding parameters used in constructing the silicon and SiO_2 Bethe lattices. The SiO_2 parameters differ from those in Ref. 20 in that the oxygen-oxygen interactions are chosen to reproduce the ordering of the degeneracies of the pseudopotential band structure for α-quartz (see Ref. 12, p. 55). This substitution does not affect the interface results in any significant way. As we have chosen a Hamiltonian without orbital overlaps, the silicon levels are different in the two Bethe lattices. The silicon self-energies are scaled linearly for interface atoms having less than four oxygen neighbors. Values are given in eV.

SiO_2	
$\langle Si_s \mid H \mid Si_s \rangle = -0.55$	$\langle O_s \mid H \mid O_s \rangle = -19.6$
$\langle Si_p \mid H \mid Si_p \rangle = 6.55$	$\langle O_p \mid H \mid O_p \rangle = -6.80$
	$\langle O' \mid H \mid O \rangle$:
$\langle Si_s \mid H \mid O_s \rangle = -2.85$	$V_1^0 = 0.574$
$\langle Si_p \mid H \mid O_s \rangle = -9.5$	$V_2^0 = -0.274$
$\langle Si_s \mid H \mid O_p \rangle = -5.4$	$V_3^0 = -0.574$
$\langle Si_p \mid H \mid O_p \rangle_\sigma = 5.4$	$V_4^0 = -0.45$
$\langle Si_p \mid H \mid O_p \rangle_\pi = -1.4$	$V_5^0 = -0.45$
Si	
$\langle Si_s \mid H \mid Si_s \rangle = -5.05$	
$\langle Si_p \mid H \mid Si_p \rangle = 1.2$	
$\langle Si_s \mid H \mid Si_s \rangle = -1.94$	
$\langle Si_p \mid H \mid Si_s \rangle = -1.75$	
$\langle Si_p \mid H \mid Si_p \rangle_\sigma = 3.05$	
$\langle Si_p \mid H \mid Si_p \rangle_\pi = 1.075$	

The computation of the interface electronic structure using this technique is as follows. One first places a local coordinate system[21] on every silicon atom in such a way that bonds lie along the [111] directions. In the silicon, local coordinates of adjacent atoms are related by a parity operation, while in the SiO_2 they are related by four complicated rotations,[21] M_ν, characteristic of α quartz. These coordinates induce a natural basis for p orbitals. The parameters in Table I are then used to construct 4×4 matrices representing the on-site and nearest-neighbor Hamiltonian. For the silicon, these include the self-energy matrix E_s and four interaction matrices D_ν. For the SiO_2, they include the silicon self-energy E_s', the oxygen self-energy E_o', four silicon-oxygen interactions D'_ν, and twelve oxygen-oxygen interactions $C'_{\nu\mu}$. The oxygen coordinates for D'_ν and $C'_{\nu\mu}$ are those of the common silicon. Using these matrices, one then constructs "effective fields"[15] for the two Bethe lattices, related simply to the Green's function of a Bethe lattice with a broken bond. In silicon, for example, if one breaks a ν_0 bond and denotes the 4×4 submatrix of the Green's function connecting the terminal atom to itself by $G_{00}^{\nu_0}$, the submatrix connecting this atom with its neighbor in the ν_1 direction by $G_{10}^{\nu_1}$, etc., then one has

$$(E-H)G=1\,, \tag{1}$$

$$(E-E_s)G_{00}^{\nu_0}=1+\sum_{\nu_1\neq\nu_0}D_{\nu_1}G_{10}^{\nu_0\nu_1}\,, \tag{2}$$

$$\cdots$$

$$(E-E_s)G_{n0}^{\nu_0,\ldots,\nu_n}=D_{\nu_n}^{T}G_{n-1\,0}^{\nu_0,\ldots,\nu_{n-1}}+\sum_{\nu_{n+1}\neq\nu_n}D_{\nu_{n+1}}G_{n+1\,0}^{\nu_0,\ldots,\nu_{n+1}}\,. \tag{3}$$

These equations have a solution of the form

$$G_{n+1\,0}^{\nu_0,\ldots,\nu_{n+1}}=\Phi_{\nu_{n+1}}G_{n0}^{\nu_0,\ldots,\nu_n}\,, \tag{4}$$

provided the four fields Φ_ν satisfy

$$(E-E_s)\Phi_\nu=D_\nu^{T}+\sum_{\mu\neq\nu}D_\mu\Phi_\mu\Phi_\nu\,. \tag{5}$$

These may be computed by iterating the continued fraction

$$\Phi_\nu=\left(E-E_s-\sum_{\mu\neq\nu}D_\mu\Phi_\mu\right)^{-1}D_\nu^{T}\,, \tag{6}$$

starting from $\Phi_\nu=0$. Substituting into (2), one obtains

$$\Phi_\nu=G_{00}^{\nu}D_\nu^{T}\,. \tag{7}$$

One may find the fields for SiO_2 by transforming the problem into one of interacting SiO_4 tetrahedra.[20] One first splits each oxygen orbital in half, and then formally bonds them together again via a large interaction V. For the self-energy of a tetrahedron, one then has

$$A=\begin{pmatrix} E_s' & \sqrt{2}D_1' & \sqrt{2}D_2' & \sqrt{2}D_3' & \sqrt{2}D_4' \\ \sqrt{2}D_1'^{T} & E_0'-V & 2C_{12}' & 2C_{13}' & 2C_{14}' \\ \sqrt{2}D_2'^{T} & 2C_{21}' & E_0'-V & 2C_{23}' & 2C_{24}' \\ \sqrt{2}D_3'^{T} & 2C_{31}' & 2C_{32}' & E_0'-V & 2C_{34}' \\ \sqrt{2}D_4'^{T} & 2C_{41}' & 2C_{42}' & 2C_{43}' & E_0'-V \end{pmatrix}. \tag{8}$$

Adopting a bond-matching convention characterized by a permutation p,

$$M_\nu=M_{p(\nu)}^{T}\,, \tag{9}$$

one has for the interaction matrices between adjacent tetrahedra

$$B_\nu = \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \overset{p(\nu)^{\text{th}}\ \text{column}}{\overbrace{}} \\ \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ 0 & \cdots & VM_\nu & \cdots & 0 \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \end{matrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \nu^{\text{th}}\ \text{row}. \tag{10}$$

Severing the Bethe lattice between two halves of an oxygen atom, with $p(\nu_0)$ denoting the direction of the severed bond in the base tetrahedron, and letting $g_{n0}^{\nu_0,\ldots,\nu_n}$ be the 20×20 analog of $G_{n0}^{\nu_0,\ldots,\nu_n}$, one obtains

$$(E-A)g_{00}^{\nu_0} = 1 + \sum_{\nu_1 \neq p(\nu_1)} B_{\nu_1} g_{10}^{\nu_0\nu_1}, \tag{11}$$

$$\cdots$$

$$(E-A)g_{n0}^{\nu_0,\ldots,\nu_n} = B_\nu^T g_{n-1\,0}^{\nu_0,\ldots,\nu_{n-1}} + \sum_{\nu_{n+1}\neq p(\nu_n)} B_{\nu_{n+1}} g_{n+1\,0}^{\nu_0,\ldots,\nu_{n+1}}. \tag{12}$$

One then has

$$\Psi_\nu = \left(E - A - \sum_{\mu\neq p(\nu)} B_\mu \Psi_\mu\right)^{-1} B_\nu^T \tag{13}$$

and

$$\Psi_\nu = g_{00}^\nu B_\nu^T. \tag{14}$$

The silicon and SiO_2 fields are used to compute the Green's function for the type-I model interface as follows. For the interface Si–O molecule, one has a self-energy of the form

$$E_\nu^0 = \begin{pmatrix} E_s & \sqrt{2}D_1' \\ \sqrt{2}D_1'^T & E_0' - V \end{pmatrix}, \tag{15}$$

and interactions with nearest neighbors of the form

$$D_\nu^0 = \begin{pmatrix} D_\nu \\ 0 \end{pmatrix} \tag{16}$$

and

$$B_\nu'^0 = \begin{matrix} \overset{p(\nu)^{\text{th}}\ \text{column}}{\overbrace{}} \\ \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & VM_\nu & \cdots & 0 \end{pmatrix} \end{matrix}. \tag{17}$$

This leads to equations for the interface Green's function $\mathcal{G}$ of the form

$$(E - E_{\nu_0}^0)\mathcal{G}_{00}^{\nu_0} = 1 + \sum_{\nu_1\neq\nu_0} D_{\nu_1}^0 \mathcal{G}_{10}^{(\mathrm{Si})\nu_0\nu_1} + B_{\nu_0}^0 \mathcal{G}_{10}^{(\mathrm{SiO_2})\nu_0}, \tag{18}$$

$$(E - E_s)\mathcal{G}_{10}^{(\mathrm{Si})\nu_0\nu_1} = D_{\nu_1}^{0T}\mathcal{G}_{00}^{\nu_0} + \sum_{\nu_2\neq\nu_1} D_{\nu_2}\Phi_{\nu_2}\mathcal{G}_{10}^{(\mathrm{Si})\nu_0\nu_1}, \tag{19}$$

$$(E - A)\mathcal{G}_{10}^{(\mathrm{SiO_2})\nu_0} = B_{\nu_0}^T \mathcal{G}_{00}^{\nu_0} + \sum_{\nu_1\neq p(\nu_0)} B_{\nu_1}\Psi_{\nu_1}\mathcal{G}_{10}^{(\mathrm{SiO_2})\nu_0}. \tag{20}$$

The Green's function on the molecule is thus

$$\mathcal{G}_{00}^{\nu_0} = E - E_{\nu_0}^0 - \sum_{\nu_1\neq\nu_0} (D_{\nu_1}^0 G_{00}^{\nu_1} D_{\nu_0}^{0T} + B_{\nu_0}^0 g_{00}^{\nu_0} B_{\nu_1}^{0T})^{-1}. \tag{21}$$

The remaining diagonal Green's function matrix elements may be generated from $\mathcal{G}_{00}^{\nu_0}$ as follows. One has

$$(E - E_s)\mathcal{G}_{11}^{(\mathrm{Si})\nu_0\nu_1} = 1 + \sum_{\nu_2\neq\nu_1} \mathcal{G}_{21}^{(\mathrm{Si})\nu_0\nu_1\nu_2} + D_{\nu_1}^{0T}\mathcal{G}_{01}^{\nu_0} \tag{22}$$

and

$$\mathcal{G}_{01}^{\nu_0} = \mathcal{G}_{10}^{\nu_0 T}. \tag{23}$$

Thus

$$\mathcal{G}_{11}^{(\mathrm{Si})\nu_0\nu_1} = G_{00}^{\nu_1} + (G_{00}^{\nu_1}D_{\nu_1}^{0T})\mathcal{G}_{00}^{\nu_0}(G_{00}^{\nu_1}D_{\nu_1}^{0T})^T, \tag{24}$$

and thereafter,

$$\mathcal{G}_{nn}^{(\mathrm{Si})\nu_0\cdots\nu_n} = G_{00}^{\nu_n} + \Phi_{\nu_n}\mathcal{G}_{n-1,n-1}^{\nu_0\cdots\nu_{n-1}}\Phi_{\nu_n}^T. \tag{25}$$

Likewise for the SiO_2 side, one has

$$\mathcal{G}_{11}^{(\mathrm{SiO_2})\nu_0} = g_{00}^{\nu_0} + (g_{00}^{\nu_0}B_{\nu_0}^{0T})\mathcal{G}_{00}^{\nu_0}(g_{00}^{\nu_0}B_{\nu_0}^{0T})^T, \tag{26}$$

and thereafter,

$$\mathcal{G}_{nn}^{(\mathrm{SiO_2})\nu_1\cdots\nu_{n-1}} = g_{00}^{\nu_{n-1}} + \Psi_{\nu_{n-1}}G_{n-1,n-1}^{\nu_0\cdots\nu_{n-2}}\Psi_{\nu_{n-1}}^T. \tag{27}$$

The local density of states[13] associated with a degree of freedom $|n\rangle$ is given in terms of a diagonal Green's-function matrix element by

$$\rho_n(E) = -\frac{1}{\pi}\,\mathrm{Im}\langle n|G(E)|n\rangle. \tag{28}$$

III. CHARGE TRANSFER AT THE INTERFACE

A weakness of any non-self-consistent theory of the interface is its failure to deal with the charge transfer across the Si–O bond. This is usually[22] estimated to be $0.5e^-$, which is sufficient, in principle, to generate large fluctuations in the Hartree potential and thus have a large effect on orbital self-energies near the interface. The effects of intrinsic Coulomb fluctuations are currently not known, although it has been speculated[23] that they give rise to interface states. There are, however,

two experimental findings which indicate that self-consistency may not be necessary. Firstly, the energy needed to excite an electron from the silicon valence band to the oxide conduction band[24] is 4.4 eV. The parameters in Table I, which were fit to accomplish this alignment in Fig. 1, show a Si $3p$–O $2p$ splitting (8.0 eV), only slightly larger than that of neutral atoms.[25] Thus, it appears that a tight-binding description based on atomic levels can account for the correct interface dipole. This is consistent with the observation[26] in heteropolar semiconductors that Hamiltonians based on atomic levels tend to produce reliable surface states. Secondly, chemisorption of up to a monolayer of oxygen on silicon (111) induces features[11] in the density of states corresponding to and lying within 2 eV of features of SiO_2 in mechanical contact. As shown in Fig. 2, this is exactly what one would predict from the Hamiltonian in Table I. Figure 2 shows local densities of states near a silicon (111) surface, as modeled by a Bethe lattice with a dangling bond, both with and without a chemisorbed oxygen *atom*. On the clean surface, one finds the well-known[7-10] midgap state (0 eV) and resonances (arrows) obtained by more realistic[10] calculations. (Because this model surface is not periodic, all surface features coincident in energy with the bulk bands are resonances.) On the oxidized surface, the midgap state has been replaced by three precursors to the valence bands of the oxide: an oxygen $2s$ state (−25 eV), an Si–O bond (−13 eV), and a nonbonding oxygen p level (−8 eV). The dashed line in Fig. 2 is the difference between ultraviolet ($\hbar\omega = 40.8$ eV) photoemission spectra of oxidized and unoxidized silicon (111) taken from Ref. 11. Theory and experiment have been aligned to be consistent with Fig. 1. The

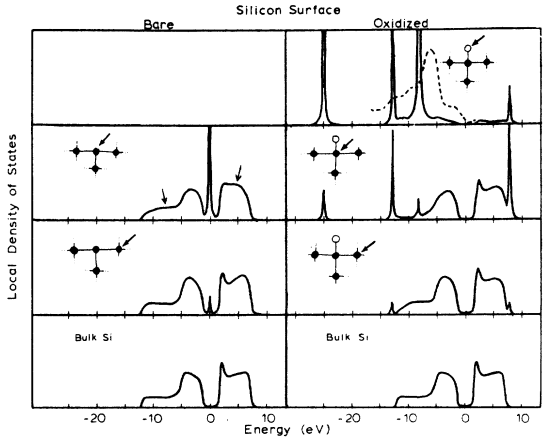

FIG. 2. Bare and oxidized silicon surface as modeled by a Bethe lattice with a dangling bond. The dashed curve is the difference between ultraviolet-photoemission spectra of these two systems taken from Ref. 11.

results in Fig. 2 are similar to those obtained recently by Batra,[27] who has performed a more realistic, but also non-self-consistent calculation. The initial stages of oxidation of silicon are still controversial,[28-30] many researchers[11,14] feeling that oxygen initially chemisorbs as O_2 or as a bridge between adjacent dangling bonds. In the event oxygen chemisorbs as neutral[14] O_2, the experimental peaks at −9 and −6 eV in Fig. 2 must be identified with the π and π^* levels of the O_2 molecule,[14] in which case the positioning of the nonbonding O $2p$ level in the theory is correct. If monatomic oxygen is on the surface, on the other hand, then the peak at −6 eV must be identified with the nonbonding level. Regardless of which of these pictures is correct, therefore, the tight-binding oxygen $2p$ levels must lie approximately 7.5 eV below the silicon valence-band maximum. The fact that the tight-binding prediction works for the surface is somewhat surprising. Previous calculations[31] for nonbridging oxygen chemisorption on GaAs have shown that occupancy of all the single-particle levels below the Fermi level results in a net charge on the oxygen atom (molecule) in excess of $1e^-$. The proves, according to one suggestion,[14,31] that the single-particle picture is inadequate for understanding the initial stages of oxidation. The nonbonding p orbitals of the oxygen atom or molecules are presumed[14] to be sufficiently localized so that they accommodate only an integral number of electrons. Since occupying all the states below the Fermi level would charge the atom and raise some of these above Fermi level, which is impossible, one p orbital must contain a hole. The oxidized surface should thus have spins. (It apparently does not.[5]) One might also speculate that the oxygen levels are doubly occupied but sufficiently screened by the silicon so that the net charge on the oxygen does not matter. Oxygen atoms of SiO_2 in mechanical contact with silicon have a comparable charge, yet do not have their self-energies raised. Accordingly, a compensating charge at one bond length should be sufficient to counteract the oxygen charging, and this might be accomplished by the silicon if it screens sufficiently strongly. The oxidized surface would, of course, have no spins in this case. The strongest piece of evidence supporting such an explanation is the apparent absence of a correlation energy in the self-consistent solution[8] to 2×1 reconstructed silicon (111). We estimate the Coulomb integral between two electrons in a silicon pseudohybrid to be 1.2 Ry, while the compensating potential due to $1e^-$ at 3.82 Å is only 0.28 Ry. Thus, the self-consistent results are sensible[32] only if the silicon screens surface states strongly, and, in light of the short Si–O

bond length, it is not unreasonable that it should similarly screen oxygen orbitals.

IV. IDEAL INTERFACES

An important question to ask about the interface is whether localized states in the silicon gap can arise simply from the presence of a boundary. This question may be explored by performing calculations for a number of model interfaces which are ideal in the sense of having no broken bonds or distorted bond angles.

In Fig. 3, we show calculations for an interface (type I) in which the oxygen atom in Fig. 2 has been replaced by SiO_2. Each frame is a local density of states, starting at the top with an oxygen atom in the oxide and proceeding atom-by-atom through the interface and into the silicon. The upper and lower frames are nearly the bulk densities of states of Fig. 1. The oxide density of states shows three major bands corresponding[20] to three of the surface features in Fig. 2: (i) the strong-bonding band (−18 to −13 eV), composed of silicon hybrids bonding with oxygen p, (ii) the lone-pair-like band (−11 to −6 eV) composed of nonbonding O $2p$, and (iii) the conduction bands (4 eV and above) composed of silicon hybrids antibonding with oxygen. As one approaches the interface from below, one sees a smooth, continuous transition to the oxide density of states in which the region of the silicon gap remains clean. Two mild interface resonances in the valence states can be seen: an enhancement near −5 eV decaying, both into the oxide and into the silicon, and an enhancement near −9 eV behaving similarly. These are analogs of the silicon surface resonances marked by arrows in Fig. 2. Each arises from the effective undercoordination of the bond orbitals[20] compromising the band, and is back-bonding in character. The dashed lines in Fig. 3 refer to a calculation in which the self-energy of the atom in the fourth panel from the top has been raised by approximately 3 eV. This shows (i) that the results of Fig. 3 are relatively insensitive to this self-energy, which is difficult to estimate,[20] and (ii) that *intrinsic* Coulomb fluctuations probably do not localize states in the gap. Because of efficient silicon screening, oxide-induced potential fluctuations should be severe only in the first layer. As the interface resonance marked with an arrow in Fig. 3 has a substantial magnitude on the silicon atom below, it should pop out into the gap if a number of neighboring sites are likewise affected. It seems unlikely, however, that such a large potential fluctuation could be sustained over large distances if it is due simply to disorder in the SiO_2.

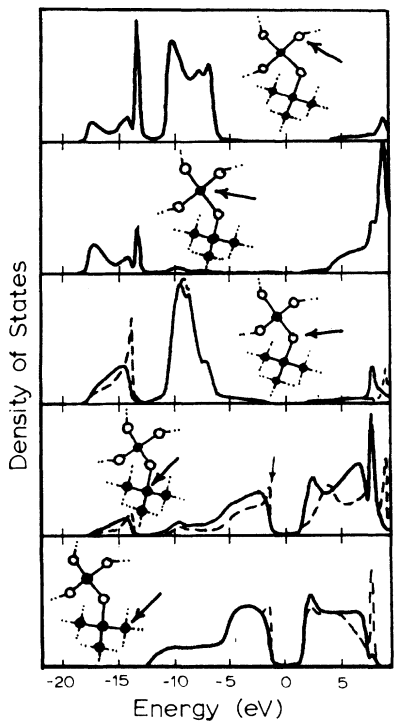

FIG. 3. Solid lines: local densities of states vs energy for a type-I ideal interface. The top frame refers to an oxygen atom in the oxide, the second frame to the atom below it, and so on, down through the interface and into the silicon. Dashed lines: identical calculation performed for an interface in which the self-energy of the atom in frame four is raised by 3.4 eV.

In Fig. 4, we show an ideal interface (type II) in which the surface atom is doubly coordinated with oxygen. This might, for example, be the predominant bonding configuration on silicon (100). There is very little qualitative difference between these results and those in Fig. 3. The strong-bonding resonance at −15 eV is broader than that in Fig. 3, due to the interaction of the two terminating bond orbitals with one another. The weak-bonding resonance is similar in the two figures. In Fig. 5, we have taken this evolution one step further and surrounded the interface atom on three sides by SiO_2 (type III). The two SiO_2 resonances are even less pronounced in Fig. 5, and there is a small peak at −2 eV in the center panel. A silicon atom at the interface can be bonded to SiO_2 either through an oxygen atom or through another silicon. Thus the interface of Fig. 5 may be thought of, like that of Fig. 3, as silicon (111) with SiO_2 attached. Attaching SiO_2 to (100) silicon in this latter manner is shown in Fig. 6 (type IV). The upper two panels of Fig. 6 are virtually indistinguishable from those of Fig. 5. The third panel, however, shows a pronounced resonance at −2 eV, which comes about in the following manner. As this atom is surrounded on four sides by silicon atoms, each of its hybrids

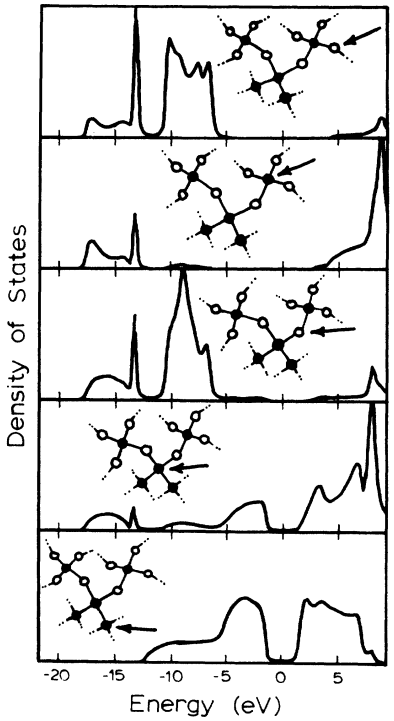

FIG. 4. Local densities of states vs energy for a type-II interface.

interacts with that of its neighbor to form a bonding-antibonding pair. These then split via the tetrahedral interaction into one s and three p states. When the atom is embedded in silicon, these four states broaden out into the silicon valence and conduction bands. When the atom is walled in by oxide, however, waves originating from the atom cannot propagate in two of the four available directions, and thus half as much broadening occurs. The conduction-band resonance at 4 eV is the antibonding analog of the resonance at

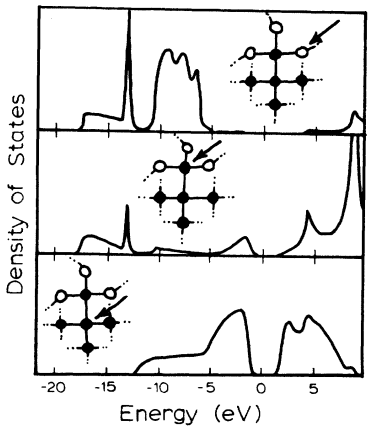

FIG. 5. Local densities of states vs energy for a type-III interface.

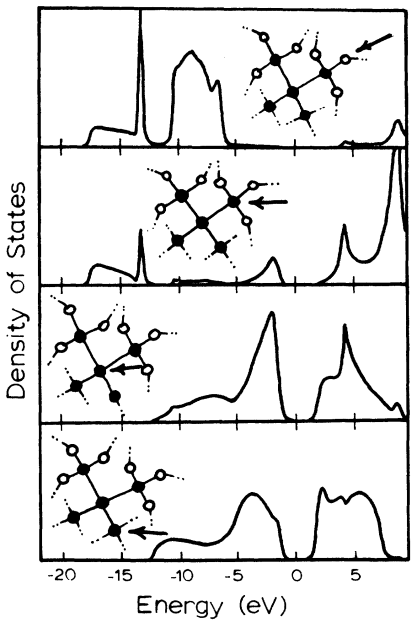

FIG. 6. Local densities of states vs energy for a type-IV interface. Note the characteristic resonances at −2 and 4.5 eV.

−2 eV. Both of these are partially broadened p states. The s-like resonance in the valence band lies near −7 eV. Since isolation of the interface atom is responsible for these resonances, one would expect further isolation to enhance the effect. This is shown to be the case in Fig. 7 (type V). In this arrangement, which is a peninsula of silicon

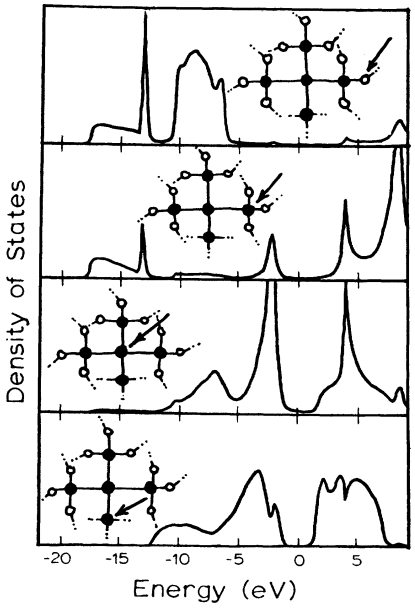

FIG. 7. Local densities of states vs energy for a type-V interface (silicon peninsula).

protruding into the oxide, the resonance at −2 eV has become extremely large and narrow, and the s-like resonance at −7 eV has become a well-defined peak.

The results displayed in Figs. 3–7 suggest that interface states do not arise simply from the presence of a boundary. The closest one can come to producing interface states with topography is the silicon peninsula, and even this produces only a deeply bound resonance.

V. DISORDER

The density of gap states at the interface, as determined from field-effect measurements, has a shape[33,34] reminiscent of that of hydrogenated amorphous silicon[35] (tails of states at the band edges, the one near the valence band significantly larger, possibly accompanied by midgap structure). Conductivity measurements[36] indicate that states deep in the gap are localized and there is some evidence[37] of a mobility edge.[38] A peak appearing near 0.35 eV above the valence-band edge has been observed[34] to disappear reversibly with hydrogen annealing and to have photoionization[39] properties distinct from those of the tail states.

There is currently great uncertainty[23] as to the origin of interface states. While the list of possible explanations is extensive, there appear to us to be two main categories—those attributing interface states to strain (for example, distortions, dislocations, or broken bonds) and those attributing them to charged centers. There are a number of reasons we favor the former. Firstly, similar behavior is seen in amorphous silicon,[35] in which charged centers presumably do not exist. Secondly, the tailing appears to be quite extensive, whereas the binding energy of donor or acceptor levels in silicon is the order of 0.05 eV. Although overlap of near-neighbor defect wave functions[33] could conceivably bond these to midgap, we find this unlikely in light of (i) the large density (10^{12} cm^{-2}) required to force the centers within one effective Bohr radius of one another, (ii) the fact that tailing occurs from both edges (indicating defect potentials of either sign), and (iii) the likelihood that binding energies are smaller than in bulk silicon due to the inability of the defect wave function to penetrate the oxide. We cite the report[40] of observation of a defect band 0.02 eV below the conduction-band edge for a surface density of 0.5×10^{12} cm^{-2} extrinsic Coulomb centers, compared to a theoretical estimate of 0.03 eV. Thirdly, while charged centers may exist at the interface, strain must exist, and it can account naturally for both band-tailing and midgap structure.

It is a simple procedure[10,41] to investigate the effects of strain within the framework of a tight-binding Hamiltonian. There are three reasons such calculations should be physically meaningful in this case: (i) Surface results show that tight binding correctly describes broken bonds,[10] the most serious distortion of all. (ii) Trends observed in pseudopotential calculations[13] for complicated allotropes of silicon having severely distorted bond angles are reproduced[19] in tight binding. (iii) Surface results indicate[8,10] that correlation energies in localized states should be small. This is corroborated by recent, independent,[42,43] self-consistent results for the silicon vacancy, in which a correlation energy of approximately 0.1 eV has been found. While polaronic effects could be significant, we feel that they are probably not, since the 2×1 reconstruction, which affects the surface states minimally[8,10] is presumably[9] a Jahn-Teller distortion and will not occur to lowest order when no degeneracy is lifted. Thus it is reasonable to assume that single-particle predictions are meaningful, and that any localized state will be to a good approximation both a donor and an acceptor.

Both bond-angle and bond-length distortions can produce localized states in the gap if they are sufficiently severe. For example, in Fig. 8 we have

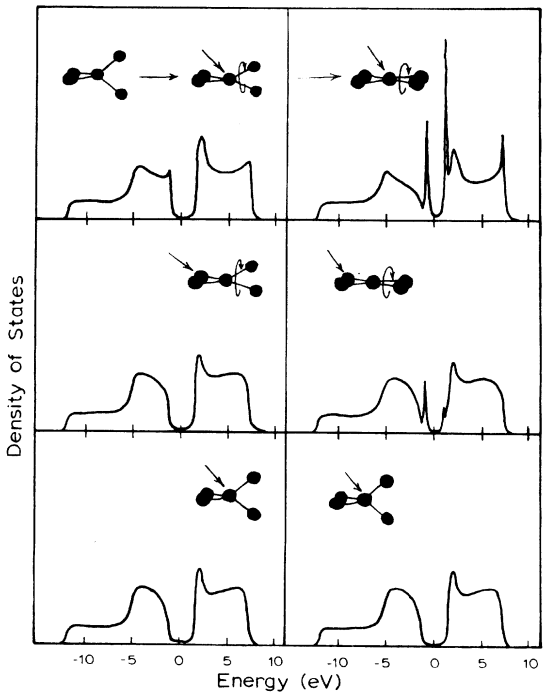

FIG. 8. Local densities of states vs energy for a distorted tetrahedron in the silicon Bethe lattice. Left and right columns refer to 30° and 60° twists, respectively. The top frame refers to the central atom, the middle frame to one of its neighbors, and the bottom frame to the bulk.

subjected one tetrahedron in the silicon Bethe lattice to "E" distortions[19] (which involve rotating one pair of bonds into the plane of the other so as to flatten the tetrahedron) of 30° and 60°. The localized states come about as follows. In a planar (90°) configuration, the $d(+-+-)$ combination of nearest-neighbor hybrids is prevented by symmetry from interacting with the atom at the center, and thus forms a state at the hybrid level. The p state normal to the bonding plane interacts via V_π with a symmetric combination of nearest-neighbor p orbitals to form bonding (1^- eV) and antibonding (7^- eV) levels. As the tetrahedron is untwisted, the gap states bond apart, merging into the bulk bands at about 50°. In Fig. 9, we have similarly subjected a bond in the Bethe lattice to stretches of 1 and 1.9 Å using the Pandey-Phillips rule[10] of scaling all nearest-neighbor interactions as $e^{-\Delta d/d_0}$. The states in the gap in this case arise from ordinary unbonding. As the atoms are brought together from infinity, the two dangling-bond states at the hybrid level bond slowly apart and eventually draw back into the valence and conduction bands.

Because of uncertainty in the repulsive part of the total energy,[41] it is difficult to estimate reliably whether bond-angle or bond-length distortions of this severity should be more prevalent. On one hand, ST-12, which exists in nature, has bond-angle distortions of up to 20° with virtually no bond-length distortion.[13] On the other hand, distorting an angle sufficiently to localize or state in the gap amounts to unbonding without changing the internuclear distances. As the total energy is the first moment of the density of states plus a short-range repulsion,[10] one could argue that *less* energy would be required to stretch a bond in severe cases. The likely situation in real interfaces is that both exist. Assuming this to be true, one would expect localized states in the gap consistent with the observed[33,34] U-shaped continuum of intrinsic interface states, in that (i) they occur at both band edges; (ii) they can penetrate deep into the gap; (iii) they should be more prevalent at the valence-band edge (*Small* bond-angle distortions are definitely[13] preferred over small bond-length distortions, i.e., as the distortion angle is increased, the valence-band state appears well before the conduction-band state); and (iv) half of the states should be insensitive to hydrogen annealing.[34] If a bond is stretched, interstitial hydrogen can bond with the symmetric state at the valence-band edge but not with the antisymmetric state. Likewise, if a tetrahedron is twisted, interstitial hydrogen can band with the dehybridized p state at the conduction-band edge but not with the d state. Thus, the presence of both kinds of strain is necessary to account for the annealing behavior.

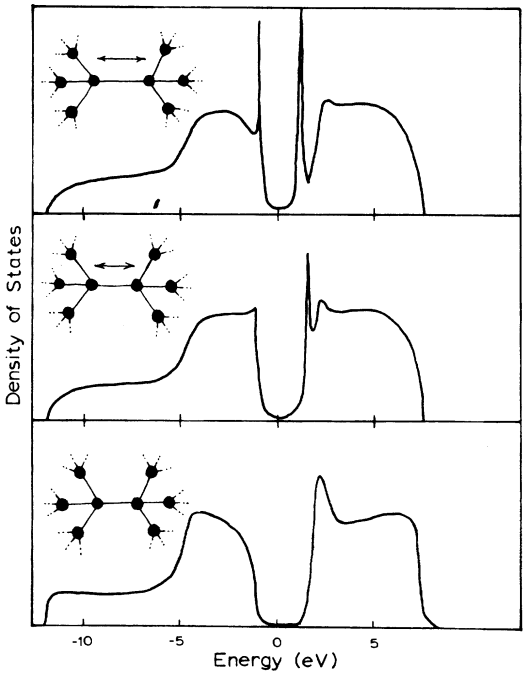

FIG. 9. Local densities of states vs energy for an atom with a stretched bond. Top: $d/d_0=1.80$; middle: $d/d_0=1.45$; bottom: bulk.

Dehybridizing distortions can also occur immediately at the interface and can have similar effects, as shown in Fig. 10. Here we have subjected the tetrahedral unit of a type-II interface to a 30° twist. The valence states show the same trends on the distorted atom as they do in Fig. 8, but the conduction states do not. This is mostly a result of the raised orbital self-energies we have used for this atom, but is also a consequence of undercoordination. In this case, the antisymmetric combination of nearest-neighbor hybrids is responsible for the state near the valence band, and the p orbital normal to the bonding plane interacts with two silicon neighbors, rather than four, to form a bonding-antibonding pair. Because of uncertainty in the orbital self-energies at the interface and the strength of bonding parameters, it is difficult to tell whether the conduction states in Fig. 10 are meaningful. However, the tendency of a state to appear at the valence-band edge is certainly correct.

Not all kinds of bond-angle distortion at the interface will produce localized states. Dihedral-angle changes, for example, are known[21] to have a minimal effect on the Bethe-lattice density of states. Also, in Fig. 11, we show a type-I interface in which the Si–O–Si angle at the boundary has been distorted ±10°. While noticeable changes

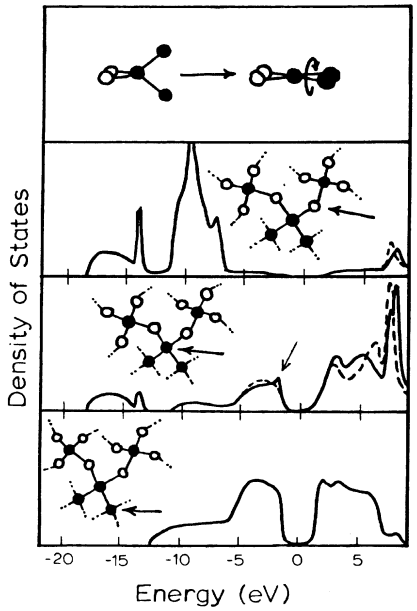

FIG. 10. Type-II interface in which the tetrahedron at the interface is twisted as shown by 30°. The interface resonance, which results, is marked with an arrow.

occur in the oxide valence bands, the region of the silicon gap remains relatively unperturbed. The largest differences occur near −13 eV. A decrease of the angle causes the strong-bonding and weak-bonding bond orbitals on the distorted site to draw together, thus producing resonances at the band edges.

We find it likely that dangling bonds, if they are present at the interface, should produce an inhomogeneously broadened discrete level at midgap. Dangling-bond states on silicon surfaces tend to be localized sufficiently so that their energies are influenced significantly only by fluctuations in the back bonds and the Hartree potential at the site. At the interface, the localization of such states should not invalidate the one-electron approximation, because as for localized states in the tails, we would expect small correlation and relaxation effects. Thus, each state should be both a donor and an acceptor to lowest order. The tight-binding prediction for a broken bond at a type-I interface is shown in Fig. 12. We emphasize that the energy of a dangling bond at the interface is difficult to estimate. A likely guess, however, would be that it lies near the center of the dangling-bond surface band on unrelaxed[8] silicon (111) (roughly 0.4 eV

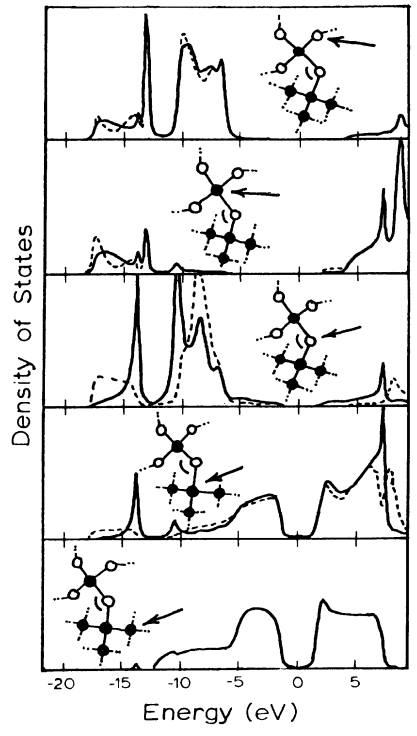

FIG. 11. Type-I interface in which the Si-O-Si angle marked in the figure is closed (solid line) or opened (dashed line) by 10°.

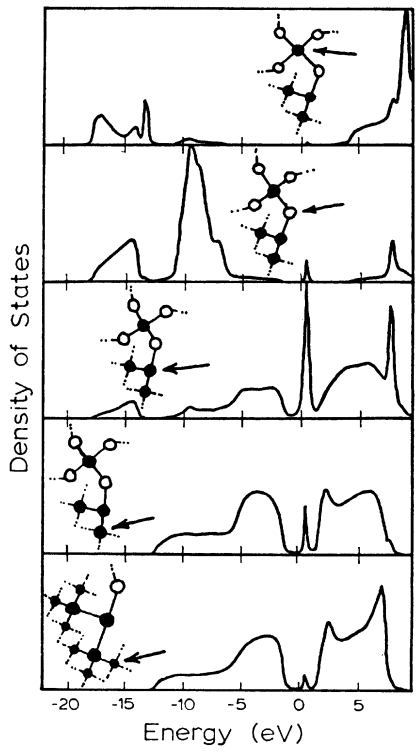

FIG. 12. Type-I interface in which a bond has been broken on the silicon side of the interface.

above the valence-band maximum). This guess correlates well with the discrete level at that energy reported[34] to disappear reversibly with hydrogen annealing. One serious objection to identifying this feature as a dangling bond is that no state at this energy is observed[35] in amorphous silicon. This does not conclusively rule out the dangling-bond explanation, however, because (i) the environment of the dangling bond at the interface is crystalline and thus might be more constrained and (ii) because of Fermi-level pinning, the density of gap states in amorphous silicon can be measured only in hydrogenated samples, in which dangling bonds may not exist in appreciable quantities. Dangling bonds deeper in the oxide are also potentially important, but because of the nature of screening in SiO_2 (ϵ_1 is about $\frac{1}{3}$ that of silicon and is due largely to phonons) we find it unlikely that a single-particle picture is applicable to them. Preliminary[44] indications are that dangling silicon and oxygen bonds should act as donors and acceptors, respectively, which might spontaneously ionize with the Fermi level pinned through mechanical contact with the silicon. However, considerably more work needs to be done before a reliable statement about oxide defect can be made.

VI. SUMMARY

The theory of the Si–SiO_2 interface presented in this paper is based on two fundamental approximations necessitated by the problem of lattice mismatch: (i) Tight-binding Hamiltonians fit to bulk silicon and SiO_2 can be interpolated across the interface and (ii) surface two dimensionality does not affect the density of states significantly, so that Bethe lattices may be used to model the silicon and SiO_2 structures. There are three major points of the theory:

(1) The interface dipole is consistent with ionization energies of neutral silicon and oxygen atoms. It thus appears as though the negative charge on the oxygen atoms in SiO_2 is largely compensated by positive silicon neighbors and by exchange.

(2) Interface states do not arise simply from the presence of a boundary.

(3) Strain, rather than charged centers, is the probable cause of the U-shaped continuum.[34]

We have, in addition, found the following: (a) Weak resonances can occur at the interface which have analogs in SiO_2 and silicon surfaces. (b) Interface topography can induce strong interface resonances. (c) Bond-angle fluctuations in SiO_2 do not make interface states near the silicon gap, although bond-angle fluctuations in the silicon can. Finally, we have made two important speculations: (i) There are no trapped holes in oxygen chemisorption, and (ii) the midgap state in metal-oxide-semiconductor devices is a dangling bond.

ACKNOWLEDGMENTS

This research was supported in part by the U. S. Navy, ONR Grant No. N0014-77-C-0132. Two of us (R.B.L. and J.D.J.) would like to acknowledge receipts of IBM and Alfred P. Sloan Fellowships, respectively.

[1]D. E. Eastman and W. D. Grobman, Phys. Rev. Lett. 28, 1378 (1972).
[2]L. F. Wagner and W. E. Spicer, Phys. Rev. Lett. 28, 1381 (1972).
[3]J. E. Rowe and H. Ibach, Phys. Rev. Lett. 32, 421 (1974).
[4]J. E. Rowe and H. Ibach, Phys. Rev. Lett. 31, 102 (1973).
[5]D. Haneman, Phys. Rev. 170, 705 (1968).
[6]G. Chiarotti, S. Nannarone, R. Pastore, and P. Chiaradia, Phys. Rev. B 4, 3398 (1971).
[7]J. A. Appelbaum and D. R. Hamann, Phys. Rev. Lett. 31, 106 (1973); 32, 1433 (1974).
[8]M. Schlüter, J. R. Chelikowsky, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. 34, 1385 (1975); Phys. Rev. B 12, 4200 (1975).
[9]M. Schluter, *Physics of Semiconductors* (Tipografia Marves, Rome, 1976), p. 646.
[10]K. C. Pandey and J. C. Phillips, Phys. Rev. Lett. 32, 1433 (1974).
[11]H. Ibach and J. E. Rowe, Phys. Rev. B 10, 710 (1974).
[12]F. Herman and I. P. Batra, *The Physics of SiO_2 and its Interfaces* (Pergamon, New York, 1978), p. 333.
[13]J. D. Joannopoulos and M. L. Cohen, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1976), Vol. 31.
[14]W. A. Goddard III, A. Redondo, and T. C. McGill, Solid State Commun. 18, 981 (1976).
[15]J. D. Joannopoulos, J. Non-Cryst. Solids 32, 241 (1979).
[16]C. Domb, Adv. Phys. 9, 145 (1960).
[17]L. Ley, S. Kowalczyk, R. Pollak, and D. Shirley, Phys. Rev. Lett. 29, 1088 (1972).
[18]B. Fischer, R. A. Pollak, T. H. DiStefano, and W. D. Grobman, Phys. Rev. B 15, 3193 (1977).
[19]J. D. Joannopoulos, Phys. Rev. B 16, 2764 (1977).
[20]R. B. Laughlin, J. D. Joannopoulos, and D. J. Chadi, Phys. Rev. B 20, 5228 (1979).
[21]R. B. Laughlin and J. D. Joannopoulos, Phys. Rev. B 16, 2942 (1977); 17, 2490 (1978).
[22]T. L. Gilbert, W. J. Stevens, H. Schrenk, M. Yoshimine, and P. S. Bagus, Phys. Rev. B 8, 5988 (1973).
[23]Y. C. Cheng, Prog. Surf. Sci. 8, 181 (1977).
[24]T. H. DiStefano, J. Vac. Sci. Technol. 13, 856 (1976).

[25]C. E. Moore, *Atomic Energy Levels*, National Bureau of Standards, Circular No. 467 (U.S. Government Printing Office, Washington, D.C., 1949).
[26]E. J. Mele and J. D. Joannopoulos, Surf. Sci. 66, 38 (1974).
[27]I. P. Batra, Proceedings of the 6th PCSI (to be published).
[28]J. E. Rowe, G. Margaritondo, H. Ibach, and H. Froitzheim, Solid State Commun. 20, 277 (1976), and references therein.
[29]C. M. Garner, I. Lindau, C. Y. Su, P. Pianetta, and W. E. Spicer, Phys. Rev. B 19, 3944 (1979).
[30]R. S. Bauer, J. C. McMenamin, R. Z. Bachrach, A. Bianconi, L. Johansson, and H. Petersen, *Physics of Semiconductors* (Institute of Physics, London, 1978), p. 797.
[31]E. J. Mele and J. D. Joannopoulos, Phys. Rev. B 18, 6999 (1978).
[32]M. Schlüter reports that the self-consistent calculation for the reconstructed surface was very unstable.
[33]A. Goetzberger, E. Klausmann, and M. J. Schulz, CRC Critical Review 1, (1976).
[34]N. M. Johnson, D. J. Bartelink, and M. Schulz, Ref. 12, p. 421.
[35]A. Madan, P. LeComber, and W. E. Spear, J. Non-Cryst. Solids 20, 239 (1976).
[36]F. F. Fang and A. B. Fowler, Phys. Rev. 169, 619 (1968).
[37]F. Stern, Phys. Rev. B 9, 2762 (1974).
[38]N. F. Mott, Adv. Phys. 16, 49 (1967).
[39]E. Kamieniecki and R. Nitecki, Ref. 12, p. 417.
[40]A. Hartstein and A. B. Fowler, Surf. Sci. 73, 19 (1978).
[41]D. J. Chadi, Phys. Rev. Lett. 41, 1062 (1978).
[42]S. T. Pantelides, private communication.
[43]M. Schlüter, private communication.
[44]R. B. Laughlin, J. D. Joannopoulos, and D. J. Chadi, Ref. 12, p. 325.