Theoretical study of the Si(100) surface reconstruction

A. Ramstad, G. Brocks, and P. J. Kelly
Philips Research Laboratories, 5656 AA Eindhoven, The Netherlands
(Received 30 August 1994)

We present the results of a systematic study of the reconstruction of the Si(100) surface based upon total energies calculated within the framework of the local density approximation. We focus on the extent to which total energy *differences* may be calculated reliably by examining these differences for the ideal surface and four proposed reconstructions: $p(2 \times 1)$ symmetric, $p(2 \times 1)$ asymmetric, $p(2 \times 2)$, and $c(4 \times 2)$. The calculations were performed using norm-conserving pseudopotentials and a plane-wave basis. The convergence of the total energy differences was assessed by varying the energy cutoff used to truncate the plane-wave basis and the number of sampling points used to perform Brillouin zone (BZ) integrals over a large range. The effect of optimizing atomic geometries as a function of the energy cutoff and density of BZ sampling points was determined. With the exception of the $p(2 \times 2)$ and $c(4 \times 2)$ reconstructions, whose energies only differ by 3 meV per dimer, we are able to unambiguously determine the energy ordering of the five systems studied. Disagreements between previous calculations can be largely understood in terms of the different energy cutoffs and BZ samplings used. The electronic structures of the different reconstructions are calculated and compared.

I. INTRODUCTION

In recent work we have addressed the adsorption and subsequent migration of single adatoms on the clean Si(100) surface as well as the energetics of pairs of adatoms on the same surface.[1–3] These studies were based on first-principles calculations of total energies where norm-conserving pseudopotentials were used to represent the interaction of the valence electrons with the ion cores and the wave functions were expanded in plane waves. In such calculations it is rarely possible to obtain absolute values of the energy accurately and the usefulness of the total energy approach depends on there being a systematic cancellation of errors for energies calculated for similar systems so that total energy *differences* can be determined reliably. The starting point for our adsorption studies was the reconstruction of the ideal surface, which had been investigated extensively both experimentally[4] and theoretically.[5–21] In spite of all this attention, there was no consensus in the literature as to the nature of the lowest energy reconstruction. A number of different proposed reconstructions for the Si(100) (shown schematically in Fig. 1) lead to relatively small surface unit cells and have atomic configurations which are closely related; this should have made the theoretical determination of the corresponding energy differences straightforward. The lack of agreement between different calculations, however, led us to conclude that the systematic cancellation of errors was not as systematic as had been widely assumed. In order to determine to what extent total energy differences could be calculated using the standard machinery of the local-density approximation (LDA), norm-conserving pseudopotentials, repeated slabs, and plane-wave bases, we performed a systematic study of a number of these reconstructions [ideal, $p(2\times1)$ symmetric, $p(2 \times 1)$ asymmetric, and $p(2 \times 2)$]. The $p(2 \times 2)$ symmetry reconstruction was found to have the lowest energy[1] and formed the basis for our studies of adsorption. Since then, we have extended this unpublished work to include reconstructions with $c(4 \times 2)$ symmetry and in the following we will present the results of this extensive study of the influence of a number of different factors on the energy *differences* between the four different reconstructions shown in Fig. 1. Particular attention will be devoted to the influence of the number of plane waves used in the expansion of the wave functions and to the number of sampling points used in integrals over the Brillouin zone (BZ). Frequently, geometries are optimized using a limited basis set and BZ sampling. Improved energies are then calculated by increasing the size of the basis set and/or the number of sampling points but without reoptimizing the geometry. The assumption is made that total energy differences would not change much if the geometries were relaxed. Sometimes a frozen-potential or "force-theorem" approach is used so as to avoid reoptimizing the charge density. We examine the validity of a number of such approximations.

Silicon surfaces have been studied in great detail over the past three decades with a large variety of experimental techniques.[4] The (100) surface has received particular attention, partly because most silicon devices are formed on this surface and partly because its reconstructions are simple compared to those of other surfaces. Much of the interest has focused on understanding such basic properties of the clean surface as how growth occurs, how foreign atoms are adsorbed, etc. Our own interest in the Si(100) stems from growth studies performed using scanning tunneling microscopy both in our own[22] and in other laboratories.[23]

In a low energy electron diffraction (LEED) experiment in 1959, Schlier and Farnsworth[24] observed a (2×1) periodicity on the Si(100) surface. They argued that

0163-1829/95/51(20)/14504(20)/$06.00

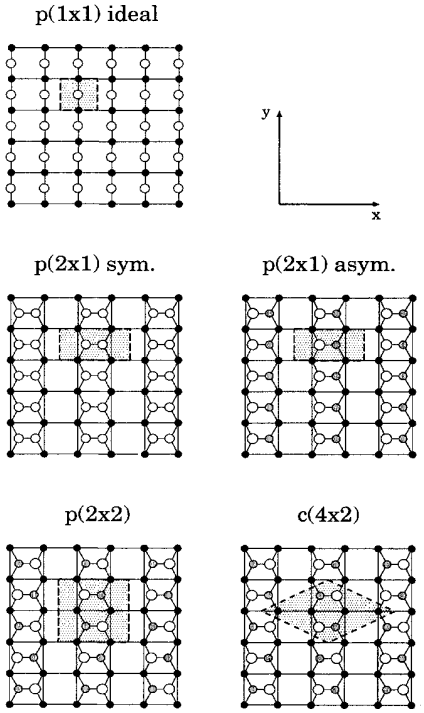

FIG. 1. Top view of the unreconstructed (ideal) surface and the four reconstructions to be considered in this study. Smaller and darker circles represent deeper atoms. Thus the small black circles are second layer atoms and the larger grey and white circles are surface atoms. In the buckled dimer reconstructions the large white circles protrude further out of the surface than the grey circles. The dashed lines and shaded areas show the surface unit cells used in our calculations.

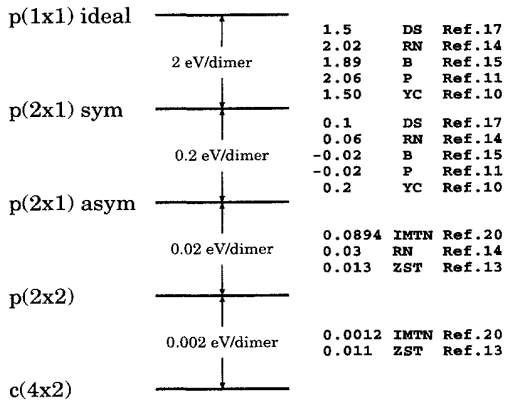

FIG. 2. A schematic illustration of the order of magnitude of energy differences between different reconstructions of the Si(100). On the right-hand side we summarize the results of a number of calculations of the corresponding total energy differences.

this was caused by surface atoms moving together in pairs to form dimers. The formation of surface dimers would remove one of the two dangling bonds per surface atom without necessarily changing any bond lengths by more than a few percent. Dimerization as the basic reconstruction has been used in interpreting LEED,[25] photoemission,[26] ion channeling,[27] optical absorption,[28] electron energy loss,[29] and core-level spectroscopy[30] experiments. Dimers have been observed directly by scanning tunneling microscopy (STM).[31,32] This basic reconstruction lowers[10,11,14,15,17] the energy of the surface by approximately 2 eV per dimer; see Fig. 2.

One of the first calculations for the dimerized surface was that of Appelbaum *et al.*[6] In their model, the dimer bond length was slightly shorter than the bulk bond length. They also showed that the dimerization is accompanied by significant subsurface distortion extending 4–5 layers into the bulk.[7] Their model predicts a metallic surface, whereas experimentally[33] the surface is semiconducting. Chadi[8] then proposed an asymmetric dimer model, where the dimers buckle out of the plane of the surface. On the basis of an empirical tight-binding calculation, he predicted that such a surface reconstruction should be semiconducting. Support for dimer buckling came from ion scattering measurements,[27,34–36] grazing incidence x-ray diffraction,[37] transmission electron diffraction,[38] and from STM experiments.[31,32] After a period of uncertainty where different theoretical calculations supported both symmetric[11,15] and asymmetric dimers,[10] all recent total energy calculations[14,17–19] indicate that buckled dimers are more favorable than symmetric ones by typically 0.1 eV per dimer.

Further complications in the understanding of the surface structure came with the discovery of $p(2 \times 2)$ (observed by STM[31,32]) and $c(4 \times 2)$ [evidenced by LEED[39] and angle-resolved ultraviolet photoemission spectroscopy (ARUPS)[40] at low temperatures] surface periodicities. Since the dimer buckling can have two orientations, these reconstructions can be explained in terms of different arrangements of alternately buckled dimers. Distinguishing between $p(2 \times 1)$ asymmetric buckled dimers and alternately buckled dimers requires going down yet another order of magnitude on the energy scale (Fig. 2). Roberts and Needs[14] calculated the energy difference between the $p(2 \times 2)$ and the $p(2 \times 1)$a (asymmetric) reconstruction to be 0.03 eV per dimer, favoring the $p(2 \times 2)$. Inoue *et al.*[20] found this difference to be 0.09 eV per dimer.

The smallest energy difference of all is that between the $p(2 \times 2)$ and the $c(4 \times 2)$ reconstructions. This was also calculated by Inoue *et al.*,[20] who reported an energy difference of about 1 meV favoring the $c(4 \times 2)$.

The spread in results shown in Fig. 2 is very considerable. One of the main issues that we want to address is the origin of the discrepancies between different calculations and to indicate globally what needs to be done in order to calculate various energy differences with some degree of reliability. We will conclude that the present state of the art is not capable of resolving energy differences of the order of 1 meV/surface atom and that it is not possible to resolve the energy difference between the $p(2 \times 2)$ and the $c(4 \times 2)$ without improving the accuracy of the best calculations carried out so far by about an order of magnitude. We identify a number of aspects of

the calculations which will have to be improved before it can be decided what the local-density approximation yields for the ground state of the Si(100) surface.

The paper is organized as follows. In Sec. II we describe briefly the methods used in our calculation. A detailed investigation of the accuracy of the calculations is performed in Sec. III. Our results are presented and discussed in Sec. IV. In Sec. V we discuss a number of ways in which the calculations could be extended and some conclusions are drawn in Sec. VI.

II. DESCRIPTION OF THE METHOD

The framework for our calculations is the density-functional theory with the local-density approximation[41,42] for the exchange and correlation. In all of the calculations which will be discussed below, the results of Ceperley and Alder,[43] as parametrized by Perdew and Zunger,[44] are used. The interaction of the valence electrons with the nucleus and the core electrons is treated by using norm-conserving pseudopotentials[45] with Kleinman-Bylander nonlocality including s and p terms.[46] In order to assess the behavior of this pseudopotential we have calculated a number of bulk properties, the results of which are discussed in the Appendix. To relax the electronic degrees of freedom, a conjugate gradient minimization of the energy functional[47] is used. The ionic degrees of freedom are relaxed with a similar algorithm and, as in the Car-Parrinello scheme,[48] ions and electrons are treated simultaneously. A supercell containing a slab of silicon 12 layers thick and the equivalent of six atomic layers of vacuum (= 9.5 Å) is used to represent the surface region. The five outermost layers on each side of the slab are allowed to relax, while the atoms of the two central layers are kept in their bulk positions. We make use of inversion symmetry about the center of the slab, keeping the two surfaces identical and thus avoiding instabilities such as charge transfer from one surface to the other. The experimental bulk lattice constant of 5.43 Å is used throughout. Our calculated bulk lattice constant of 5.37 Å is only very slightly smaller. In cases where the discrepancy between theory and experiment is larger, there is no consensus as to whether it is better to use the theoretical or the experimental lattice constant and both choices are found in the literature. Wave functions are expanded in a basis set comprising plane waves with kinetic energy up to a certain cutoff. Eventually we will choose a cutoff of 16 Ry to carry out all geometry optimizations but the results of tests using cutoffs as large as 24 Ry will be presented. An important result of our calculations is that all the dimer reconstructions are found to have a mirror plane containing the dimer bonds (the xz plane in Fig. 1). With time reversal symmetry, the irreducible part of the Brillouin zone then becomes 1/4 of the full BZ. In order to study how calculated energies converge as the number of points used to evaluate BZ integrals is increased, we will use a wide range of BZ sampling densities. It is advantageous if the sampling grid can be made finer by adding new **k** points while keeping the old ones from the coarser grid. This is possible if we use the sets of points shown in Fig. 3 rather than following the more common procedure of Monkhorst and Pack.[49]

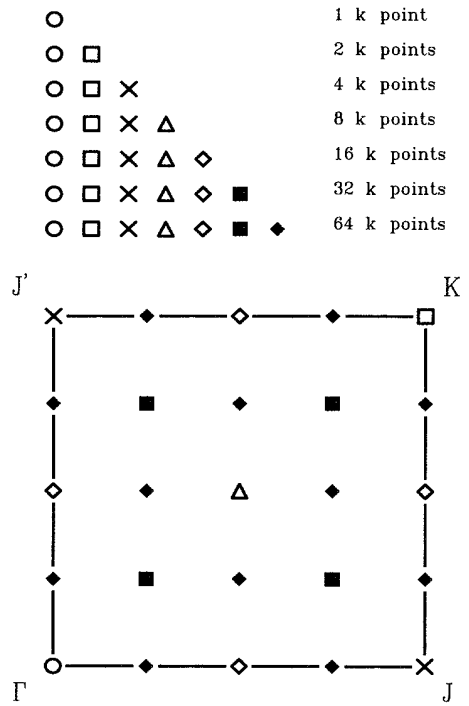

FIG. 3. A quarter part of the $p(2 \times 2)$ Brillouin zone showing the **k** points used in the various grids. The coarsest grid contains only the point marked with a circle. To double the density, the point marked with an open square is added. Another doubling is achieved by also including the points marked by ×'s, and so on. The numbers of **k** points refer to the full $p(2 \times 2)$ zone.

III. CONVERGENCE TESTS

As mentioned in the Introduction, numerous calculations have been carried out for the Si(100) surface. The spread in the reported results is very large; see Fig. 2. All of the calculations referred to in the figure are first-principles calculations that contain no parameters. Furthermore, they are all density-functional calculations based on norm-conserving pseudopotentials and plane-wave basis sets. The discrepancies must then be accounted for either by a difference such as in the choice of exchange-correlation potential or by one or more of the restrictions which must be imposed in order to limit the size of the calculations. Factors that will affect the results include the basis set cutoff energy, the number of **k** points used in the Brillouin zone summation, the slab thickness, the number of layers in the slab that are allowed to relax, the vacuum spacing, and the pseudopotentials. In principle, one could simply use a very high kinetic energy cutoff together with a large number of **k** points in the Brillouin zone, a very thick slab, and a large separation between

neighboring slabs, etc. In practice, this is of course not possible and the usual procedure is to optimize the geometry with some plane-wave cutoff, BZ sampling, number of layers in the slab, etc., and to check these parameters in a manner which is more often than not dictated by the amount of computer time and storage available. The cutoffs and **k** point samplings used for the optimization of the ionic and electronic degrees of freedom are frequently different. The choice of norm-conserving pseudopotential and the form of the exchange and correlation potential will also influence the results. None of the commonly used exchange-correlation potentials gives systematically better agreement with experiment.

The two factors which vary most in earlier plane-wave pseudopotential calculations and which are most likely to be responsible for the discrepancies between them are the cutoff energy and the number of **k** points used in the BZ sampling. In the present section we will pay particular attention to finding out how energy *differences* depend on these two factors. In Sec. III A we will perform calculations where only the cutoff energy is varied and the **k** point sampling is kept fixed. This will show how the energy differences converge with increasing cutoff. The same procedure will be applied to the **k** point sampling in Sec. III B. In Sec. III C we address the question of whether the convergence with respect to **k** point sampling and cutoff energy can be treated independently. In Sec. III D we will briefly mention the effect on total energy differences of varying a number of other parameters such as the slab thickness, vacuum spacing, etc.

The surfaces which we will study are the ideal $p(1\times1)$ surface and the $p(2\times1)$s, $p(2\times1)$a, $p(2\times2)$, and $c(4\times2)$ reconstructions (see Fig. 1). The starting point for the convergence tests to be described are atomic configurations which were obtained by optimizing a geometry for each reconstruction with a kinetic energy cutoff of 16 Ry and a sampling density of four **k** points in the full $p(2\times2)$ Brillouin zone.[50] In all energy minimizations, we iterated the conjugate gradient algorithms until the total energy had converged to within 0.1 meV per dimer and the forces were as small as 2 meV/Å.

A. Plane-wave basis cutoff energy

The convergence of total energy differences is expected to be better than the convergence of the individual energies. Owing to the variational principle, any kind of optimization of the wave functions will lower the electronic energies and any geometry optimization will also, by definition, lower the total energy. Thus in calculating relative energies there will be a cancellation which is expected to be more complete if the two systems being compared are more similar. The cancellation will be particularly large when the optimization is an increase in the cutoff energy used to select the plane waves in the basis. This is because increasing the cutoff means adding plane waves with higher kinetic energies and shorter wavelengths. Thus increasing the cutoff leads to an increased sampling density in real space and thus to an improved description of the short length-scale features of the wave functions which are most important in the core regions. Since bonding characteristics are independent of the wave functions close to the core, improving these will lower the total energy by the same amount for all geometries.

In this section we will examine the effect that the choice of basis set cutoff energy has on total energies and energy differences. This cutoff is a factor which varies a great deal from one publication to the next, and there is an obvious trend towards higher cutoffs as more computer power and storage becomes available. Our final results will be calculated with a cutoff of 16 Ry, which is, to the best of our knowledge, the highest cutoff used for the Si(100) surface so far. We will show that this is still not high enough to resolve energy differences of a few meV; for this an energy cutoff in excess of 24 Ry would be required. (The precise choice depends on the pseudopotential used.) Geometries are frequently optimized with a relatively low cutoff energy, under the assumption that further changes in geometry resulting from an increased cutoff will translate into much smaller energy differences when two similar configurations are compared. We will examine the validity of this assumption explicitly by first looking at the convergence as a function of cutoff energy keeping the ionic positions fixed in Sec. III A 1. Then, in Sec. III A 2, the geometry will be optimized as a function of the cutoff energy and the resulting energy differences compared.

1. Frozen geometries

We calculate the total energy for the reconstructions sketched in Fig. 1 using values of the cutoff energy ranging from 8 to 24 Ry. These calculations are carried out with the geometries which were obtained by energy minimization using a 16 Ry cutoff and four **k** points. It should be noted that some of the energies will change substantially when the number of **k** points is increased and the results to be presented here only represent an intermediate step.

The total energy is shown as a function of the cutoff energy in Fig. 4 for the $p(2\times1)$ symmetric and the $p(2\times1)$ asymmetric reconstructions. We see that using a 24 Ry cutoff is only good enough to achieve an accuracy of about 0.5 eV. The energy *difference* shown in Fig. 5 is seen to converge much better and is within 70 meV of the converged value when a cutoff of 8 Ry is used. This is more than two orders of magnitude better than the convergence of the absolute energies. The same is true for a cutoff of 24 Ry where the convergence of the energy difference is $\sim$ 1 meV. The cutoff dependence of the other energy differences shown in Fig. 2 are shown in Figs. 6–8. Because the $p(2\times1)$ symmetric dimer reconstruction is frequently used as a reference, we also show explicitly the energy difference between it and the $p(2\times2)$ reconstruction in Fig. 9. The squares joined by solid lines in Figs. 5–9 show the convergence of the energy differences calculated using frozen geometries.

A number of observations can be made about the figures. First, Fig. 4 and Figs. 5–9 make explicit the systematic cancellation of errors. Between 8 and 24 Ry, the

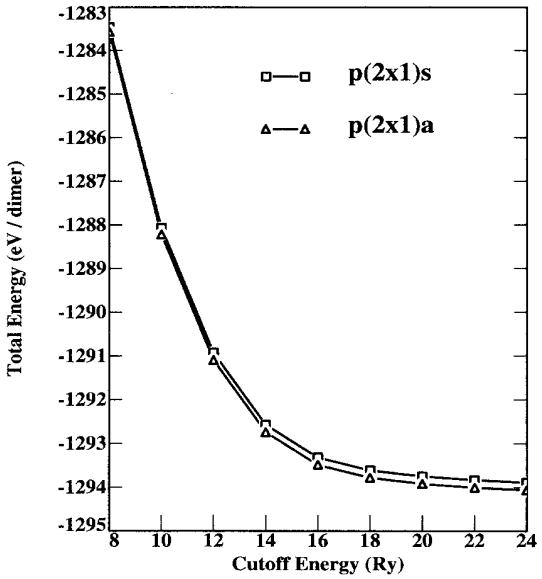

FIG. 4. Total energy as a function of the cutoff for the $p(2\times1)$s (square markers) and the $p(2\times1)$a (triangular markers). In each case the energies were calculated for a fixed geometry which was determined by energy minimization using four **k** points and a 16 Ry plane-wave cutoff. All calculations were carried out with a sampling density of four **k** points in the full $p(2\times2)$ Brillouin zone.

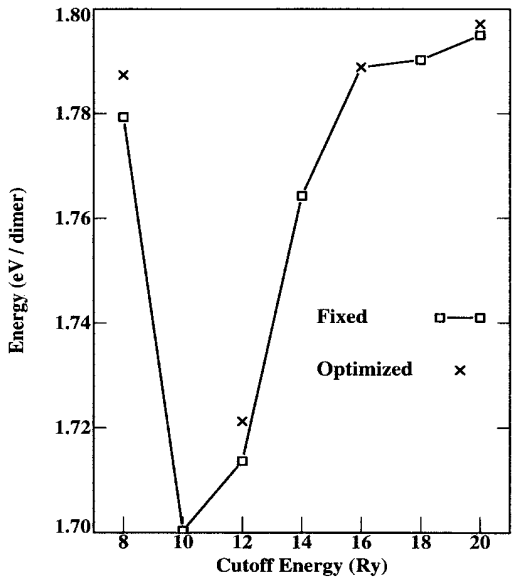

FIG. 6. Convergence of the dimerization energy $E_{\text{Ideal}} - E_{p(2\times1)\text{s}}$ as a function of the cutoff energy. The squares joined by the solid line are the results of calculations performed with fixed (16 Ry) geometries; the $\times$'s denote calculations where the geometry of the $p(2\times1)$s has been optimized at the given cutoff. A dimerization energy of 1.92 eV at 6 Ry is not plotted since it would dominate the figure. All calculations were carried out with a sampling density of four **k** points in the full $p(2\times2)$ Brillouin zone.

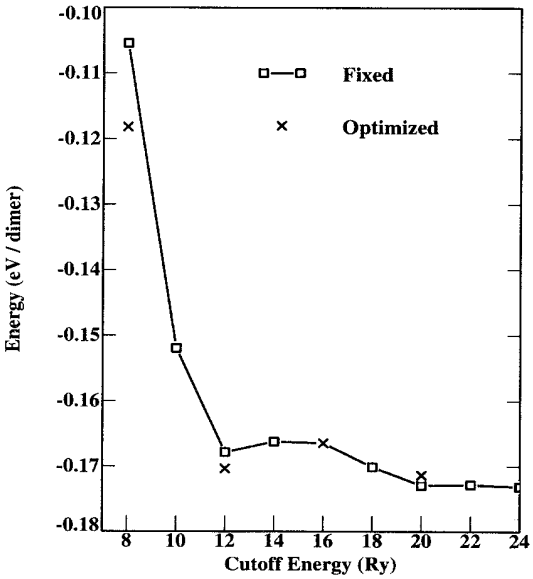

FIG. 5. Convergence of $E_{p(2\times1)\text{a}} - E_{p(2\times1)\text{s}}$ as a function of cutoff. Squares joined by the solid line are the results of calculations performed with fixed (16 Ry) geometries; the $\times$'s represent calculations with optimized geometries. All calculations were carried out with a sampling density of four **k** points in the full $p(2\times2)$ Brillouin zone.

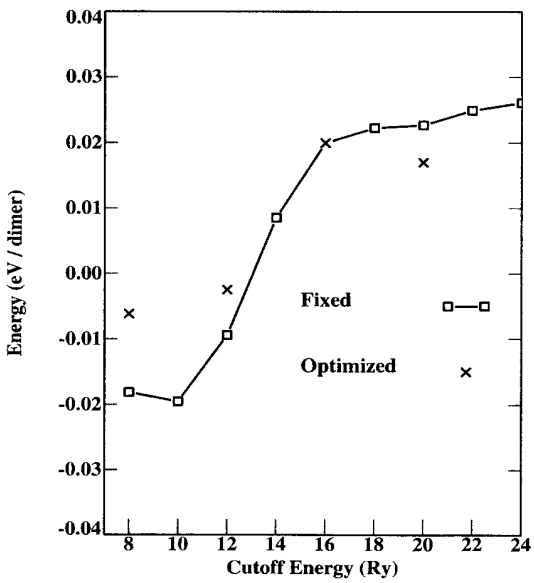

FIG. 7. Convergence of $E_{p(2\times2)} - E_{p(2\times1)\text{a}}$ as a function of cutoff. Squares joined by the solid line are the results of calculations performed with fixed (16 Ry) geometries; the $\times$'s represent calculations with optimized geometries. All calculations were carried out with a sampling density of four **k** points in the full $p(2\times2)$ Brillouin zone.

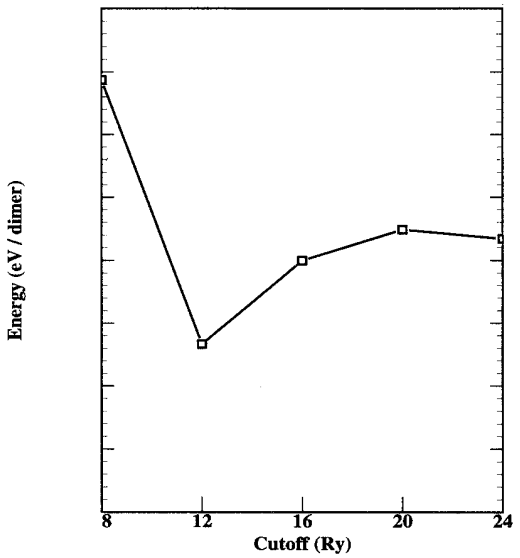

FIG. 8. Convergence of $E_{c(4\times2)} - E_{p(2\times2)}$ as a function of cutoff. All calculations are performed with fixed 16 Ry geometries. Because the four **k**-point sets used for the two reconstructions are not the same, the absolute values of the energy differences are not meaningful. Here we are only interested in seeing how the energy difference converges as a function of the cutoff energy. We therefore omit an absolute energy scale on the y axis. The separation between two large tick marks corresponds to 10 meV.

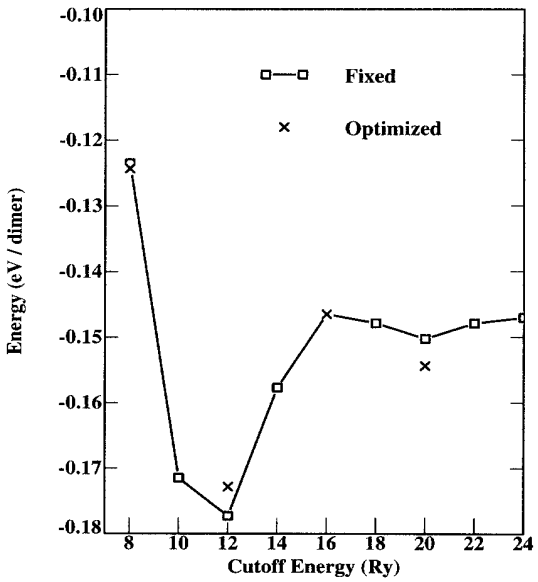

FIG. 9. Convergence of $E_{p(2\times2)} - E_{p(2\times1)s}$. Squares joined by the solid line are the results of calculations performed with fixed (16 Ry) geometries; the ×'s represent calculations with optimized geometries. All calculations were carried out with a sampling density of four **k** points in the full $p(2\times2)$ Brillouin zone.

total energies change by almost 10 eV; the energy differences on the other hand, change by at most 100 meV. Thus the systematic cancellation of errors amounts to an improvement in the calculation of energy differences by roughly a factor of 100. Second, to obtain convergence of an energy difference to within 1 meV, a cutoff of 24 Ry or higher is needed. This would require a very substantial computational effort if the number of **k** points which is needed for 1 meV accuracy were also used. We will return to this point later. While a cutoff of 12 Ry is sufficient to converge the energy difference between the $p(2\times1)$s and the $p(2\times1)$a to better than 10 meV (see Fig. 5) this is clearly seen not to be true in general (see Figs. 6, 7, and 9). To achieve a convergence of 10 meV per dimer for energy differences, a cutoff of at least 16 Ry seems to be required. Third, we see that whereas the total energy must decrease monotonically as the cutoff is increased because of the variational principle, this does not apply to total energy differences which can display an oscillatory behavior. The results show that because of the oscillatory behavior, calculations with a substantially lower cutoff can result in energy differences very close to the converged values. Unfortunately this must be regarded as being purely accidental and there is no way of knowing beforehand whether or not it will occur.

2. Relaxed geometries

All the above results were calculated using fixed geometries, obtained from a 16 Ry calculation. Optimizing the geometries at each cutoff will lower the energies and when energy differences are calculated, there will again be a partial cancellation. To investigate this effect, we optimized geometries for cutoffs of 8, 12, 16, and 20 Ry for the $p(2\times1)$s, $p(2\times1)$a, and $p(2\times2)$ reconstructions. The results of these calculations are shown as ×'s in Figs. 5–9. By construction the crosses coincide with the solid curve at 16 Ry. We see that qualitatively, the convergence shows the same characteristics with or without geometry optimization. Compared to the calculations with frozen geometries, there is no systematic variation; the crosses lie above the solid line for one cutoff and below it for another. There is also no indication that the correction to the energy differences on relaxing the geometries decreases rapidly with increasing cutoff. The deviations of the crosses from the solid curve at 12 Ry and 20 Ry have opposite signs but the same magnitude. The magnitude of the change is about 5 meV and this is the measure we will use of the error incurred by keeping the geometries fixed.

B. Brillouin zone sampling

The sampling densities used in previous *ab initio* calculations for the Si(100) surface vary from a few to 64 **k** points in the (2×2) BZ. We now present the results of calculations with samplings in this range. This will give us an idea of how large a part of the discrepancies between earlier published calculations is due to the dif-

ferences in the **k** point samplings. For every given **k** point sampling, the energy should be reoptimized by relaxing the charge density and the geometry. This procedure has the disadvantage of masking the different effects of geometry optimization and charge density optimization. To separate out the different effects, we will begin by performing calculations using a variety of different **k** point samplings, keeping the charge density and the ionic positions fixed. This will show how energy differences converge with increasing sampling densities. After this we will study the effect of optimizing the charge density when the **k** point sampling is increased and examine the extent to which frozen-potential or force-theorem type approximations[51–54] can be used. Lastly, we will examine the importance of reoptimizing the geometry for different BZ samplings.

1. *Frozen potentials*

When, for a given cutoff and **k** point sampling, self-consistency is reached, the total energy is

$$E[\rho,\{\mathbf{R}_I\}] = \sum_{i\mathbf{k}} \varepsilon_{i\mathbf{k}} - E_H[\rho(\mathbf{r})] + E_{\rm xc}[\rho(\mathbf{r})] - \int d\mathbf{r}\,\rho(\mathbf{r})V_{\rm xc}[\rho(\mathbf{r})] + E_{II}, \quad (1)$$

where $\varepsilon_{i\mathbf{k}}$ is the Kohn-Sham eigenvalue in the ith band at the point **k** in reciprocal space, E_H is the Hartree energy, $E_{\rm xc}$ is the exchange-correlation energy, $V_{\rm xc}$ is the exchange-correlation potential, and E_{II} is the ion-ion interaction energy. The total energy is thus a functional of the charge density $\rho(\mathbf{r})$ and the set of ionic coordinates $\{\mathbf{R}_I\}$. In general, the ρ and $\{\mathbf{R}_I\}$ which minimize the energy functional will depend on the **k** point sampling. In this section we are interested in the effect of the **k** point sampling on the total energy when the charge density and the ionic coordinates are kept fixed. In this case the expression for the total energy reduces to

$$E[\rho^0,\{\mathbf{R}_I^0\}] = \sum_{i\mathbf{k}} \varepsilon_{i\mathbf{k}} + \text{const}, \quad (2)$$

where ρ^0 and $\{\mathbf{R}_I^0\}$ might have been obtained from some self-consistent calculation employing a coarser **k** point sampling. Thus, for the moment we neglect all self-consistency cycling and simply use the potential defined by some ρ^0 and $\{\mathbf{R}_I^0\}$ to generate eigenvalues at new **k** points. The effect of letting the charge density relax will be studied in the next section.

We now take the ρ^0 and $\{\mathbf{R}_I^0\}$ from our 16 Ry, 4 **k**-point self-consistent calculations and using the expression (2) to calculate the energy, double the **k** point sampling three times, reaching a final sampling of 32 **k** points in the full (2×2) Brillouin zone. Figure 10 shows the convergence of the energies of the $p(2\times 1)$a, $p(2\times 2)$, and $c(4\times 2)$ relative to the energy of the $p(2\times 1)$s as a function of the number of **k** points.[55] We see that with 32 **k** points the energy difference between the $p(2\times 1)$a and $p(2\times 1)$s reconstructions seems to be converged to within 2–3 meV per dimer, whereas for the $p(2\times 2)$ and the $c(4\times 2)$ the convergence is no better than about 5–10 meV per dimer. It seems, however, that the energy of the $p(2\times 2)$ relative to that of the $c(4\times 2)$ has converged to a higher level of accuracy. Since this energy difference is of particular interest and is so small, we perform yet another doubling of the sampling for these two reconstructions. The resulting energy difference is shown in Fig. 11 for samplings ranging from 4 to 64 **k** points in the (2×2) BZ. We see that meV convergence is finally reached after the last doubling of the sampling. If anything, the energy difference is increasing, which indicates that increasing the sampling further will only make the $c(4\times 2)$ more favorable.

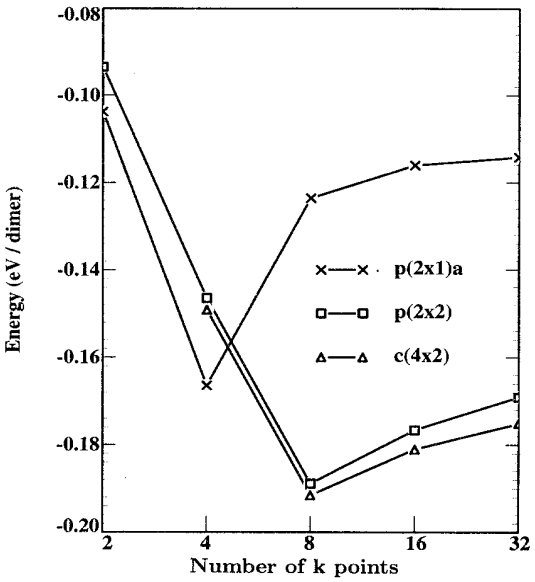

FIG. 10. Convergence of the energy of the $p(2\times 1)$a, $p(2\times 2)$, and the $c(4\times 2)$ relative to the $p(2\times 1)$s as a function of the number of **k** points. A cutoff of 16 Ry is used. The same **k** points are used for the $c(4\times 2)$ as for the other reconstructions so that the energy differences involving this reconstruction are significant.

The differences of sums of single-particle eigenvalues shown in Figs. 10 and 11 will not converge to the fully self-consistent energies, but differ from those by the error introduced by keeping the potential frozen. Nonetheless, we expect the rate of convergence to be essentially the same as that of a fully self-consistent calculation. In other words, the error made in a self-consistent calculation using a given sampling will be of approximately the same size as that found in Figs. 10 and 11.

2. *Charge density optimization*

To examine the convergence of energy differences as the **k**-point sampling was increased, we used the expression in Eq. (1), but with a fixed charge density obtained using

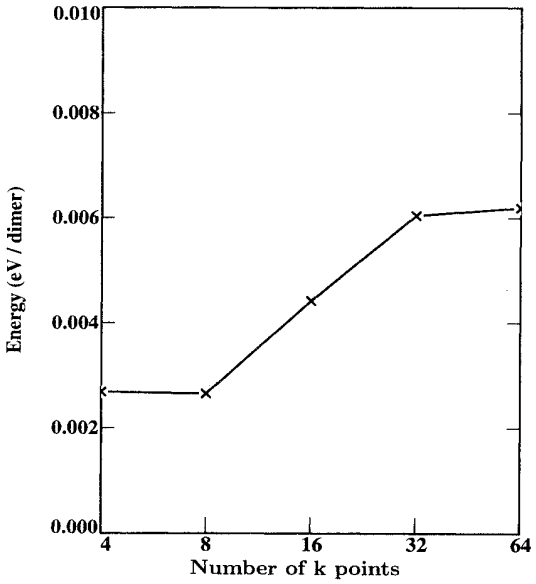

FIG. 11. Convergence of $E_{p(2\times2)} - E_{c(4\times2)}$ as a function of the number of **k** points. The same **k** points are used for both reconstructions. The energy cutoff used is 16 Ry.

four **k** points. In this section we examine the effects of optimizing the charge density as the **k** point sampling is increased.

If a charge density obtained from a coarse sampling of reciprocal space is used to calculate the energy corresponding to a finer sampling (by sampling only the $\sum \varepsilon_{i\mathbf{k}}$ term on a finer mesh), an error $\delta\rho = \rho - \rho^0$ is introduced in the charge density, where ρ^0 and ρ are the self-consistent charge densities corresponding to the coarse and fine samplings, respectively. Harris[52] has shown that the energy expression (1) only contains errors of order $(\delta\rho)^2$.

To find the actual size of the errors, we compare energy changes estimated using Eq. (2) to self-consistent results. When the sampling grid is made finer the total energy changes and, in the frozen-potential approximation, this change is given according to Eq. (2) by

$$\Delta E \approx \sum_{i\mathbf{k}}{}' \varepsilon_{i\mathbf{k}} - \sum_{i\mathbf{k}} \varepsilon_{i\mathbf{k}}, \tag{3}$$

where primed and unprimed sums run over the **k** points of the fine and coarse grids, respectively. Both sums are generated using the self-consistent charge density of the coarse grid (ρ^0). ΔE can be added to the self-consistent total energy of the coarse grid to approximate the total energy upon increasing the density of sampling points. Table I shows ΔE along with the errors made from using the approximation (3) to the total energy, as compared with the fully self-consistent energy. We see that, apart from the results based on the 1 **k**-point Hamiltonian, the self-consistency effect is less than 15 meV/dimer. Furthermore, it is always negative; thus when calculating relative energies we will again find a cancellation. This was also noted by Robertson and Payne,[54] and proven by Zaremba[56] under certain mild conditions. Such a cancellation will, of course, be most useful if the errors not only have the same sign, but are also close in absolute value. If we look at the bottom row of Table I, which contains the most relevant information, we see that if we neglect self-consistency on going from 4 to 8 point sampling, we make an error of at most a few meV per dimer in the absolute energy and of the order of 1 meV per dimer in energy differences.

Table I also shows that the changes in the energy come in "steps"; increasing the sampling from one to two **k** points results in a large change of energy, while increasing from two to four **k** points results in a small change. The next doubling in sampling from four to eight **k** points results in a somewhat larger energy change. This steplike convergence can be seen more clearly in Fig. 12, which shows the convergence of the sum of single particle eigenvalues as the **k**-point sampling is increased from 1 to 64 sampling points for the $p(2 \times 2)$ reconstruction. This behavior can be understood in terms of the dispersion of the surface bands (see Sec. IV D). The dispersion is large in directions parallel to the dimer rows (the $\Gamma J'$ direction in Fig. 3) and small in directions perpendicular to that. The projection of the sampling grid onto the $\Gamma J'$ direction is only changed when the number of **k** points is

TABLE I. The effect of self-consistency on total energies when the BZ sampling is changed. Results for three different reconstructions are given in columns 2–7. The first column shows the change in **k**-point sampling. The change in total energy, ΔE, resulting from this change in sampling, as estimated using the frozen-potential approximation of Eq. (3), is given in columns 2, 4, and 6. The error in this estimate, "Error," is given in columns 3, 5, and 7 and is the difference between the total energies calculated self-consistently for two different samplings and using the frozen-potential approximation. All energies are given in eV/dimer. The calculations are performed with an 8 Ry cutoff.

	Reconstruction					
	$p(2\times1)$s		$p(2\times1)$a		$p(2\times2)$	
Change in sampling	ΔE	Error	ΔE	Error	ΔE	Error
1 → 2 **k** points	-3.4230	-0.0795	-3.4739	-0.0523	-3.4665	-0.0387
1 → 4 **k** points	-3.4126	-0.0538	-3.5054	-0.0558	-3.5079	-0.0449
2 → 4 **k** points	-0.0249	-0.0096	-0.0291	-0.0011	-0.0365	-0.0014
2 → 8 **k** points	-0.2915	-0.0132	-0.2521	-0.0049	-0.3430	-0.0054
4 → 8 **k** points	-0.2669	-0.0038	-0.2174	-0.0033	-0.3051	-0.0027

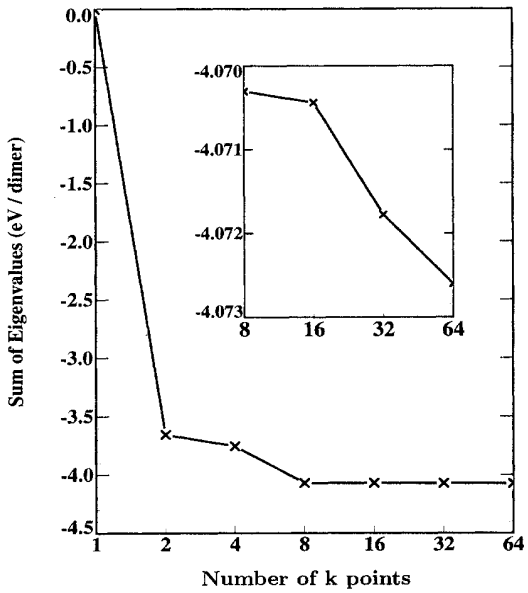

FIG. 12. The sum of single-particle eigenvalues for the $p(2 \times 2)$ reconstruction as a function of the **k**-point sampling. A 4 **k**-point, 16 Ry charge density is used to generate the eigenvalues. The reference energy is chosen arbitrarily. The eigenvalue sum calculated with 8–64 points is shown on an expanded scale in the inset.

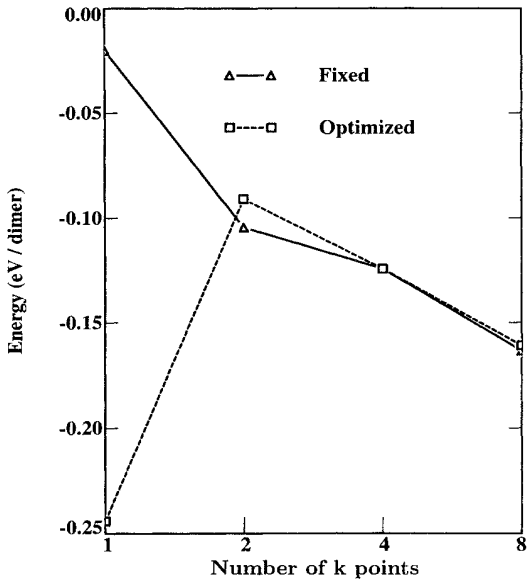

FIG. 13. Convergence of $E_{p(2\times2)} - E_{p(2\times1)s}$ as a function of the number of **k** points. The triangles indicate results obtained with a fixed geometry (optimized using four **k** points and a cutoff of 8 Ry). The squares show the results of calculations where the geometries were optimized for each **k**-point sampling. The lines joining the symbols are guides to the eye. A cutoff of 8 Ry is used for all calculations.

changed from one to two, four to eight, etc. (see Fig. 3). Thus these changes in the sampling should result in the largest changes in energy, which is what we observe. The above tests were all carried out using a cutoff of 8 Ry. To check that the conclusions do not depend on the cutoff, we repeated a number of tests using a 16 Ry cutoff. The results are shown in Table II. By comparing Table II to the corresponding row of Table I, we see that the self-consistency effect changes by 1 meV or less when the cutoff is increased from 8 to 16 Ry. The conclusions drawn above thus hold for a 16 Ry cutoff as well.

We conclude that we can perform self-consistent calculations with a limited number of **k** points (four for the systems we are interested in here), then use the frozen-potential approximation to improve the **k**-point sampling and calculate energy differences without introducing errors greater than ~ 1 meV.

3. Geometry optimization

We now allow the geometry to relax as the **k**-point sampling is increased. Using the **k**-point sets of Fig. 3 and a cutoff of 8 Ry, self-consistent calculations with one, two, four, and eight **k** points in the full (2×2) zone were performed for the $p(2 \times 1)$s and $p(2 \times 2)$ reconstructions. Figure 13 shows how the energy difference between the $p(2 \times 2)$ and the $p(2 \times 1)$s reconstructions converges as a function of the **k**-point sampling. The values indicated as triangles and joined by a solid line were calculated with ions fixed in their 8 Ry, 4 **k**-point optimized positions, while the values indicated as squares and joined by a dashed line result from optimizing the geometry for each particular **k**-point sampling. We see that the effect of the geometry optimization on the energy difference converges rapidly as a function of the number of **k** points. At eight **k** points, it is as low as 2.5 meV per dimer.

C. Independence of the variation of the cutoff energy and BZ sampling

We have implicitly assumed that we can study the convergence of energy differences as a function of the Brillouin zone sampling and the plane-wave cutoff energy independently. Our investigation of the convergence of energy differences with increasing cutoff energy was performed using a rather coarse BZ sampling. In Sec. III B

TABLE II. Self-consistency effect calculated with a cutoff of 16 Ry. Conventions are the same as in Table I.

	Reconstruction					
	$p(2 \times 1)$s		$p(2 \times 1)$a		$p(2 \times 2)$	
Change in sampling	ΔE	Error	ΔE	Error	ΔE	Error
$2 \rightarrow 4$ **k** points	-0.0665	-0.0100	-0.1117	-0.0013	-0.1042	-0.0024

it was shown that a much denser sampling is needed in order to reduce the error bar arising from the finite sampling of the BZ to an acceptable level. We therefore need to check whether our conclusions concerning the cutoff (Sec. III A) also hold for other samplings.

To find out how the convergence as a function of cutoff depends on the **k**-point sampling, we took the geometries corresponding to the crosses in Figs. 5 and 9 and, without allowing for any further relaxation, recalculated the energies (self-consistently) using only two **k** points in the (2 × 2) BZ. The results are shown in Figs. 14 and 15 as squares. The values shown as crosses correspond to the crosses in Figs. 5 and 9. We see that the qualitative behavior of the convergence does not change when the number of **k** points is doubled. The deviation from a rigid shift becomes smaller as the cutoff is increased and for cutoffs of 16 and 20 Ry it is approximately 3 meV or less. It thus appears that, for sufficiently high cutoffs, the quantitative convergence behavior is unchanged by an increase in sampling. This allows us to take the results from Sec. III B 1, which were obtained with a dense sampling and a cutoff of 16 Ry, and apply a correction for a higher cutoff to them. Performing such a correction will introduce an error no greater than 3 meV in the relative energies.

D. Other factors

In order to calculate energy differences on a meV per dimer scale the influence of the thickness of the repeated slab and of the vacuum spacing should be checked. In our studies of adsorption[1] we looked at the effect of changing the thickness of the vacuum region from 9.5 Å (equivalent to six atomic layers) to 13 Å. This led to changes in energy differences of only a few meV. The binding of the adatom to the surface should be more sensitive to the separation between the slabs than the surface reconstructions since the adatom sticks out from the surface and is closer to the opposite slab. We thus expect ~ 1 meV per dimer to be an upper limit on the sensitivity of energy differences between different reconstructions to increasing the vacuum thickness.

We have not made a detailed estimate of the effect of increasing the number of atomic layers in the slab. Our slab thickness of 12 layers is already larger than in most of the recent calculations and we do not expect this to play an important role. Nevertheless, this is clearly a test which should be carried out if one is primarily interested in the energy difference between the $p(2 \times 2)$ and the $c(4 \times 2)$. We will see in the next section that the error bars imposed by our maximum BZ sampling and plane-wave cutoff are so large that we will be unable to make a definitive statement as to which of these reconstructions is lowest in energy. All the other energy differences are sufficiently large that the uncertainty arising from the slab thickness is unlikely to change the results to be presented below.

The choice of exchange-correlation potential may well

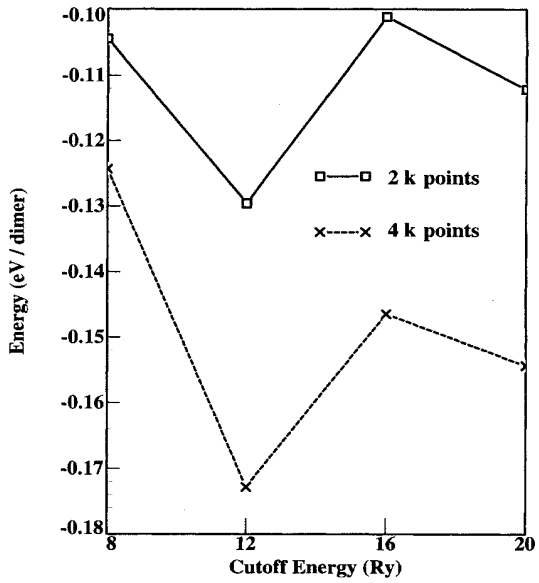

FIG. 15. Convergence of $E_{p(2\times2)} - E_{p(2\times1)s}$ as a function of cutoff, using two and four **k** points. The crosses joined by dashed lines represent calculations performed with four **k** points (the same calculations as the crosses in Fig. 9). The squares joined by solid lines represent calculations performed using two **k** points. The two **k**-point calculations use the same geometries as the corresponding 4 **k**-point calculations.

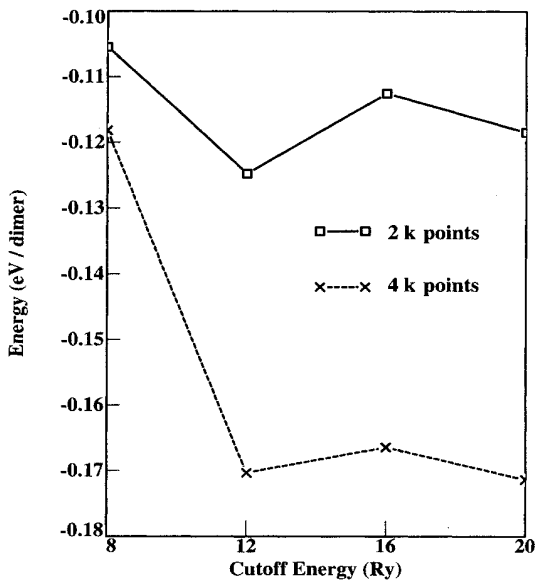

FIG. 14. Convergence of $E_{p(2\times1)a} - E_{p(2\times1)s}$ as a function of cutoff, for two different sets of **k** points. The crosses joined by dashed lines represent calculations performed with four **k** points (the same calculations as the crosses in Fig. 5). The squares joined by solid lines represent calculations performed using two **k** points. The 2 **k**-point calculations use the same geometries as the corresponding 4 **k**-point calculations.

be important on a meV energy scale. For iron, the energy difference between the bcc and fcc structures varies by as much as 25 meV per Fe atom if different potentials are used.[57] The effect may be much smaller for non-transition-metal elements such as Si but we have not carried out any systematic investigation. However, if the results should depend on the choice of exchange-correlation potential, there is no strong preference for one potential over another since no one potential consistently outperforms the others when tested for a range of materials and properties.

IV. RESULTS

We can now collect the results presented above to obtain our best estimates of the energy differences between the different reconstructions. We also combine the different error bars we have discussed to determine whether or not the calculated energy differences are significant. These results are given in Sec. IV A and compared to the results of other calculations in Sec. IV B. Structural data is given in Sec. IV C and band structures are presented and discussed in Sec. IV D.

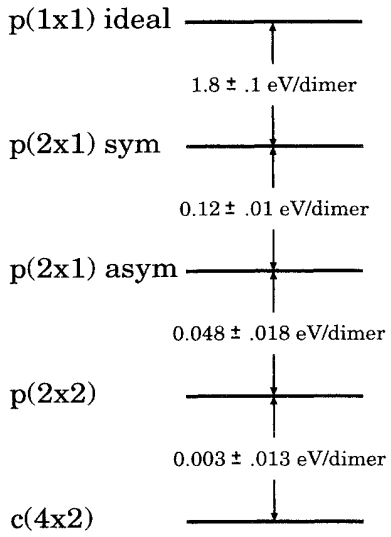

FIG. 16. Schematic illustration of the relative energies and error bars found in our calculations. The error bars are worst case values.

A. Energy differences

To obtain our final estimate of the energy differences shown in Fig. 2, we start with the 16 Ry, 32 **k**-point results obtained using the frozen-potential approximation which are shown in Fig. 10. For the $p(2\times2)-c(4\times2)$ energy difference we use the 16 Ry, 64 **k**-point results shown in Fig. 11. We then estimate the change in these energy differences when the cutoff is increased from 16 Ry to 24 Ry using the results given in Figs. 5–9. The resulting values form our best estimate of the energy differences. They contain errors resulting from using:

a finite sampling (evaluated in Sec. III B 1);
a finite cutoff of 20 or 24 Ry (Sec. III A);
a geometry optimized with a lower, 16 Ry cutoff;
the frozen potential approximation;
and the "cutoff correction" (Sec. III C).

Following the discussion in Secs. III A 2 and III B 3, we will take the error bar resulting from not optimizing the geometry with the largest possible cutoff and number of **k** points to be 5 meV.[58] The error bar resulting from using the frozen potential is taken to be 2 meV following the discussion in Sec. III B 2. The cutoff correction contribution to the error bar will be taken to be 3 meV as discussed in Sec. III C. In addition, there are errors coming from the factors mentioned in Sec. III D which we will neglect, assuming them to be of order 1 meV. Our final results are summarized in Fig. 16. It should be noted that the error bars indicated are worst case errors, where we have taken the arithmetic sum of the contributions to the error listed above.

The largest energy is the dimerization energy, the energy difference between the ideal surface and the $p(2\times1)$s, 1.8 ± 0.1 eV. Because it is so large, we have not taken the trouble to reduce the error bar below 0.1 eV, which is an order of magnitude larger than the other error bars in the figure. The energy difference of 1.8 eV was calculated using only four **k** points and a cutoff of 20 Ry (see Fig. 6). The error bar of 0.1 eV is determined by the small number of **k** points used and we expect it to be an upper limit if we look at how energy differences vary when the sampling is increased from 4 to 32 **k** points in Fig. 10.

We estimate the energy difference between the $p(2\times1)$s and $p(2\times1)$a reconstructions, the buckling energy, to be 0.12 ± 0.01 eV. This value was obtained from the energy of 0.114 eV in Fig. 10 and corrected for the cutoff by 0.007 eV from Fig. 5. The error bar is $13 = 2+1+5+2+3$ meV. The first contribution to the error, 2 meV, is estimated from the behavior of the last two points in Fig. 10 as the number of **k** points is increased from 16 to 32. We assume that the change in going from 32 sampling points to convergence will be less than the change in going from 16 to 32 points. The second contribution, 1 meV, is determined from Fig. 5. The third contribution of 5 meV is the correction we allow for the effect of additional geometry optimization. Although the contribution to this term from Fig. 5 is clearly much smaller, there is in addition a contribution of 2.5 meV discussed in Sec. III B 3. The last two values of 2 meV and 3 meV are the correction for the frozen potential approximation and the cutoff correction, respectively.

The energy difference between the $p(2\times1)$a and $p(2\times2)$ reconstructions is estimated to be 48 ± 18 meV. This value was obtained from the energy of 54 meV in Fig. 10 and corrected for the cutoff by -6 meV from Fig. 7. The error bar is $18 = 5+3+5+2+3$ meV, where the last three contributions are the same values as taken above. The first contribution to the error, 5 meV, is estimated from the behavior of the last two points in Fig. 10. The second contribution, 3 meV, is determined from Fig. 7. The error bar for this energy difference is relatively large as a

result of slower convergence both as a function of cutoff and **k**-point sampling than the other energy differences. If the effect of geometry optimization for a 20 Ry cutoff (see Fig. 7) were included in the energy difference rather than in the error,[58] then 48 ± 18 meV would become 54 ± 15 meV .

The smallest energy difference is that between the $p(2\times2)$ and $c(4\times2)$ reconstructions. We estimate it to be 3 ± 13 meV. This value was obtained from the energy of 6 meV in Fig. 11 and corrected for the cutoff by -3 meV from Fig. 8. The error bar is $13 = 1 + 2 + 5 + 2 + 3$ meV where we use the same values for the last three terms as above. The first contribution to the error, 1 meV, is estimated from Fig. 11, and the second contribution, 2 meV, from Fig. 8. Because we have not carried out an explicit estimation of the correction for the geometry optimization to the error, we have had to assume the "standard" value of 5 meV. This may well be an overestimate but to reduce it significantly would be a costly exercise.

With the exception of the last energy difference, we are able to arrange the different reconstructions on an energy scale in the order sketched in Fig. 16. A less conservative estimate of the error bars, e.g., a reduction by a factor of 2 would not change the ordering we find and would still not allow the sign of the $p(2 \times 2) - c(4 \times 2)$ energy difference to be determined unambiguously. The small size of this energy difference not only makes it difficult to make a definitive statement about the ground state from a theoretical (local-density approximation) point of view, but is probably also responsible for the corresponding experimental difficulties.[32]

B. Comparison with other calculations

The energy differences from the various calculations cited in Fig. 2 are presented in the four panels of Fig. 17, where they are compared with our own results. As we remarked in the Introduction, there is a considerable spread in the values obtained by different authors. In order to assess the (lack of) agreement between the various calculations, including our own, we have to estimate the uncertainty associated with the points in Fig. 17. We can assign rigorous error bars to our own results as discussed in the preceding section, but the information needed to make a similar analysis for the other points is incomplete. In Sec. III we studied the dependence of various energy differences on the energy cutoff. These curves contain information about the errors made when using a given cutoff, provided the same pseudopotential has been used. Since most authors used a pseudopotential identical or very similar to the one we have used,[45] these curves can serve to assign uncertainties to the results cited in Fig. 2 resulting from the cutoff only. Each of the vertical dashed lines in Fig. 17 is constructed by taking the value of the cutoff energy cited by an author and projecting the part of the energy convergence curves (Figs. 6–9) to the right of that value onto the energy axis. This shows the interval in which the energy differences will vary when the cutoff energy is increased. The lines constructed in this fashion are merely intended to give a rough estimate of the uncertainties which arise by using the energy cutoffs quoted in the references of Fig. 2. There are other possible reasons for discrepancies between different calculations such as the use of different BZ sampling points or exchange-correlation parametrizations, which will add to the uncertainties shown in Fig. 17. In most cases, the uncertainties resulting from using the cited cutoff energy are larger than those attributable to different BZ sampling.

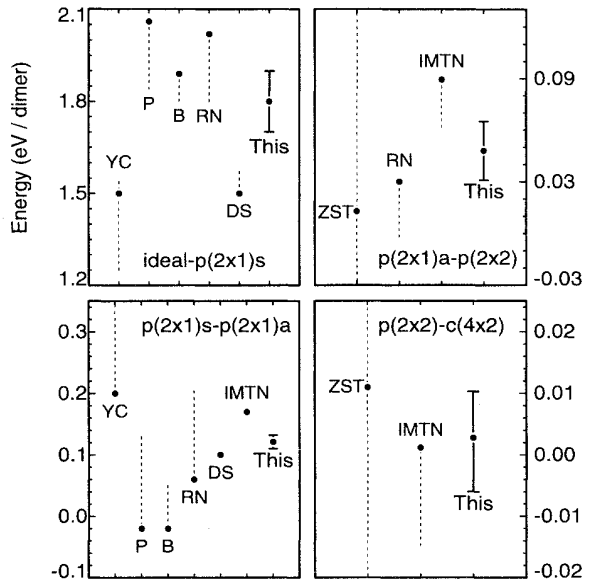

FIG. 17. The energy differences from the various calculations cited in Fig. 2 are presented in the four panels. "This" denotes the present work; the error bar has been discussed in Sec. IV A. The letters labeling the other data points are the same abbreviations as used in Fig. 2. The dashed lines denote the systematic trends observed when increasing the cutoff used by the various authors, as discussed in Sec. IV B. Along the x axes the references are ordered according to the publication year of the calculations.

The dimerization energies are shown in the upper left-hand panel of Fig. 17. Most of the results are in agreement once account is taken of the different cutoff energies used but there is a large discrepancy with the results of Dąbrowski and Scheffler[17] (DS). This cannot be explained in terms of cutoff energy only since the 12 Ry cutoff these authors use should not lead to an uncertainty larger than 0.1 eV. The discrepancy is not likely to be the result of a difference in BZ sampling either since their sampling (which corresponds to the solid squares in Fig. 3) is the same as the sampling used, for instance, by Roberts and Needs (RN).[14]

As far as the results of Yin and Cohen[10] (YC) are concerned, it seems that the discrepancy is also unresolved. However, we have probably underestimated the uncertainty in this calculation as a cutoff of only 4.3 Ry was used and we have not performed any tests for cutoffs lower than 6 Ry. There is an indication that the geometry might deteriorate rapidly for such low cutoffs: Yin and Cohen find a buckling angle for the $p(2 \times 1)$a of only about 8°, while we find that it increases from about 16° to 19° as the cutoff is increased from 8 to 20 Ry. Other

sources of error which we have not taken into account are their incomplete geometry optimization and their use of an eight layer slab and a different exchange-correlation potential.[59]

The next energy we consider is the $p(2\times1)$s $-p(2\times1)$a energy difference. The results quoted in Fig. 2 are shown in the lower left-hand panel of Fig. 17. Again there is a significant discrepancy with YC, presumably for the same reasons as given above. Note that for this energy difference, our result is in quite reasonable agreement with that of DS. DS optimized their geometry with a cutoff of 8 Ry and from Fig. 5 it can be seen that this can introduce an uncertainty of order 0.01 eV, which is sufficient to account for the small remaining discrepancy. Both Pandey[11] (P) and RN used a 6 Ry cutoff; because the lowest cutoff used in Fig. 5 was 8 Ry, we estimate the 6 Ry cutoff value using the results of Dąbrowski and Scheffler. The remaining discrepancy with the result of Batra[15] (B) can, in our opinion, be explained by the fact that his geometry optimization is far from complete. The only discrepancy which we cannot explain is with the result of Inoue, Morikawa, Terakura, and Nakayama[20] (IMTN). These authors used a 12.25 Ry cutoff and a k-point sampling which corresponds to the solid squares in Fig. 3. This should lead to an uncertainty of no more than 0.01 eV. If one were to attempt an explanation for it anyway, it might be mentioned that IMTN use the (less commonly used) Wigner interpolation formula[59] for the exchange and correlation energy. For completeness we should mention the results of Krüger and Pollmann,[18] who found the energy of the asymmetric dimer to be 0.14 eV per dimer lower than that of the symmetric dimer. Although these calculations were based on norm-conserving pseudopotentials and the local-density approximation, they are not included in Fig. 17 because we have no means of estimating an uncertainty for their Gaussian orbital basis.

For the $p(2\times1)$a $-p(2\times2)$ energy differences shown in the upper right-hand panel of Fig. 17, discrepancies can (just about) be accounted for by the choice of cutoff. This time, there seems to be no gap between our error bar and the work by IMTN.[20] The uncertainty associated with the calculation by Zhu, Shima, and Tsukada[13] (ZST) goes off the energy scale of the figure because they have used a cutoff of only 2.4 Ry. We have not attempted to estimate the uncertainty due to lack of cutoff convergence more precisely but merely indicate that it must be considerably larger than the scale of the figure.

Finally, we come to the energy difference between the $p(2\times2)$ and $c(4\times2)$ reconstructions. This is the only energy whose sign remains uncertain (see the lower right-hand panel of Fig. 17). Although our calculation comes fairly close to resolving this energy difference, a very substantial effort would be needed to reduce the error bar to 1 meV or less.

C. Geometries

A complete list of atomic displacements from the ideal (unreconstructed) positions is given in Cartesian coordinates in Tables III and IV and the corresponding bond lengths are summarized in Fig. 18. The geometries are those which were obtained using a 16 Ry cutoff and four **k** points. For the two $p(2\times1)$ reconstructions, displacements in the y direction (perpendicular to the dimer bonds) were only of order 10^{-5} Å and are therefore not given. The calculated bond lengths are estimated to be converged to within 0.01 Å, the bond angles to within 1°.

For the $p(2\times1)$s we find a large relaxation of the outermost layers into the bulk material; see Table III. The displacement of the topmost dimerized atoms in the z direction is 0.524 Å, which may be compared with the value of 0.33 Å reported in Ref. 14. We find that the size of this relaxation depends directly on the energy cutoff used; a low energy cutoff such as the one used by Roberts and Needs[14] leads to a smaller relaxation in the z direction. In spite of this, our calculated bond lengths are very similar to theirs. We find a dimer bond length of 2.23 Å as they do, and back bonds (bonds between dimer atoms and second layer atoms) of 2.27 Å compared to their value of 2.29 Å. For comparison, the bond length in bulk Si is 2.35 Å. The bond lengths between the second and third layer atoms are 2.33–2.34 Å where Roberts and Needs find values between 2.36 and 2.37 Å. The vertical relaxation of 0.524 Å is distributed over the five layers which are allowed to relax.

For the $p(2\times1)$a we find a very large buckling angle of 18.3° compared to values of 6.9° reported by Roberts and Needs[14] and 15° reported by Dąbrowski and Scheffler.[17] We attribute the larger buckling angle to general improvements in the calculation: more efficient geometry optimization, better basis set, etc. A similar increase in surface buckling has been found for the Si(111) $p(2\times1)$ reconstruction as the calculations were improved.[60,61] The $p(2\times1)$a dimer bond length of 2.26 Å is slightly longer than that found for the symmetric case and also somewhat larger than the 2.21 Å found in Ref. 14. The bond lengths corresponding to the back bonds from the "up" and "down" atoms of the buckled dimer to the second layer atoms are 2.34 Å and 2.29 Å, respectively (Roberts and Needs: 2.30 and 2.28 Å). Krüger and Pollmann ob-

TABLE III. Displacements (in Å) from ideal positions for the $p(2\times1)$s and $p(2\times1)$a. Ideal positions are given by $\mathbf{R}_{klm}=(k\times\sqrt{2},l\times\sqrt{2},m)\frac{a}{4}$, where the bulk lattice constant $a=5.431$ Å. The axes are defined in Fig. 1. Displacements in the y direction are all zero.

Layer	Ideal pos.	$p(2\times1)$s		$p(2\times1)$a	
	(k,l,m)	Δx	Δz	Δx	Δz
1	(0, 0, 0)	0.805	−0.524	1.162	−0.921
1	(2, 0, 0)	−0.805	−0.524	−0.534	−0.213
2	(0, 1, −1)	0.075	−0.141	0.066	−0.141
2	(2, 1, −1)	−0.075	−0.141	−0.099	−0.112
3	(1, 1, −2)	0.000	−0.216	0.031	−0.240
3	(3, 1, −2)	0.000	0.005	−0.025	−0.003
4	(1, 0, −3)	0.000	−0.139	−0.013	−0.155
4	(3, 0, −3)	0.000	0.002	−0.005	0.002
5	(0, 0, −4)	−0.022	−0.040	−0.042	−0.044
5	(2, 0, −4)	0.022	−0.040	0.022	−0.040

TABLE IV. Displacements (in Å) from ideal positions for the $p(2 \times 2)$ and $c(4 \times 2)$. Coordinate frame as defined in Table III.

Layer	Ideal pos.	$p(2 \times 2)$			$c(4 \times 2)$		
	(k,l,m)	Δx	Δy	Δz	Δx	Δy	Δz
1	(0, 0, 0)	0.992	0.000	−0.832	0.989	0.000	−0.789
1	(2, 0, 0)	−0.688	0.000	−0.094	−0.685	0.000	−0.055
1	(0, 2, 0)	0.675	0.000	−0.076	0.675	0.000	−0.045
1	(2, 2, 0)	−1.010	0.000	−0.829	−1.001	0.000	−0.788
2	(0, 1, −1)	0.105	0.119	−0.101	0.108	0.120	−0.079
2	(2, 1, −1)	−0.118	−0.112	−0.109	−0.120	−0.117	−0.086
2	(0, 3, −1)	0.105	−0.118	−0.101	0.108	−0.119	−0.079
2	(2, 3, −1)	−0.118	0.113	−0.109	−0.120	0.117	−0.086
3	(1, 1, −2)	−0.011	0.001	−0.237	−0.009	0.001	−0.223
3	(3, 1, −2)	−0.003	0.002	0.050	−0.003	0.020	0.066
3	(1, 3, −2)	−0.011	0.000	−0.237	−0.009	0.000	−0.223
3	(3, 3, −2)	−0.003	−0.002	0.050	−0.003	−0.020	0.066
4	(1, 0, −3)	0.024	0.000	−0.160	0.006	0.000	−0.153
4	(3, 0, −3)	0.037	0.000	0.037	−0.005	0.000	0.069
4	(1, 2, −3)	−0.031	0.000	−0.164	−0.011	0.000	−0.155
4	(3, 2, −3)	−0.048	0.000	0.034	−0.006	0.000	0.028
5	(0, 0, −4)	−0.012	0.000	−0.039	−0.041	0.000	−0.023
5	(2, 0, −4)	0.066	0.000	−0.030	0.039	0.000	−0.040
5	(0, 2, −4)	−0.074	0.000	−0.031	−0.045	0.000	−0.038
5	(2, 2, −4)	0.007	0.000	−0.041	0.037	0.000	−0.023

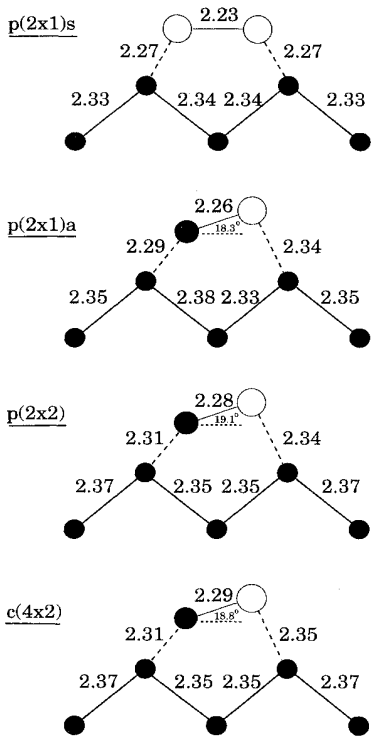

FIG. 18. Side view of surface dimers (larger circles) and the second and third layer atoms (smaller filled circles) for the four reconstructions discussed in the text. The bonds indicated as dashed lines are not in a plane parallel to the plane of the drawing. The buckling angles given for the $p(2 \times 2)$ and $c(4 \times 2)$ reconstructions are the averages of two slightly different angles for the two dimers. All bond lengths are given in Å.

tain an asymmetric dimer bond length of 2.25 Å and a buckling angle of 19° with their Gaussian orbital basis Green's function calculations.[18]

The calculations were carried out without imposing any symmetry (apart from the inversion symmetry about the center of the slab) so that the two surface dimers in the $p(2 \times 2)$ and $c(4 \times 2)$ reconstructions need not and, indeed, are not found to be exactly equivalent (see Table IV). However, the energy lowering associated with the symmetry breaking is so small that it is doubtful whether it is significant. For the $p(2 \times 2)$ reconstruction, surface dimer buckling angles of 18.9° and 19.3° were found; for the $c(4 \times 2)$ the buckling angles were 18.7° and 18.9°. A similar inequivalence was found for the $p(2 \times 2)$ reconstruction by Roberts and Needs, who, however, found smaller values of 11.6° and 12.25°. A buckling angle of 17.5° was found by Northrup in a recent study of the $c(4 \times 2)$.[19] In both reconstructions we find that the various bond lengths corresponding to the two dimers are the same within 0.002 Å or better so that only one dimer is shown in each case in Fig. 18. As might be expected from the very small energy difference between the $p(2 \times 2)$ and $c(4 \times 2)$ reconstructions, the bond lengths are virtually identical. We also note that the asymmetry in the bonds between second and third layer atoms found for the $p(2 \times 1)$a is not present for the alternating asymmetric dimer reconstructions. Table IV shows that, with the exception of the second layer atoms, the displacements in the y direction are very small. The y displacements of the second layer atoms are quite substantial and are such that these atoms are always displaced towards the up atom of the surface dimer. This was found originally for Ge by Needels[21] and since then has also been found for the Si $c(4 \times 2)$[13,19] and Si $p(2 \times 2)$.[14] To an excellent approximation the xz planes which contain the surface

dimers are mirror planes; the zero entries in Table IV do not result from imposing any symmetry.

The calculations which we have discussed until now were performed assuming that the temperature is zero. In view of the small energies separating different reconstructions (Fig.16) the calculated geometries should only be compared with geometries determined experimentally at low temperature. The proposal by Hamers *et al.*,[31] that the symmetric dimers seen by STM at room temperature may actually be asymmetric dimers which are oscillating rapidly to and fro between two degenerate states so that they appear to be symmetric on the time scale of an STM experiment, has been supported by a number of subsequent experiments.[39,32] On studying the geometry as a function of the temperature, the number of alternating buckled dimers was seen to increase as the temperature was reduced. We have found further support for this suggestion of dynamic buckling in a recent *ab initio* molecular dynamics simulation.[62] We refrain from making a detailed comparison of our geometries with those determined recently in transmission electron diffraction (TED) measurements[38] and by grazing incidence x-ray diffraction[37] since these experiments were carried out at room temperature. The results of both the TED experiment and the x-ray diffraction experiment were fitted assuming a uniform structure with $p(2 \times 1)$ periodicity. The buckling angles extracted from this analysis of the TED and x-ray diffraction experiments are about 5° and 7°, respectively. These values are substantially smaller than the 19° which we calculate for zero temperature.

Both room-temperature[31] and low-temperature[32] STM experiments observed that intrinsic surface defects (presumably missing dimers) could pin an alternating asymmetric dimer structure. It is not obvious that the detailed structure of such pinned asymmetric dimers should be the same as the intrinsic buckling which would be observed on a defect-free surface at low temperature. Since the number of intrinsic defects appears to be very large even in the best of cases, care will have to be taken in future low-temperature experiments in identifying buckling angles with pinned or unpinned dimers.

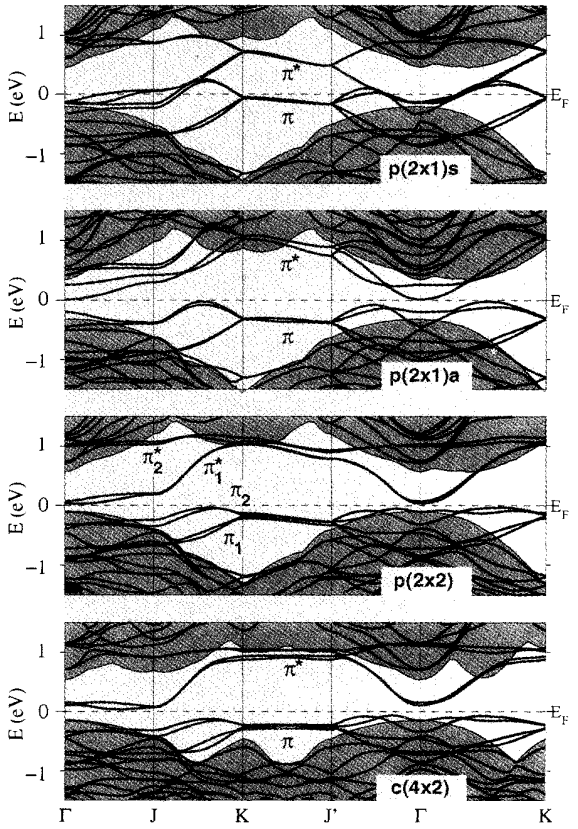

FIG. 19. Band structures for the four reconstructions, plotted along the directions in reciprocal space corresponding to the edges and diagonal in Fig. 3. The $p(2 \times 1)$s and $p(2 \times 1)$a bands were calculated in a $p(2 \times 2)$ unit cell so that the surface unit cell contains four dangling bonds in all cases. The shaded areas are the corresponding projections of the bulk bands in the $p(2 \times 2)$ surface unit cell, except in the figure at the bottom, where the bulk bands are projected in the $c(4 \times 2)$ unit cell. The reference energy, marked by E_F, is chosen to be in the middle of the surface band gap if such a gap exists. For the $p(2 \times 1)$s, E_F is chosen "by hand" and is not a calculated Fermi energy. π labels the occupied surface states and π^* labels the unoccupied surface states; the lower and upper branches are labeled with subscripts 1 and 2, respectively.

D. Band structures

The original motivation for introducing asymmetric dimers was because band structure calculations which assumed a symmetric dimer model invariably resulted in a metallic surface[8] which was in disagreement with the results of photoemission experiments.[33] In the meantime, we know that the local-density approximation systematically underestimates band gaps[63] so that the failure to find a gap in a LDA calculation would no longer be a strong argument against a symmetric dimer model. Nevertheless, it is still interesting to examine and compare the band structures around the Fermi energy for the different reconstructions. The reason for doing this is because of the relationship between structural stability and the existence of band gaps or regions of low densities of states in the single-particle eigenvalue spectrum.[64]

The band structures calculated for the four reconstructions using an 8 Ry cutoff and a 4 **k**-point charge density are shown in Fig. 19 along a number of lines in reciprocal space. These lines correspond to the edges and diagonal in Fig. 3. The shaded areas are the corresponding projections of the bulk bands. To facilitate comparison of the band structures, the $p(2 \times 1)$s and $p(2 \times 1)$a bands were calculated in a $p(2 \times 2)$ unit cell so that there are four surface atoms in all cases. The first thing to notice is that all bands come in pairs. This is a result of using a slab geometry whereby there are actually two surfaces. Because we impose inversion symmetry, all pairs of states form an even and an odd linear combination with respect to the inversion symmetry. If the slab were infinitely thick, these pairs would be degenerate, but because of the finite thickness, a small energy splitting remains.[65] In what fol-

lows, we will refer to each such pair of bands as a single band.

Corresponding to the four surface atoms per unit cell, there are four surface states in the bulk gap for all four reconstructions. These correspond to the four dangling bonds which remain after dimerization has occcured. These dangling bonds form two π bonding (π_1, π_2) and two π^* antibonding (π_1^*, π_2^*) states. For the $p(2 \times 1)$s reconstruction, these bands overlap, resulting in a metallic surface.

For the $p(2 \times 1)$a reconstruction a gap is formed between the π and the π^* states. By transferring charge from the down to the up atom a partly ionic π bond is formed between the two atoms, which results in a downward shift of the π and an upward shift of the π^* band. The stabilization of the asymmetric dimers is associated with the corresponding energy gain. While the dispersion of the π states is essentially the same for the $p(2 \times 1)$s and $p(2 \times 1)$a reconstructions, there are large changes in the dispersion of the π^* states (mainly in the π_2^* band) in the direction parallel to the rows of dimers. The overall dispersion of both reconstructions is in good qualitative agreement with the results of other first-principles calculations, e.g., Ref. 18.

The lower energy of the $p(2 \times 2)$ and $c(4 \times 2)$ reconstructions may be associated with a further increase of the gap. The separation of up (or down) atoms is increased in the alternating dimer reconstructions and this leads to a decrease in the bandwidths of the π bands from about 0.9 eV to 0.7 eV. Again, the biggest changes are seen in the π_2^* band, which has become almost dispersionless. The π_1^* band displays the greatest dispersion for directions parallel to the dimer chains.

The $p(2 \times 2)$ and $c(4 \times 2)$ reconstructions differ in the way asymmetric surface dimers are arranged on neighboring dimer rows. Changes in the environments of surface dimer atoms first occur at a distance of 5.5 Å, which corresponds to a fourth nearest neighbor separation in the perfect crystal. Within a tight-binding or linear combination of atomic orbitals framework, only very minor differences in the band structures are to be expected, and indeed this is so. The bandwidths are determined by hopping along the dimer chains and this is the same for the $p(2 \times 2)$ and $c(4 \times 2)$ reconstructions. The most obvious change in the corresponding band structures in Fig. 19 is the absence of dispersion of both π and π^* bands in directions perpendicular to the dimer chains in the $c(4 \times 2)$ case.

The decreasing change in the band gaps in the series $p(2 \times 1)$s $\rightarrow$ $p(2 \times 1)$a $\rightarrow$ $p(2 \times 2)$ $\rightarrow$ $c(4 \times 2)$ is consistent with the decreasing energy separation of the same reconstructions shown in Fig. 16. No significance should be attached to the absolute value of the gap, which decreases as the cutoff energy is increased and all the reconstructions become metallic at sufficiently high cutoff with the exception of the $c(4 \times 2)$. Its indirect gap persists up to at least a cutoff of 16 Ry, where it is of the order of a few hundredths of an eV. At this cutoff the gap of the $p(2 \times 2)$ has disappeared and there is an overlap of the valence and conduction bands of a few hundredths of an eV. Calculations of the quasiparticle states using the *GW* approximation, which result in very good agreement with experiment for the band gaps of inorganic semiconductors,[63] show that the occupied and unoccupied LDA states of silicon are almost rigidly shifted with respect to one another by 0.5 eV. This has been verified by explicit calculation for the $c(4 \times 2)$ by Northrup.[19]

V. DISCUSSION

The main purpose of this paper has been to identify the factors which must be addressed in order to determine total energy differences with a given accuracy. We have restricted ourselves to calculations performed within the local-density approximation using norm-conserving pseudopotentials in combination with a plane-wave basis where one might expect the most systematic cancellation of errors. The calculation of total energy differences becomes even more difficult when use is made of a localized orbital basis, nonfrozen cores, a cluster representation, or combinations of these. From our study of the reconstructions of the Si(100), we conclude that the systematic cancellation of errors has been overstated in the past. While the evaluation of an error bar is very time consuming, if done critically it would make the repetition of the same calculations by different groups largely unnecessary.

A. Improving the calculations

A well defined aim of electronic structure calculations is to solve the Kohn-Sham equations within the local-density approximation. To do this in practice, some approximations must be made. The most basic approximation at this level is the choice of exchange-correlation potential, of which a number exist which are commonly used.[43,44,59,66,67] As discussed in Sec. III D, this choice may well lead to changes in total energy differences of the order of meV per atom and even larger. This should be checked. However, we note that if the results are found to depend on this choice, there is no criterion for preferring one potential to another since no one potential consistently outperforms the others when tested for a range of materials and properties. Eventual scatter in results must then be attributed to the intrinsic uncertainty of the LDA.

Other approximations may depend on the physical system being studied. To study the electronic structure of a surface, the semi-infinite system is frequently modeled with a finite slab (two free surfaces, periodic in two dimensions), a repeated slab (periodic in three dimensions) or with a finite cluster. The use of a repeated slab configuration and norm-conserving pseudopotentials makes it possible to use a plane-wave basis. The advantages of this are that (i) the treatment of nonspherical charge densities and potentials is straightforward (no shape approximations), (ii) the calculation of Hellmann-Feynman forces is greatly simplified so that extensive geometry optimization is possible, (iii) the basis is spatially unbiased with no preference for a particular arrangement of atoms, and

(iv) the basis can be systematically improved in a very simple way. One of the disadvantages of this approach is that a frozen-core approximation is made whereby the core electron charge density does not relax in response to changes in the valence charge density. Although such effects are not significant under normal circumstances, it is certainly possible that the frozen-core approximation may affect energy differences on the meV level.

The slab approximation can be tested systematically by increasing the slab thickness. A more elegant way of treating semi-infinite systems makes use of Green's functions and such a method has been implemented for semiconductor surface calculations based on norm-conserving pseudopotentials.[18] The advantage of this method is that the correct boundary conditions for the semi-infinite system are obeyed and by formulating the problem in terms of the *change* made to the bulk by introducing a surface, surface states and resonant states can be identified without the ambiguity encountered in slab calculations. The disadvantage of the method is that a localized orbital basis must be used and such bases cannot be improved systematically. We have not attempted a critical evaluation of the Green's functions localized orbital calculations of Krüger and Pollmann.[18] Fortunately there are no obvious discrepancies between their results and our own.

Common to all of the calculations discussed so far is the use of norm-conserving pseudopotentials. It certainly would be interesting to know to what extent this approximation limits our ability to calculate total energy differences. The all-electron localized orbital cluster calculations of Tang *et al.*[16] are interesting in that they do not make such an approximation. Unfortunately, it is not possible to draw any conclusions about the importance of the use of pseudopotentials, because other features of these calculations, namely, the use of finite clusters to represent infinite systems and the use of a localized orbital basis, make a detailed comparison impossible.

One way of improving the pseudopotential approximation is to take account of the (nonlinear) contribution of the core electron density to the exchange-correlation energy of the valence electrons using a scheme such as that proposed by Louie, Froyen, and Cohen.[68] This so-called nonlinear exchange-correlation correction is apparently important for elements with extended core states such as the alkaki elements. To the best of our knowledge, none of the calculations for the clean Si surface make use of this correction and we have not attempted to do so ourselves. Such an improvement could lead to changes in total energy differences on a meV scale.

In Secs. III and IV as well as in this section we have discussed a number of improvements which should be made in order to calculate energy differences reliably at the meV level. The effort required to do this would be very substantial. It is very unlikely that anything about the reconstruction of the Si(100) could be learned from such a calculation which would justify the effort.

B. Improving the physical model

So far we have been concerned with the predictions of the LDA assuming implicitly that it is a model capable of describing the real physical system in which we are interested. In this context we should point out a number of other limitations of the calculations here.

First, it should be remembered that the calculations discussed so far assume that the temperature is zero whereas almost all of the experimental work which has been discussed was carried out at room temperature. While this temperature may not be significant on the energy scale of the electrons, it must certainly be taken into account for the ions in view of the small energy differences between different reconstructions. Finite temperature simulations can in principle be carried out using the Car-Parrinello method,[48,60] and a simulation of the Si(100) surface is possible.[62] However, in order to make direct contact with experiment possible it would be necessary to study a system containing an impossibly large number of atoms (to avoid spurious correlations due to the periodic boundary conditions) for an impossibly long time (in order to accumulate enough statistics). In view of the discussion in previous sections, this simulation would have to be carried out with a plane-wave kinetic energy cutoff and Brillouin zone sampling far in excess of anything which has yet been attempted. An alternative is to create a parametrized model which is sufficiently simple that a finite temperature treatment is possible and to evaluate the parameters in the model using the results of electronic structure calculations. This approach was introduced to study the $c(4\times2)$ to $p(2\times1)$ phase transition of the Si(100) surface, using empirical tight-binding calculations[69] to determine the parameters. A Monte Carlo simulation by Saxena *et al.* led to the prediction[70] of a transition temperature of 250 K which was subsequently observed in temperature dependent LEED experiments.[39] In recent work Inoue *et al.* have evaluated the parameters on the basis of first-principles calculations and elaborated the Monte Carlo simulations by considering the effect of pinning by defects.[20] A weak point in all of these approaches is the mapping onto an Ising spin Hamiltonian which assumes that the transition is of the order-disorder type, whereas STM clearly indicates that the $p(2\times1)$ structure at room temperature does not consist of static disordered buckled dimers. The low-temperature LEED[39] and STM[32] studies both indicate that the ground state has $c(4\times2)$ symmetry.

A second possible inadequacy of the LDA calculations is that spin ordering on the surface is neglected. This point was made by Artacho and Yndurain[12] in an attempt to reconcile theoretical predictions which favored an asymmetric dimer ground state[8,10] with experiments which they interpreted as favoring a symmetric dimer ground state. These authors proposed that symmetric dimers could be stabilized by antiferromagnetic ordering of the spins. They estimated that such ordering could lower the surface energy by as much as 1 eV per dimer, which is almost an order of magnitude larger than the energies favoring buckling that we have been considering up till now. It is difficult to assess the reliability of this conclusion as it was drawn on the basis of a combination of *ab initio* Hartree-Fock calculations using very small clusters of Si atoms to model the surface and tight-binding Hub-

TABLE V. Lattice constant a_0, bulk modulus B, and its pressure derivative dB/dP for a number of cutoff energies and **k**-point samplings. The **k** points are derived from the Γ point of an equivalent unit cell which contains the number of atoms given in the second column. The values given in the table were determined by fitting the energies calculated for a number of volumes ranging from 85% to 110% of the equilibrium volume to Murnaghan's expression for the equation of state Ref. 74. The cohesive energy $E_{\rm coh}$, which is obtained by comparing the bulk energy with the energy of a single (non-spin-polarized) Si atom calculated with the same pseudopotential/plane-wave formalism and the same cutoff energy, contains a spin polarization energy of 0.612 eV obtained from a standard atomic calculation. It is not corrected for zero-point motion.

$E_{\rm cut}$ (Ry)	No. atoms	a_0 (Å)	B (Mbar)	dB/dP	$E_{\rm coh}$ (eV)
12	64	5.418	0.95	3.3	5.35
16	64	5.388	1.02	3.6	5.34
24	64	5.384	0.98	4.1	5.34
24	512	5.368	0.98	4.4	5.44
Expt.		5.429[a]	0.99[b]	4.2	4.63[c]

[a]Reference 80, 0 K.
[b]Reference 81, 77 K.
[c]Reference 75, 0 K.

bard Hamiltonians. Their proposal has been supported by similar studies by Vinchon *et al.*[71] We note here that this issue can be addressed by extending the LDA to take account of spin ordering. The corresponding local-*spin*-density approximation (LSDA)[72,67] has been by and large very successful in describing the ground state properties of transition-metal itinerant electrons such as Fe, Co, and Ni. Indeed, the possibility of spin-ordering on a semiconductor surface has already been studied within the LSDA for the Si(111).[73] Using the plane-wave basis, norm-conserving pseudopotential formalism, Northrup *et al.* found that the surface energy could be lowered by 40 meV per dimer if spin ordering was allowed. This value is a factor 25 smaller than the estimate of Artacho and Yndurain. Even if spin ordering did occur on the surface, it would not necessarily lead to a reordering in Fig. 16. Although there is no experimental evidence for spin ordering on the Si(100), LSDA calculations should be carried out to resolve this issue.

VI. CONCLUSIONS

In this paper we have presented a numerical study of the extent to which total energy differences can be calculated within the local-density approximation using norm-conserving pseudopotentials in combination with a plane-wave basis. Because of our interest in adsorption on the Si(100) we chose the reconstructions of this surface as the object of our study. We showed that the ideal surface and four reconstructions which are generally considered as possible ground states of the surface form a sequence of states which are separated by energies which range from eV's to meV's. With the exception of the lowest energy reconstructions, which are separated by an energy of only a few meV per surface dimer, we were able to order the different reconstructions unambiguously according to their energies. By carrying out the calculations over a large range of cutoff energy and BZ sampling density, we are able to evaluate previous calculations which were carried out with the same methods and resolve most of the discrepancies.

APPENDIX

As stated in Sec. II, we use the norm-conserving pseudopotential for Si as tabulated by Bachelet *et al.*[45] and modified according to the prescription given by Kleinman and Bylander.[46] We use the d component as the reference potential for $l > 2$, which means that the potential contains s and p nonlocal terms. The results of convergence tests on the properties of bulk Si are given in Table V. The lattice constant, bulk modulus, and its pressure derivatives were obtained by fitting the energies

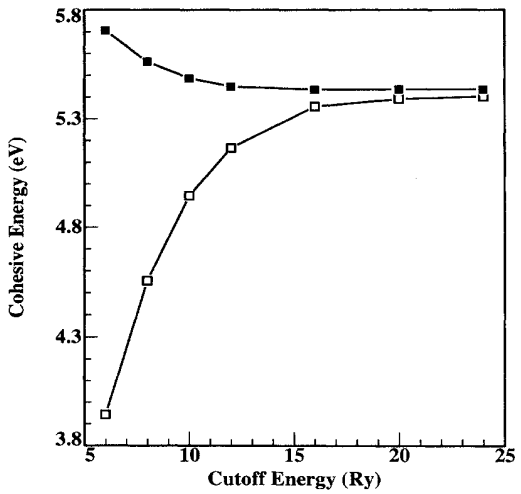

FIG. 20. Total energy (lower curve) and cohesive energy (upper curve) of bulk Si in eV/atom as a function of the plane-wave kinetic energy cutoff used in the calculation. The total energy is referred to the (converged) energy of a Si atom; the cohesive energy is obtained by comparing to a Si atom calculated with the same energy cutoff. The latter is seen to converge much faster as a function of the energy cutoff. The **k**-point sampling used in these calculations is equivalent to using a 512 atom unit cell and only the Γ point. A correction of 0.612 eV for the spin polarization energy of the atom is included.

calculated for a number of volumes, ranging from 85% to 110% of the equilibrium volume, to Murnaghan's expression for the equation of state.[74] As in our surface calculations, we have used **k**-point sets derived from Γ. Each such set is equivalent to using a larger unit cell and only the Γ point; in the table, we characterize the **k**-point set by giving the number of atoms in the larger unit cell. The lattice parameter and bulk modulus for a cutoff energy of 24 Ry and a 512 atom unit cell are approximately 1% smaller than the experimental values. This agrees with other pseudopotential/plane-wave calculations[75–78] and also with all-electron calculations,[79] within the uncertainties present in such calculations. These values are not extremely sensitive to the basis set used. For a cutoff energy of 16 Ry and a 64 atom unit cell, the lattice constant changes by 0.4% and the bulk modulus by 2%. The pressure derivative of the bulk modulus, and to a lesser extent the bulk modulus itself, is more sensitive to the basis set. However, the intrinsic uncertainty due to using the Murnaghan expression is also much larger.

The cohesive energy given in Table V is obtained by comparing the energy of the crystal with the energy of a single (non-spin-polarized) Si atom calculated with the same pseudopotential/plane-wave formalism and the same cutoff energy. A spin polarization energy of 0.612 eV was obtained from a standard atomic calculation. The cohesive energy is not corrected for the zero-point motion, which would reduce it slightly. The convergence with respect to cutoff energy is fast; for 16 Ry, the cohesive energy is converged within 0.01 eV. This is in marked contrast with the convergence of the total energy of the crystal, which is at least an order of magnitude less converged at this cutoff. The different behavior of total and cohesive energy as a function of the cutoff is shown in Fig. 20 and nicely illustrates the systematic cancellation of errors in pseudopotential/plane-wave calculations.

[1] G. Brocks, P.J. Kelly, and R. Car, Phys. Rev. Lett. **66,** 1729 (1991).

[2] G. Brocks, P.J. Kelly, and R. Car, Surf. Sci. **269/270,** 860 (1992).

[3] G. Brocks, P.J. Kelly, and R. Car, Phys. Rev. Lett. **70,** 2786 (1993).

[4] There are too many experimental studies of the Si(100) surface to be considered in detail here. References to work which will not be considered explicitly below can be found in J.E. Griffith and G.P. Kochanski, Crit. Rev. Solid State Mat. Sci. **16,** 255 (1990); R.I.G. Uhrberg and G.V. Hansson, *ibid.* **17,** 133 (1991).

[5] It is also not possible to give a comprehensive list of references to theoretical work dealing with the reconstruction of the Si(100) surface. We will restrict ourselves to a consideration of recent first-principles studies and only refer to a small number of other theoretical works. More extensive references can be found in J.P. Lafemina, Surf. Sci. Rep. **16,** 133 (1992).

[6] J.A. Appelbaum, G.A. Baraff, and D.R. Hamann, Phys. Rev. B **14,** 588 (1976); **15,** 2408 (1977).

[7] J.A. Appelbaum and D.R. Hamann, Surf. Sci. **74,** 21 (1978).

[8] D.J. Chadi, Phys. Rev. Lett. **43,** 43 (1979).

[9] J. Ihm, M.L. Cohen, and D.J. Chadi, Phys. Rev. B **21,** 4592 (1980).

[10] M.T. Yin and M.L. Cohen, Phys. Rev. B **24,** 2303 (1981).

[11] K.C. Pandey, in *Proceedings of the 17th International Conference on the Physics of Semiconductors,* edited by D.J. Chadi and W.A. Harrison (Springer, New York, 1985), p. 55.

[12] E. Artacho and F. Yndurain, Phys. Rev. Lett. **62,** 2491 (1989); Phys. Rev. B **42,** 11 310 (1990).

[13] Z. Zhu, N. Shima, and M. Tsukada, Phys. Rev. B **40,** 11 868 (1989).

[14] N. Roberts and R.J. Needs, Surf. Sci. **236,** 112 (1990).

[15] I.P. Batra, Phys. Rev. B **41,** 5048 (1990).

[16] S. Tang, A.J. Freeman, and B. Delley, Phys. Rev. B **45,** 1776 (1992).

[17] J. Dąbrowski and M. Scheffler, Appl. Surf. Sci. **56,** 15 (1992).

[18] P. Krüger and J. Pollmann, Phys. Rev. B **38,** 10 578 (1988); **47,** 1898 (1993).

[19] J.E. Northrup, Phys. Rev. B **47,** 10 032 (1993).

[20] K. Inoue, Y. Morikawa, K. Terakura, and M. Nakayama, in *Proceedings of the 15th Taniguchi Symposium,* edited by K. Terakura and H. Akai (Springer, Berlin, 1993), p. 77; Phys. Rev. B **49,** 14 774 (1994).

[21] Analogous studies have been made of the Ge surface. See M. Needels, M.C. Payne, and J.D. Joannopoulos, Phys. Rev. Lett. **58,** 1765 (1987); Phys. Rev. B **38,** 5543 (1988).

[22] D. Dijkkamp, E.J. van Loenen, and H.B. Elswijk, in *Proceedings of the 3rd NEC Symposium on Fundamental Approach to New Material Phases,* edited by A. Yoshimori, T. Shinjo, and H. Watanabe, Springer Series in Materials Science Vol. 17 (Springer, Berlin, 1992).

[23] Y.-W. Mo, J. Kleiner, M.B. Webb, and M.G. Lagally, Surf. Sci. **268,** 275 (1992)

[24] R.E. Schlier and H.E. Farnsworth, J. Chem. Phys. **30,** 917 (1959).

[25] W.S. Yang, F. Jona, and P.M. Marcus, Phys. Rev. B **28,** 1049 (1983).

[26] R.I.G. Uhrberg, G.V. Hansson, J.M. Nicholls, and S.A. Flodström, Phys. Rev. B **24,** 4684 (1981).

[27] R.M. Tromp, R.G. Smeenk, F.W. Saris, and D.J. Chadi, Surf. Sci. **133,** 137 (1983).

[28] Y.J. Chabal, S.B. Christman, E.E. Chaban, and M.T. Yin, J. Vac. Sci. Technol. A **1,** 1141 (1983).

[29] H.H. Farrell, F. Stucki, J. Anderson, D.J. Frankel, G.J. Lapeyre, and M. Levinson, Phys. Rev. B **30,** 721 (1984).

[30] F.J. Himpsel, P. Heimann, T.C. Chiang, and D.E. Eastman, Phys. Rev. Lett. **45,** 1112 (1980).

[31] R.M. Tromp, R.J. Hamers, and J.E. Demuth, Phys. Rev. Lett. **55,** 1303 (1985); R.J. Hamers, R.M. Tromp, and J.E. Demuth, Phys. Rev. B **34,** 5343 (1986).

[32] R.A. Wolkow, Phys. Rev. Lett. **68,** 2636 (1992).

[33] F.J. Himpsel and D.E. Eastman, J. Vac. Sci. Technol. **16,** 1297 (1979).

[34] M. Aono, Y. Hou, C. Oshima, and Y. Ishizawa, Phys. Rev. Lett. **49,** 567 (1982).

[35] I. Stensgaard, L.C. Feldman, and P.J. Silverman, Surf. Sci. **102,** 1 (1981).

[36] R.M. Tromp, R.G. Smeenk, and F.W. Saris, Phys. Rev. Lett. **46,** 9392 (1981).
[37] N. Jedrecy, M. Sauvage-Simkin, R. Pinchaux, J. Massies, N. Greiser, and V. H. Etgens, Surf. Sci. **230,** 197 (1990).
[38] G. Jayaram, P. Xu, and L.D. Marks, Phys. Rev. Lett. **71,** 3489 (1993).
[39] T. Tabata, T. Aruga, and Y. Murata, Surf. Sci. **179,** L63 (1987).
[40] Y. Enta, S. Suzuki, and S. Kono, Phys. Rev. Lett. **65,** 2704 (1990).
[41] P. Hohenberg and W. Kohn, Phys. Rev. **136,** B864 (1964).
[42] W. Kohn and L.J. Sham, Phys. Rev. **140,** A1133 (1965).
[43] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45,** 566 (1980).
[44] J. Perdew and A. Zunger, Phys. Rev. B **23,** 5048 (1981).
[45] G.B. Bachelet, D.R. Hamann, and M. Schlüter, Phys. Rev. B **26,** 4199 (1982).
[46] L. Kleinman and D.M. Bylander, Phys. Rev. Lett. **48,** 1425 (1982).
[47] I. Štich, R. Car, M. Parrinello, and S. Baroni, Phys. Rev. B **39,** 4997 (1989).
[48] R. Car and M. Parrinello, Phys. Rev. Lett. **55,** 2471 (1985).
[49] H.J. Monkhorst and J.D. Pack, Phys. Rev. B **13,** 5188 (1976). In this scheme, only special **k** points in the interior of the Brillouin zone are used. This scheme has the disadvantage that each sampling density is obtained by using a unique set of **k** points. That is, none of the **k** points used in one grid can be kept when upgrading to a finer grid.
[50] Although the sampling densities are equal, the four points chosen for the $c(4 \times 2)$ surface are not the same as for the other reconstructions. The reason for not using identical sampling points for the two unit cells from the start is that we want to begin with special symmetry points that are mapped onto themselves by the symmetry operations of time reversal and lattice translation. For these points, one can make use of the symmetry so that only half the number of expansion coefficients for the wave functions are needed, thus gaining a factor 4 in computation speed and making this scheme as efficient as the Monkhorst-Pack scheme (Ref. 50). These are the only points that will be treated self-consistently and must therefore be calculated first. For the $p(2 \times 2)$ reconstruction these points are Γ, J, K, and J' and for the $c(4 \times 2)$ reconstruction, Γ, J, and the point midway between J' and K. After we have completed our convergence tests and want to compare absolute energies, we shall take care to use identical samplings.
[51] A.R. Mackintosh and O.K. Andersen, in *Electrons at the Fermi Surface*, edited by M. Springford (Cambridge University Press, Cambridge, 1980); M. Weinert, R.E. Watson and J.W. Davenport, Phys. Rev. B **32,** 1215 (1985).
[52] J. Harris, Phys. Rev. B **31,** 1770 (1985).
[53] P.J.H. Denteneer, Ph.D. thesis, Eindhoven Technical University, 1987; P.J.H. Denteneer and W. van Haeringen, Solid State Commun. **59,** 829 (1986).
[54] I.J. Robertson and M.C. Payne, J. Phys. Condens. Matter **2,** 9837 (1990).
[55] One should always compare energies calculated with exactly the same **k** points. As noted earlier, the four **k** points calculated self-consistently are not the same for the $p(2\times2)$ and $c(4 \times 2)$ unit cells. To calculate the appropriate eigenvalue sums, we regroup the points so that the 4 **k** point energy for the $c(4 \times 2)$ reconstruction was calculated using the same sampling points as for the other reconstructions.
[56] E. Zaremba, J. Phys. Condens. Matter **2,** 2479 (1990).
[57] H.J.F. Jansen and S.S. Peng, Phys. Rev. B **37,** 2689 (1988).
[58] Where the effect of the geometry optimization has been calculated explicitly with a higher cutoff energy, the corresponding correction could be made to the energy difference shown in Fig. 16. Because we have not calculated it for cutoffs higher than 16 Ry for all the reconstructions, we choose to include the correction as part of the error.
[59] E. Wigner, Phys. Rev. **46,** 1002 (1934).
[60] F. Ancilotto, W. Andreoni, A. Selloni, R. Car and M. Parrinello, Phys. Rev. Lett. **65,** 3148 (1990).
[61] J.E. Northrup, M.S. Hybertsen, and S.G. Louie, Phys. Rev. Lett. **66,** 500 (1991).
[62] G. Brocks and P.J. Kelly, (unpublished).
[63] M.S. Hybertsen and S.G. Louie, Phys. Rev. Lett. **55,** 1418 (1985); Phys. Rev. B **34,** 5390 (1986); R.W. Godby, M. Schlüter, and L.J. Sham, Phys. Rev. Lett. **56,** 2415 (1986).
[64] This relationship is largely empirical. In general, the total energy is not given by a sum of single-particle eigenvalues. The usual approach is to relate total energy differences to differences between sums of single-particle eigenvalues using a force-theorem or frozen-potential-type approximation; see Ref. 51. We are not aware of any attempts to apply such ideas within the framework of plane-wave pseudopotential calculations.
[65] This splitting must occur in all calculations which make use of a slab representation (Refs. 9–11,13–15,17,19,20). The simplest way to remove this blemish is to take the average of the split pairs, which is presumably the solution adopted by most of the above authors. A more difficult problem is that of identifying resonant states. We are not aware of any simple way of doing this unambiguously for a finite slab. This problem is, at least formally, best treated with Greens function methods. See Ref. 18 and references therein.
[66] L. Hedin and B.I. Lundqvist, J. Phys. C **4,** 2064 (1971).
[67] O. Gunnarsson and B.I. Lundquist, Phys. Rev. B **13,** 4274 (1976).
[68] S.G. Louie, S. Froyen, and M.L. Cohen, Phys. Rev. B **26,** 1738 (1982).
[69] J. Ihm, D.H. Lee, J.D. Joannopoulos, and J.J. Xiong, Phys. Rev. Lett. **51,** 1872 (1983).
[70] A. Saxena, E.T. Gawlinski, and J.D. Gunton, Surf. Sci. **160,** 618 (1985).
[71] T. Vinchon, M.C. Desjonquères, A.M. Oleś, and D. Spanjaard, Phys. Rev. B **48,** 8190 (1993).
[72] U. von Barth and L. Hedin, J. Phys. C **5,** 1629 (1972).
[73] J.E. Northrup, J. Ihm, and M.L. Cohen, Phys. Rev. Lett. **47,** 1910 (1981).
[74] F.D. Murnaghan, Proc. Natl. Acad. Sci. U.S.A. **30,** 244 (1944).
[75] M.T. Yin and M.L. Cohen, Phys. Rev. B **26,** 5668 (1982).
[76] M.S. Hybertsen and S.G. Louie, Phys. Rev. B **30,** 5777 (1982).
[77] O.H. Nielsen and R.M. Martin, Phys. Rev. B **32,** 3792 (1985).
[78] V. Milman, M.H. Lee, and M.C. Payne, Phys. Rev. B **49,** 16 300 (1994).
[79] M. Methfessel, C.O. Rodriguez, and O.K. Andersen, Phys. Rev. B **40,** 2009 (1989).
[80] R.W.G. Wyckoff, *Crystal Structures* (Wiley, New York, 1965), Vol. 1.
[81] H.J. McSkimm and P. Andreatch, Jr., J. Appl. Phys. **35,** 2161 (1964).

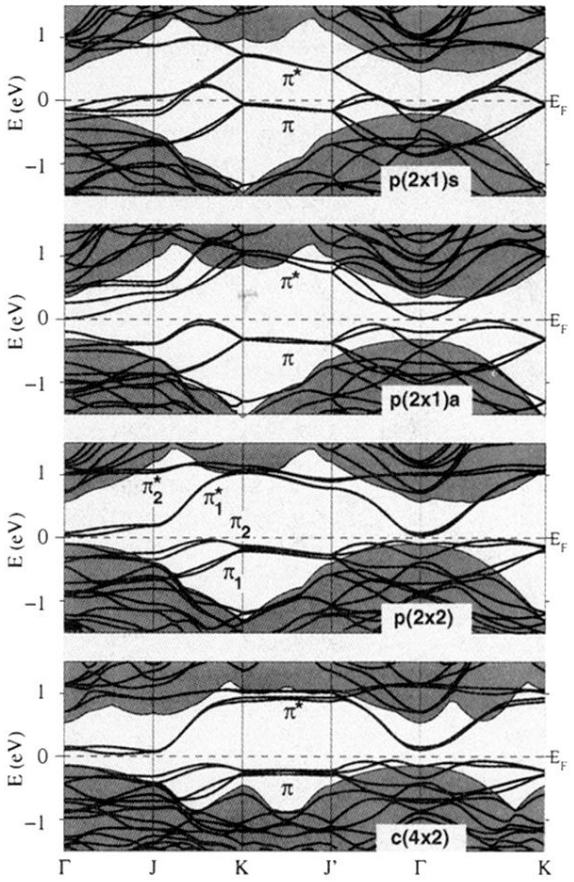

FIG. 19. Band structures for the four reconstructions, plotted along the directions in reciprocal space corresponding to the edges and diagonal in Fig. 3. The $p(2 \times 1)$s and $p(2 \times 1)$a bands were calculated in a $p(2 \times 2)$ unit cell so that the surface unit cell contains four dangling bonds in all cases. The shaded areas are the corresponding projections of the bulk bands in the $p(2 \times 2)$ surface unit cell, except in the figure at the bottom, where the bulk bands are projected in the $c(4 \times 2)$ unit cell. The reference energy, marked by E_F, is chosen to be in the middle of the surface band gap if such a gap exists. For the $p(2 \times 1)$s, E_F is chosen "by hand" and is not a calculated Fermi energy. π labels the occupied surface states and π^* labels the unoccupied surface states; the lower and upper branches are labeled with subscripts 1 and 2, respectively.