Ab Initio Total Energy Study of Adsorption and Diffusion on the Si(100) Surface

ARTICLE *in* THIN SOLID FILMS · JANUARY 1996

Impact Factor: 1.76 · DOI: 10.1016/0040-6090(95)06960-7

3 AUTHORS, INCLUDING:

Victor Milman
BIOVIA

164 PUBLICATIONS **3,351** CITATIONS

SEE PROFILE

ELSEVIER

Thin Solid Films 272 (1996) 375–385

thin solid films

Ab initio total energy study of adsorption and diffusion on the Si(100) surface

V. Milman [a], S.J. Pennycook [b], D.E. Jesson [b]

[a] *Molecular Simulations, The Quorum, Barnwell Road, Cambridge CB5 8RE, UK*

[b] *Solid State Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA*

Abstract

A review is given of the pseudopotential total energy method and of its applications to the study of adsorption and diffusion processes of single adatoms on the Si(100) surface. The subjects covered include the description of the computer code CASTEP/CETEP used for the simulations, results on the equilibrium structure and reconstructions of the clean Si(100) surface, the potential energy surface for a single Ge adatom on the Si(100) surface, investigation of the possibility of the exchange mechanism for the Ge diffusion on this surface, the influence of the adatom on the buckling of Si dimers, and the structure of the S_B rebonded step.

Keywords: Adsorption; Surface diffusion; Silicon; Computer simulation

1. Introduction

The atomistic study of the epitaxial growth of semiconductors has become a rapidly growing field of surface physics over the last few years. The interest in growth processes has its roots in both fundamental scientific problems and in technological questions posed by the semiconductor industry. The atomistic information currently provided by simulations is certainly comparable with that accessible by the experimental techniques. This trend, which is based on the constantly increasing computing power and an ongoing improvement of theoretical methods, suggests that in the near future modelling will assume an increasingly important role in surface science.

There are two main approaches to atomistic simulations in materials science. They represent two extremes with respect to overall accuracy and computational cost of the simulation. By far the more popular technique is based on the use of empirical or semi-empirical interaction potentials to describe bonding between atoms. These potentials might be of different degrees of sophistication (e.g. two-body or many-body potentials), but they always rely on a specific parametrization of the total energy of the system as a function of atomic positions. Parameters for such schemes are usually selected so as to reproduce chosen experimental properties of the material in question. This approach has the obvious advantage of computational simplicity which allows large-scale computations to be performed at a relatively low cost. The disadvantages, however, are significant since empirical schemes are neither flexible nor transferable. An example will be presented illustrating the failure of the best available potentials for silicon [1,2] in reproducing the ground-state structure of the Si(100) surface and the migration path of either Si or Ge adatoms on it [3,4]. The problem can be traced back to the fact that the potentials are constructed to favour four-fold diamond structure coordination and the accuracy of modelling a system with dangling bonds using these potentials was never guaranteed. Another aspect of the same problem is an intrinsic difficulty in finding a suitable parametrization for systems containing different atomic species. A possible solution is to create a database of energies for a set of structures using first-principles techniques and to fit an empirical potential to these data. Still one can never fully rely on empirical potential as a predictive modelling tool. The chosen potential is always fitted to a restricted set of data and there is no assurance that it will produce accurate results for a completely different set of properties.

The best way to reproduce the rich variety of experimentally observed effects in the simulation is to use the quantum mechanical description of the ion–electron interactions in a solid. The advantages and drawbacks of this technique are directly opposite to those of empirical potential schemes. Ab initio methods are computationally expensive which has until recently delayed their appearance as a major modelling tool. As a consequence, the definition of a large scale study differs from that for empirical methods. The largest systems studied so far from first principles involve geometry optimization for

SSDI 0040-6090(95)06960-7

supercells with up to 600–800 atoms using parallel supercomputers [5]. Empirical potentials allow one to perform routine molecular dynamics (MD) simulations of much bigger systems on a workstation. However, the advantage of being robust and reliable outweighs the higher computational cost of ab initio methods.

This paper is arranged as follows. Section 2 is devoted to the particular ab initio technique we use, namely to the pseudopotential total energy method [6]. This section gives an introduction to the technique and includes a brief description of the features that distinguish our approach from other methods. A review of results on the structure and energetics of the clean Si(100) surface is given in Section 3. The results on adsorption and surface diffusion of the Ge adatom on the same surface are presented in Section 4, where both hopping and exchange mechanisms for diffusion are considered. Section 5 deals with the structure of rebonded S_B step on the Si(100) surface and with the Si adatom diffusion across the step edge. Finally, discussion of the prospects of ab initio modelling in surface physics is given in the final section.

2. Pseudopotential total energy method

The calculations presented in the following sections were performed within the density functional theory (DFT) formalism using the computer code CASTEP/CETEP developed in the TCM Group in the Cavendish Laboratory, University of Cambridge, for performing ab initio total energy calculations. These calculations solve the quantum mechanical equations for the electronic states for systems containing arbitrary arrangements of atoms so that the total energy, the electronic charge density and many other physical properties can be computed. The only input to the calculations are the atomic numbers of the constituent atoms in the system. The calculations are performed on periodically repeated unit cells usually referred to as supercells. The ionic potentials are replaced by pseudopotentials which are effective potentials that act only on valence electrons in the system. The electronic wave functions are expanded in terms of a plane-wave basis set and electron–electron interactions are included through the use of DFT.

One of the most important physical quantities that can be computed from a total energy pseudopotential calculation is the force on each atom. The combination of the use of pseudopotentials and plane-wave basis sets makes the calculation of the forces on all the atoms in the system straightforward [6]. This is a critical point when performing calculations on complex systems since optimization of the atomic configuration is a crucial step and it only can be carried out efficiently if the forces are easily determined.

The key ingredient of the method is the use of an iterative minimization technique, the conjugate gradient scheme [6], to solve the DFT one-electron equations. The underlying idea of the DFT approach is that the total energy of the system is a function of its charge density which has a global minimum when the system is in its ground state. A practical consequence of this general statement is that the electronic ground state can be found from the minimization of the total energy. The code performs iterative minimization by updating a single wave function at a time [7], although it is also possible to update all wave functions simultaneously [8] at the cost of using more computer memory than in the band-by-band scheme.

Considerable effort has gone into improving the pseudopotentials used in total energy calculations. One drawback of pseudopotentials is that they are angular-momentum dependent, or non-local. Since it is computationally very expensive to determine the angular momentum components of each electronic wave function around every atom, special techniques were developed for reducing the cost of this operation. The scheme we use here is based on the Kleinman–Bylander representation [9] that reduces the number of non-local projections that have to be performed. Further reduction in computational cost is achieved by performing these projections in real space and exploiting the finite range of the non-locality around every atom [10]. Finally, in most cases it is desirable to optimize pseudopotentials in order to reduce the size of the plane-wave basis set needed in the calculations. This procedure is not required for Si where the original Kerker scheme [11] generates a soft and rapidly convergent potential. However, it was important to apply the so-called three-term optimization scheme [12] to improve convergence properties of the Ge potential [4]. It might be noted that the optimization methods allow one to use pseudopotentials successfully for studies of transition metals or first row elements.

The total energy calculation includes a technical step of integrating the electronic contribution over the Brillouin zone. The special points scheme used here allows the replacement of an expensive three-dimensional integration by a summation over a few specially selected points. The complex systems typically have large supercells and a Brillouin zone is so small that just one special point is sufficient. A rule of thumb suggests using $100/N$ sampling points for non-metallic systems, so in the calculations described below we typically employ a single Γ point.

A specific feature of the CASTEP code is the use of efficient minimizer to find the equilibrium atomic structure simultaneously with finding the electronic ground state. This particular technique performs a conjugate gradient optimization with the second-order polynomial line search in the combined space of atomic coordinates and electronic wave functions. The gradient of the total energy with respect to the atomic coordinates is given simply by the Hellmann–Feinman forces, and the gradient in the electronic subspace can be calculated as a product of the Hamiltonian and a wave function. The standard optimization techniques used converged electronic wave functions from the previous iteration as a starting point for the electronic self-consistency loop at the updated atomic coordinates. The combined space search described above results in a faster structural minimization

largely due to the use of more accurate wave functions at the beginning of each self-consistency loop.

The CASTEP technique allows efficient structural minimization and even MD modelling of systems with few hundred atoms, which involves electronic minimization for about 1 000 wave functions, each being expanded into up to 50 000 plane waves. This system size required massively parallel computers, and most of the results below were obtained using the Intel iPSC/860 and Paragon XP/S 35 at the ORNL Center for Computational Science. Technical details of the parallel implementation are given in Ref. [13].

The following sections illustrate the power of the total energy method when applied to important surface science problems.

3. Properties of the clean Si(100) surface

The structure of the clean Si(100) surface has been a subject of numerous studies employing various techniques. This provides an easy way to compare the results of ab initio and empirical calculations. The empirical description of covalent bonding on the Si surface has a number of fundamental difficulties resulting in the fact that even the lowest energy structure cannot be predicted correctly.

The most pronounced feature of the Si(100) surface is the (2×1) reconstruction that involves formation of Si dimers. The creation of symmetric dimers lowers the energy by about 2 eV per dimer as a result of reduction in the number of dangling bonds on the surface (this energy gain has been calculated from first principles by different authors producing very similar results of 2.18 [4], 1.96 [14], 2.10 [15], and 2.02 [16] eV per dimer). The symmetric dimer structure was for some time considered to be the lowest energy structure as supported by room-temperature scanning tunnelling microscopy (STM) results [17] and by some first principles calculations [18]. However, low-temperature STM experiments demonstrated unambiguously that the dimers at equilibrium are asymmetric (buckled) [19], and the room-temperature observations refer to dimers rapidly flicking between the two possible buckled orientations. Further accurate ab initio studies [20] showed that the symmetric ground state structure was an artefact of too small a basis set used in earlier calculations [18]. Empirical potentials are unable to predict the asymmetric structure in view of a rather small energy difference between the two structures (about 0.1 eV per dimer) and electronic charge density redistribution that is involved in the buckling process.

The main goal in modelling the processes on the Si(100) surface is to gain physical insight into fundamental aspects of crystal growth, so ultimately one is interested in simulations of large systems over substantial time intervals. At present such studies invariably require empirical description of atomic interactions in view of computational demands, so it is rather unsatisfactory not to be able to predict even the ground state of the surface in question. We show below that the discrepancies between the first principles and empirical results are substantially more serious for calculations involving surface adatoms. Interestingly, the same empirical potentials [1,2] produce reliable structural data and the essentially correct energetics for the edge dislocation dipole in silicon [5], a system with complex bond rearrangements but without dangling bonds. This indicates that there is an intrinsic error in empirical treatment of dangling bonds, and large-scale empirical modelling should in some way rely on ab initio results.

The clean Si(100) surface represents a convenient test object for any technique in view of the wealth of already available information. Let us briefly review the known structural features and compare them with our results.

Accurate local density approximation (LDA) results of the DFT were reported recently by Northrup using a total-energy pseudopotential scheme [21]. He found the $c(4\times2)$ structure to be lower in energy than the symmetric dimer surface by $\Delta E_{\text{asym}}=0.14$ eV per dimer, with the surface energy being 1.4 ± 0.1 eV per (1×1) cell. The dimer buckling is very pronounced, the up and down atoms of the dimer are separated by a $\Delta z=0.69$ Å. The dimer bond length, d, was found to be 2.29 Å which is approximately 2.6% shorter than the bulk bond length of 2.35 Å. The calculated angle α between the dimer and the horizontal plane was 17.5°. A previous LDA study of the same $c(4\times2)$ structure with a smaller basis set [22] gave slightly smaller buckling with $\Delta z=0.54$ Å, $d=2.27$ Å and $\alpha=13.4°$. This is consistent with the tendency of a reduced buckling driving force with decreasing basis set size [20]. Previous tight binding calculation results are $\Delta z=0.63$ Å, $d=2.35$ Å, and $\alpha=15.6°$ [23], and this geometry was shown to be consistent with the medium energy ion scattering data. However, other theoretical studies do not support the idea of the dimer length being the same as the bulk bond length.

A number of studies have been done on other buckled structures such as $p(2\times2)$, $p(2\times1)$, etc. The $p(2\times2)$ structure illustrated in Fig. 1 is only slightly higher in energy than $c(4\times2)$. The main difference between the two structures is the absence of antiferroelectric ordering between dimers in the two parallel dimer rows. Using the results from Refs. [16] and [21] we can find the following hierarchy of interaction energies. The main energy gain comes from the dimerization, and symmetric dimers are about 2 eV per dimer lower in energy than the unreconstructed surface. The (2×1) buckling which involves all dimers being tilted in the same direction adds $\Delta E_1=0.06$ eV per dimer. The alternating ordering of dimers within a row creates an extra energy gain of $\Delta E_2=0.03$ eV per dimer which adds up to the 0.09 eV per dimer energy difference between the $p(2\times2)$ and (2×1) symmetric structure. The consequent ordering between rows which is responsible for stabilisation of the $c(4\times2)$ structure further reduces the energy by $\Delta E_3=0.05$ eV per dimer. Since this effect is already dangerously close to the accuracy of ab initio methods, and the main effect of dimerization itself is

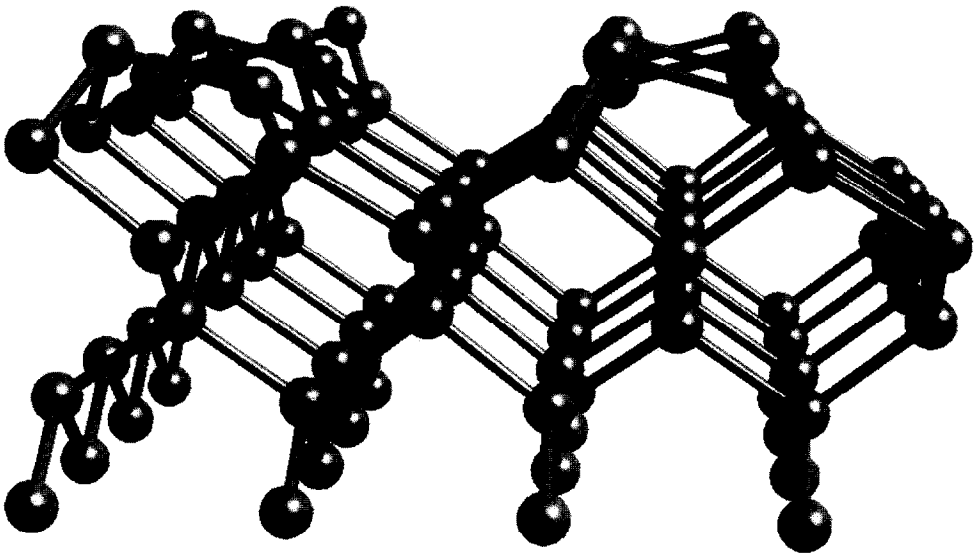

Fig. 1. Perspective side view of the p(2×2) structure of the Si(100) surface.

much stronger, we consider p(2×2) surface as the starting point in most of our calculations.

Results similar to those cited above were reported by different authors. Wang et al. [14] employed the pseudopotential technique and found asymmetric (2×1) dimers to be lower in energy than the symmetric ones, with $d=2.27$ Å. A similar study with the well-converged basis set [16] found the (2×1) structure to have $\Delta E_1=0.09$ eV per dimer and $\alpha=15°$. The angle calculated using smaller basis set was found to be 12° which is consistent with our previous comments on the relation between the basis set and the degree of buckling. A number of calculations were carried out using the cluster approach. For example, Tang et al. [24] employed an all-electron total energy code DMOL to study clusters containing up to 63 atoms. This approach does not seem to be promising since long-range relaxations are present at the Si surface and it is far from clear what inaccuracy is introduced by modelling the surface as a hydrogen-terminated cluster. Consequently the calculated energy gain for buckling $\Delta E_1=0.02$ Å per dimer is too small, and the buckling amplitude of 0.36 eV does not agree with other ab initio results.

We modelled the p(2×2) surface structure using a supercell containing 12 layers of silicon atoms, two k points for the Brillouin zone sampling and a 10 Ry cutoff energy for the plane waves. The buckling amplitude was found as 0.66 Å, with the dimer bond length of 2.27 Å so that the angle was 16.9°. These parameters are in obvious agreement with other pseudopotential studies mentioned earlier. The energies involved into the buckling process were $\Delta E_1=0.05$ eV per dimer and $\Delta E_2=0.04$ eV per dimer. We also performed a calculation for the c(4×2) surface where only one k point was used. This gave $\Delta E_3=0.03$ eV per dimer, so that the total energy difference between the ground-state asymmetric configuration and the symmetric model was $\Delta E_{asym}=0.12$ eV per dimer. The buckling parameters were very close to those of the p(2×2) structure, i.e. $\Delta z=0.68$ Å, $d=2.28$ Å, and $\alpha=17.4°$. These results demonstrate that a large body of independent total-energy pseudopotential calculations produces the same answers for the structure and energetics of the Si(100) surface when sufficient care is taken of the computational convergence issue.

It is always tempting to compare theoretical results with the experimental data in order to verify the accuracy of the chosen computational scheme. However, in this case there is no direct measurement available to give even such basic information as the dimer bond length directly. Experimental errors are in many cases bigger than those of theoretical results since there is a number of assumptions involved in the interpretation of measurements. For example, elastic low-energy electron diffraction (LEED) data give $d=2.47$ Å, but the authors quote even as low value as 2.13 Å as being compatible with their results [25]. The same study gave a buckling amplitude of 0.36 Å which cannot even be compared directly with the calculated values since experiment was not performed at low enough temperatures to freeze rapidly flicking dimers. Another dynamic LEED study quotes the dimer bond length as 2.45±0.1 eV [26], while the medium energy ion scattering results produce $d=2.36$ Å [23] and a recent photoemission EXAFS (PEXAFS) study gives 2.20±0.04 Å [27]. Our conclusion is that the computational experiment definitely agrees with the "real" measurements within the uncertainty of the latter, and at present the theoretical approach appears to be more powerful in terms of the amount of information and the depth of physical insight gained.

4. Adsorption and surface diffusion of the Ge adatom on the Si(100) surface

The study of atomistic processes in the Si–Ge epitaxial growth is of great scientific and technological interest [28–30]. In particular, a number of fundamental problems related

to interface sharpness [28] or ordering during alloy growth [29] have to be explained on a microscopic level. The technological interest in this system is related to the recent progress in the fabrication of the molecular beam epitaxy (MBE) grown Si/Ge superlattices aimed at the direct-gap silicon-based material for optoelectronic applications. One of the most fundamental questions is the identification of the binding sites for Ge adatoms on the Si (100) surface and the determination of the activation energy for surface diffusion. Recent STM observations suggest that the behaviour of a Ge adatom is essentially similar to that of a Si adatom. The Ge islands have the same anisotropy of 1000:1 at typical growth temperatures as the Si islands, and the diffusion barrier in the fast direction is estimated to be 0.62 eV [30] as compared with 0.67 eV reported for Si surface diffusion [31].

Such STM measurements provide an excellent starting point for subsequent theoretical study which can provide new insight into atomistic diffusion processes. The STM data on surface diffusion barriers are derived from an indirect experiment where the temperature dependence of the diffusivity is deduced from the island growth rate. These measurement on Si surfaces are always performed in a rather narrow temperature range which introduces a statistical error into the procedure of fitting experimental data to the Arrhenius dependence. Furthermore, these results give no information whatsoever on the atomistic processes associated with surface diffusion.

There exists a number of theoretical studies of the adatom diffusion on the Si(100) surface, mainly performed using empirical potentials. The scarcity of ab initio data on single adatom behaviour can be explained by the necessity to introduce the periodic boundary conditions in the system. The corresponding supercells need to be rather large in order to avoid artificial interactions between periodic images. As a result, studies of clean surfaces or of a single monolayer adsorbate coverage became a favourite object for the first principle investigations. Applications of ab initio techniques to the single adatom case have utilized at most $p(4\times4)$ surface cells [4], in the worst case scenario cells as small as $p(2\times1)$ have been used [32] which corresponds to 50% coverage rather than to the isolated adatom case.

Empirical schemes easily overcome the system size problem, but the accuracy of the description of the adatom–surface interaction leaves much to be desired. The best illustration of the problem is a compilation of the results for the Si adatom diffusion obtained using empirical potentials of the Stillinger–Weber (SW) [1] or Tersoff (T) [2] type. The calculated Si adatom diffusion barrier vary appreciably as 0.67 eV (SW) [33], 0.75 (T) [34], 0.24 eV (SW) [35], 1.1 eV (T) [36], 0.6 eV (modified T) [36], 0.33 eV (modified SW) [37], or 1.2 eV (SW) [38]. Furthermore, these studies predict different binding sites for the Si adatom, different migration paths (on top of dimers, in the interdimer trough, or a zig-zag path), and different degrees of the surface diffusion anisotropy.

One can expect that the Si–Ge interaction is described by empirical potentials even less accurately since it is difficult to find a reasonable fitting scheme for the parameters of the Si–Ge interaction. These potentials predict the barrier for fast migration of Ge on Si to be 0.73 eV (T) [39] or 0.64 eV (SW) [40]. The barrier for diffusion across the dimer rows is calculated to be 1.17 eV and 0.80 eV, respectively. These barriers correspond to the anisotropy of the surface diffusion at 500 K of either 20 000:1 [39] or 50:1 [40] if we assume that the pre-exponential coefficient is the same in both directions. The experimentally observed anisotropy is of the order of 1 000:1 [30] at this temperature and is not reproduced correctly by either of these empirical potentials.

An ab initio study of the surface diffusion appears to provide a more reliable approach. The experimentally observed anisotropic diffusion of Si adatoms was successfully explained in the pseudopotential total energy study [3] which involved the calculation of the potential energy surface for the surface cell with 8–16 atoms (98 atoms in the supercell). The diffusion barriers were found to be 0.6 eV in the direction along dimer rows and 1.0 eV in the perpendicular direction, with the estimated accuracy of 0.1 eV. These numbers are in good agreement with the experimental observations with respect to both the fast diffusion barrier and the diffusion anisotropy [31]. The binding geometry and the migration path are, however, completely different from those predicted using empirical potentials.

We performed a similar study for Ge adatoms using a slightly bigger supercell [4]. It was constructed in the (100) slab geometry and included 12 layers of Si with the inversion symmetry imposed and 8 layers of vacuum. This supercell with the $c(4\times4)$ surface cell contained 192 Si atoms and two Ge adatoms (one on each side of the slab). Only the atoms in the two innermost layers were kept fixed, the coordinates of the rest of the atoms were optimised during the structural relaxation. The energy surface was mapped out by calculating the total energy for positions (x, y) of the adatom in the irreducible quarter of the $p(2\times1)$ surface cell shown in Fig. 2. These positions form an equidistant grid with a spacing between the grid points of 0.96 Å in either direction. We start by placing the adatom 2.5 Å above the surface, and proceed with the simultaneous total energy minimization in the space of ionic coordinates (including the adatom's z coordinate) and wave-function expansion coefficients. The local minima were found by allowing the (x, y) coordinates of the adatom to relax starting from the nearby grid point. The energy surface was further refined by mapping out the regions around the local minima on a finer grid. The final energy surface (Fig. 3) was constructed using triangulation which utilises the gradient values, i.e. the force on the adatom [38].

We found only two binding positions for the adatom. The global minimum is located at point M on top of the second layer atom, while the pedestal site H is only 0.06 eV higher in energy. However, we do not expect a high occupancy for the pedestal site due to the small volume of the configuration space associated with it. Contrary to the case of Si adsorption,

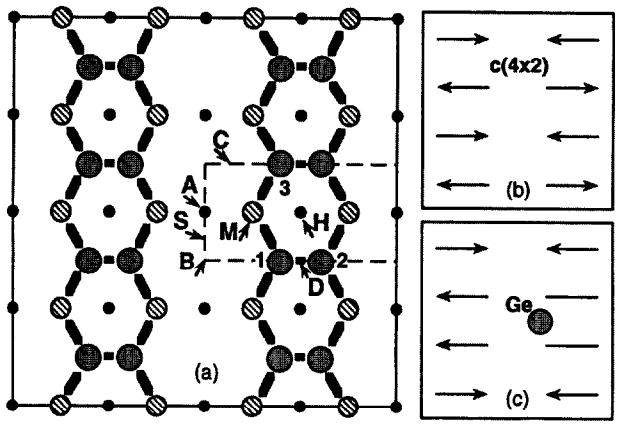

Fig. 2. (a) Top view of the three top layers of the c(4×2) reconstructed Si(100) surface. The (4×4) surface cell is shown by the solid line, p(2×1) by the dashed line. Letters correspond to critical points on the energy surface (see text). The buckling sequence is represented schematically for the clean surface (b) and in presence of the adatom at site M (c). The arrows point toward a higher atom in the dimer.

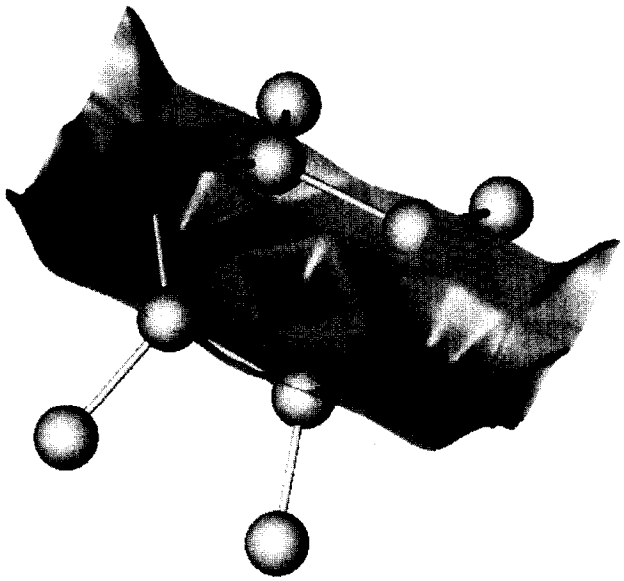

Fig. 3. The energy surface for the Ge adatom (from Ref. [4]). The first and second layer Si atoms from the p(2×1) cell are shown. The central minimum corresponds to the site H, two equivalent side minima correspond to the binding site M. The highest maximum is at the long bridge site B.

the point C that was found to be a local minimum [3] is now only a saddle point. No binding site was found in the channel between the dimer rows, similar to ab initio results for the Si adsorption [3] and contrary to the predictions of the empirical potential simulations [33,38]. The long-bridge site B corresponds to the absolute maximum and is 1.90 eV higher in energy than the binding site M. The cave site A represents another maximum with the height of 1.21 eV. These results contrast sharply with empirical simulations where the long-bridge site is the absolute minimum [39,40] and the fast diffusion path is in the middle of the interdimer channel, B–A–B. The same discrepancy exists between ab initio [3] and empirical simulation results [33–38] for the Si adatom. We have performed an extensive search of the lower-energy configuration for the B-site adsorption in order to resolve this controversy. We found that our original relaxation strategy produced only one of the two possible B-site configurations with the adatom being 0.5 Å higher than the dimers, B*. The other configuration, B_0, is obtained by placing the adatom below the surface. The equilibrium position is then 0.9 Å below the dimers level, and this configuration is more stable by 0.18 eV than the B*. A similar double-well energy profile was reported in the SW simulations of the Si adatom moving along the z axis at the B site [33,35]. However, there is a barrier between the two configurations so one does not expect the B_0 configuration to be accessible during growth, and furthermore it is still much higher in energy than the predicted binding site and is not even a local minimum on the potential energy surface.

The diffusion path in the direction along the dimer rows is found [4] to be a zig-zag between the point M and the dimer bridge site D which is the saddle point of the trajectory, and no visit to the H point is necessary, a fundamental difference from the case of the Si adatom [3]. The barrier height is 0.62 eV in excellent agreement with the experimental estimate [30]. There are two consecutive inequivalent saddle points on the trajectory for the slow diffusion across the dimer rows as can be seen from Fig. 3. The experimentally observed barrier is determined by the highest of the two which corresponds to the hopping across the channel at point S and has a height of 0.95 eV. These results lead to a diffusion anisotropy of 2 000:1 at 500 K which agrees well with the STM results [30]. Thus it is possible to conclude that the experimental data are consistent with Ge adatom diffusion that takes place on top of the dimers and not in the interdimer channel. Note that the SW potential predicts the barrier for such zig-zag path of only 0.24 eV [40] in complete disagreement with the experiment.

The bonding geometry seems to be the most controversial issue in the comparison of ab initio and empirical results. Our data indicate that the bond length between the Ge adatom and the surface Si atoms is normally close to the Si dimer bond length. In the equilibrium adsorption geometry, the Ge adatom at site M is bonded to two dimers from the same row, and all bond lengths are about 2.39 Å. The distance from the adatom to the second-layer Si atom is very similar (2.46 Å), but there is no directional bond to this atom according to the analysis of the charge density distribution. This explains why empirical potentials do not recognise the site M as the absolute minimum. Both SW and T potentials give strong repulsive three-body interactions for this configuration, while in fact there is no third bond involved. Adsorption at the pedestal site H is characterised by the formation of four long bonds of 2.48 Å and by the dimer bond stretching to 2.52 Å. The latter effect is more pronounced than for the Si adsorption [3] due to the bigger size of the adatom.

We find that the adatom causes maximum strain on the dimer bond when it is located at the saddle point S for the hopping across the trough or in its vicinity. In this configuration the distance between the dimer atoms becomes 3.3 Å

which is already closer to the bulk distance of 3.84 Å than to the dimer bond at the clean surface, 2.30 Å. According to the charge density distribution we consider this configuration as a broken dimer. No such effect induced by the Si adatom was reported by Brocks et al. [3]. It is interesting to compare these results with the empirical potential modelling data on the adatom-induced dimer opening. This issue has been studied in some length as a possible mechanism for Si layer by layer growth on the Si(100) surface. The SW potential predicts the dimer opening with the Si adatom at any position along the H–D line [35]. The results obtained with the T potential are essentially similar, the only difference is that the absorption at H site or in its immediate vicinity does not lead to the dimer opening [34]. The T potential predicts qualitatively the same picture for Ge adsorption as for the Si adatoms: dimers always open when there is an adatom on the H–D line except at the H site itself [39]. It is interesting that the silicon dimer stretching for the H-point adsorption as found with T potential is the same as in the present work. However, we have seen no signs of the dimer opening for adatom moving from H to D site. The bond length of the dimer comprised of atoms 1 and 2 (Fig. 2) decreases monotonically from 2.52 to 2.39 Å while the adatom advances from H to D.

The role of the dimer buckling on the energetics of adsorption and surface diffusion was addressed for the first time in our study of Ge adsorption. This problem has not been considered previously for a number of reasons. In ab initio investigations it was dismissed mainly because of the small energy effect involved, but also because of the prohibitively large size of the corresponding supercell [3]. Empirical simulations, on the other hand, cannot approach this problem as they predict symmetric reconstruction as the ground state of the Si(100) surface.

As mentioned earlier, symmetric dimers were used as a starting point for the energy surface calculations. However, during atomic relaxation we always observe spontaneous dimer buckling initiated by numerical noise in the calculated ionic forces. The character of this buckling depends on the adatom position. With the adatom at point H the second dimer row in the unit cell is only slightly perturbed compared with the clean asymmetric (2×1) reconstructed surface. The dimer bond length in that row is 2.28 Å and the buckling amplitude is 0.70 Å. Buckling is entirely absent within the four dimers of the row that contains the adatom, although we have no doubt that with the longer cell one would observe a buckled dimer row with a localised symmetric part around the adatom. The picture is qualitatively different in the vicinity of the Ge adatom at the binding site M (Fig. 2(c)). In this case both rows are affected. The dimers bonded to the adatom are slightly tilted (as is the case for the Si adatom [3]) by 0.16 Å. The buckling amplitude grows to 0.43 Å for the next dimers in the same row, so we can be certain that the influence of the single adatom on buckling is localised within approximately four dimers. A new effect can be seen as a result of the relative proximity of the Ge adatom to the next dimer row. Strong in-phase buckling (0.62 Å) is induced in the two dimers nearest to the adatom. Obviously, there is no direct bonding between these dimers and the Ge atom, and the driving force for this configuration should come from the second-layer atom displacements. In this case elastic strain caused by the presence of the adatom prevails over the energy gain related to the antiferroelectric ordering of asymmetric dimers.

The resulting configuration of two adjacent dimers buckled in the same direction has been observed indeed on Si (100) at low temperatures [19,41,42]. It corresponds to the so-called defect "C" [42] which is essentially a fault in the stacking of alternately buckled Si dimers. Such defects are seen at low temperatures, when all dimers are buckled, but they also persist at higher temperatures when the rest of the surface is seen as symmetrically reconstructed. Since it would be expected that a Si adatom at site M produces qualitatively the same structure as a Ge adatom, and Si adatoms are typically present in the STM studies of vicinal surfaces, we suggest that these defects observed in STM [19,41,42] might correspond to single Si adatoms. It is important for creation of the "C" defect that the Si (or Ge) adatom is located offside of the dimer row and on top of it (and not in the middle of the trough). Other adatoms are known to occupy more symmetric positions (e.g. sites H, A or B) which are unlikely candidates to be responsible for the "C" defect creation [43]. Thus imaging of these defects potentially gives the only experimental information on the binding site of the Si or Ge adatom and could be used in atomistic STM study of surface diffusion. It should be noted that the discovery of the adatom influence on buckling is due to the fact that we used the supercell twice as large as the one employed in the previous study of a Si adatom [3].

The study of surface diffusion would be incomplete if only surface-hopping processes were considered. However, important alternative mechanisms might also occur such as an exchange diffusion process which is theoretically possible but has so far been proven to be operational only in diffusion on fcc metal surfaces [44–46]. Exchange diffusion in such systems was observed experimentally using field ion microscopy [44], which was followed by the semi-empirical embedded atom simulation [45] and finally by the ab initio pseudopotential study [46]. The mechanism of the exchange diffusion is simple—a diffusing adatom pushes out the nearest surface atom and takes its place, while the latter becomes an adatom travelling around the surface. This might be the preferred diffusion mechanism if the barrier for exchange is lower than for hopping provided that the energy of the surface with one substituted atom is not much higher than that of the ideal surface. Of course, in the case of surface self-diffusion the total energy argument disappears and one has simply to compare diffusion barriers.

The exchange process might be potentially important for explanation of a number of effects observed during Si–Ge alloy growth, and it is reminiscent of the so-called pump mechanism responsible for alloy ordering [29]. However, the calculation of the barrier for exchange diffusion on a flat

Si(100) surface is more difficult than in the case of fcc metal surfaces [46]. The problem is to find a plausible path for such process and a well-defined saddle point for diffusion which is non-trivial in view of the complex surface structure. Attempts to move an adatom in the direction of the nearest surface atom create enormous strain in the silicon framework and result in a very high barrier. On the other hand, comparison of the total energies for the Si surface with Ge adatom and for the same surface with one Si atom interchanged with Ge shows that the latter is only 0.11 eV per adatom higher in energy than the former. This energy difference is sufficiently low to expect that the preferential diffusion mechanism (hopping vs. exchange) is defined by kinetic arguments rather than by the thermodynamic ones.

We attempted two different approaches in the study of exchange diffusion of Ge on the Si(100) surface. First of all, we performed a molecular dynamics run in order to visualise any possible exchange events. The supercell for this simulation was smaller than the one used for static calculations as we had to decrease the number of layers from 12 to 8 in order to make the simulation computationally feasible. The MD simulation included 1 ps of initial equilibration followed by about 10 ps of production run during which the temperature was gradually increased from 700 K to 1 200 K. During this simulation we observed effects that could be interpreted as surface melting of Si, e.g. the height of the first peak in the radial distribution function for surface silicon atoms at 1 200 K is eight times smaller than that at 700 K. However, the distribution of Ge–Si bonds stays largely unchanged and centred around 2.38 Å. No exchange events were recorded during this run. Current computational restrictions do not allow longer ab initio simulations for a system like this which contained 130 atoms and a vacuum region. Consequently we tried another approach based on static total energy calculations.

First we attempted simple displacement of the adatom towards one of the Si atoms it is bonded to, marked as atom 1 in Fig. 2. As we mentioned before, this approach was unsuccessful, so we tried a different strategy. We started from the geometry obtained during the potential surface calculation, where the adatom was located approximately midway on its zig-zag trajectory from the ground state M to the dimer bridge site D. The adatom in this starting geometry has just undergone a bond switching and is bonded only to two silicon atoms from one dimer (atoms 1 and 2) and not to two different dimers (atoms 1 and 3) (see Fig. 2). This configuration is still rather low in energy, being only 0.22 eV higher than the ground state and 0.40 eV lower than the saddle point. Of course, this configuration has unbalanced forces on the adatom in the x–y plane. The next step is to change gradually the height of the adatom with respect to the surface and to allow all coordinates to relax at each fixed value of the adatom's height. We anticipated that such path might eventually lead to bond breaking between the silicon atom 1 (see Fig. 2) and the second layer atoms, which would result in Ge replacing atom 1 and pushing it to the M site in the next (2×1) cell. However, we were able to force adatom so deep into the surface that it was 1 Å below the surface atoms with still no trace of the expected bond breaking. The total energy went up by nearly 2 eV per adatom during this process, which indicates that the chosen path cannot lead to a successful exchange process.

The results of our dynamical and static modelling strongly suggest that there is no convenient exchange path for the Ge adatom surface diffusion on the Si(100) surface. More advanced techniques of the transition-state search might be applied to find the exact value of the barrier for this process, but it is qualitatively evident that it would be substantially higher than the hopping barrier in any direction. We expect the situation to be different in the presence of steps which is the case for fcc metals [7,45] so the present negative result does not directly affect assumptions used in developing the pump mechanism concept [29].

To conclude this section, a total-energy pseudopotential study allowed us to identify the geometry and energetics of Ge binding and migration on the Si(100) surface. We showed that surface hopping is a preferred diffusion mechanism and exchange events on a flat surface are highly unlikely. Finally, buckling defects observed in STM were interpreted as being due to the presence of an adatom bonded to the adjacent dimer row.

5. Surface diffusion in the presence of steps

The geometry of adatom migration on a flat surface and associated energy barriers described in the previous section provide the most basic level of information required for phenomenological modelling of crystal growth. However, all realistic growth models should take into account an interaction between diffusing adatom and surface steps. As early as in 1960s the concept of the step barrier, the so-called Ehrlich–Schwoebel barrier, was developed [47]. Its value represents the difference between the energy barrier for descending diffusion across a step edge and the barrier for diffusion on a flat terrace. On the simplest level, the positive step barrier implies that islanding is a plausible growth mechanism, while a negative step barrier should correspond to a layer by layer growth since it is easy for adatoms to descend to the lower terrace during growth. A number of sophisticated modelling approaches have been recently developed that rely on numerical estimates of the Ehrlich–Schwoebel barrier [48–50]. Consequently the question of accurate determination of the barrier becomes very important in view of the obvious interest to the epitaxial growth modelling.

The Si(001) surface has a number of different kinds of steps, and here we restrict ourselves to the study of Si adatom diffusion across the so-called S_B rebonded step [51] which is known to play an important role in kinetics of epitaxial growth. The equilibrium structure of this step with symmetric dimer reconstruction is shown in Fig. 4. The dimer rows on the upper terrace are perpendicular to those on the lower

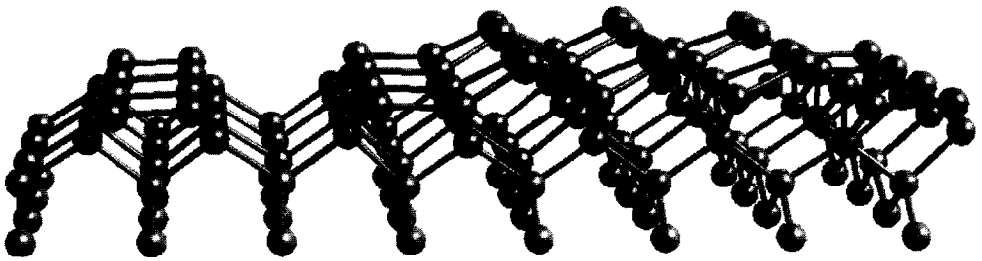

Fig. 4. Relaxed structure of the S_B rebonded step on the symmetrically reconstructed Si(100) surface.

terrace, and the term "rebonded" simply means that there are bonds at the step edge between the upper and the lower terrace dimers. Diffusion across such step has been studied only using empirical potentials [52,53] since the required supercell is quite big which prohibits routine ab initio studies.

Existing results of empirical calculations of the energetics of the Si adatom diffusion across the S_B rebonded step are as contradictory as they are for a flat surface. The energy surface obtained using a SW potential predicts a descending path along the trough on the upper terrace with the barrier of 0.76 eV to overcome at the step edge. The same potential was found to give a 0.67 eV barrier for fast diffusion on a flat surface [33], which means that a positive Ehrlich–Schwoebel barrier of 0.09 eV is predicted. A totally different geometry is predicted using the Tersoff potential [53]. To begin with, in the studies of a flat surface this potential predicted the dimer bridge site D to be nearly as low in energy as the long-bridge site B [34]. (One cannot help but notice that empirical potentials find all maxima of the ab initio potential energy surface to be binding sites.) In the presence of the step edge the dimer bridge site was found to be most stable on the upper terrace, but the descending trajectory from it would have a very high barrier of 2.05 eV. The barrier for diffusion on top of the dimer row on a flat surface was previously found to be 1.03 eV [34], the barrier for the trough diffusion was 0.75 eV, so that a positive Ehrlich–Schwoebel barrier of 1.02–1.30 eV can be deduced. A more sophisticated descent route is possible that involves two jumps and gives the Ehrlich–Schwoebel barrier of 0.2–0.5 eV. The conclusion from this analysis is that the diffusion characteristics at the S_B rebonded step depend strongly on the choice of the interatomic potentials, but in both reported calculations the barrier for descending path was estimated to be greater than that for the terrace diffusion. It is slightly surprising since experimentally no development of large scale mounds have been found for the Si(100) homoepitaxy.

We report here preliminary results of the ab initio study of the descending diffusion trajectory for Si adatom at the S_B rebonded step. Half of the supercell used is shown in Fig. 4. It contains 282 atoms altogether, with six dimers on the upper terrace (two rows three dimers long), four dimers (one row) and another eight rebonded dimers (two rows of four) on the lower terrace. We start by placing the adatom at the site M on the upper terrace and allowing for complete relaxation (which involves substantial dimer buckling). The adatom is further forced towards the step with all coordinates allowed to relax except for the diffusion coordinate. The diffusion trajectory is illustrated in Fig. 5 which shows a familiar zigzag path from the site M through the dimer bridge site and further to the binding site on top of the rebonded dimer row. This trajectory has a metastable site on it which is approximately 0.2 eV higher in energy than the binding site on the upper terrace. This site is encountered after the adatom passes through the dimer bridge site at the edge of the step, and its existence reflects the difficulty of breaking a bond to the step edge dimer. The overall barrier for this trajectory is 0.42 eV, while the value we calculated for the flat surface was 0.63 eV (close to another ab initio result of 0.6 eV [3]). This means that the total energy pseudopotential scheme predicts a qualitatively different behaviour from that given by empirical potentials. We find a negative Ehrlich–Schwoebel barrier of 0.21 eV and the energy gain of about 0.12 eV due to a higher binding energy next to the step edge compared with that on the flat surface. We believe these results are consistent with the experimentally observed features of the epitaxial growth on the Si(100) surface.

6. Discussion

The results analyzed in this article amply illustrate possible applications of the total energy pseudopotential method relevant to surface physics studies. It is instructive, though, to look at this technique in its historical perspective. The basic methodology was developed during the 1960s and 1970s. It was used by a small number of groups in the early 1980s but their calculations were restricted to semiconductor systems containing of the order of 10 atoms in a supercell (Fig. 6). A generally held opinion at this time was that the pseudopotential technique could not be applied to first row elements or to transition metals. These limitations were due to the fact that at the time the electronic states were determined by diagonalising the Hamiltonian matrix. The largest matrix size that could be diagonalised was of the order of 1 000. As at least 100 plane waves per atom are needed to represent the electronic orbitals in a total energy calculation, this matrix size just corresponded to a supercell with 10 atoms which is definitely insufficient for calculations on realistic systems.

Although restrictions on memory and computational speed have eased with modern computers, the reason that the total

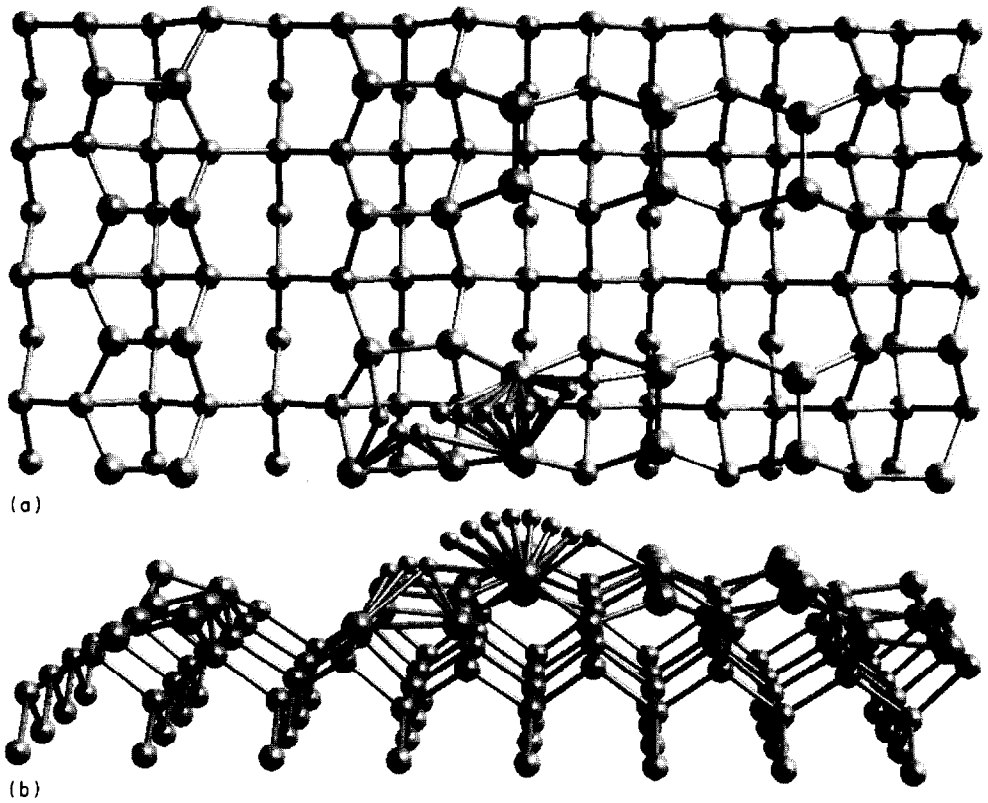

Fig. 5. Diffusion trajectory for Si adatom across the S_B rebonded step: (a) top view; (b) side view.

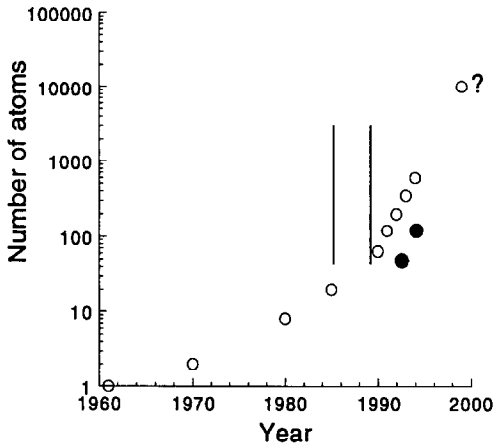

Fig. 6. System size accessible for ab initio total energy study as a function of time. Open circles correspond to static geometry minimization, filled circles to ab initio molecular dynamics studies. Data after year 1990 mainly refers to CASTEP/CETEP applications. Vertical lines correspond to publications of Refs. [7] and [54].

energy pseudopotential method has become so powerful is primarily due to a complete change in the numerical methods used to solve the DFT equations. These changes were initiated in 1985 by Car and Parrinello [54] who introduced a number of ideas that significantly increased the speed of the calculations. Their method allowed calculations to be performed on systems containing many tens of atoms and also allowed the first ever dynamical simulation to be performed where atomic forces were correctly described by including the quantum mechanics of the electronic states. However, as can be seen from Fig. 6, the real breakthrough in computational performance is due to the development of new iterative schemes [7,8] that provide a faster way of solving one-electron equations than the original Car–Parrinello technique. The combination of a number of algorithmic improvements described in Section 2 makes it possible to carry out the total energy calculations for systems containing up to 100 atoms using desktop workstations and for cells with up to 1 000 atoms using massively parallel supercomputers. The exponential growth in the computational performance over the last few years as illustrated by Fig. 6 projects to the 10 000 atom supercells by the year 2000. More importantly, given the continuing increase in performance at all levels of computers and further algorithmic improvements, it is likely that the calculations for a thousand atoms in the supercell will be routinely performed on desktop machines over the next decade.

Applications of the total energy pseudopotential method proved to be invaluable in complementing experiments and providing otherwise inaccessible physical insight into fundamental atomic scale processes. We are confident that this technique will have great future in surface physics in general and in atomistic studies of the growth processes in particular.

Acknowledgements

We are grateful to M.C. Payne, I. Stich and M.H. Lee for numerous illuminating discussions and for their contribution to the study of Ge adsorption. We would like to thank T.

Kaplan and M. Mostoller for useful discussions on empirical methods and A. Cline for providing a "Faultpack" package (we used the "Faultpack" package by A. Cline and R. Renka to perform triangulation on an irregular grid using the values of the gradient). This research was sponsored by the Division of Materials Sciences, US Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc. This research is part of the Grand Challenge project on first principles simulation of materials properties, funded under the High-Performance Computing and Communications Initiative. Analysis of valence charge densities and of optimised structures was performed using the interface to CASTEP developed by Molecular Simulations.

References

[1] F. Stillinger and T.A. Weber, *Phys. Rev., B36* (1987) 1208.
[2] J. Tersoff, *Phys. Rev. Lett., 56* (1986) 632.
[3] G. Brocks, P.J. Kelly and R. Car, *Phys. Rev. Lett., 66* (1991) 1729.
[4] V. Milman, D.E. Jesson, S.J. Pennycook, M.C. Payne, M.H. Lee and I. Stich, *Phys. Rev., B50* (1994) 2663.
[5] F. Liu, M. Mostoller, V. Milman, M.F. Chisholm and T. Kaplan, *Phys. Rev., B51* (1995) 17192.
[6] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias and J.D. Joannopoulos, *Rev. Mod. Phys., 64* (1992) 1045.
[7] M.P. Teter, M.C. Payne and D.C. Allan, *Phys. Rev., B40* (1989) 12255.
[8] M.J. Gillan, *J. Phys.: Cond. Matter, 1* (1989) 689.
[9] L. Kleinman and D.M. Bylander, *Phys. Rev. Lett., 48* (1982) 1425.
[10] R.D. King-Smith, M.C. Payne and J.S. Lin, *Phys. Rev., B44* (1991) 13063.
[11] G.P. Kerker, *J. Phys., C13* (1980) L189.
[12] L.S. Lin, A. Qteish, M.C. Payne and V. Heine, *Phys. Rev., B47* (1993) 4174.
[13] L.J. Clarke, I. Stich and M.C. Payne, *Comp. Phys. Commun., 72* (1992) 14.
[14] J. Wang, T.A. Arias and J.D. Joannopoulos, *Phys. Rev., B47* (1993) 10497.
[15] S. Ihara, S.L. Ho, T. Uda and M. Hirao, *Phys. Rev. Lett., 65* (1990) 1909.
[16] N. Roberts and R.J. Needs, *Surf. Sci., 236* (1991) 112.
[17] J. Knall and J.B. Pethica, *Surf. Sci., 265* (1992) 156; F. Iwawaki, M. Tomitori and O. Nishikawa, *Ultramicroscopy, 42–44* (1992) 902.
[18] I.P. Batra, *Phys. Rev., B41* (1990) 5 048.
[19] R.A. Wolkow, *Phys. Rev. Lett., 68* (1992) 2 636.
[20] J. Dabrowski and M. Scheffler, *Appl. Surf. Sci., 56–58* (1992) 15.
[21] J.E. Northrup, *Phys. Rev., B47* (1993) 10 032.
[22] Z. Zhu, N. Shima and M. Tsukada, *Phys. Rev., B40* (1989) 11 868.
[23] R.M. Tromp, R.G. Scmeenk, F.W. Saris and D.J. Chadi, *Surf. Sci., 133* (1983) 137.
[24] S. Tang, A.J. Freeman and B. Delley, *Phys. Rev., B45* (1992) 1 776.
[25] B.W. Holland, C.B. Duke and A. Paton, *Surf. Sci., 140* (1984) L269.
[26] C.M. Wei, H. Huang, S.Y. Tong, G.D.S. Glander and M.B. Webb, *Phys. Rev., B42* (1990) 11 284.
[27] P.S. Mangat et al., *Phys. Rev., B47* (1993) 16 311.
[28] D.E. Jesson, S.J. Pennycook and J.-M. Baribeau, *Phys. Rev. Lett., 66* (1991) 750.
[29] D.E. Jesson, S.J. Pennycook, J.-M. Baribeau and D.C. Houghton, *Phys. Rev. Lett., 68* (1992) 2 062.
[30] M.G. Lagally, *Jpn. J. Appl. Phys., 32* (1993) 1 493; Y.-W. Mo and M.G. Lagally, *Surf. Sci., 248* (1991) 313; M.G. Lagally, private communication, 1993.
[31] Y.W. Mo, J. Kleiner, M.B. Webb and M.G. Lagally, *Phys. Rev. Lett., 66* (1991) 1 998.
[32] T. Miyazaki, H. Hiramoto and M. Okazaki, *Jpn. J. Appl. Phys., 29* (1990) L1 165.
[33] C. Roland and G.H. Gilmer, *Phys. Rev., B46* (1992) 13 428.
[34] D. Srivastava and B.J. Garrison, *J. Chem. Phys., 95* (1991) 6 885.
[35] Y.-T. Lu, Z.Y. Zhang and H. Metiu, *Surf. Sci., 257* (1991) 199.
[36] J. Wang and A. Rockett, *Phys. Rev., B43* (1991) 12 571.
[37] C.P. Toh and C.K. Ong, *Phys. Rev., B45* (1992) 11 120.
[38] Z.-H. Huang and R.E. Allen, *J. Vac. Sci. Technol., A9* (1991) 876.
[39] D. Srivastava and B.J. Garrison, *Phys. Rev., B46* (1992) 1 472.
[40] C. Roland and G.H. Gilmer, *Phys. Rev., B47* (1993) 16 286.
[41] D.J. Hamers and U.K. Kohler, *J. Vac. Sci. Technol. A, 7* (1989) 2 854.
[42] M. Chander, Y.Z. Li, J.C. Partin and J.H. Weaver, *Phys. Rev., B48* (1993) 2 493.
[43] G. Brocks, P.J. Kelly and R. Car, *Phys. Rev. Lett., 70* (1993) 2 786; Y. Morikawa, K. Kobayashi and K. Terakura, *Surf. Sci., 283* (1993) 377.
[44] G.L. Kellogg and P.J. Feibelman, *Phys. Rev. Lett., 64* (1990) 3 143; C. Chen and T.T. Tsong, *Phys. Rev. Lett., 64* (1990) 3 147.
[45] C.L. Liu, J.M. Cohen, J.B. Adams and A.F. Voter, *Surf. Sci., 253* (1991) 334; C.L. Liu and J.B. Adams, *Surf. Sci., 294* (1993) 197.
[46] R. Stumpf and M. Scheffler, *Phys. Rev. Lett., 72* (1994) 254; R. Stumpf and M. Scheffler, *Surf. Sci., 309* (1994) 501.
[47] G. Ehrlich, *J. Chem. Phys., 44* (1966) 1 039; R.L. Schwoebel, *J. Appl. Phys., 40* (1969) 614.
[48] M.D. Johnson et al., *Phys. Rev. Lett., 72* (1994) 116.
[49] P. Smilauer, M.R. Wilby and D.D. Vvedensky, *Phys. Rev., B47* (1993) 4 119.
[50] J.E. Vannostrand, S.J. Chey, M.A. Hasan, D.G. Cahill and J.E. Green, *Phys. Rev. Lett., 74* (1995) 1 127.
[51] D.J. Chadi, *Phys. Rev. Lett., 59* (1987) 1 691.
[52] C. Roland and G.H. Gilmer, *Phys. Rev., B46* (1992) 13437.
[53] D. Srivastava and B.J. Garrison, *Phys. Rev., B47* (1993) 4 464.
[54] R. Car and M. Parrinello, *Phys. Rev. Lett., 55* (1985) 2 471.