HARVARD UNIVERSITY

THE GRADUATE SCHOOL OF ARTS AND SCIENCES

THESIS ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Applied Sciences

Department

Committee

have examined a thesis entitled

"A First-Principles Theoretical Study of Microstructural Transformation of Solids."

presented by Yu-min Juan

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature .....

Typed name Professor E. Kaxiras

Signature .....

Typed name Professor J. Hutchinson

Signature .....

Typed name Professor H. Ehrenreich

Date May 18, 1995

A First-principles Theoretical Study of Microstructural Deformations of Solids

A thesis presented

by

Yu-Min Juan

to

The Division of Applied Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Applied Physics

Harvard University
Cambridge, Massachusetts
May 1995

Abstract

The development of density functional theory (DFT) and advances in computational capabilities have made it possible to obtain reliable information on the energetics of structural transformations in solids. However, limitations do exist due to the use of the local density approximations (LDA) to the exchange-correlation functionals. Recently, there have been efforts to try to go beyond LDA by including gradient corrections, which are referred to as the generalized gradient approximation (GGA). To obtain an accurate assessment of these GGA functionals, we investigated in details the applicability of two recently proposed functionals. The effects on the atomic charge distribution were studied. We also examined different approaches to construct pseudopotentials self-consistently in the context of gradient-corrected functionals. We concluded that although GGA functional produce improved results in some cases, in general LDA remains a reliable choice for the energetic calculations in solids.

We considered a range of phenomena related to structural transformations of a prototypical covalent solid, namely silicon. We first studied the energetics for the high-pressure plastic flow of silicon by performing DFT total energy calculations for structural transformations which might correspond to mass flow. To explore the phase space efficiently, the magic strain concept was used. Entropy effects were taken into account with the use of Vineyard's transition state theory. An upper bound for the energy barrier was obtained from our study and has been found to

be lower than the melting point of silicon, suggesting that such transformations may be possible under indentation.

As a second application, we have obtained accurate generalized stacking fault (GSF) energy surfaces for both the (111) and (100) planes of silicon with first-principles calculations. The importance of this GSF energy surface is that it can be used to identify the value of the unstable stacking fault energy γ_{us}. This quantity has been shown by J. Rice and collaborators to be the solid state parameter controlling dislocation nucleation near a crack tip. We also combined the GSF energy surfaces with the Peierls-Nabarro model to study the dislocation properties of silicon.

Finally, in a third application, we performed DFT calculations to simulate the finite size block shearing process within the supercell approximation. The results of this calculation can be used to account for surface energy effects which were neglected in Rice's original theory for dislocation nucleation. We obtained an estimate for the energy associated with the surface creation during dislocation emission. This energy factor was incorporated into Rice's original theory within the framework of continuum mechanics by assuming an evanescent surface stress as suggested by Y. Sun. An estimate of the surface effects on the critical loading was also obtained.

Acknowledgements

This thesis would not have been possible without the help of many people. First of all, I would like to thank my advisor, Professor Efthimios Kaxiras, for his guidance and support over the years. His enthusiasm for science and physical insight into problems has enabled me to learn much from him. I would like to thank Professor Ehrenreich for teaching me the basics of solid state physics and also for many helpful comments and suggestions on this thesis. My sincere thanks go to Professor James Rice for the help and comments during the course of this work, as well as serving on my committee. I also like to thank Professor John Hutchinson for kindly serving on my committee.

I owe many thanks to Professor Alex Antonelli at Brazil for many beneficial discussions and assistance with computer programs. I also have to thank Linda Zeger for a lot help both inside and outside physics. I am indebted to Dr. Yuemin Sun for a lot useful discussions. The discussions with Normand Modine, Dr. Gil Zumbach, Hang Xu, and Lung-Han Peng enabled me to have more understanding on the subject. I would like to thank my fellow graduate students who have made the life at Harvard more interesting and enjoyable.

I would like to thank Kao scholarship for the financial support during my second year in graduate school. My research was also supported in part from ONR grants.

Last but not least, I thank my family and old friends for their support.

To my parents and my wife ching-i,

who have given me much needed emotional

and finicial support over the years.

Contents

1 Introduction

The microstructural deformation of solids is one of the most important problems in materials science due to its scientific interest as well as its importance in high technology applications. The development of density functional theory (DFT) [1] and recent advances in computational capabilities have made it possible to study a variety of problems of this nature using first-principles theoretical calculations. However, an approximation has to be adopted for the exchange-correlation functional in performing this kind of calculations. The local density approximation (LDA) [2] is the most widely used approximation which renders DFT computations feasible. But even with its well recognized success, LDA does have some limitations [3]. Therefore, there have been extensive efforts recently to go beyond the LDA to improve the exchange-correlation functional. One of the promising schemes is to include gradient corrections. This approach is referred to as the generalized gradient approximation (GGA) [4, 5, 6, 7]. Before this kind of functional can be really adopted for large scale calculations, the question whether such functionals really represent an improvement over LDA has to be addressed. Thus, we performed detailed investigations to compare two recently proposed GGA functionals for a wide range of systems. To understand the origin of the differences between LDA and GGA, we examined the effects on the distribution of charge density with the use of different functionals. The way in which GGA functional should be implemented in the context of pseudopotential calculations consistently

was also investigated. Based on our analysis, we concluded that LDA is as reliable a choice as any GGA scheme, as far as the energetics of solids is concerned. Therefore, we performed all our solid calculations within LDA.

As our first application, we studies the following problem: It has been observed during micro-indentation experiment [8] that silicon exhibits plastic flow under high pressure conditions. We attempted to understand the energetics associated with this process by performing DFT calculations for homogeneous structure transformations of silicon which might correspond to mass flow. The basic idea is to try to find a structural transformation path which has the lowest possible energy barrier. The magic strain concept which has been very efficient for identifying low energy structures in phase space was applied here in the search for a possible transformation path. With a unit cell consisting of only 2 atoms, we have been able to determine an upper bound for the energy barrier. The effects due to entropy were taken into account with the use of Vineyard's transition state theory [9] to give a more complete thermodynamical picture.

As a second application, we calculated the quantities relating to the dislocation properties of silicon. Since this problem is related to interfaces in solids, a supercell approach has to be adopted. We performed extensive calculations to obtain the generalized stacking fault (GSF) energy surfaces for various planes of silicon. A GSF energy surface is the energy surface obtained by moving one half of an infinite crystal with respect to the other half by a general displacement vector. The importance of the GSF energy surfaces is as follows: (1) By combining the GSF

energy surface with the Peierls-Nabarro model [10], the detailed properties of dislocations can be investigated. (2) Recent studies by Rice and collaborators of the dislocation nucleation problem, based on the Peierls stress concept, have identified the controlling parameter for dislocation emission to be the unstable stacking fault energy, γ_{us} [11], which is a saddle point in the GSF energy surface. In order to get a complete thermodynamical picture, both entropy effects and pressure effects were considered.

Since the effects due to the surface creation during dislocation nucleation were neglected in Rice's original theory, one might wonder how important these effects are in a realistic situation. We attempted to answer this question by performing large scale supercell calculations to simulate the finite size block shearing process in silicon. A supercell as big as 144 silicon atoms were used for these calculations. We obtained the energy needed for ledge creation. This energy was then incorporated into Rice's original theory within the framework of continuum mechanics. An evanescent law was assumed for the surface stress associated with the energy factor for ledge creation, following a suggestion by Y. Sun. The critical loading for dislocation nucleation was determined by solving an equation based on the force balancing concept.

This thesis is organized as follows: Chapter 2 is devoted to the discussion of two recently proposed GGA functionals. We discuss the energetics for the high-pressure plastic flow of silicon in Chapter 3. In Chapter 4 the GSF energy surfaces of silicon are investigated in detail, as well as an application to the study

of the dislocation profiles. Chapter 5 presents the results of our study on the surface effects during dislocation nucleation. We give general conclusions in the final chapter. The first section of each chapter provides a specific introduction for that chapter. Computational details and particular methods used during the calculations are discussed in the second section of each chapter. Following sections contain results and analysis. Each chapter concludes with a brief summary. The references are listed at the end of each chapter.

References

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. Sham, Phys. Rev. **140**, A1133 (1965).

[2] J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1984).

[3] R.O. Jones and O. Gunnarson, Rev. Mod. Phys. **61**, 689 (1984).

[4] J.P. Perdew, Phys. Rev. Lett. **55**, 1665 (1985); J.P. Perdew and Y. Wang, Phys. Rev. B **33**, 8800 (1986).

[5] A.D. Becke, Phys. Rev. A **38**, 3098 (1988); F.W. Kutzler and G.S. Painter, Phys. Rev. Lett. **59**, 1285 (1987).

[6] J.P. Perdew, in *Electronic Structure of Solids '91,* edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991).

[7] D.J. Lacks, and R.G. Gordon, Phys. Rev. A **47**, 4861 (1993).

[8] G.M. Pharr, W.C. Oliver, and D.S. Harding, J. Mater. Res. **6**, 1129 (1991).

[9] G.H. Vineyard, J. Phys. Chem. Solids **3**, 121 (1957).

[10] J.P. Hirth and J. Lothe, Theory of Dislocations, 2nd ed. (Wiley, New York, 1982)

[11] Rice, J. Mech. Phys. Solids **40**, 239 (1992).

2 Application of the GGA Functionals

2.1 Introduction

Density functional theory (DFT) [1] has proved to be a versatile and valuable tool in modern computational physics. The local density approximation (LDA) [2] is the most widely used implementation of DFT, which renders applications to a variety of problems computationally feasible. [3] Recent developments in computational capabilities have made it possible to study complex problems relating to microstructural deformations in solids. However, even with its well recognized success, LDA has several limitations (see e.g. Ref. [3]). Accordingly, improvements to LDA have been sought for many years. Ortiz and Ballone [4] pointed out that gradient corrections constitute a promising approach from both a computational and a conceptual point of view. In their calculation for homonuclear dimers, such corrections gave an improvement over LDA results in every respect. It was also reported by Bagno et al. [5] that by applying gradient corrections, the solid properties of third row elements are predicted more accurately. Despite these encouraging indications, a more thorough test for finite systems (atoms, molecules, clusters) as well as infinite systems (crystalline solids) is necessary before adopting these corrections for large scale computations. Recently some tests of this type have been reported by Perdew et al. [6], Ortiz [7], García et al. [8], and Körling and Häglund [9].

In this chapter, we present a detailed study of the application of gradient cor-

rections to DFT for a broad range of atoms and solids. In the present calculation we used the gradient-corrected exchange-correlation functionals which were introduced by Perdew and Wang [10, 11] and Lacks and Gordon [12] in the spin-unpolarized form [these two functionals will be referred to as PW(91) and LG respectively]. For the convenience of the reader the functional form of PW(91) and LG has been reproduced in Appendix A of this chapter. The reasons for choosing these two functionals are: (1) They are among the most recent additions to the list of proposed gradient corrections and are intended to give better results compared to earlier attempts. (2) The two functionals have similar expressions which simplifies computational implementation. (3) The Lacks-Gordon exchange functional is produced by fitting to exact results, while the PW91 functional is derived from first principles; comparison between these two functionals may provide insight for a more accurate approach. We found that the use of GGA functionals improves the calculation of atomic energies over LDA results. However, we found that by simply combining the PW91 functional with the pseudopotential approach leads to lattice constants for solids, such as simple metals and semiconductors, that are larger than experiment, and the percentage errors are significantly larger than those obtained from LDA. In order to understand the origin of these errors, we examined the charge distribution of individual atoms. An increase in the nonlinearity of the exchange-correlation functional in GGA calculations was found. In order to perform the calculations consistently, we demonstrated that it is necessary to take into account the nonlinear coupling between the core and valence electrons

within the pseudopotential approach.

The remaining of this chapter is organized as follows: we discusses the application of PW(91) to atoms in Section 2.2. The results for several types of solids, including metals, semiconductors and insulators with the simple combination of GGA functionals and the pseudopotential approach are presented in Section 2.3. The effects on the charge distribution of individual atoms are examined in Section 2.4. The results obtained with the partial core corrections, which takes into account the nonlinear character of the exchange-correlation functional, are given in Section 2.5. We conclude with some remarks on the usefulness of gradient corrections.

2.2 Application to Atoms

We first consider the application of the gradient corrections to scalar non-relativistic all-electron calculations for atoms. The elements we considered range from alkali metals to inert gases. Specifically, they include the 2nd, 3rd and 4th rows of the Periodic Table, except for the transition metals, for which some results have already been reported in the literature.[4] Since the results obtained with PW(91) and LG are very similar, only the results from PW(91) are given. The results of these calculations, which employed the standard logarithmic grid for the radial variable, [13] are given in Table 1. The total energy, exchange energy, and correlation energy are compared to experimental measurements where available and to LDA results.

Table 1: Atomic total energy, exchange energy and correlation energy calculated from LDA and PW(91). Experimental data for the total energy are from Ref. [14]. Experimental data for exchange and correlation energy are from Ref. [15], except for Ar; the correlation energy for Ar is from Ref. [16]. All values are in atomic units.

	X		
	$-E_{tot}$	$-E_x$	$-E_c$
LDA calculation			
PW(91) calculation			
Experiment			

Li			Be			B			C			N			O			F			Ne		
7.334	1.651	0.161	14.446	2.514	0.226	24.343	3.522	0.297	37.424	4.727	0.376	54.023	6.137	0.461	74.469	7.764	0.552	99.094	9.617	0.648	128.227	11.703	0.749
7.480	1.782	0.058	14.673	2.724	0.087	24.656	3.812	0.119	37.828	5.101	0.158	54.524	6.601	0.201	75.072	8.324	0.249	99.804	10.278	0.300	129.049	12.472	0.354
7.479	1.781	0.045	14.667	2.67	0.094	24.654		0.125	37.844		0.156	54.587		0.189	75.063			99.725			128.928	12.110	0.387
Na			**Mg**			**Al**			**Si**			**P**			**S**			**Cl**			**Ar**		
161.433	13.516	0.819	199.133	15.448	0.900	241.309	19.547	0.980	288.192	19.547	1.067	339.940	21.778	1.158	396.710	24.134	1.253	458.657	26.618	1.350	525.938	29.233	1.450
162.379	14.413	0.382	200.198	16.463	0.419	242.498	20.804	0.458	289.506	20.804	0.504	341.382	23.161	0.554	398.282	25.646	0.608	460.364	28.263	0.664	527.782	31.015	0.723
162.245			200.043		0.426	242.336			289.348			341.240			398.100			460.149			527.549	30.18	0.731
K			**Ca**			**Ga**			**Ge**			**As**			**Se**			**Br**			**Kr**		
598.192	31.640	1.525	675.733	34.133	1.608	1921.831	71.903	2.818	2073.792	75.729	2.922	2232.520	79.627	3.029	2398.097	83.602	3.137	2570.605	87.659	3.247	2750.132	91.802	3.359
600.178	33.568	0.755	677.855	36.192	0.796	1925.508	75.484	1.487	2077.618	79.458	1.546	2236.498	83.503	1.608	2402.227	87.628	1.673	2574.890	91.838	1.739	2754.573	96.135	1.807
																						93.900	

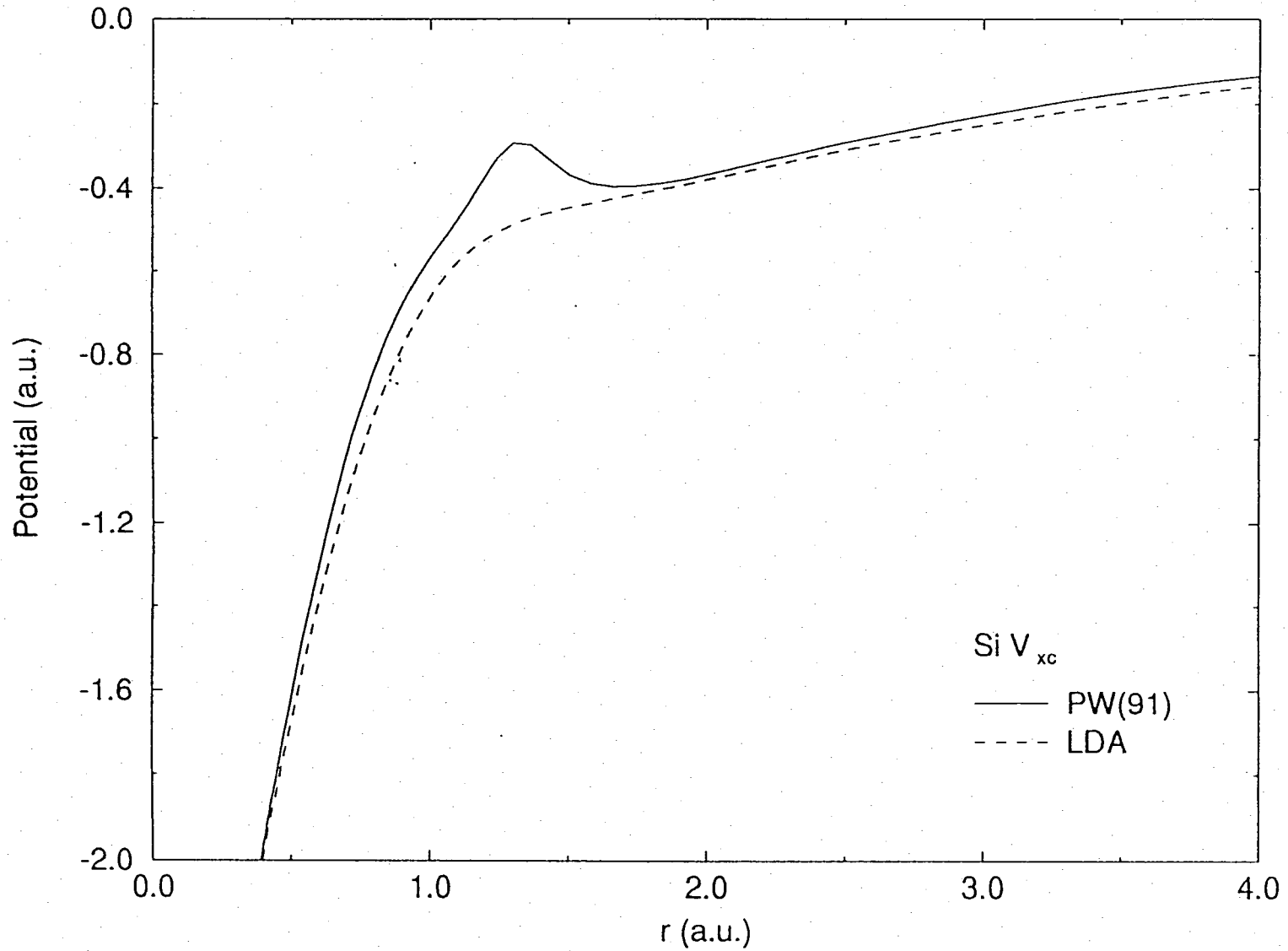
By using the PW(91) functional, one obtains a lower total energy and exchange energy but a higher correlation energy, all in better agreement with experiment than LDA results. A noticeable point is that for heavier elements, the agreement between calculation and experiment seems to worsen (although still better than LDA, where comparison to experiment is possible). The reason is most likely due to the neglect of relativistic effects which become more important for the heavier atoms. PW(91) appears to over-correct the LDA results in some cases, even though the difference with experiment is small. For example, the total energy of a Si atom is -288.192 from LDA and -289.506 from PW(91), while the experimental result is -289.348 (all numbers are in units of Hartree = 27.2116 eV).

For a more detailed look at the effect of PW(91) on the atomic calculations, we present an example in Figure 1. In this figure, we display the exchange correlation potential for Si as obtained by LDA and PW(91). In general, LDA gives a more negative exchange-correlation potential than PW(91). Except for the shoulder region, where the large difference is due to the gradient correction term which arises from a rapid change in the density, the difference between LDA and PW(91) is actually quite small. Nevertheless, this small difference in the exchange-correlation potential gives significant differences in the energies (see Table 1).

2.3 Application to Solids

We first apply these functionals to the study of solids by directly incorporating them into the pseudopotential framework. In order to perform consistent first-

Figure 1: The all-electron exchange-correlation potential for the Si atom as given by the LDA and PW(91) functionals. The shoulder region at $r \simeq 1.3\, a.u.$ is due to sharp variations in the density.


0.0
-0.4
-0.8
-1.2
-1.6
-2.0
Potential (a.u.)
0.0
1.0
2.0
3.0
4.0
r (a.u.)
Si V xc
PW(91)
LDA

principles pseudopotential calculations for solids, it is necessary to construct new pseudopotentials. The reasons are both conceptual and practical. Pseudopotentials that are available in the literature, such as the ones by Bachelet et al. [17] (BHS) were constructed from LDA all-electron calculations. If these potentials were to be used in the PW(91) solid calculations, then the screening effect for core electrons would be approximated by LDA while the screening effect for valence electrons will be approximated by the PW(91) functional, which evidently is not a self-consistent approach. Moreover, it was found by Ortiz and Ballone [4] that a self-consistent pseudopotential that incorporates gradient corrections gives different results from a BHS pseudopotential. The details of the construction of the pseudopotentials used in the present calculation are given in Appendix B.

We performed total energy pseudopotential calculations for two simple metals (Na and Mg), one *sp*-bonded metal (Al), two elemental semiconductors (Si and Ge), one compound semiconductor (GaAs) and one inert gas crystal (Ar). The calculations were done by using a plane wave basis. For the LDA calculations, we used the exchange-correlation potential of Ceperley and Alder as parametrized by Perdew and Zunger. [16] The use of a plane-wave basis is advantageous in that the gradient and the laplacian of the density needed in PW(91), can be obtained easily by Fast Fourier Transforms. In the PW(91) calculation only 9 additional FFT's are needed, which results in a very modest increase of computational time (less than 3%).

The ground state properties of each solid were obtained by minimizing the en-

ergy with respect to the lattice constant. In our present calculation, we fitted the results to the Murnaghan equation of state, [18] which then gives the equilibrium lattice constant and bulk modulus. In order to obtain a reliable cohesive energy one needs to calculate the energy of the solid with the same degree of accuracy as that of the isolated atom. The latter is obtained by integration of the Schrödinger equation and is essentially exact (within the limits of DFT). The energy of the solid converges variationally with the size of the plane-wave basis. For this reason, we first calculate the ground state energy corresponding to the calculated equilibrium lattice constant for several values of the cutoff energy E_c, which is the highest plane-wave kinetic energy and determines the size of the basis. The cutoff energy ranges from 12 to 20 Ry, except for Ar, in which case E_c ranges from 16 to 32 Ry. Then, we fitted the calculated total energies to $E = E_0 + \lambda/E_c^\gamma$ in order to obtain the asymptotic value of the total energy E_0 (λ and γ are fitting parameters; the optimal value of γ is 2.5). We find that the error introduced by using this formula is less than 0.1% (this was estimated by comparing the fitted result to the calculated result for the highest cutoff energy). In order to include spin-polarization effects, we used the empirical formula $\Delta E_p = -0.18 \times n_p^2$ eV ($n_p = n_\uparrow - n_\downarrow$, $n_\uparrow$ being the number of electrons with spin up and $n_\downarrow$ the number of electrons with spin down) to estimate the shift in the total energy of the pseudo-atoms in spin-polarized configurations. Spin-polarized calculations for non-magnetic solids are expected to have negligible effect to the total energy E_0. The cohesive energy is then determined by the difference between E_0 and the total atomic energy calculated for the

pseudo-atom.

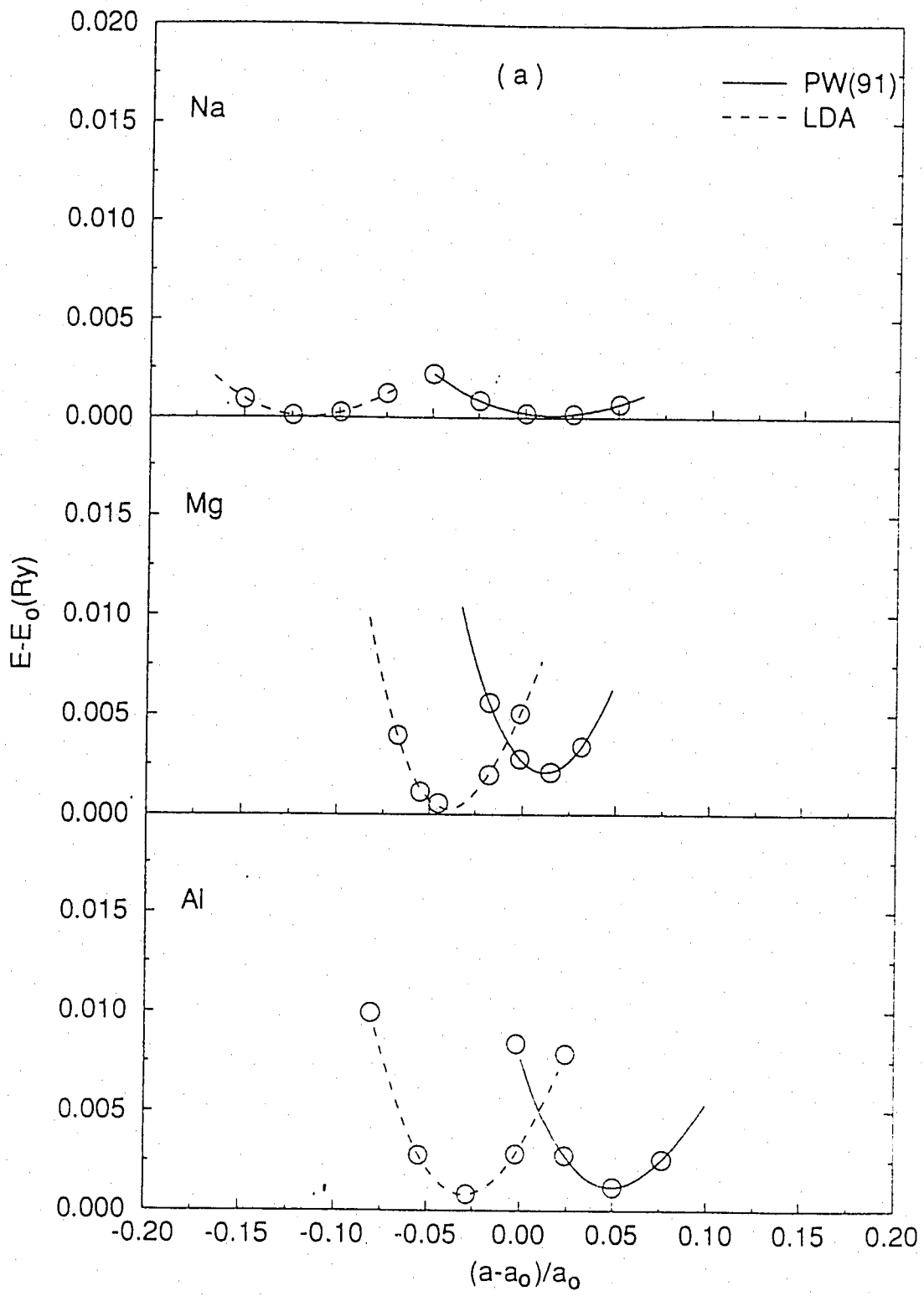
The total energy vs. lattice constant curves from LDA and PW(91) for Na, Mg, and Al are shown in Figure 2(a) and for Si, Ge, GaAs, and Ar in Figure 2(b). The curves from the PW(91) calculations are flatter and they are shifted to the right compared to the LDA curves. Thus, the PW(91) functional gives a larger lattice constant and smaller bulk modulus than LDA. The latter approximation overestimates the cohesive energy in most cases, and gives too small lattice constants compared to experiment. The results for the calculations of ground state properties of solids are summarized in Table 2 . Except for the simple metals Na and Mg, the results from PW(91) seem to over-correct the LDA results. Specifically, the PW(91) results give a lattice constant which is bigger than experiment by a significant amount. For example, the PW(91) results give lattice constants which are higher than experiment by 3%, 5%, and 5% for Si, Ge, and GaAs, respectively, whereas LDA results give lattice constants which are smaller than experiment by 1%, 2%, and 2% for the same solids. For the bulk modulus, the results from the PW(91) calculations are better for Na, Mg, Al, and Ar, since the LDA values are too high for these three elements. But for Si, Ge, and GaAs, the bulk modulus from the PW(91) calculations are worse than the LDA results, which were already underestimates of the experimental values. For the cohesive energy, the PW(91) calculations give an improvement over the LDA results in most cases, even though the PW(91) cohesive energies are still larger (in absolute value) than experiment.

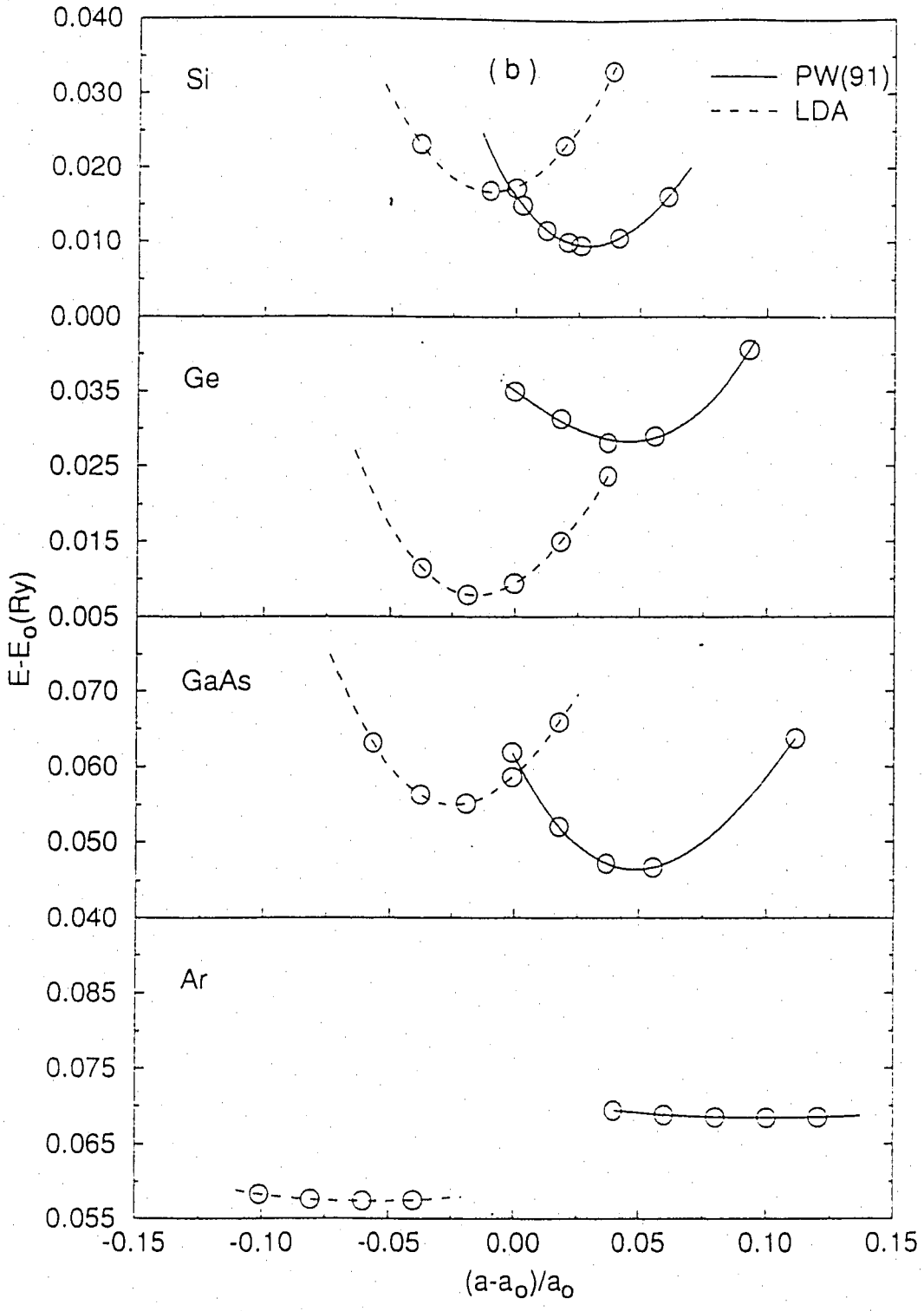
Table 2: Lattice constant, bulk modulus, and cohesive energy calculated from LDA and PW(91). The experimental data for lattice constants are taken from Ref. [19], except for GaAs, which is from Ref. [20] The experimental data for bulk modulus and cohesive energy are from Ref. [21], except for GaAs, which are from Ref. [22] and Ref. [23]. The unit for lattice constant is Å; the unit for bulk modulus is $10^{10} dyne/cm^2$; the unit for cohesive energy is eV.

	X		
	lattice constant	bulk modulus	cohesive energy
LDA calculation			
PW(91) calculation			
Experiment			

Na			Mg			Al			Si			Ge			GaAs			Ar		
3.74	10.65	1.50	3.08	41.77	2.00	3.93	87.65	4.13	5.38	96.57	5.34	5.57	74.72	4.49	5.51	77.09	8.48	4.94	6.46	0.419
4.29	5.68	1.17	3.25	37.70	2.06	4.25	81.07	3.21	5.59	85.22	4.60	5.92	58.12	3.67	5.92	54.59	6.33	5.83	2.75	0.066
4.23	6.80	1.11	3.21	35.40	1.51	4.05	72.20	3.39	5.43	98.80	4.63	5.66	77.20	3.85	5.65	75.57	6.52	5.26	1.60	0.080

Figure 2: (a) Comparsion of total-energy vs. lattice constant curves for Na, Mg, and Al as given by the LDA and PW(91) functionals. The dots are calculated values, the lines are fits to Murnaghan's equation of state. a_0 is the experimental lattice constant, E_0 is the asymptotic value of the solid's total energy (see text). (b) Same as in (a), for Si, Ge, GaAs and Ar.


(a)
PW(91)
LDA
Na
Mg
Al
0.020
0.015
0.010
0.005
0.000
E-E$_o$(Ry)
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
(a-a$_o$)/a$_o$


Si
(b)
PW(91)
LDA
Ge
GaAs
Ar
0.040
0.030
0.020
0.010
0.000
0.035
0.025
0.015
0.005
0.070
0.060
0.050
0.040
0.085
0.075
0.065
0.055
E-E$_o$(Ry)
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
(a-a$_o$)/a$_o$

For Ar, we note that the energy vs. volume curve is much flatter than other solids, making the prediction of the equilibrium lattice constant and bulk modulus more difficult. In summary, we find that use of the PW(91) functional gives a weaker binding between atoms in a solid, resulting in larger lattice constants and smaller bulk moduli. The deviation from experiment depends somewhat on the character of the solid, but in many cases it is larger than the deviation of LDA results from experiment.

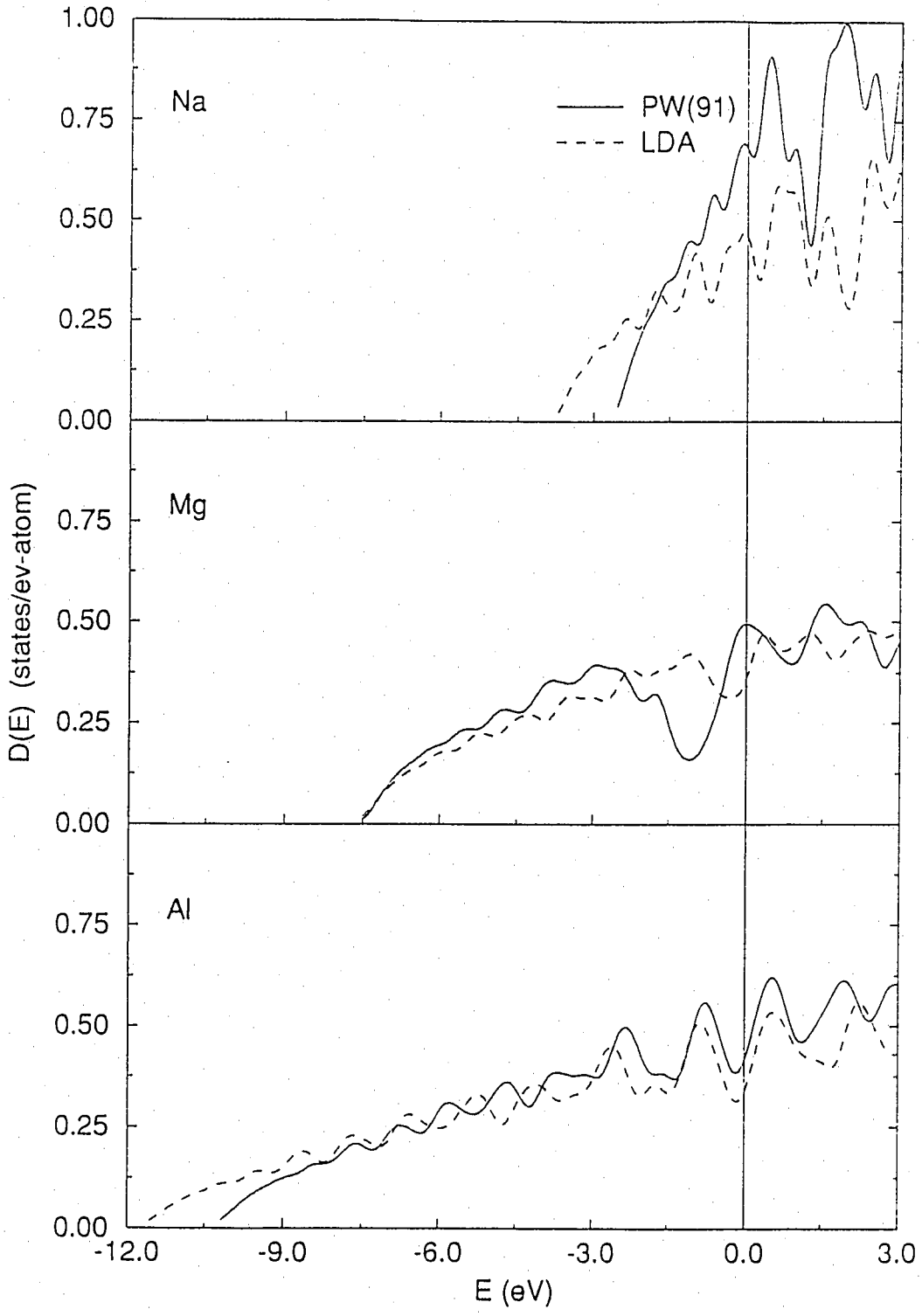
The calculated band gaps for the semiconductors and the insulator we have considered, at both the theoretically predicted equilibrium lattice constant and the experimentally measured one, are given in Table 3. At the experimentally measured lattice constant, PW(91) gives consistently a better band gap than LDA. At the theoretically predicted lattice constant, however, it cannot be said conclusively whether PW(91) is better or worse than LDA in predicting band gaps. This is to be expected since the band gap depends on the lattice constant, and PW(91) and LDA give significantly different equilibrium lattice constants. In Figure 3 we compare the density of states for the three metals we have considered, calculated from LDA and PW(91). As can be seen from this figure, the PW(91) calculation gives a narrower band width and a larger density of states at the Fermi level. This is because PW(91) gives a "softer" solid than LDA, and the band width scales with the strength of the interaction between atoms.

Table 3: Band gap calculated from LDA and PW(91). a_0^{th} is the equilibrium lattice constant predicted by theory. a_0^{exp} is the equilibrium lattice constant measured from experiment. The experimental data are taken from Ref. [24], except for Ar, which is from Ref. [25]. The unit is eV.

	X	
	at a_0^{th}	at a_0^{exp}
LDA calculation		
PW(91) calculation		
Experiment		

Si		Ge		GaAs		Ar	
0.49	0.54	0.30	0.02	1.03	0.58	8.41	8.33
0.81	0.69	0.05	0.61	0.62	1.43	8.25	8.45
	1.17		0.79		1.52		14.30

Figure 3: Density of states (including spin degeneracy) for Na, Mg, and Al calculated from LDA and PW(91). The Fermi energy is taken to be zero.


Na
PW(91)
LDA
Mg
Al
D(E) (states/ev-atom)
E (eV)
1.00
0.75
0.50
0.25
0.00
-12.0
-9.0
-6.0
-3.0
0.0
3.0

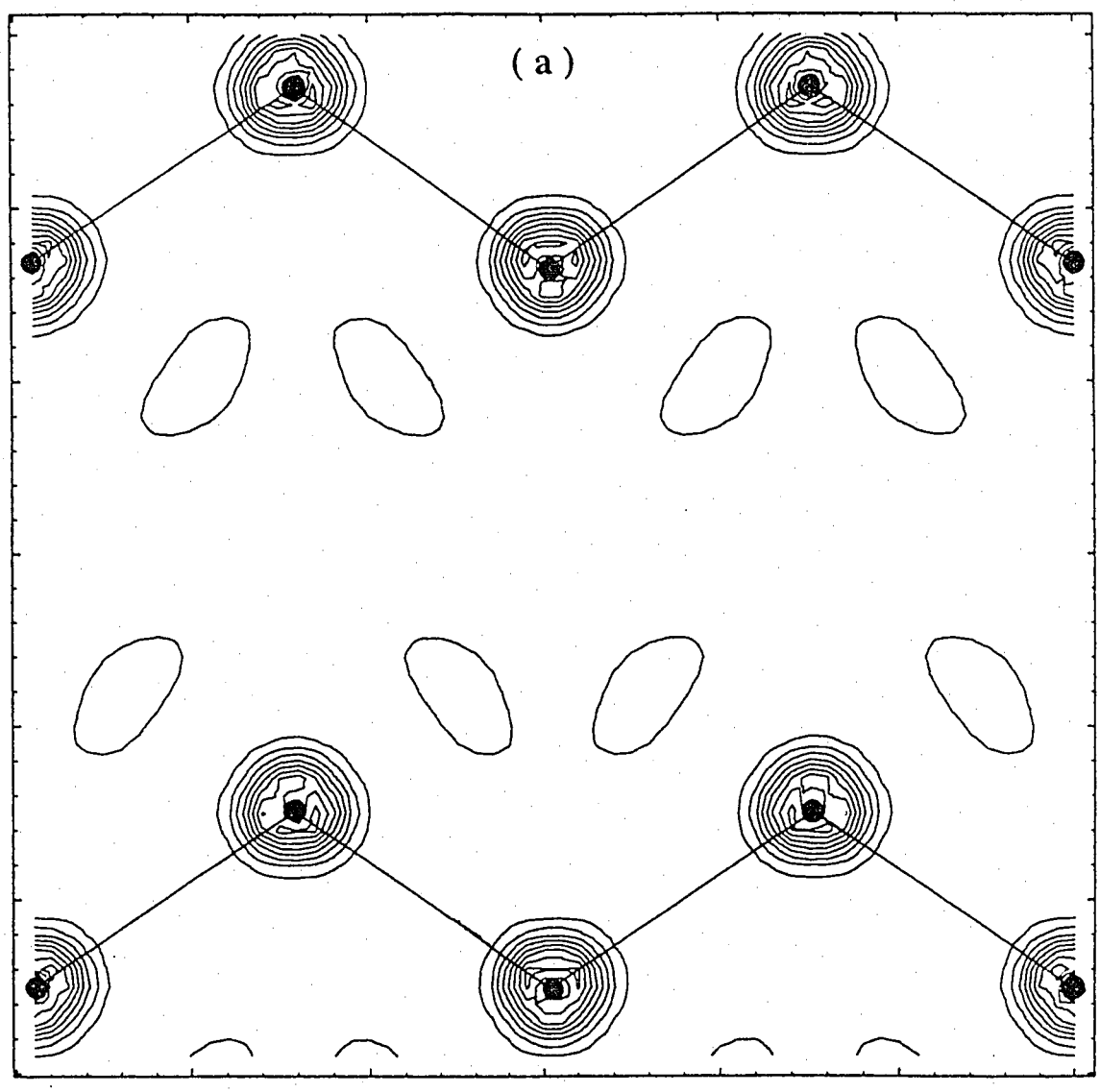
For a more detailed look at the effect of gradient corrections, we show in Figure 4(a) the difference between the PW(91) and LDA exchange-correlation energy functionals for Si. Figure 4(b) presents a comparison between PW(91) and LDA exchange-correlation potentials, which are essentially the functional derivatives of the energy with respect to density. In both cases the most important difference between LDA and PW(91) occurs near the atomic nuclei. This is not surprising, since the density varies significantly inside this region, giving large gradient corrections. In the present calculations the use of a pseudopotential results in vanishingly small, but rapidly changing density near the atomic nuclei. The only other region where the PW(91) exchange-correlation potential shows large difference from the LDA potential is between the chains of covalent bonds where the electronic density is again very small.

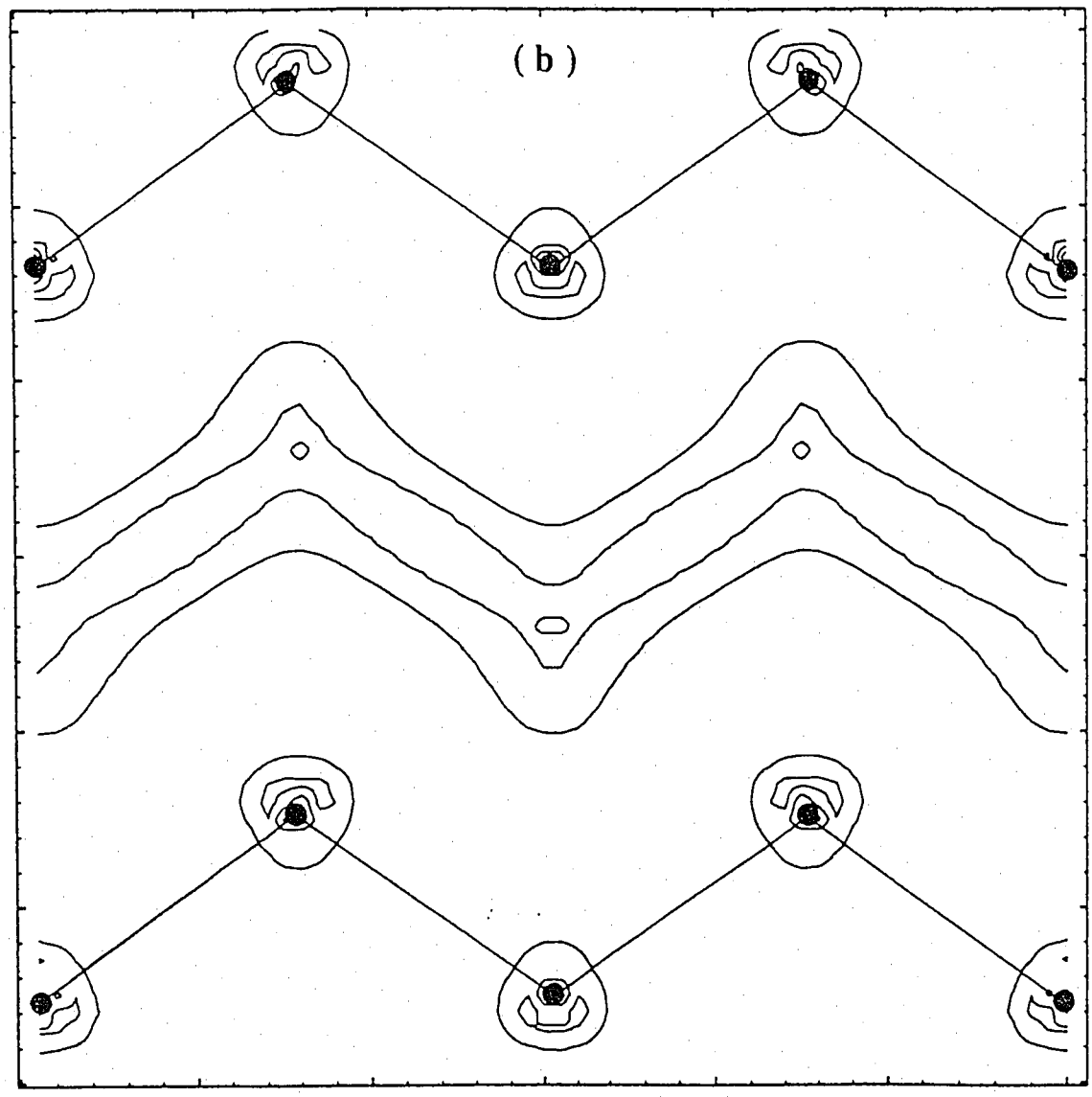
2.4 The effects on Atomic Charge Distribution

The above discussion points to the fact that in order to understand the origin of the softening of solids with the use of GGA functionals, it is necessary to examine the effects of GGA functionals on the charge distribution of individual atoms. In studying the charge distributions of individual atoms, we have found it instructive to evaluate the following quantity

$$C_{nl}(R) = 4\pi \int_0^R r^2 dr \left[\left| \phi_{nl}^{LDA}(r) \right|^2 - \left| \phi_{nl}^{GGA}(r) \right|^2 \right] \qquad (1)$$

Figure 4: (a) Difference between PW(91) and LDA exchange-correlation energy functionals for Si on the [110] plane of the diamond lattice. Dark symbols indicate the atomic positions. (b) Same as in (a), for the exchange-correlation potentials.

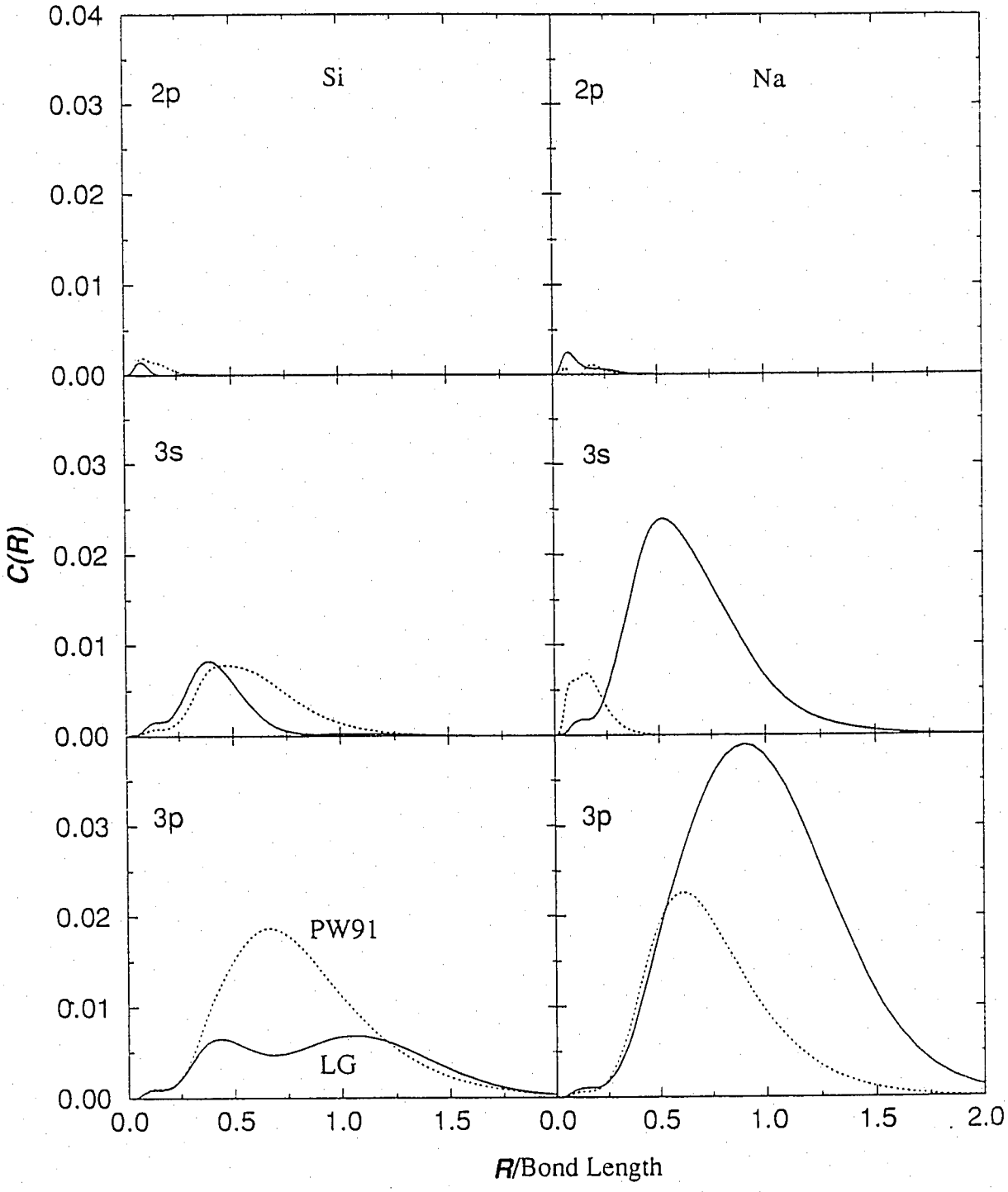

(a)


(b)

$C_{nl}(R)$ is the difference between the charge enclosed within a sphere of radius R around the nucleus calculated from the LDA and the GGA functional for each single electron orbital nl. In Figure 5 we display $C_{nl}(R)$ for the $nl = 2p$, $3s$ and $3p$ orbitals of Si, and the $nl = 2p$, $3s$, $3p$ orbitals of Na. The more positive $C_{nl}(R)$ is the more charge has been pushed outside the region contained by the sphere of radius R in the GGA calculation compared to the charge obtained by LDA.

In order to see the difference in the physically important range, we used the experimentally measured bond length as the unit along the x-axis in Figure 5. As is obvious from this figure, there is almost no difference in the charge distribution between LDA and GGA results for the $2p$ core orbital of Si. For the $3s$ and $3p$ valence orbitals of Si, in the neighborhood of the bonding region substantial charge has been pushed away from the nucleus in both the PW91 and the LG calculations, relative to the LDA results. A similar situation occurs in the case of Na as is shown in Figure 5. These comparisons indicate a weaker interaction between the valence electrons of the atom and the ion when using GGA functionals as opposed to LDA, due to the spreading out of the valence charge in GGA calculations. In an approach which simply replaces the effects due to the ion and the core electrons with a pseudo-core which is constructed to reproduce the results of the all-electron calculation, this will unavoidably lead to a weaker interaction between the valence electrons and the pseudo-core. Since in the pseudopotential framework the properties of solids are determined by the interaction between valence electrons and the pseudo-core, the tendency for valence charge to be pushed away from

Figure 5: Integrated charge difference $C_{nl}(R)$ [see Eq. (1)] calculated with PW91 (so lid lines), and LG (dashed lines) for nl =2p, 3s, 3p orbitals of Si, and nl =2p, 3s, 3p orbitals of Na. The results for the 3p orbital of Na have been divided by a factor 2 so that they can be displayed on the same scale as the results for Si


Si
Na
2p
2p
3s
3s
3p
3p
PW91
LG
C(R)
R/Bond Length
0.04
0.03
0.02
0.01
0.00
0.0
0.5
1.0
1.5
2.0

the nucleus when GGA functionals are used leads to a softer solid, characterized by larger equilibrium lattice constant and smaller bulk modulus than LDA.

2.5 Nonlinearity of the GGA functional

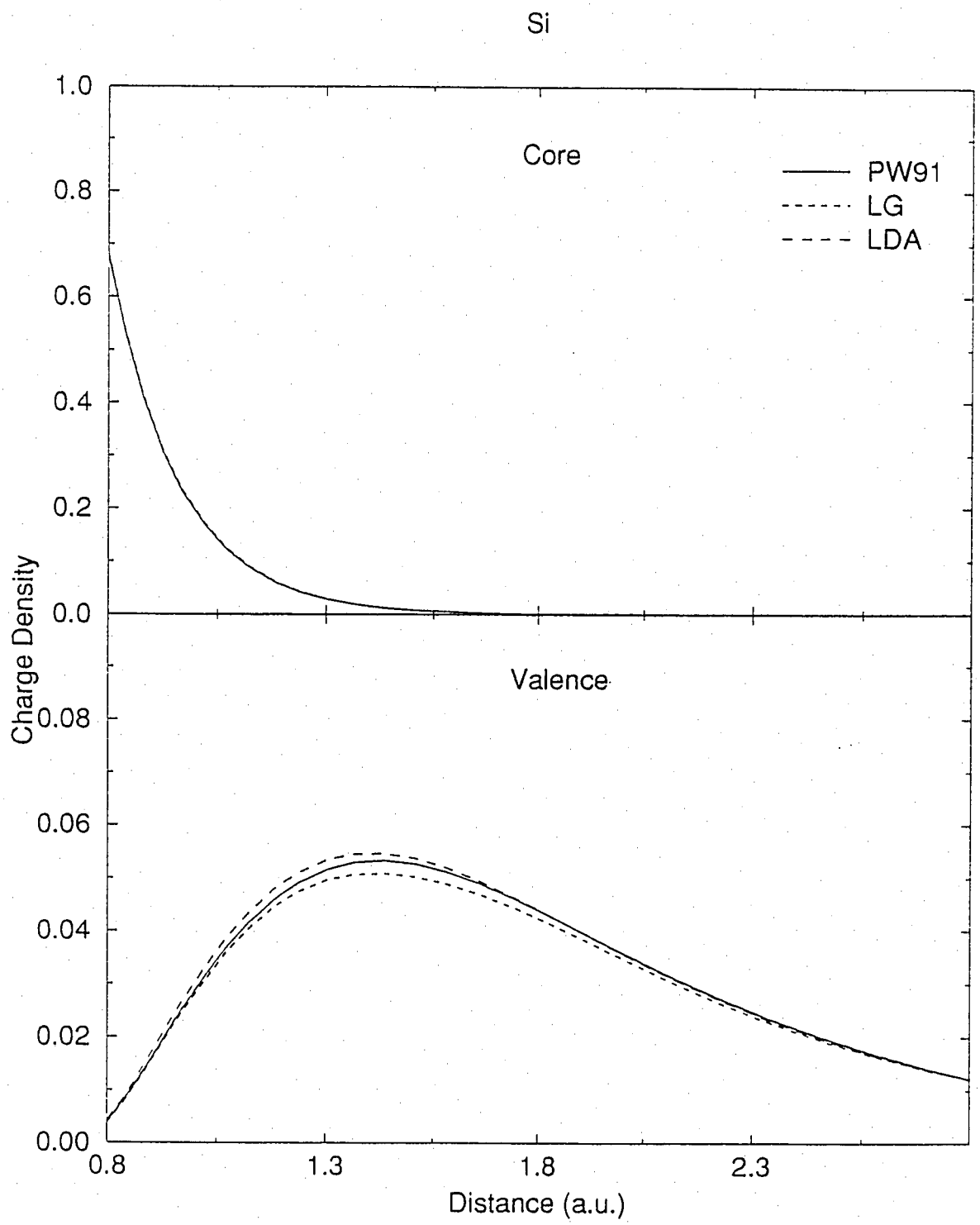
This observation points to the necessity of properly taking into account the non-linear coupling between the valence and core electrons in the exchange correlation functional within the pseudopotential approach, when GGA functionals are used. To examine the effects due to the inclusion of gradient corrections, we first consider the behavior of the quantity s which is used in the definition of GGA functionals in addition to the charge density n. s is defined as a function of the charge density n and its gradient ∇n:

$$s = |\nabla n| / 2(3\pi^2)^{1/3} n^{4/3} \tag{2}$$

Although the charge density n can be separated into the core charge density n_c and the valence charge density n_v so that $n = n_c + n_v$, such a separation can not be written for s due to the nonlinearity of the expression of Eq. 2. We use Si as an example to further illustrate this point. We display in Figure 6 both the core and valence charge density for the Si atom, calculated with the LDA and the GGA functionals respectively. There is evidently no significant difference between the results of these two calculations as far as the overlap between charge density is concerned. Figure 7 (a) displays the quantity

$$\Delta V_{xc} = V_{xc}[n_c] + V_{xc}[n_v] - V_{xc}[(n_c + n_v)] \tag{3}$$

Figure 6: The core (upper panel) and valence (lower panel) charge density of the silicon atom calculated with LDA and GGA functionals as a function of distance from the nucleus. Notice the different density scales in the two panels.


Si
Core
PW91
LG
LDA
Valence
Charge Density
Distance (a.u.)
1.0
0.8
0.6
0.4
0.2
0.0
0.08
0.06
0.04
0.02
0.00
0.8
1.3
1.8
2.3

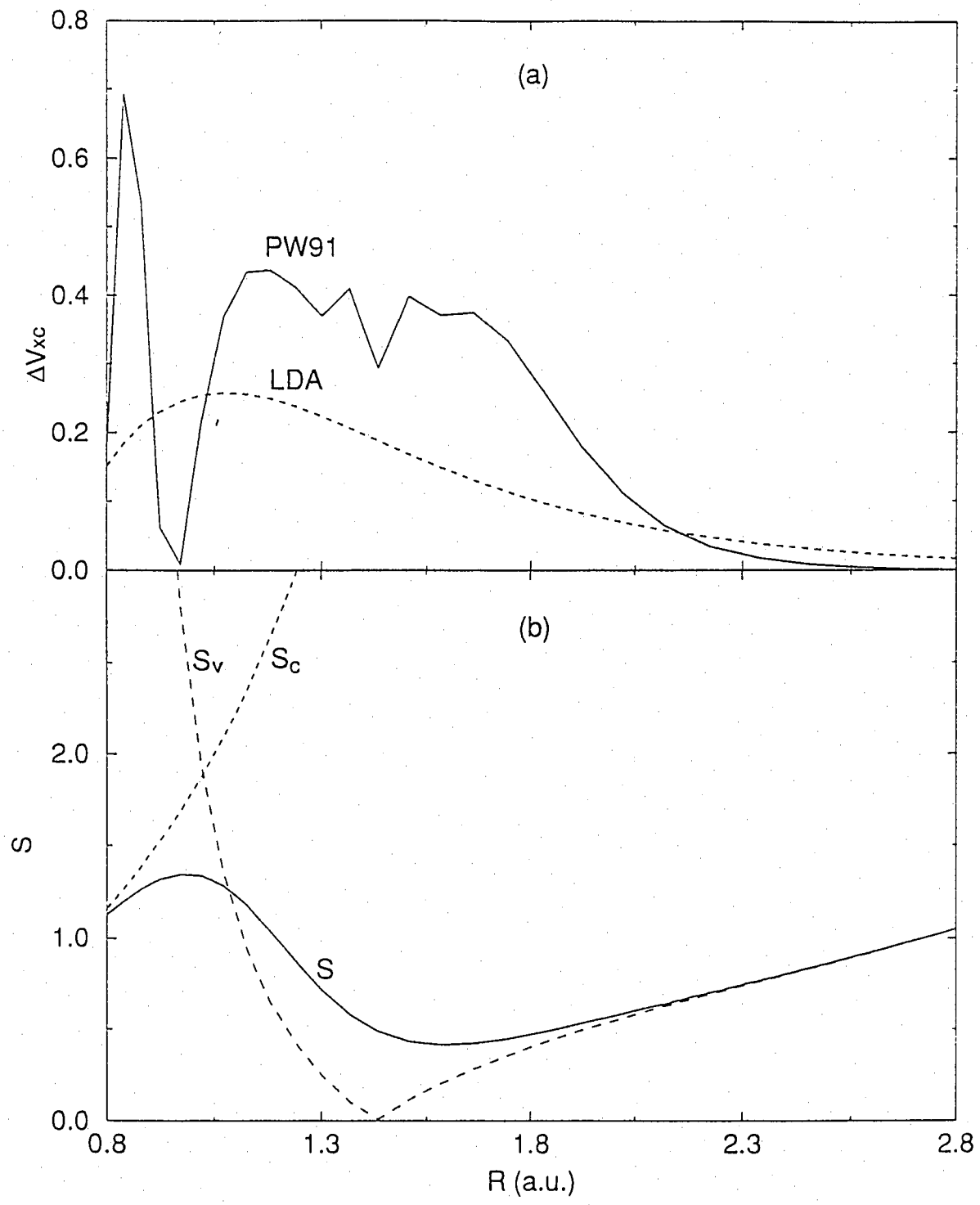
which is a measure of the nonlinearity of the exchange-correlation potential. There is a substantial increase in the nonlinearity of the exchange-correlation potential in the region where the overlap between the core and valence charge is not negligible (compare with Figure 6). This increase in nonlinearity is due to the fact that the variables used in the expression for the GGA functionals are not separable in terms of the valence and core parts. For example, we show in Figure 7 (b) the values of s_c, s_v and s which correspond to the value of the quantity defined in Eq. 2 calculated with the core, valence, and total charge density respectively. Therefore, the inclusion of gradient corrections seems to increase the effects of coupling between the valence and core electrons. As was discussed in the previous paragraphs, the simple unscreening of the pseudopotential by

$$V_{ion}^{ps} = V_{screened}^{ps} - V_H[n_v] - V_{xc}[n_v] \tag{4}$$

where V_H and V_{xc} are the Hartree and the exchange-correlation potentials respectively, does not give satisfactory results for the properties of solids. This approximation implicitly assumes the linearization of the exchange-correlation functional. As was pointed out by Louie el al. [26], a more consistent approach is to include the core charge density in the unscreening. That is, instead of taking out only the $V_{xc}[n_v]$ part in the unscreening procedure as in Eq. 4, V_{ion}^{ps} should be defined instead as:

$$V_{ion}^{ps} = V_{screened}^{ps} - V_H[n_v] - V_{xc}[(n_c + n_v)] \tag{5}$$

Figure 7: (a) The quantity ΔV_{xc} [see Eq. (3)] for the silicon atom calculated with LDA and PW91 as a function of distance from the nucleus. (b)The values of s [see Eq. [2]] calculated from the core (s_c), the valence (s_v), and the total charge density (s) for the silicon atom as a function of distance from the nucleus.


(a)
PW91
LDA
ΔVxc
0.8
0.6
0.4
0.2
0.0
(b)
Sv
Sc
S
2.0
1.0
0.0
0.8
1.3
1.8
2.3
2.8
R (a.u.)

where n_c is a rigid core charge density, constructed from a reference atomic system. Since the core electrons are not included in the pseudopotential calculation, whenever the exchange-correlation energy and potential are needed, the full charge density $n = n_c + n_v$ must be used. This procedure is exact within the rigid core approximation, but it would require a very large number of plane waves to describe the core charge density accurately, and one loses the advantages of using the pseudopotential formalism by adopting this approach. So even though it is theoretically correct, it is not practical from a computational point of view. Therefore, it is necessary to make some approximation in order to obtain a practical computational scheme.

We follow the partial core prescription proposed in Ref. [26]. The full core charge density is replaced with an artificial core charge density $\tilde{n}_c$ defined as:

$$\begin{aligned} \tilde{n}_c &= \frac{A \sin Br}{r}, r \leq r_c \\ \tilde{n}_c &= n_c, r > r_c \end{aligned} \tag{6}$$

The parameters A and B are determined by the requirement that the value of $\tilde{n}_c$ and its derivative with respect to the radius r be exactly the same as those of the real core charge density n_c at the cutoff radius r_c. We have found that in order to capture the nonlinear coupling between the core and the valence electrons for the case of the GGA functionals, it is necessary to use r_c smaller than what was suggested in Ref. [26] for LDA calculations. In our calculations, r_c is chosen as the radius where the core charge density n_c is 6-7 times larger than the valence

charge density n_v. In Ref. [26] (which dealt with LDA calculations), r_c was chosen as the radius where the core charge density n_c was 2-3 times larger than the valence charge density. It is worthwhile mentioning that the pathological oscillatory behavior of the PW91 exchange-correlation potential near the nuclei, which causes problems in creating smooth pseudopotentials during the unscreening procedure is automatically eliminated by using the partial core correction.

As an illustration of how this approach works, we reconsider the four *sp*-bonded solids in their ground state phase: Si (diamond), Ge (diamond), GaAs (zincblende), and Al (fcc). The calculated ground state properties using LDA and PW91 are summarized and compared to experimental data and to the results without partial core corrections, in Table 4. The improvement in cohesive energy basically does not change with the inclusion of the partial core correction. For the lattice constant, we found that the results from the GGA functional are still consistently larger than the LDA results. But in terms of the magnitude of increase, it is substantially smaller after the nonlinear behavior of the exchange-correlation functional is taken into account. In the case of Al, this makes the value obtained from PW91 closer to experiment than the LDA result. For Si, Ge, and GaAs, the results obtained from LDA and PW91 are of the same accuracy compared to experiment. PW91 tends to overcorrect the LDA results and gives an overestimate for the equilibrium lattice constant of these systems. For the bulk modulus, the values obtained from PW91 are smaller than the LDA results. While this leads to a better result for Al, the bulk moduli we obtained for the three semiconductors are significantly

underestimated (by -12 % to -25 % compared to experiment). Finally, we compare our results to recent all-electron, linearized augamented planewave (LAPW) [27] total energy and electronic structure calculations, with the same GGA functional as in our work. It is obvious from the comparison of Table 4 that the present pseudopotential calculation results with the partial core correction represent a significant improvement over the results without the partial core correction, and agree very well with the all-electron LAPW calculation results. The remaining discrepancy is probably due to the relaxation of the core electrons which is not allowed in the pseudopotential calculation.

2.6 Summary

In conclusion, we showed that it is essential to take into account the core-valence coupling in the pseudopotential calculations when using GGA exchange-correlation functionals. To this end, we have found that the partial core prescription of Louie et al. [26] is most appropriate when using a plane-wave basis. We considered the structural properties of a variety of solids using both LDA and GGA. We found that PW91 gives consistently better cohesive energies than LDA. We suspect that this is due to a better description for the pseudo-atom properties in analogy with the improvement in all-electron atomic calculations (see Section 2.2). We also demonstrated that for the lattice constant the same accuracy as in LDA can be obtained with GGA, as long as the nonlinearity of the gradient corrected functional is properly taken into account. Even though the PW91 functional does give a

Table 4: Lattice constant (a_0 in Å), bulk modulus (B in GPa), cohesive energy (E_{coh} in eV/atom), calculated from LDA and PW91. For comparison, we also list the results obtained from pseudopotential calculation without partial core correction (PW91$\star$) from Table 2 and the results from all-electron linearized augmented planewave (LAPW) calculation [27]. The values in parentheses are percent differences from the experimental values.

	Si			Al		
	a_0	B	E_{coh}	a_0	B	E_{coh}
LDA	5.38(-0.9%)	96.57(-2%)	5.38(+15.9%)	3.93(-2.9%)	87.65(+21%)	4.14(+22.1%)
PW91	5.45(+0.4%)	83.51(-15%)	4.40(-5.1%)	4.03(-0.5%)	79.25(+10%)	3.45(+1.7%)
PW91⋆	5.59(+2.9%)	85.22(-14%)	4.64(+0.2%)	4.25(+4.9%)	61.07(-15%)	3.22(-5.0%)
PW91(LAPW)	5.50(+1.3%)	83.00(-16%)				
Expt.	5.43	98.80	4.63	4.05	72.20	3.39
	Ge			GaAs		
	a_0	B	E_{coh}	a_0	B	E_{coh}
LDA	5.57(-1.6%)	74.72(-3%)	4.53(+17.7%)	5.51(-2.5%)	77.09(+2%)	8.58(+31.6%)
PW91	5.73(+1.2%)	58.04(-24%)	3.41(-11.4%)	5.76(+1.9%)	66.28(-12%)	6.00(-8.6%)
PW91⋆	5.92(+4.7%)	58.12(-25%)	3.71(-3.6%)	5.92(+4.8%)	54.59(-28%)	6.43(-1.4%)
PW91(LAPW)	5.75(+1.6%)	61.00(-21%)		5.74(+1.5%)	65.00(-14%)	
Expt.	5.66	77.20	3.85	5.65	75.57	6.52

better description for the equilibrium properties of all the metals we have considered. For the semiconductors, the bulk moduli obtained with the use of GGA functionals represent significant underestimates of the experimental results. Therefore, further search may be needed for an exchange-correlation functional which is consistently better for all solids. We believe that LDA remains to be the most reliable choice for energetic comparison of solids for the present time.

In view of the above results, one may inquire what are the physical situations in which the use of GGA functionals can provide significant improvements over LDA results. Recently, calculations have been reported for H_2 dissociation on a Cu(111) surface with the LDA and the GGA [28, 29]. The GGA results for this system represent significant improvements over the LDA results. It has also been demonstrated that the GGA gives results in better agreement with experiments than the LDA for finite systems (atoms and molecules) and metallic surfaces [4, 6, 30]. It is therefore expected that the GGA will give, in general, a better description for the interaction between molecules and other molecules or solid surfaces. The reason that the GGA should give better results for these interactions can be attributed to the fact that substantial part of the interactions in the these systems are related to the tails of the electronic wave functions, where the GGA gives a more accurate description than the LDA.

Appendix A

The PW(91) exchange energy functional used in the present calculation is:

$$E_x[n] = A_x \int d^3r\, F_x(s) n^{4/3} \tag{7}$$

where $n = n(\vec{r})$ is the electron density, k_F is the Fermi momentum which is related to the density by $k_F = (3\pi^2 n)^{\frac{1}{3}}$, $A_x = -\frac{3}{4}(\frac{3}{\pi})^{1/3}$, $s = \frac{|\nabla n|}{2k_F n}$, and

$$F_x(s) = \frac{1 + b_1 s y(s) + (b_2 + b_3 e^{-100s^2}) s^2}{1 + b_1 s y(s) + b_4 s^4}$$

$$y(s) = log(b_5 s + \sqrt{1 + (b_5 s)^2})$$

The numerical values of the constants are: $b_1 = 0.19645$, $b_2 = 0.27430$, $b_3 = -0.15084$, $b_4 = 0.004$, $b_5 = 7.7956$.

The LG functional has a similar functional form

$$F_x(s) = \frac{(1 + \frac{ad+0.1234}{b} s^2 + a_4 s^4 + a_6 s^6 + a_8 s^8 + a_{10} s^{10} + a_{12} s^{12})^b}{1 + a_d s^2} \tag{8}$$

with the numerical values for the constants given as follows: $a_d = 1.0^{-8}$, $a_4 = 29.790$, $a_6 = 22.417$, $a_8 = 12.119$, $a_{10} = 1570.1$, $a_{12} = 55.949$, $b = 0.024974$.

The correlation energy functional is :

$$\int d^3r(\epsilon_c(n) + H(n, s, t))n \tag{9}$$

where

$$H = \frac{\beta^2}{(2\alpha)} log(1 + 2\frac{\alpha}{\beta}(\frac{t^2 + At^4}{1 + At^2 + A^2 t^4})) + C_{c0}(C_c(n) - C_{c1}) t^2 e^{-100s^2}$$

$$A = 2\frac{\alpha}{\beta}(\frac{1}{e^{-2\alpha\epsilon_c(n)/(\beta^2)} - 1})$$

$$C_c(n) = C_1 + \frac{C_2 + C_3 r_s + C_4 r_s^2}{1 + C_5 r_s + C_6 r_s^2 + C_7 r_s^3}$$

$$\epsilon_c(n) = -a(1 + \alpha_1 r_s) log(1 + \frac{1}{a(\beta_1 r_s^{1/2} + \beta_2 r_s + \beta_3 r_s^{3/2} + \beta_4 r_s^{1+p})})$$

The density dependent variables t, k_s, and r_s are defined as: $t = \frac{|\nabla n|}{(2k_s n)}$, $k_s = (\frac{4k_F}{\pi})^{1/2}$, $r_s = (\frac{3}{4\pi n})^{1/3}$, and the numerical values of the constants are: $\alpha = 0.09$, $\beta = 0.0667263212$, $C_{c0} = 15.75592$, $C_{c1} = 0.003521$, $C_1 = 0.001667$, $C_2 = 0.002568$, $C_3 = 0.023266$, $C_4 = 7.389 \times 10^{-6}$, $C_5 = 8.723$, $C_6 = 0.472$, $C_7 = 7.389 \times 10^{-2}$, $a = 0.0621814$, $\alpha_1 = 0.21370$, $\beta_1 = 7.5957$, $\beta_2 = 3.5876$, $\beta_3 = 1.6382$, $\beta_4 = 0.49294$, $p = 1.00$. All values are in atomic units.

Appendix B

It is convenient to use the BHS pseudopotentials as a starting point to build the new gradient-corrected pseudopotentials. This results in substantial savings of computational effort without any loss of flexibility. Moreover, the widely used BHS pseudopotentials can be simply modified to obtain the gradient-corrected potentials. The basic idea is as follows: consider a valence state of angular momentum l. The Schrödinger equation for the radial part of the wave-function is :

$$\left[-\frac{1}{2m}\frac{d^2}{dr^2} + \frac{l(l+1)}{2mr^2} + V(r)\right]\phi_l(r) = \epsilon_l \phi_l(r) \qquad (10)$$

(in atomic units, with $\hbar = 1$). Both $\phi_l(r)$ and the DFT potential (including gradient corrections) $V(r)$ can be obtained from a self consistent all-electron calculation.

The pseudopotential equation is

$$\left[-\frac{1}{2m}\frac{d^2}{dr^2}+\frac{l(l+1)}{2mr^2}+V_{ps}(r)\right]\tilde{\phi}_l(r)=\epsilon_l\tilde{\phi}_l(r) \tag{11}$$

where $V_{ps}(r)$ is the screened pseudopotential which includes the gradient corrections, and outside a cutoff radius r_c , $\tilde{\phi}_l(r)=\phi_l(r)$. The Schrödinger equation corresponding to the BHS pseudopotential is

$$\left[-\frac{1}{2m}\frac{d^2}{dr^2}+\frac{l(l+1)}{2mr^2}+V_{ps}^{BHS}(r)\right]\tilde{\phi}_l^{BHS}(r)=\epsilon_l^{BHS}\tilde{\phi}_l^{BHS}(r) \tag{12}$$

We define $\Delta\tilde{\phi}_l(r)=\tilde{\phi}_l(r)-\tilde{\phi}_l^{BHS}(r)$. Outside the cutoff radius, $\Delta\tilde{\phi}_l(r)$ is known exactly, since both $\tilde{\phi}_l(r)$ and $\tilde{\phi}_l^{BHS}(r)$ can be obtained from a self-consistent calculation. Notice that outside the core $\tilde{\phi}_l(r)=\phi_l(r)$. However, since these two wavefunctions belong to potentials that include gradient corrections, they are different from $\tilde{\phi}_l^{BHS}(r)$. For $r<r_c$, we expand $\Delta\tilde{\phi}_l(r)=\sum_{n=0}^{3}c_n r^{l+2n}$. By requiring that $\tilde{\phi}_l(r)$ and its first and second derivatives are continuous at r_c, and imposing the normalization condition, we can solve for the coefficients c_n, $n=0,...,3$. In order to satisfy Eq. (4), $\Delta V_{ps}(r)=V_{ps}(r)-V_{ps}^{BHS}(r)$ must satisfy

$$\Delta V_{ps}(r)=\epsilon_l-\frac{\epsilon_l^{BHS}\tilde{\phi}_l^{BHS}(r)}{\tilde{\phi}_l^{BHS}(r)+\Delta\tilde{\phi}_l(r)}-\frac{\tilde{f}(r)}{\tilde{\phi}_l^{BHS}(r)+\Delta\tilde{\phi}_l(r)} \tag{13}$$

where

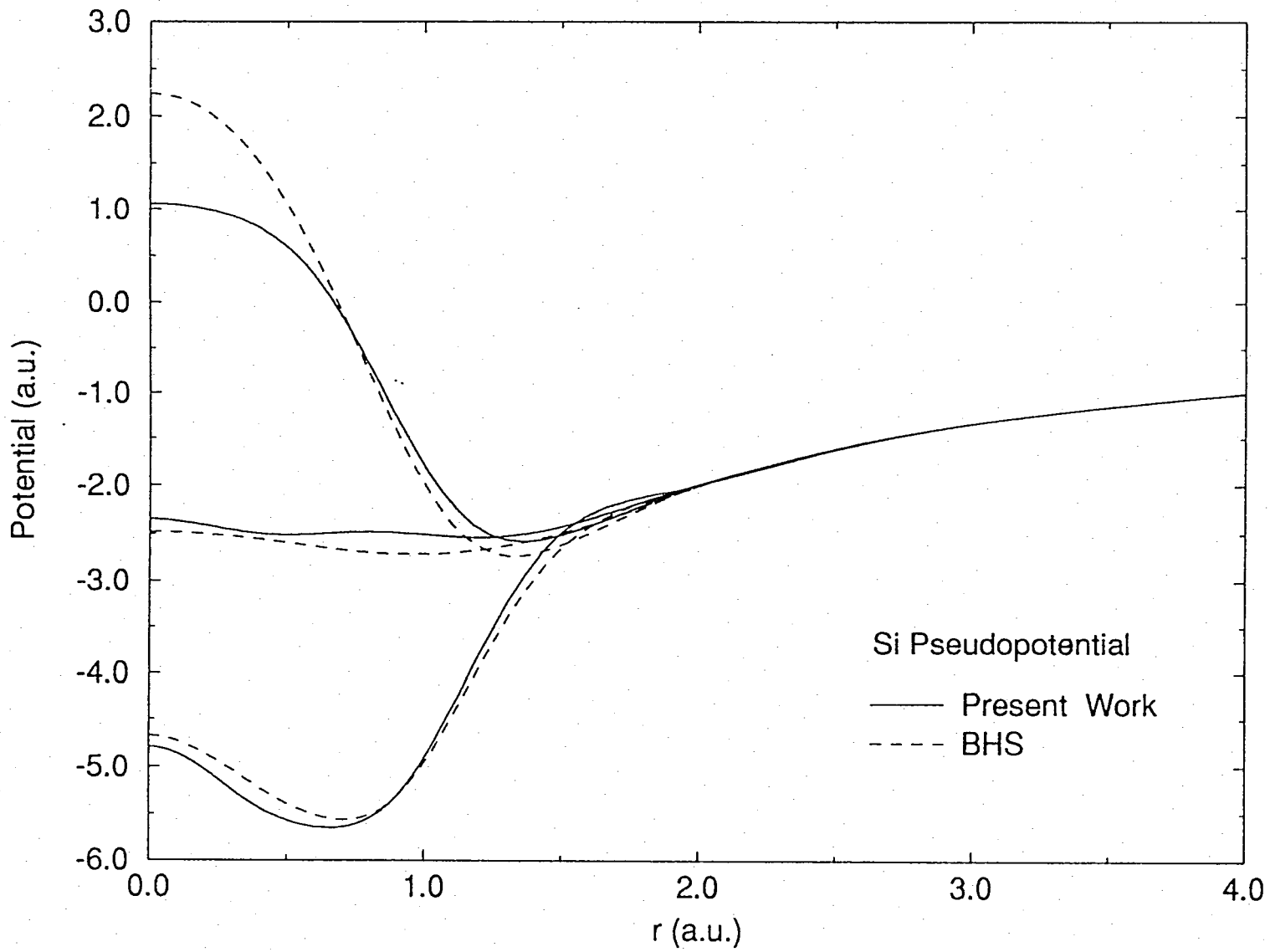
$$\tilde{f}(r)=\left[-\frac{1}{2m}\frac{d^2}{dr^2}+\frac{l(l+1)}{2mr^2}+V_{ps}^{BHS}\right]\Delta\tilde{\phi}_l(r) \tag{14}$$

Since $\Delta\tilde{\phi}_l(r)$ is exactly known everywhere, we can use Eq. (13) to obtain the screened pseudopotential $V_{ps}(r)$. By calculating the PW(91) exchange-correlation

potential and the Hartree potential from the density corresponding to the pseudo-wavefunction $\tilde{\phi}_l(r)$, we unscreen the pseudopotential and get the bare pseudopotential $V_{ps}^{ion}(r)$ to be used in the solid calculation. We then fit the numerical result for the new pseudopotential to the same functional form as in the BHS pseudopotential. For simplicity, we only refit the coefficients of the short range l-dependent part. This procedure offers the important advantage of avoiding highly nonlinear fits. In Figure 8, we compare the new pseudopotential and the BHS pseudopotential for Si. As is seen from this figure, the new pseudopotential is comparable to the BHS pseudopotential in terms of well depth and curvature.

Here we should like to point out that the PW(91) exchange-correlation potential corresponding to the pseudo-wavefunction exhibits pathological behavior near the origin (oscillations of increasing amplitude), which makes the creation of a smooth pseudopotential impossible. For this reason, we introduce a cutoff function $1/(1+e^{(r_{c0}-r)/\delta})$, where r_{c0} and δ are adjustable parameters and are both much smaller than the cutoff radius r_c. By multiplying the PW(91) exchange-correlation potential with this cutoff function, we make the potential go smoothly to zero near the origin, which makes the creation of a well behaved pseudopotential possible. This procedure is justified on grounds that the properties of the solid should not depend on features of the pseudopotential inside the core. Similar problems with the behavior of gradient corrections near the origin were noticed by Ortiz and Ballone.[4]

Figure 8: The new (gradient-corrected) and BHS pseudopotential for Si.


3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-4.0
-5.0
-6.0
Potential (a.u.)
0.0
1.0
2.0
3.0
4.0
r (a.u.)
Si Pseudopotential
Present Work
BHS

References

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[2] W. Kohn and L. Sham, Phys. Rev. **140**, A1133 (1965).

[3] For a recent review of this methodology see R.O. Jones and O. Gunnarson, Rev. Mod. Phys. **61**, 689 (1984).

[4] G. Ortiz and P. Ballone, Phys. Rev. B **43**, 6376 (1991).

[5] P. Bagno, O. Jepsen, and O. Gunnarsson, Phys. Rev. B **40**, 1997 (1989).

[6] J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, and C. Fiolhais Phys. Rev. B **46**, 6671 (1992).

[7] G. Ortiz, Phys. Rev. B **45**, 11328 (1992-I).

[8] A. Garcia, C. Elässer, J. Zhu, S.G. Louie, and M. Cohen Phys. Rev. B **46**, 9829 (1992).

[9] M. Körling and J. Häglund, Phys. Rev. B **45**, 13293 (1992).

[10] J.P. Perdew, in *Electronic Structure of Solids '91*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991)

[11] J.P. Perdew and Y. Wang, Phys. Rev. B **33**, 8800 (1986).

[12] D.J. Lacks, and R.G. Gordon, *Phys. Rev. A* **47**, 4861 (1993).

[13] D. Liberman, J.T. Waber, and D.T. Cromer, Phys. Rev. **137**, A27 (1965).

[14] M.T. Carroll, R.F.W. Bader, and S.H. Vosko, J. Phys. B **20**, 3599 (1987).

[15] R.G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989).

[16] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. **45**, 566 (1980); J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[17] G.B. Bachelet,D.R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).

[18] F.D. Murnaghan , Proc. Nat. Acad. Sci. **30**, 244 (1944).

[19] N.W. Ashcroft and N.D. Mermin, *Solid State Physics* (Rinehart and Winston, New York, 1975).

[20] M.L. Cohen and J.R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors* (Springer, New York, 1988).

[21] C. Kittel *Introduction to Solid State Physics, 4th ed.* (Wiley, New York, 1971).

[22] G. Simmons and H. Wang, *Single Crystal Elastic Constants and Calculated Aggregate Properties: A Handbook, 2nd ed.* (MIT Press, Cambridge, 1971).

[23] W.A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, 1980).

[24] O. Madelung, *Semiconductors of Group IV Elements and III-V Compounds* (Springer, New York, 1991).

[25] U. Rössler and O. Schütz, Physica Status Solidi **(b) 56**, (483) 296 (1973).

[26] S.G. Louie, S. Froyen, and M.L. Cohen, *Phys. Rev. B* **26**, 1738 (1982).

[27] C. Filippi, D.J. Singh, and C.J. Umrigar (to be published in *Phys. Rev. B*).

[28] B. Hammer, M. Scheffler, K.W. Jacobsen, And J.K. Nørskov, *Phys. Rev. Lett.* **73**, 1400 (1994).

[29] J.A. White, D.M. Bird, M.C. Payne, and I. Stich, *Phys. Rev. Lett.* **73**, 1404 (1994).

[30] B.G. Johnson, P.M.W. Gill, and J.A. Pople, J. Chem. Phys. **98** (7), 5612 (1993).

3 High-pressure Plastic Flow in Silicon

3.1 Introduction

Microindentation experiments on Si [1, 2] have shown that this material can be induced to exhibit plastic flow upon loading at room temperature. This observation is rather extraordinary since it implies that under the proper pressure conditions a normally brittle material can be transformed into a maleable, plastic one while it remains far from the molten state. [3] Consequently, the structure of the material must undergo significant changes both in the nature and in the topology of bonding during this process. In order to understand the physics underlying this process, it is necessary to study the relevant structural transformations and the ensuing changes in interatomic bonding from a microscopic point of view. A major obstacle to such studies is the lack of any information about the relevant structural changes. This demands that a very wide range of possible structures be considered and that their relevance to the phase transformation of interest be evaluated in some reliable way. In this chapter we will discuss an efficient methodology for analyzing such phenomena within the framework of first-principles theoretical calculations of the free energy.

3.2 Methods

Since it is not practical to investigate an unlimited number of possible configurations, we will rely on the magic strain concept [4, 5] to explore a limited but rather

general set of configurations, which are related to some reference structure by a mass flow transformation. The many possible choices of magic strain matrices and the versatility of their parametrization provide a convenient way for exploring configuration space. We then calculate the total energies for configurations identified by magic strains using the Local Density Approximation to Density Functional Theory [6], which is well known to provide accurate and reliable results for comparing energy differences. Finally, we use Vineyard's transition-rate theory [7] to include the entropy which provides a more complete picture of the thermodynamics associated with the process we are interested in.

Experiments [8, 9] suggest that during indentation Si assumes a metallic phase, which is believed to be described by the β-Sn structure. It is reasonable to expect the same phase to be relevant to plastic flow under microindentation. Hence, our structural studies will be based on the high-pressure β-Sn phase of Si, which is a tetragonal body-centered lattice with c/a ratio 0.522. We further assume that under the high pressure conditions in microindentation experiments, there is no memory of the ground-state diamond lattice in the material. The transformation of Si from its ground state diamond lattice to the β-Sn high-pressure phase is a fascinating phenomenon in itself which has attracted much experimental [8, 9] and theoretical attention [10], but lies beyond the scope of the present work.

Taking the β-Sn phase as the reference structure we have performed extensive investigations of deformations of Si with magic strain transformations. Let us briefly review the concept of magic strains: they are lattice transformations which

bring the crystal back to itself in many different ways. Because they are defined to be symmetric matrices, the non-relevant rotations (which correspond to inherent symmetries of the crystal and are of no interest here) are excluded. Detailed explanations of the construction and the effects of magic strains are given in Ref. [4, 5]. We will concentrate on magic strains that are likely to produce deformations corresponding to flow of atoms. Employing Van de Wall's method, [4] we generate a magic strain which connects two equivalent primitive unit cells of the body-centered tetragonal lattice given by matrices A and B:

$$A = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & (\frac{c}{2a}) \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & (\frac{c}{2a}) \end{pmatrix} \tag{15}$$

Each column of A and B defines a lattice vector in units of the lattice constant a. The two matrices differ from each other by the x-component of the third primitive lattice vector. The magic strain matrix that connects A and B is given by S_m,

$$S_m = \begin{pmatrix} 0.4832 & 0.0000 & -0.8755 \\ 0.0000 & 1.0000 & 0.0000 \\ -0.8755 & 0.0000 & 3.6553 \end{pmatrix} \tag{16}$$

Diagonalizing S_m gives the eigen-directions (-0.9684,0,-0.2495), (0.2495,0,-0.9684), and (0,1,0). Following Boyer et al. [11], we parametrize a general strain by first transforming to the coordinate system in which S_m is diagonalized, and in this frame, expressing the diagonal elements of the strain in a volume conserving form: $S_{11} = (1+f)^{-\frac{1}{3}}(1+g)^{\frac{2}{3}}$, $S_{22} = (1+g)^{-\frac{1}{3}}(1+f)^{\frac{2}{3}}$, $S_{33} = (1+f)^{-\frac{1}{3}}(1+g)^{-\frac{1}{3}}$. When

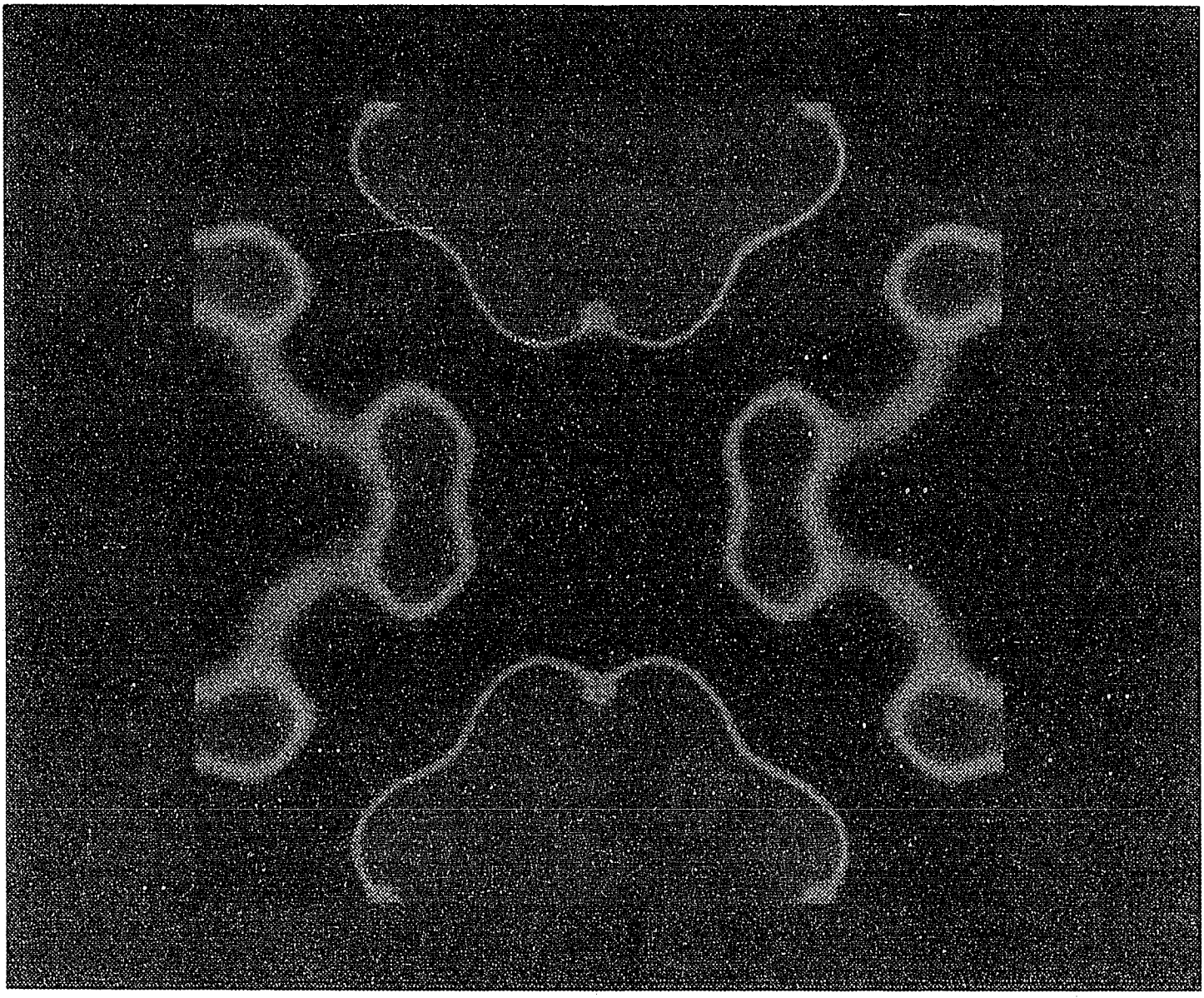
the independent variables f and g take their magic strain values $f_0 = 2.8808$ and $g_0 = -0.7423$, we obtain the magic strain matrix of eq. (2) in diagonal form. The values of f and g are parametrized by $f = f_0 \sin^2(\frac{\pi}{2}u)$ and $g = g_0 \sin^2(\frac{\pi}{2}v)$, where u, v take real values. This particular parametrization ensures that the structure will change continuously from one β-Sn lattice to another β-Sn lattice in a periodic fashion on the $\{u, v\}$ parameter plane. This formulation also facilitates the calculation of entropy, as will be evident below.

The energy for each strained lattice obtained by parametric application of S_m is then calculated within DFT/LDA. [12] The basis atoms are allowed to relax fully by using a steepest descent approach to minimize the calculated Helmann-Feynman forces. We use a plane wave expansion with a cutoff energy 16 Ry and 216 sampling points in the first Brillouin Zone for the reciprocal space integrations. With these computational parameters, the uncertainty in the calculated relative energies is of order 0.1 mRy/atom; the lowest energy barrier we have found is of order 15 mRy/atom.

3.3 Results and Discussion

Figure 9 presents a contour plot of the energy surface on the $\{u, v\}$ plane (the brighter the color, the higher the energy). The black dots represent configurations with the β-Sn structure. Energies higher than 35 mRy/atom from this reference structure are omitted to make the energy variations in the important range visible. As can be seen clearly from this plot there is a low energy path (in blue color)

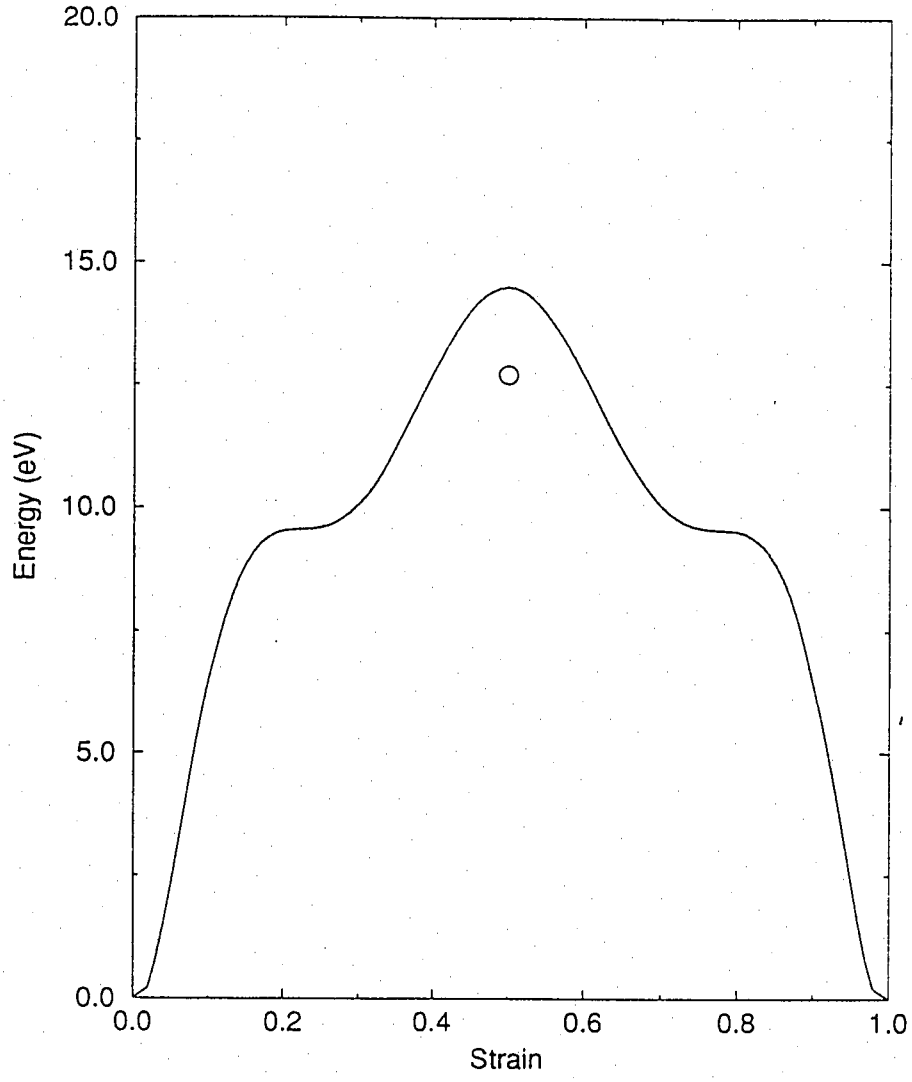
Figure 9: Contour plot of the energy surface on the $\{u, v\}$ parameter plane with u and v ranging from -1 to +1. The black area corresponds to the β-Sn structure; the blue channel is the low energy path. The color scheme denotes energy difference ΔE: black $\rightarrow 0 \leq \Delta E \leq 2.5$ mRy, blue $\rightarrow 2.5$ mRy $\leq \Delta E \leq 7$ mRy, green $\rightarrow 7$ mRy $\leq \Delta E \leq 12$ mRy, yellow $\rightarrow 12$ mRy $\leq \Delta E \leq 26$ mRy, orange $\rightarrow 26$ mRy $\leq \Delta E \leq 35$ mRy, red $\rightarrow 35$ mRy $\leq \Delta E$. The energy differences are with respect to the energy of the β-Sn phase of Si.



connecting β-Sn structures. The large variations in relative energy shown in this plot are indicative of the fact that the range of structures spanned by varying the $\{u, v\}$ parameters is very wide. The entire energy surface contains some configurations of very high energy, as high as 160 mRy/atom relative to the reference structure (these configurations lie inside the red regions of Figure 9).

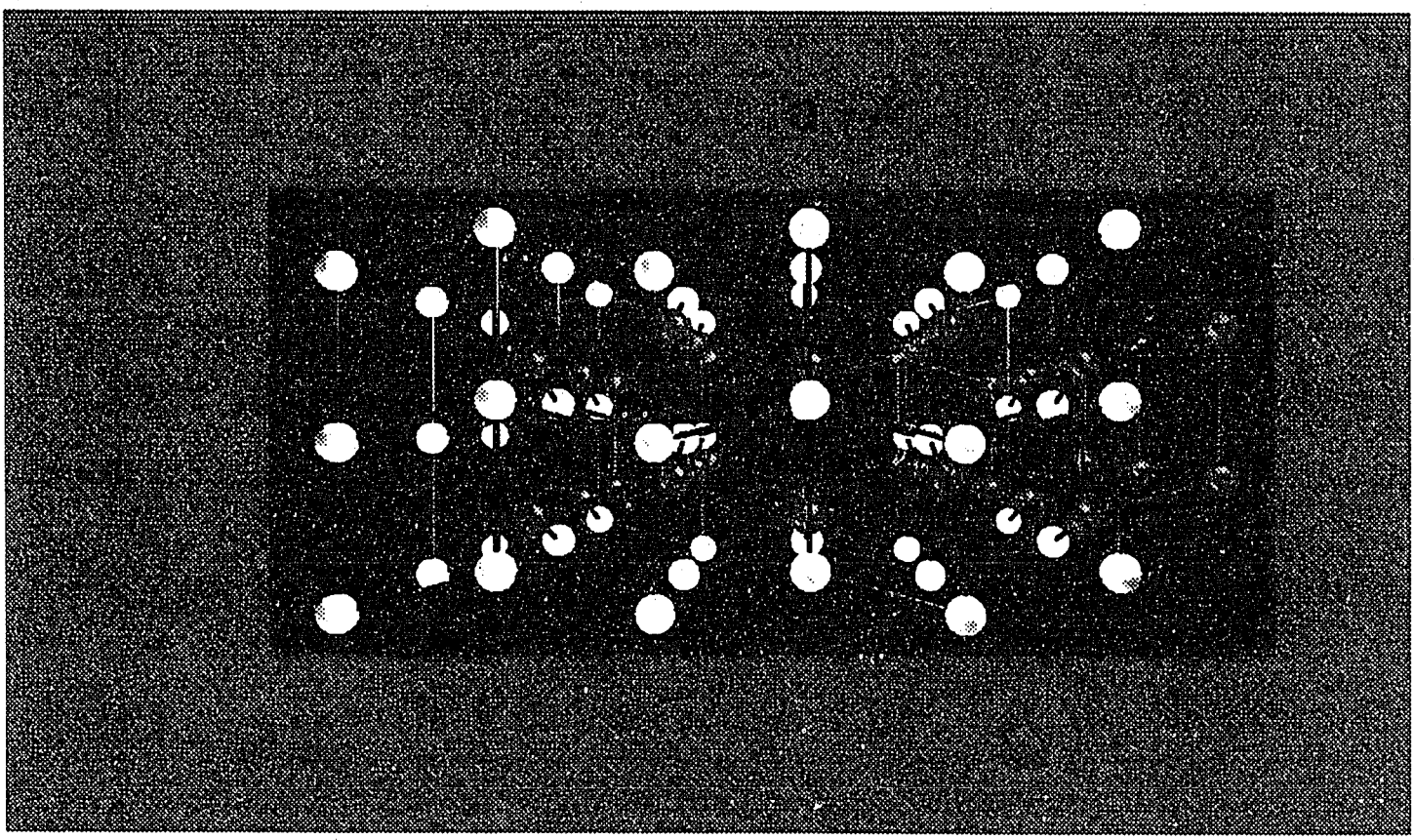
In Figure 10 we display the energy variation along the low energy path: the energy first increases quickly and almost quadratically with the strain, then it exhibits a shoulder of much smaller slope, after which it increases rapidly again until the energy barrier. The barrier is 14.5 mRy/atom. The energy is symmetric along this particular path, with respect to the barrier configuration. In our choice of the magic strain parametrization we have used a volume conserving expression. It is important to investigate how volume relaxation affects the energy barrier. We have explored this aspect for the configuration corresponding to the top of the barrier. Volume relaxation gives a decrease in the equilibrium volume of 7% and lowers the energy barrier by 1.8 mRy/atom. The reduction in the volume of the barrier configuration argues favorably for the relevance of this path to the plastic flow experiments: It is reasonable to expect that indentation will lead to compressed volume configurations, and that plastic flow will be enhanced by any structure with an equilibrium volume smaller than the reference structure. Specifically, the pressure term in the free energy ($\Delta V \cdot P$) will tend to lower the energy barrier for structures that have lower equilibrium volume than the reference structure ($\Delta V < 0$).

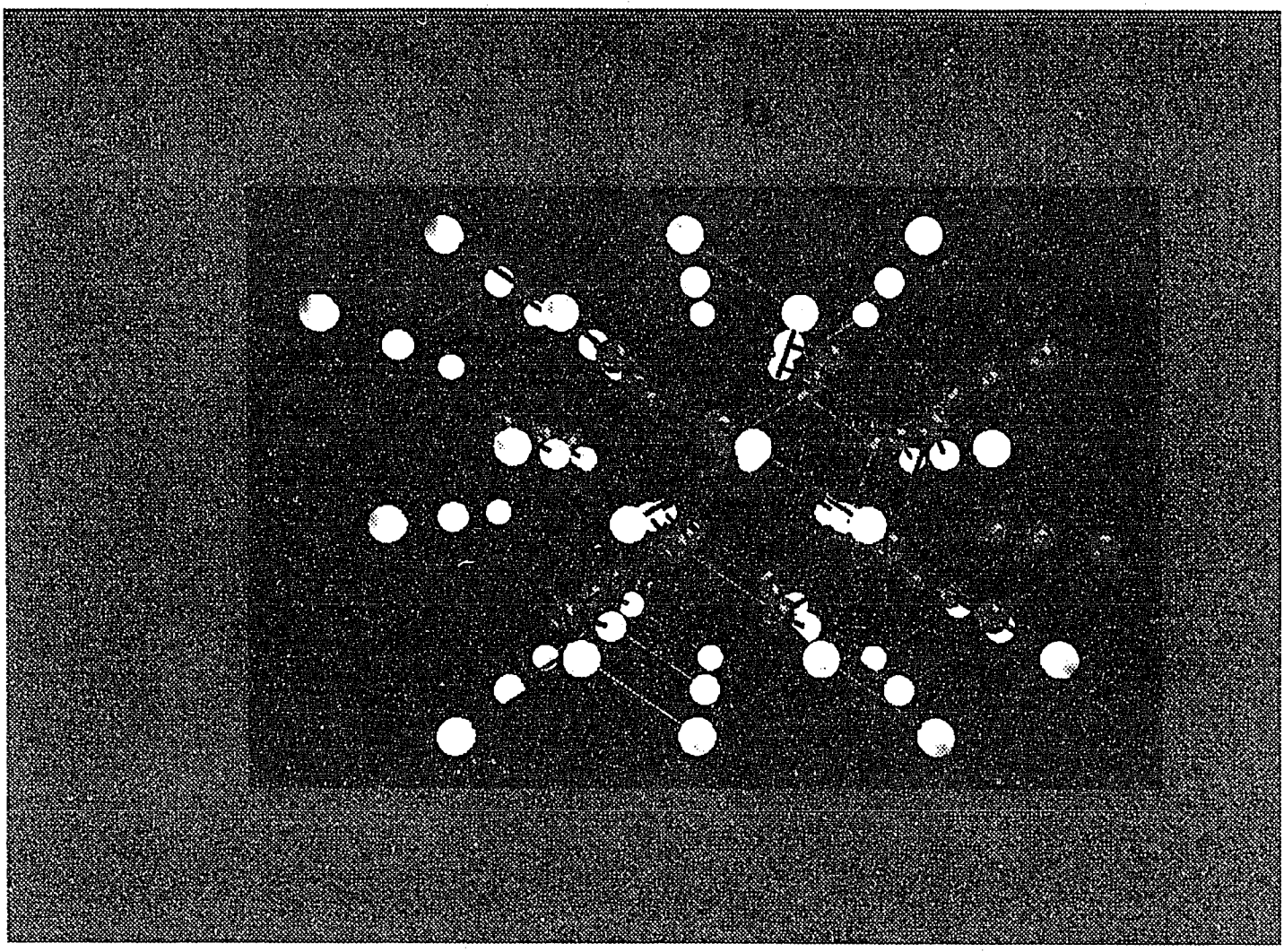
Figure 10: Energy vs. strain curve along the low energy path. The circle indicates the value of the energy after volume relaxation.

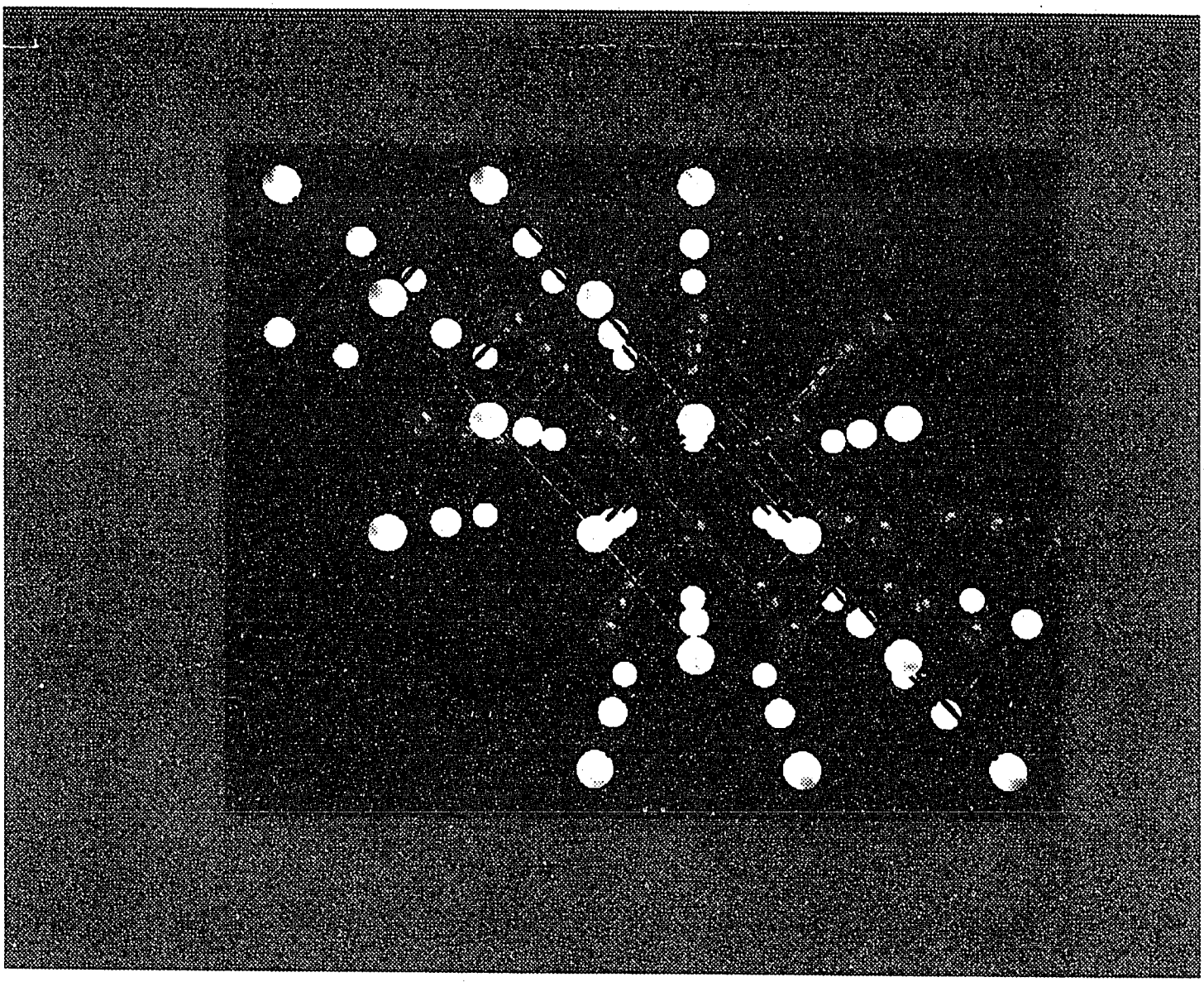

20.0
15.0
10.0
5.0
0.0
Energy (eV)
0.0
0.2
0.4
0.6
0.8
1.0
Strain

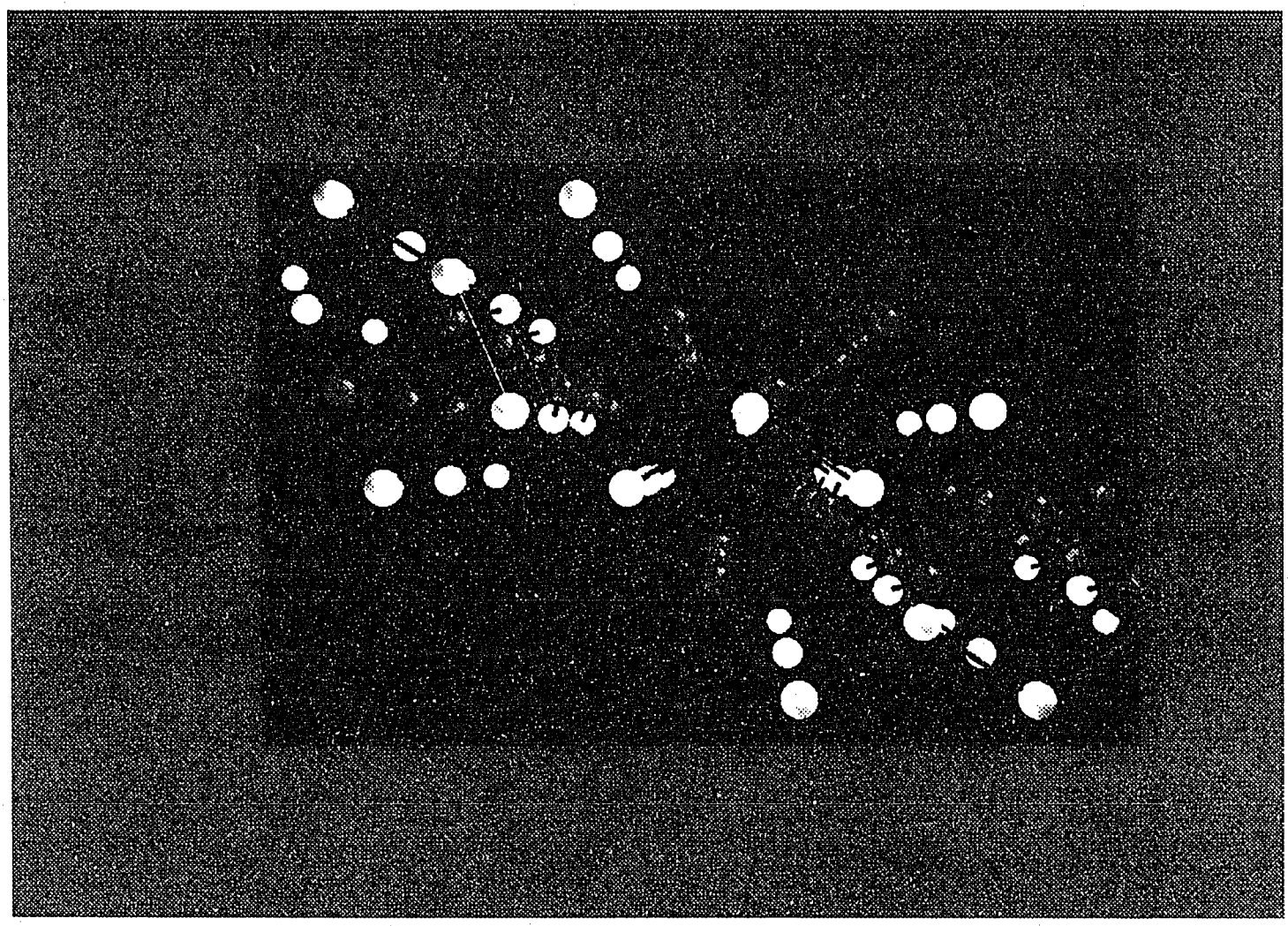
In order to visualize the structural transformation corresponding to this path, we show in Figure 11 representative configurations along the path, containing 5 layers of atoms along the c-axis of the β-Sn reference structure. Two different colors are used to identify the two interpenetrating sublattices of the body-centered tetragonal structure. Figure11(a) is the β-Sn structure of Si. In this structure each atom is six-fold coordinated with four short bonds and two longer bonds. In Figure 11(b), the pink sublattice atoms are displaced slightly to the left. At the same time, the shape of the cell is distorted. The square base on the ab plane is changed to a rectangle of smaller area, while the length along the c-axis is increased. Each atom is five-fold coordinated at this point. The distance between the two atoms in the basis (i.e. two adjacent atoms of different color) does not change very much, while their relative positions are now more aligned with the c-axis direction. This arrangement makes the angles between bonds less symmetric. Figure 11(c) is the structure corresponding to the energy barrier. The primitive cell is elongated along the c-axis even further. The bonding between different colored sublattices is broken, which facilitates the sliding of the two sublattices relative to each other. It is clear from this structure that the pink sublattice is passing through yellow sublattice. The second half the transformation is simply the mirror image of the first half with respect to the saddle point. Physically this sequence of transformations corresponds to planar flow of atoms, because an observer situated at one point in the solid will see all the atoms on the layers above his position to be moving to the left, while all the atoms below his position will be moving to the

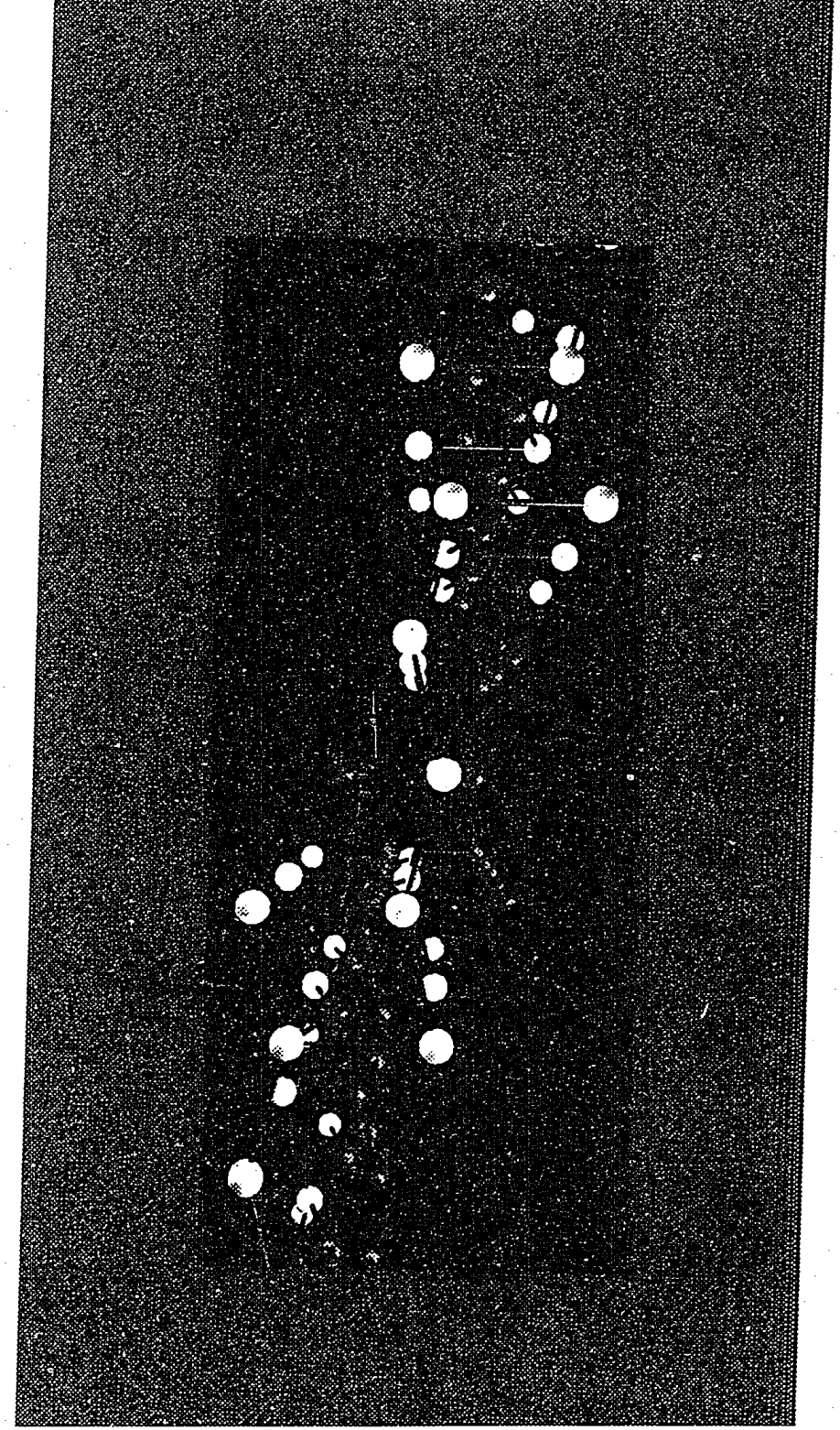
Figure 11: A sequence of structures along the low-energy path. The two colors identify the two interpenetrating sublattices of the body-centered tetragonal lattice. (a) The reference β-Sn structure. (b) A structure midway between the reference and the barrier configuration. (c) The structure corresponding to the energy barrier. (d) A structure midway between the barrier and the end-point configuration. (e) The end point configuration, which is also the β-Sn structure. Notice that in going from (a) to (e) atoms above the middle of the figure have moved to the left, whereas atoms below the middle of the figure have moved to the right, corresponding to planar flow.











right.

As can be seen from the energy surface of Figure 9, there are actually several nearby paths with comparable energy barriers to the one described above. This suggests that a more complete picture should include entropy contributions. We use Vineyard's transition rate theory to calculate the entropy associated with the energy surface, which is given by:

$$S = k_B \ln \left[\frac{\int dS \exp\left[-(E - E_a)/k_B T\right]}{\int dV \exp\left(-E/k_B T\right)} \right] \tag{17}$$

Here E_a is the activation energy (the relative energy of the barrier) and k_B the Boltzman constant; $\int dS$ is an intergal over a surface perpendicular to the transformation path around the saddle point configuration and $\int dV$ is an intergal over a volume around the equilibrium configuration. In physical terms, the intergation in the numerator of the logarithm is a measure of the number of successful crossings over the saddle point ridge while the intergation in the denominator gives the normalization, i.e. the number of available departures from the equilibrium configuration. Using this formula, we calculate the entropy associated with the present transformation at room temperature (where the experiments are performed), to be 2.9 k_B. The free energy $F = U - TS$ associated with the transformation at room temperature is therefore 7.2 mRy/atom, which corresponds to $1137^0 K$. For comparison, the melting point of Si is $1683^0 K$.

3.4 Summary

In conclusion, we have examined the possibility of plastic flow in Si under indentation. By using the magic strain concept, coupled with first-principles total energy calculations, we have been able to explore a wide range of configurations described by an energy surface in a two-dimensional strain space. We have identified a low energy path and showed that the structural transformation along this path corresponds to planar flow of atoms. We included the entropy associated with this process using Vineyard's transition rate theory. Both the entropy and the volume relaxation of the barrier configuration contribute toward lowering the free-energy cost of achieving planar flow through the path we described. The free energy cost for this particular transformation at room temperature is 7.2 mRy/atom, which is relatively low (approximately 2/3 of the melting temperature of Si). However, we caution the reader that what we have found is actually an upper bound for the energy barrier. Other paths with even lower energy barrier may exist. Such paths can in principle be found by considering magic strain transformations of lower symmetry and wider range (e.g. spanning several unit cells). The method we have presented in this chapter is quite general and should be applicable to other phenomena which involve complex structural transformations.

References

[1] G.M. Pharr, W.C. Oliver, and D.S. Harding, *J. Mater. Res.* **6**, 1129 (1991).

[2] D.L. Callahan and C. Morris, *J. Mater. Res.* **7**, 1614 (1992).

[3] It is also known that plastic flow in Si can be induced at low pressure under high temperature conditions. In this case the effects of thermal motion are probably overwhelming, requiring a completely different theoretical approach than the one described here.

[4] L.L. Boyer, *Acta Crystall.* **A45**, FC29 (1989); B.W. Van de Waal, *Acta Crystallogr.* **A 45**, FC17 (1990).

[5] E. Kaxiras and L.L. Boyer, *Modelling and Simulation in Materials Science and Engineering* **1**, 91 (1992)

[6] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964); W. Kohn and L. Sham, *Phys. Rev.* **140**, A1133 (1965).

[7] G.H. Vineyard, *J. Phys. Chem. Solids.* **3**, 121 (1957)

[8] G.M. Pharr, W.C. Oliver, and D.R. Clarke, *Scripta METALLURGICA* **23**, 1949 (1989); *J. Elec. Mater.* **19**, 881 (1990).

[9] G.M. Pharr, W.C. Oliver, R.F. Cook, P.D. Kirchner, M.C. Kroll, T.R. Dinger, and D.R. Clarke, *J. Mater. Res.* **7**, 961 (1992).

[10] M.T. Yin and Marvin L. Cohen, *Phys. Rev. B* **26**, 5668 (1982).

[11] L.L. Boyer, E. Kaxiras, J.L. Feldman, J.Q. Broughton, and M.J. Mehl, *Phys. Rev. Lett.* **67**, 715 (1991).

[12] We use the exchange-correlation functional of D.M. Ceperley and B.J. Alder [*Phys. Rev. Lett.* **45**, 566 (1980)] as parametrized by J. Perdew and A. Zunger [*Phys. Rev. B* **23**, 5048 (1981)], and norm conserving pseudopotentials from G.B. Bachelet, D.R. Hamann, and M. Schlüter, *Phys. Rev. B* **26**, 4199 (1982).

4 GSF Energy Surfaces and Structural Deformation in Silicon

4.1 Introduction

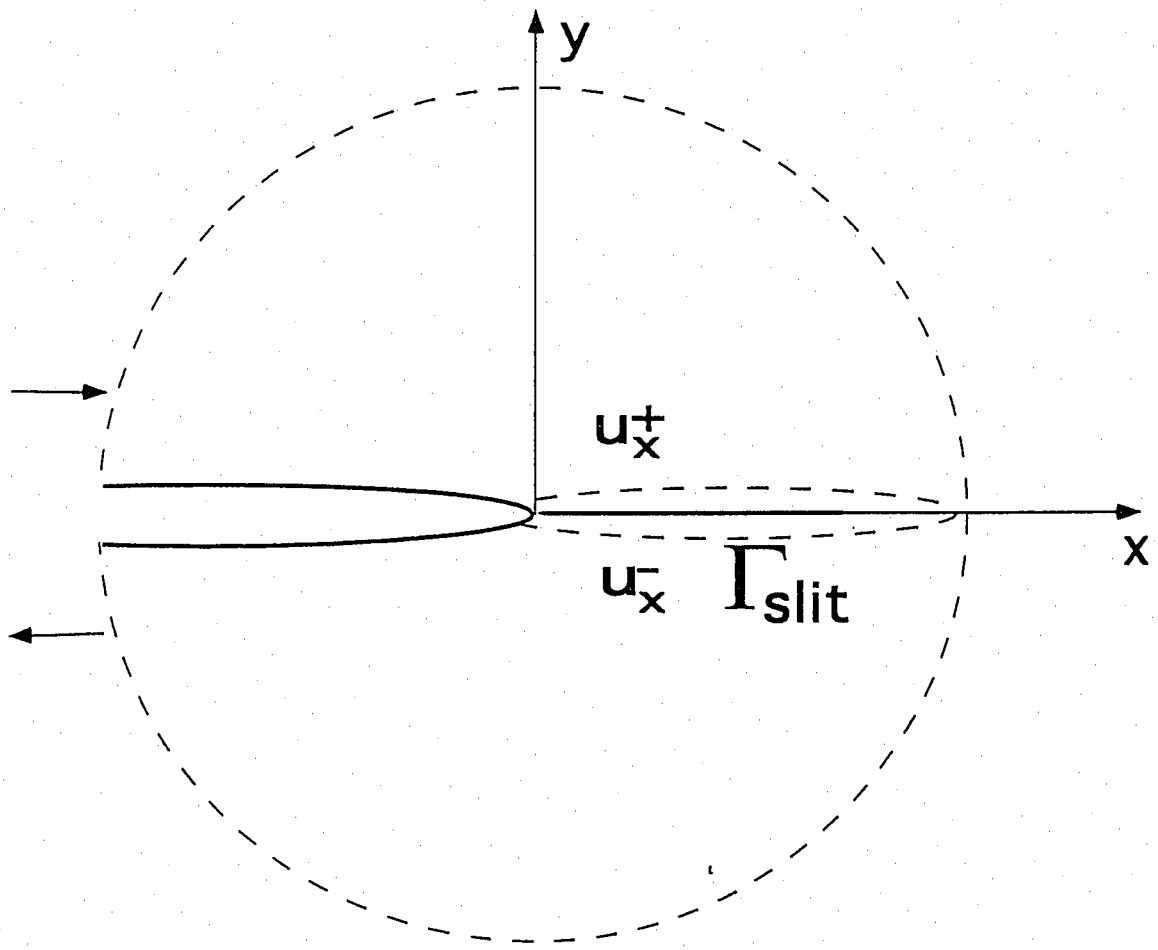
One of the most important technological problems is the issue of the intrinsic ductility of materials. To address this question, one has to consider how a crack in the solid will respond to external loading conditions. Since the dislocation nucleation from a crack tip will cause the tip to blunt, it is reasonable to associate this process with ductile behavior, while brittle behavior can be associated with crack propagation, corresponding to creation of a sharp crack tip. Therefore, one approach to analyze this problem is to compare the criterion for dislocation nucleation from the crack tip with the Griffth's criterion for cleavage [1]. Recent advances by Rice and collaborators in the modeling of dislocation nucleation at a crack tip based on the Peierls-Nabarro concept [2, 3, 5, 6, 7], identified an important solid state parameter γ_{us}, the unstable stacking fault energy, to be the controlling parameter for dislocation emission. The importance of Rice's theory is that this newly proposed dislocation emission criterion γ_{us} can be obtained from theoretical calculations without any ambiguity. By combining Rice's criterion for dislocation emission with the Griffth's criterion for crack propagation [1], the intrinsic ductility of materials can then be determined from theory. For the convenience of the reader, here we give a brief description of Rice's theory. The original papers should be

consulted for details [2, 3, 4, 5]. For simplicity, the external loading is assumed to be purely shearing along the direction of the crack front as shown in Figure 12, where $u_{\pm}$ is the displacement vector with $\{+,-\}$ referring to the two sides of the slip plane. The relative displacement across this plane δ is equal to $u_+ - u_-$. Rice's theory is based on the Pierls's stress concept: the shear stress τ on the slip plane is assumed to be a periodic function of the slip displacement δ. The energy per unit area on the slip plane associated with this displacement, $\Phi(\delta)$, is obtained with the integration $\int^{\delta} \tau(\delta')d\delta'$. The Irwin energy release rate which is related to the external loading condition can be obtained with the J integral (also called the Rice integral). For the present problem, this integral can be evaluated along the particular path Γ_{slit} as shown in Figure 12, giving

$$J = \int_0^{\infty} \tau \partial\delta/\partial x dx = \int_0^{\delta_{tip}} \tau d\delta \equiv \Phi(\delta_{tip}) \tag{18}$$

Therefore, the critical loading condition is determined by the point where the energy function Φ reaches its maximum value, γ_{us}. This value can be determined by using the same τ vs. δ relation for the block shearing process, where one half of the crystal is displaced with respect to the other half along a specific glide plane. The energy surface of Φ obtained as a function of a generalized displacement vector is called the generalized stacking fault (GSF) energy surface and γ_{us} is the lowest barrier that has to be surmonted during this block shearing process. To obtain a full thermodynamical picture of this process, we have also tried to extend Rice's original theory, which is a theory at zero temperature, to finite temperatures by

Figure 12: A schematic drawing for a crack in the solid under purely shearing external loading along the direction of the crack front, which is along the x-axis in the figure. Γ_{slit} indicates the path containing the slip zone. u_x^+ and u_x^- are the displacement vectors with the $\{+,-\}$ referring to the two sides of the slip plane respectively (after Rice, see Ref. [2]-[4]).


y
u_x^+
x
u_x^-
Γ_{slit}

including the effects due to entropy within Vineyard's transition state theory [8] (for an alternative approach to finite temperature extensions, see the work of Rice and Beltz [4]). Thus, the topology of the GSF energy surface will determine the entropy associated with this process.

Furthermore, the GSF energy surface can provide useful information on the restoring force due to the misfit of the lattice near the dislocation core. By incorporating these restoring forces into the Peierls-Nabarro model, one can calculate dislocation profiles, core energies, Peierls energies, and Peierls stresses for the dislocations on the corresponding crystal plane. This approach provides a useful description of the microscopic structural properties as well as the energetics of dislocations.

From the above discussion, it is evident that an accurate description of the GSF energy surface is desirable. Even though the most efficient method in terms of computation is the classical potential method, it is not suitable for the GSF calculations due to its inability to describe silicon accurately under different bonding situations. This limitation of the classical potential of silicon can be understood by considering the highly covalent character of its bonding where the contribution to the energy from the electronic part can not be easily separated from the ionic part. Density Functional Theory (DFT) [9], even though much more computationally demanding, has been shown to be very successful in predicting the energy differences between different structures of solids. It is also more satisfying to be able to obtain these values from a first-principles theoretical point of view, i.e.

from calculations which are free of any adjustable parameters. We have performed first-principles density functional calculations for both the shuffle and glide sets of the (111) plane as well as the (100) plane of silicon in the diamond lattice allowing full relaxation of atomic coordinates. The reason for choosing these planes is because they are the low index planes of this crystal system and therefore the natural planes to be exposed during cleavage. The effects on the GSF energy surfaces due to the relaxation of atomic coordinates are examined carefully. The possible relevance of our results to the glide-shuffle problem of dislocation in silicon [10] is discussed. The GSF energy surfaces are used with the Peierls-Nabarro model to provide detailed information on the properties of dislocations.

The remaining of the chapter is organized as follows: Section 4.2 describes the computational techniques used in our first-principles calculations for GSF energy surfaces. Section 4.3 contains our results for the GSF energy surfaces of (111) and (100) planes of silicon. The effects due to the relaxation of atomic coordinates are discussed there. In Section 4.4, we briefly describe Vineyard's transition state theory [8] and discuss how the effects due to temperature and pressure can provide some insight into the problem related to the glide-shuffle controversy of the dislocation in silicon [10]. In Section 4.5 we give a brief review of the Peierls-Nabarro model. We present the dislocation properties obtained with the calculated GSF energy surfaces and the Peierls-Nabarro model in Section 4.6. We conclude with some remarks on the usefulness of this approach based on GSF energy surfaces and Peierls stress in Section 4.7.

4.2 Methods

As discussed in Section 4.1 the GSF energy surfaces are calculated within the framework of DFT with the standard local density approximation (LDA) [11] to the exchange-correlation functional. The valence electron wavefunction is expanded in a planewave basis. The highest kinetic energy of the plane waves included in this basis is equal to to 8 Ry. The ionic potential, including the screening from core electrons, was modeled by a nonlocal norm-conserving pseudopotential [12], and the Kleinman-Bylander scheme [13] was employed to make the potential separable in Fourier space. The d component was treated as the local part of the potential with the s and p components containing the nonlocal contributions. To simulate the block shearing process, a supercell consisting of 12 atoms has been used in these calculations. For the reciprocal space integration, we have used 20 k-points in the irreducible Brillouin zone [14].

The unrelaxed energy surface is obtained by moving one part of the crystal rigidly with respect to the other half of the crystal. To obtain the relaxed energy surface, the atoms are allowed to relax following the Hellmann-Feynman forces. The four inner-most layers have been kept frozen to simulate the effects of the bulk. The structures are considered fully relaxed when the magnitude of the forces is amaller than 0.005 Ry/a.u. The Car-Parrinello [15] approach which allows simultaneous relaxation of both the ionic and the electronic degrees of freedom, has been used in the present calculations. The minimum energy was obtained with

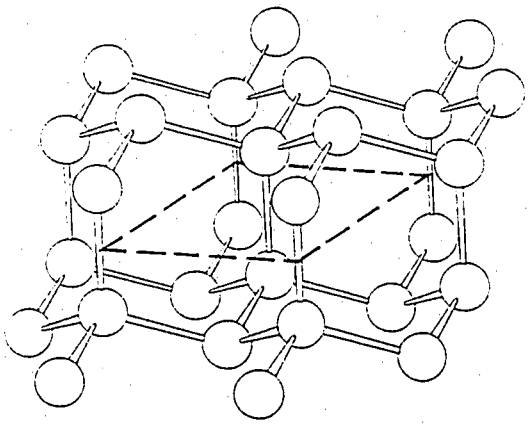
the steepest descent method. There is one subtlety needed to be clarified for the relaxation of the ionic coordinates. The GSF energy is defined as a function of the relative displacement of the two atomic planes immediately adjacent to the cut. To be consistent with this definition, we have fixed the in-plane coordinates of the ions and only allowed the atoms to relax in the direction perpendicular the glide plane within the supercell approximation.

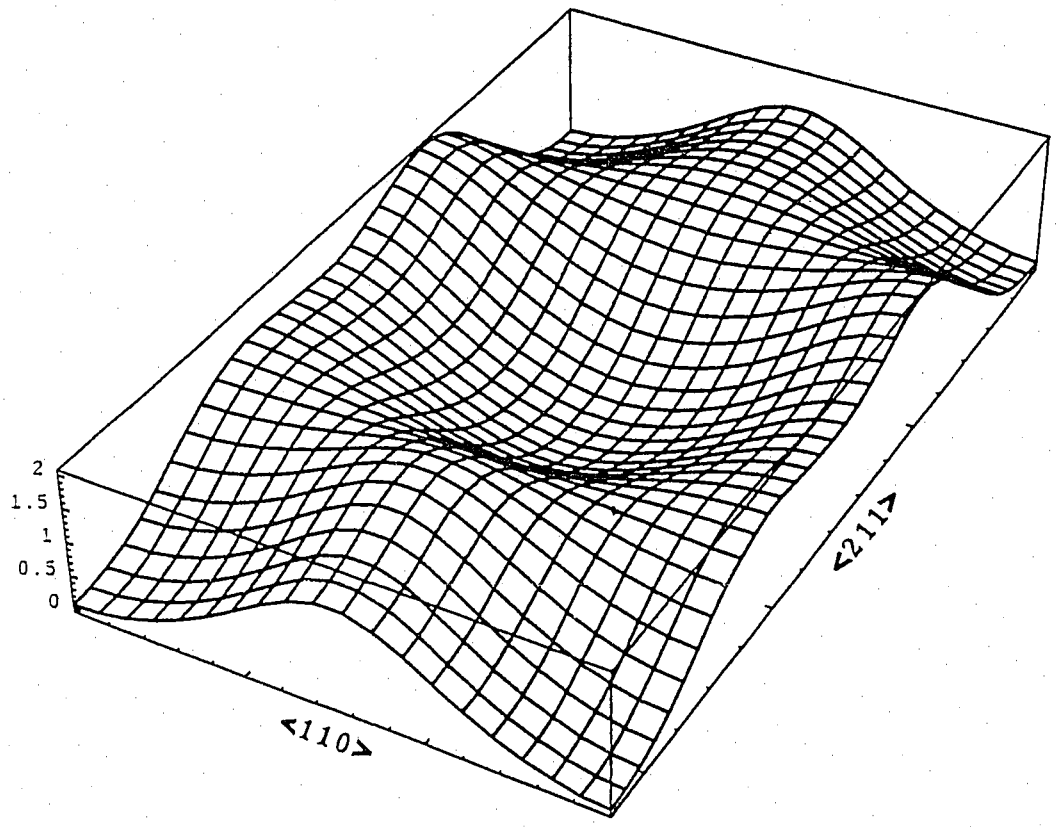
4.3 GSF Energies

We first consider the (111) plane which is the natural cleavage plane of silicon. For this plane of the diamond lattice, there are actually two different ways to cut the crystal: the shuffle set and the glide set. The atomic positions corresponding to these two sets are shown in the upper parts of Figure 13 (a) and (b) respectively. Important differences between these two ways of cutting crystal can be immediately observed by comparing these two structures. For the shuffle set, the vertical distance between the two adjacent atomic planes immediately above and below the cut is equal to the bond length. There is only one bond per atom broken during the slipping process. On the other hand, the corresponding interplanar distance for the glide set is only 1/3 of the bond length and there are three bonds per atom broken in the slipping process. The calculated GSF energy surfaces for the corresponding cuts with rigid block shearing process (the atoms in the two respective half crystals are held fixed during shearing) are displayed in the lower parts in Figure 13. The difference between the energy scales on these two energy surfaces

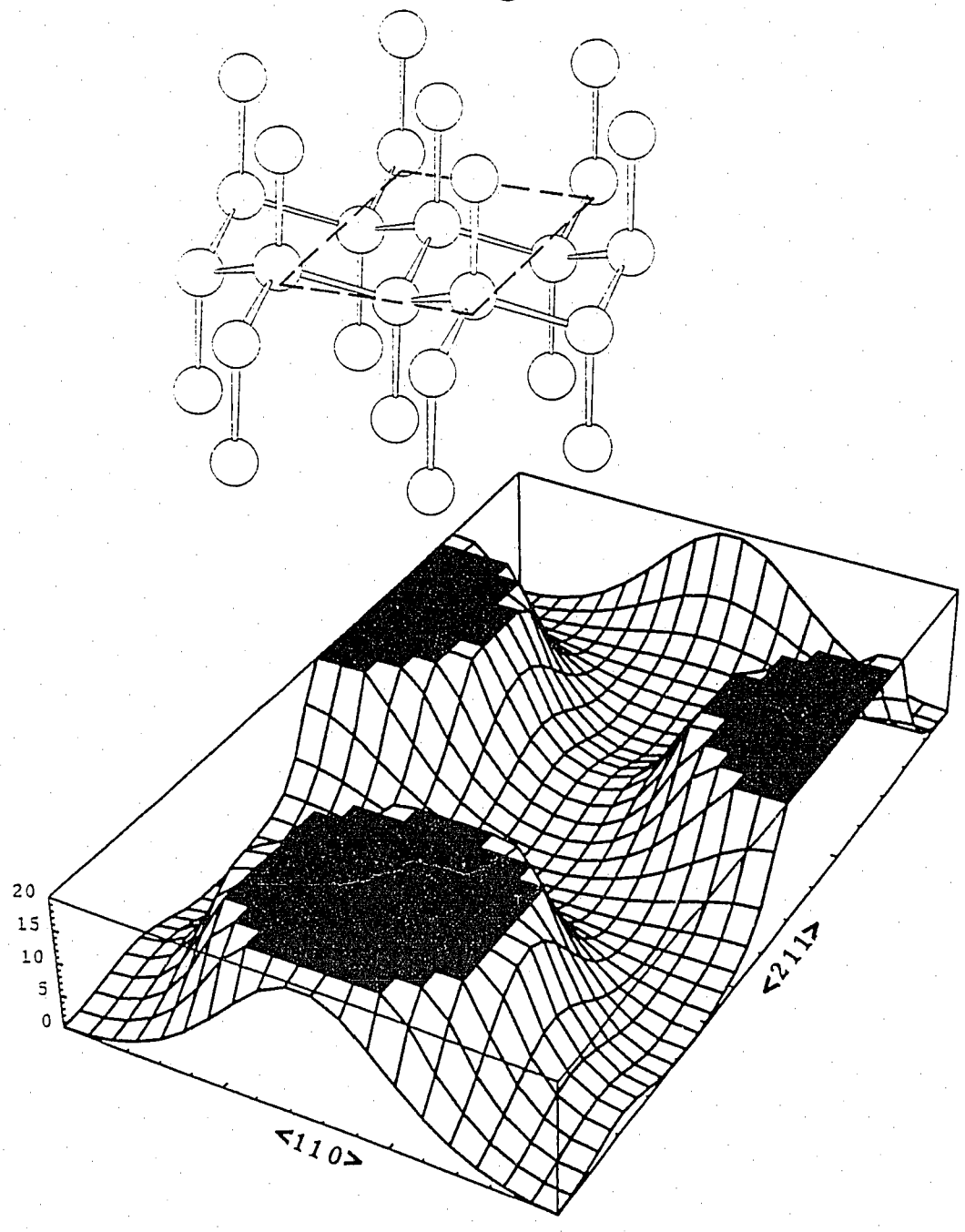
Figure 13: The calculated GSF energy surfaces with rigid block shearing for the shuffle set (a) and the glide set (b) of the (111) plane of silicon. The upper parts show the corresponding atomic structure with the dashed lines indicating the plane of the cut. The black color in the energy surface of figure (b) represents the high energy part which has been chopped off in the present figure. This figure is taken from Ref. [16] and the values of the GSF energies were reproduced during the calculations of the relaxed energy surfaces.

(a)





(b)

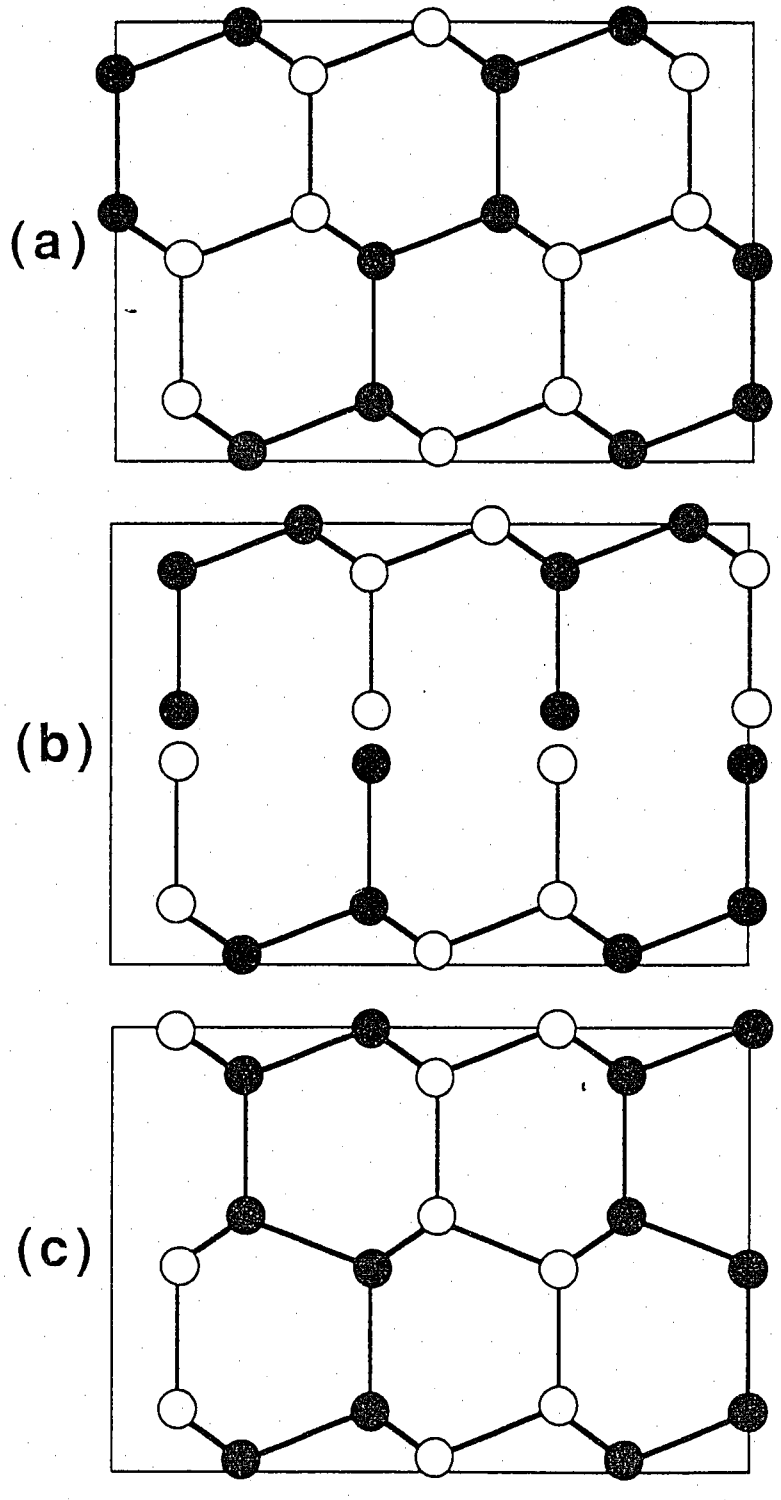


is obvious. In Figure 13, we have chopped off the part with energy higher than 20 eV/surface area, as indicated by the black color in the figure. The highest energy on the glide set is actually two orders of magnitude higher than in the shuffle set. Despite the fact that the highest energies on the two different sets are widely different, the difference between the values of γ_{us} is no more than 25%. The reason is that the lowest energy paths to slide one half of the crystal over the other half are completely different on these two cuts. On the shuffle set, the unstable stacking fault energy configuration is reached when the displacement vector is equal to one half of the primitive lattice vector in the $< 110 >$ direction, which corresponds to the Burgers vector of a full dislocation. The path is much more complicated for the case of glide set. The unstable stacking fault configuration occurs when one half of the crystal is moved with respect to the other half by 1/12 of the primitive vector in the $< 1\bar{2}1 >$ direction, The displacement vector $\frac{1}{6}[1\bar{2}1]$ corresponds to a local minimum in the energy, which represents the stacking fault configuration of silicon. The displacement vector for this stacking fault configuration is the Burgers vector of the Shockley partial dislocation. The ideal configuration can be obtained by repeating this kind of translation with a displacement vector of $\frac{1}{6}[\bar{1}\bar{1}2]$. The energy barrier along the <110> direction, which corresponds to full dislocations, is much higher on this set. To understand the origin of the high energy barriers on the glide set, we display in Figure 14 the changes of the structures when the displacement vector is along the <211> direction, where the highest energy barrier for the glide set occurs. We use two different colors here to indicate two different

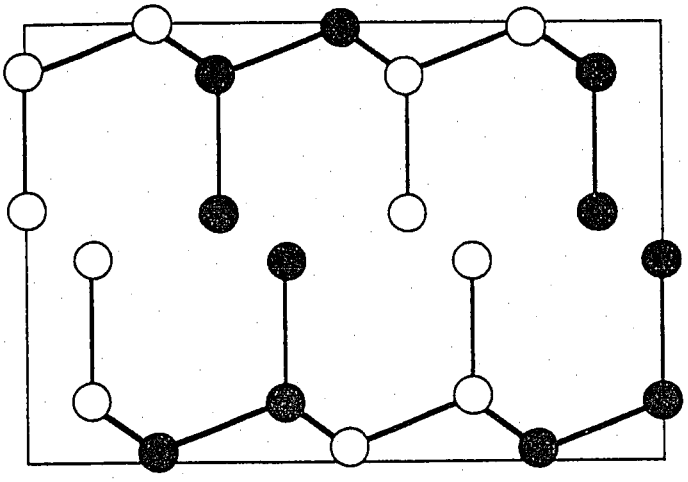
planes of atoms. The black color represents those atoms which are not on the plane of the figure, while the white atoms are on the plane of the figure. Figure 14 (a) is the original ideal structure. Figure 14 (b) is the structure for the unstable stacking fault configuration. Figure 14 (c) is the stacking fault configuration for the glide set. The reason that the energy of this structure is low is primarily due to the fact that there is essentially no difference between this structure and the original ideal structure, as far as the coordination numbers of the two sets of atoms adjacent to the cut are concerned. When the displacement is further along the <211> direction, the atoms belonging to the same color (white or black) immediately adjacent to the cut start getting very close as can be seen from Figure 14 (d) and (e). Especially in Figure 14 (e), the atoms are almost sitting on top of each other. This kind of arrangement of atoms costs a significant amount of energy due to the strong ionic interaction. Configuration 14(e) actually corresponds to the highest energy for the glide cut.

The question is then how the relaxation of the atomic coordinates may affect the energy surface. Because the distance between the two atomic planes adjacent to the cut is sufficiently large in the shuffle set, relaxation effects on this energy surface are expected to be small. Our results agree with this expectation. The energy surfaces obtained with and without relaxation are very similar. The largest decrease in the energy due to relaxation occurs at the saddle point configuration and is only 2% of the total energy. On the other hand, relaxation significantly changes the topology of the energy surface for the glide set, as is shown in Figure 15.

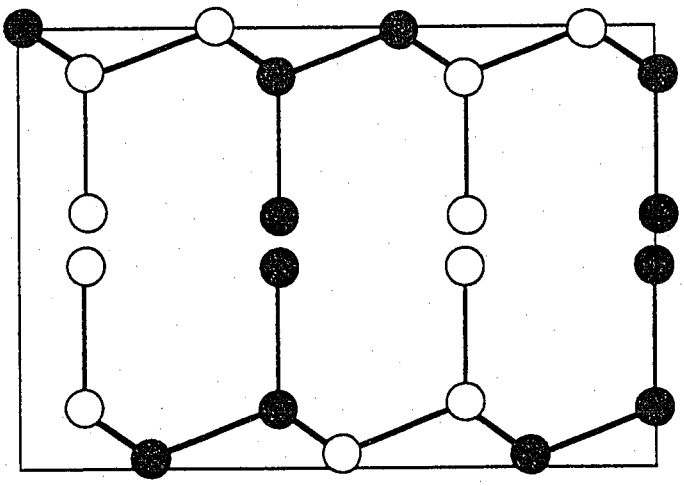
Figure 14: The change of the atomic structures when on half of the crystal is sheared rigidly with respect to the other half along the <211> direction. The two different colors represent two different planes of atoms. The white color represents atoms which are on the plane of the figure, while black atoms are not. (a) the ideal structure. (b) the unstable stacking fault configuration. (c) the stacking fault configuration. (d) and (f) are intermediate geometries, when the upper half is sheared further along the <211> direction. (e) is the geometry with the highest energy.


(a)
(b)
(c)

(d)



(e)



(f)

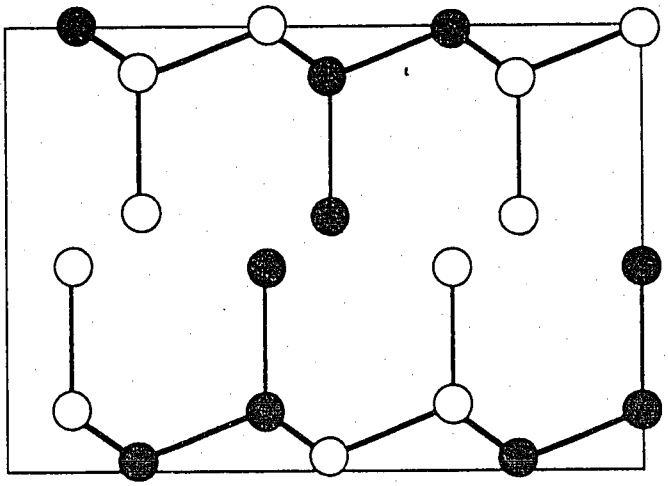
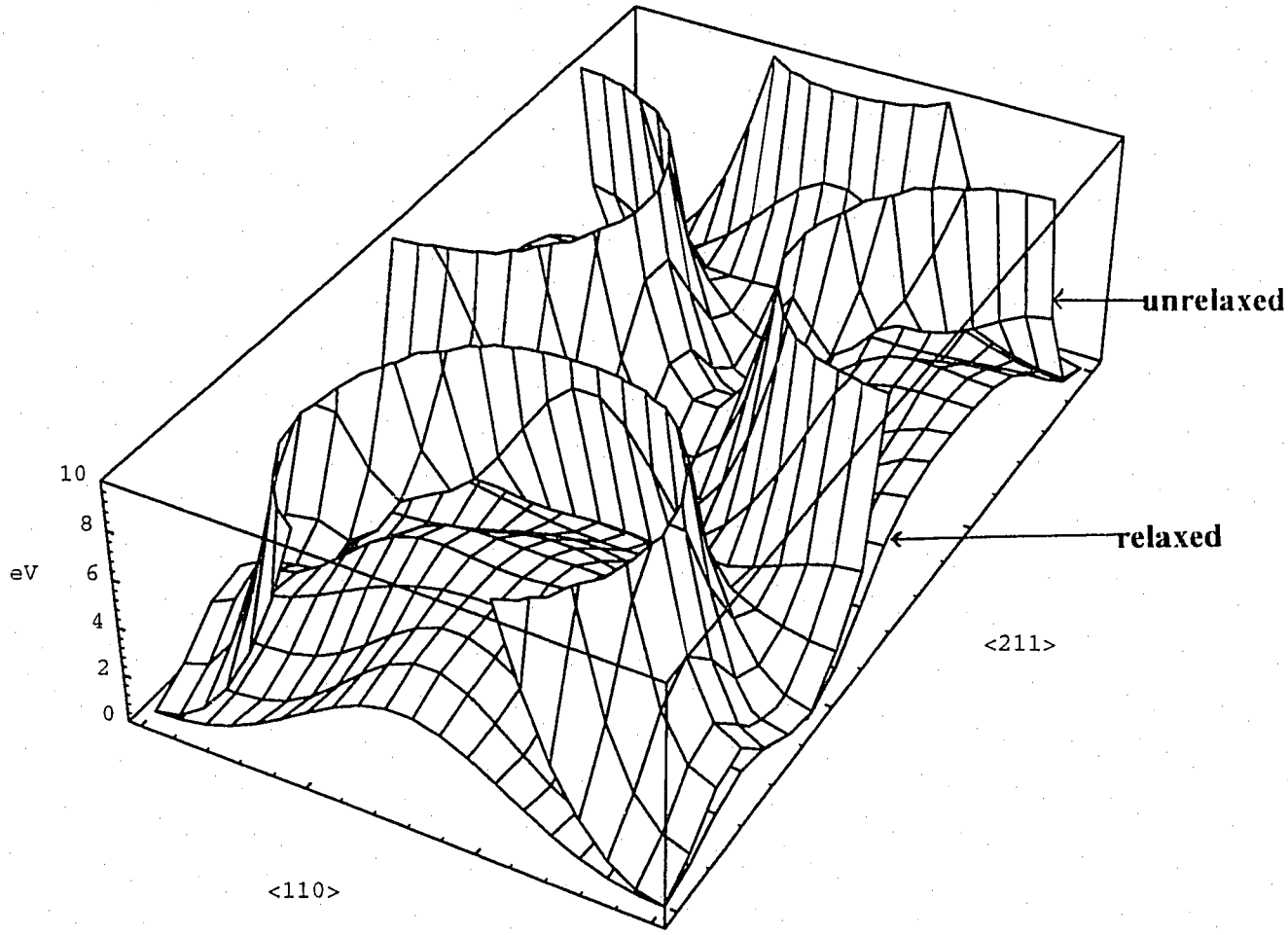


Figure 15: The relaxed GSF energy surface for the glide set is superimposed with the unrelaxed energy surface to show the effects due to the relaxation of the atomic coordinates.


unrelaxed
relaxed
eV
10
8
6
4
2
0
<110>
<211>

The high energy barrier in Figure 13 (a) has been dramatically reduced. The highest energy for the relaxed energy surface on the glide set has the same energy scale as that for the shuffle set. This makes the whole energy surface much smoother. The reason for this dramatic reduction is that through relaxation of the atomic coordinates in the direction perpendicular to the glide plane, the atoms can avoid coming very close to each other during the sliding process.

The results for the (100) plane with and without relaxation of the atomic coordinates are superimposed in Figure 16. The origin for the high energy barrier along the $<01\bar{1}>$ is similar to the situation of the glide set where atoms adjacent to the cut are very close to each other. Therefore, the relaxation of the atomic coordinates is expected to reduce the energy barrier along this direction significantly. As is shown in Figure 16, the topology of the energy surface is totally different after the positions of the atoms are allowed to relax. One point deserves further attention: $<01\bar{1}>$ which is originally the path containing the highest energy barrier has become the energetically favorable path for the gliding process on the (100) plane after the atoms are allowed to relax. The calculated values of γ_{us} both with and without the relaxation of atomic coordinates for the (111) and (100) planes of silicon are summarized in Table 5. A direct comparison based on Rice's theory indicates that it is energetically more favorable to form dislocations on the shuffle set of the (111) plane under zero temperature and zero pressure condition.

Figure 16: The GSF energy surfaces obtained for the (100) plane of silicon. For comparison, both the relaxed and unrelaxed energy surfaces are shown here.

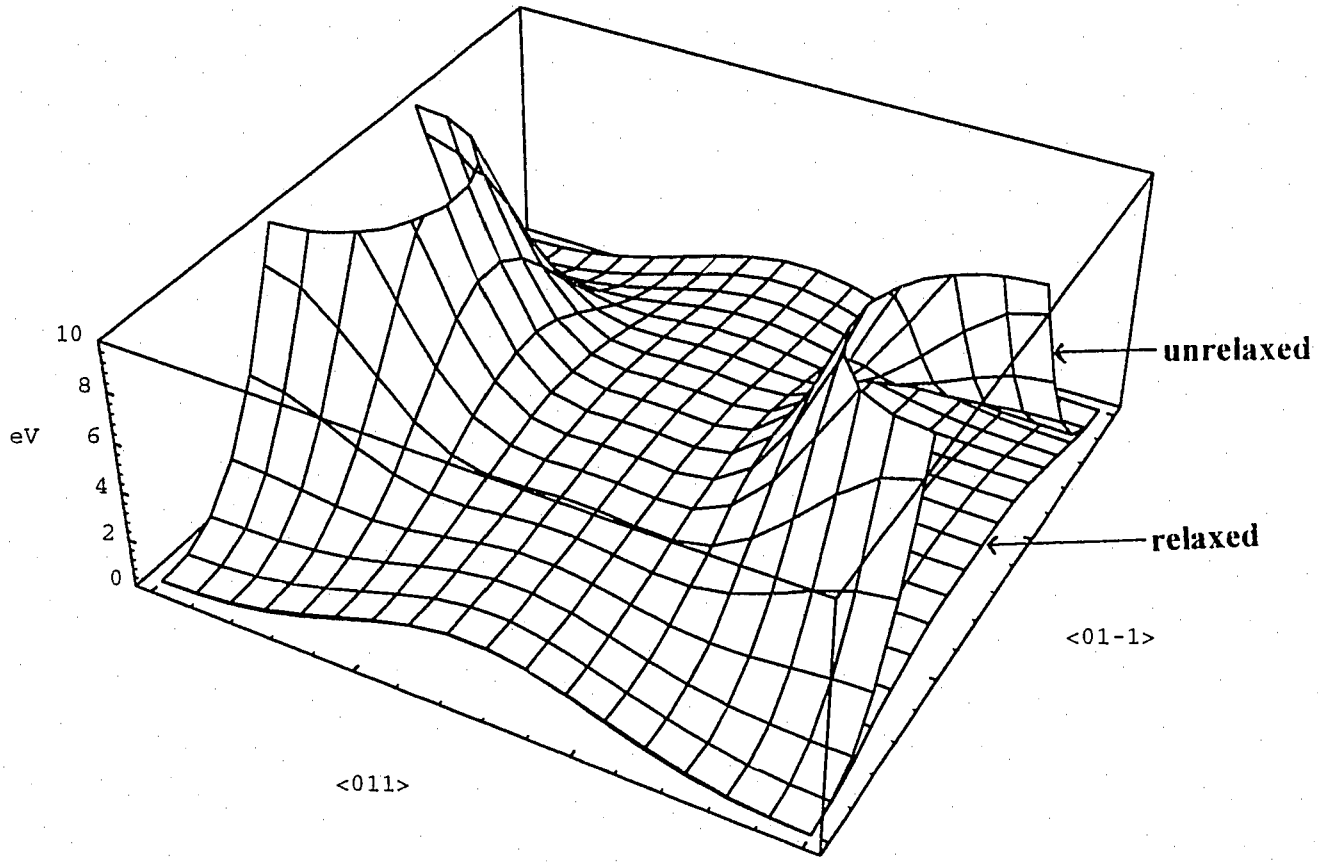

10
8
eV
6
4
2
0
unrelaxed
relaxed
<01-1>
<011>

Table 5: The unstable stacking fault energy, γ_{us}, obtained for the shuffle set and glide set of the (111) plane, as well as the (100) plane of silicon.

J/m^2	No Relaxation	Atomic Relaxation at ideal volume	Atomic relaxation and volume relaxation
Shuffle	1.84	1.81	1.67
Glide	2.51	2.02	1.91
(100)	2.97	2.15	

4.4 Temperature and Pressure Effects

Since dislocation nucleation and motion are thermodynamic processes, it would certainly be more satisfactory to have a picture which incorporates the effects due to temperature and pressure. Especially for the (111) plane, a more complete picture may shed some light into the long-standing glide-shuffle controversy [10], as was suggested by Kaxiras and Duesbery [16]. To take into account the finite temperature effects, we follow the approach given in Ref. [17]. The basic idea is that the energy surface we have obtained for this complicated process which originally involves many degrees of freedom can be thought of as the potential energy surface of the motion of a classical particle. With this mapping, we can then use Vineyard's transition state theory [8] to calculate the entropy associated with the energy surface, which is given by:

$$S = k_B \ln \left[\frac{\int dS \exp[-(E - E_a)/k_B T]}{\int dV \exp(-E/k_B T)} \right] \tag{19}$$

Here E_a is the activation energy (the unstable stacking fault energy γ_{us} for the present case) and k_B is the Boltzman constant. $\int dS$ is an intergal over a surface perpendicular to the transformation path around the saddle point configuration and $\int dV$ is an intergal over a volume around the equilibrium configuration.

In physical terms, the intergation in the numerator of the logarithm is a measure of the number of successful crossings over the saddle point ridge while the intergation in the denominator gives the normalization, i.e. the number of available departures from the equilibrium configuration. Therefore, this expression is

the logarithmic value of the probability of successful events, which is essentially the definition of entropy.

For the (111) plane, the effects due to pressure can be taken into account by comparing the volume changes at the unstable stacking fault configuration relative to the ideal structure between the shuffle and the glide slipping. The volume corresponding to the lowest energy barrier is obtained by performing first-principles calculations for several configurations whose volumes range from smaller to larger than the ideal silicon bulk crystal. The results show that the slab contracts 0.24 a.u. in the direction perpendicular the slip plane for the shuffle set, and expands 0.24 a.u. for the glide set. This is consistent with the fact that bonds are dilated in the shuffle sliding and compressed in the glide sliding.

We can write down the free energy associated with the slipping process

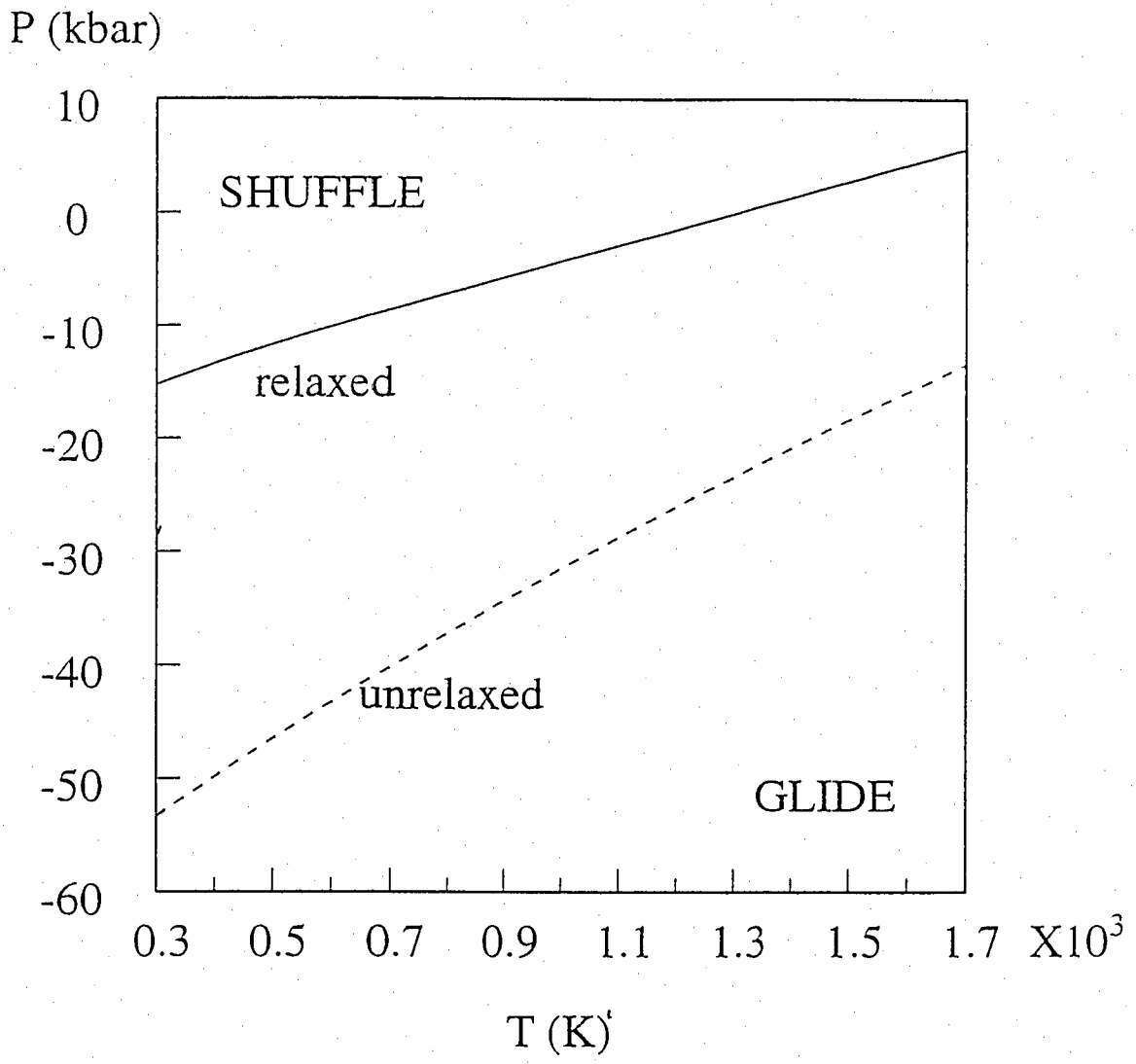
$$F^{shuffle/glide} = \gamma_{us}^{glide/shuffle} - T(\frac{S^{glide/shuffle}}{A}) + P(\frac{\Delta V}{A}) \tag{20}$$

where A indicates the surface area per atom over which the energy is calculated, and γ_{us} is the unstable stacking energy per unit area. This formula shows how the free energy for one particular slipping process depends on the thermodynamical parameters T and P. Under certain conditions, one slipping process may be more favorable compared to the other. The boundary condition over which the system changes from one mode to the other is given by the following equation,

$$F^{glide} = F^{shuffle} \tag{21}$$

The results obtained with this approach are displayed in Figure 17. The dashed

Figure 17: The shuffle set-glide set transition phase diagram on the (P,T) plane. Both results with relaxation (solid line) and without relaxation (dashed line) are shown for comparison. For (P,T) values below (above) the line, the glide (shuffle) set is more favorable for dislocation nucleation.


P (kbar)
10
0
-10
-20
-30
-40
-50
-60
SHUFFLE
relaxed
unrelaxed
GLIDE
0.3
0.5
0.7
0.9
1.1
1.3
1.5
1.7
X10^3
T (K)$^{-1}$

line is the boundary obtained with the GSF energy surface without relaxation of the atomic coordinates (see Ref. [16]), while the solid line is the one obtained with the GSF energy surface allowing full relaxation of the atomic coordinates. The glide set is more favorable for (P,T) values below the line, while the shuffle set is more favorable for (P,T) values above the line. The importance of relaxation can be observed with a direct comparison of the stress needed to change from shuffle to glide set at room temperature (300^0 K). While an external pressure of only 15 kbar is needed as calculated from the results with the relaxed energy surface, as high as 53 kbar of pressure is required when the numbers are taken from the energy surface without the relaxation of the atomic coordinates. However, we would like to caution the reader not to take these numbers literally for the following reasons: (1) During the calculation of the entropy, there are other factors (for example, the electronic degrees of freedom) which are neglected in the entropy calculation. The inclusion of these effects may further increase the entropy and reduce the necessary stress and temperature for the transition from the shuffle to the glide set. (2) This comparison is based on Rice's theory in which the surface energy needed for the creation of the ledge surface during dislocation emission is neglected. Therefore, further investigations are needed to get an estimate of the effects due to the ledge. A more detailed discussion on this subject will be given in Chapter 5. All these approximations may affect the exact numbers for the transition even though the qualitative picture will probably remain the same.

4.5 Peierls-Nabarro Model for Planar Dislocation

By combining our results for the GSF energy surfaces with the Peierls-Nabarro model, we can further analyze the properties of dislocations on these planes within the elastic model. We give a brief description of the Peierls-Nabarro model following the discussion of [18, 19, 20]. The Peierls-Nabarro model is based on the concept that a dislocation can be thought of as a continuous distribution of infinitesimal crystal misfits on the glide plane. For each point on the glide plane at a distance x' from the dislocation line, there is a corresponding infinitesimal burgers vector $df' = df/dx|_{x'}dx'$, where df/dx is defined to be the dislocation density $\rho(x)$ and $f(x)$ is the disregistry at the point x, satisfying the condition:

$$\int_{-\infty}^{+\infty} \rho(x')dx' = \int_{-\infty}^{+\infty} \frac{df(x')}{dx'}dx' = b \tag{22}$$

b is the Burgers vector. From the elastic model, the stress due to a dislocation along the direction of the Burgers vector on the slip plane is given by

$$\sigma_{bn} = \frac{K}{2\pi}\frac{b}{x} \tag{23}$$

Where x, n, and b are the distance between the dislocation and the point where the stress is evaluated, the normal to the glide plane, and the Burgers vector respectively. The constant K depends on the elastic properties of the crystal and the dislocation type. Its value for an isotropic solid has been derived [19]:

$$K = \mu\left[\frac{\sin^2(\theta)}{(1-\nu)} + \cos^2(\theta)\right] \tag{24}$$

where μ and ν are the shear modulus and Poisson ratio respectively. θ is the angle between the dislocation line and the Burgers vector. Therefore, the stress due to the existence of the dislocation for a point x on the glide plane is:

$$\frac{K}{2\pi}\int_{-\infty}^{+\infty}\frac{\rho(x')}{x-x'}dx' \tag{25}$$

Meanwhile, there is another stress which is due to the local disregistry of the crystal and tends to restore the crystal. This crystal restoring stress is a periodic function of the displacement and can be represented as $F_b(f)$. Equilibrium is attained when the stresses due to the two terms mentioned above are in balance with each other as expressed in the following equation, known as the Peierls-Nabarro equation,

$$\frac{K}{2\pi}\int_{-\infty}^{+\infty}\frac{\rho(x')}{x-x'}dx' = F_b(f(x)) \tag{26}$$

The dislocation profile $f(x)$ and $\rho(x)$ can be determined by solving this equation with the normalization condition $\int_{-\infty}^{+\infty}\rho(x)dx = b$. For the Peierls-Nabarro equation, when a simple sinusoidal form is assumed for the restoring stress,

$$F_b(f(x)) = F_{max}\sin(\frac{2\pi f(x)}{b}) \tag{27}$$

with F_{max} the maximum stress, there is an analytic solution for $f(x)$

$$f(x) = \frac{b}{\pi}\tan^{-1}\frac{x}{\zeta} + \frac{b}{2} \tag{28}$$

Here, the parameter $\zeta = Kb/4\pi F_{max}$ can be viewed as the width of the dislocation within this model, since the value of dislocation density ρ at this point is exactly one half its value at $x = 0$. For our purpose, the important thing is that the crystal

restoring force $F_b(f)$ can be obtained by simply taking the derivative of the GSF energy surface $\gamma(f)$ of our first-principles calculation results with respect to the lattice mismatch f [18, 21]. But for a general functional form of F, there is no analytic solution. We follow the recipe given by Joós et al. [19] to solve the Peierls-Nabarro equation numerically. This is done by simply expressing the disregistry vector $f(x)$ as a series

$$f(x) = \frac{b}{\pi} \sum_{i=1}^{n} \alpha_i \tan^{-1} \frac{x - x_i}{\zeta_i} + \frac{b}{2} \tag{29}$$

with the parameters α_i, ζ_i, x_i to be determined, subject to the normalization condition

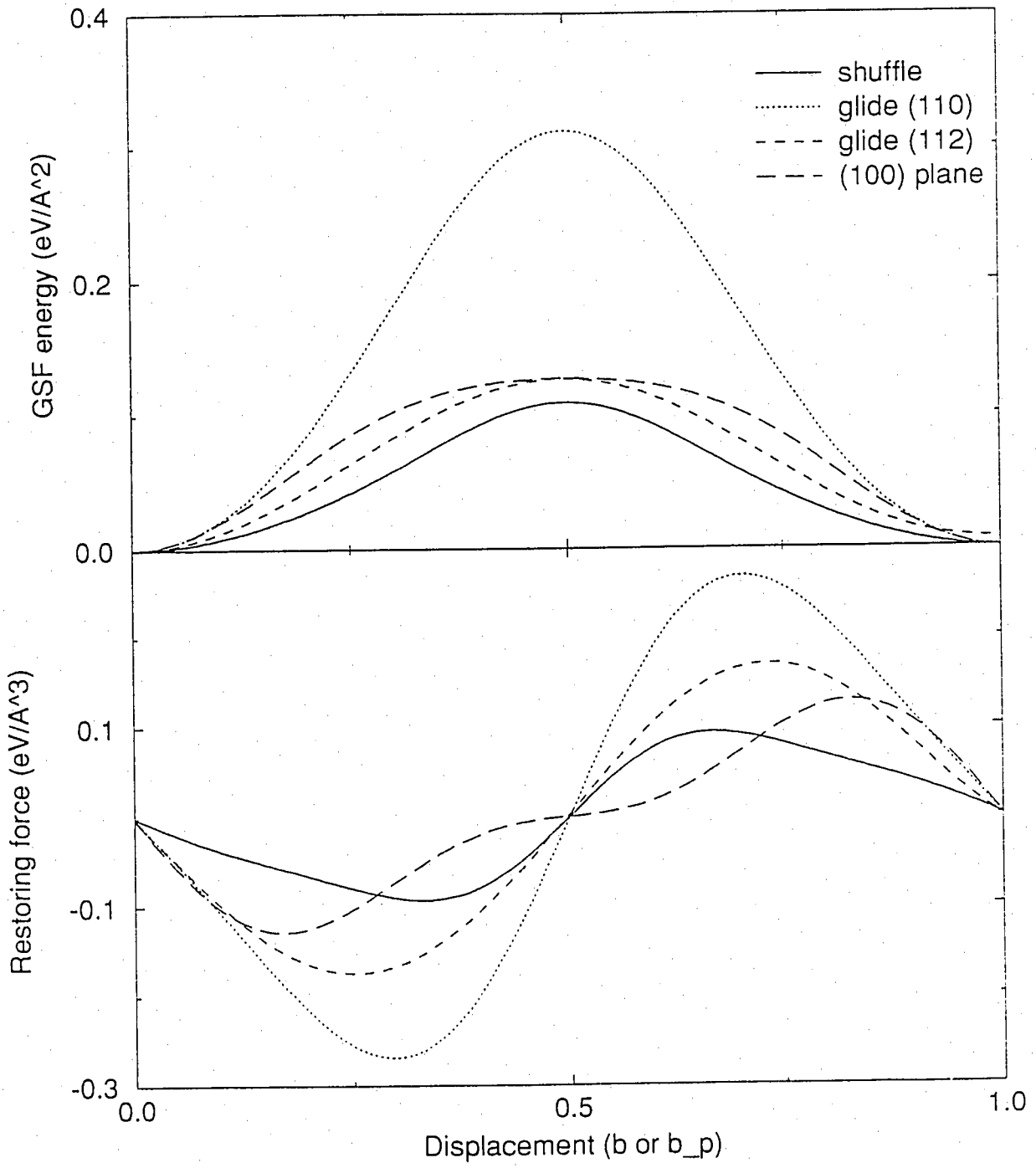
$$\sum_{i=1}^{n} \alpha_i = 1 \tag{30}$$

as the the total Burgers vector should be equal to b. By substituting the above displacement formula into the two sides of the Peierls-Nabarro equation, we get the forces $F(\alpha_i, \zeta_i, x_i, x)$ and $F'(\alpha_i, \zeta_i, x_i, x)$ corresponding to the contributions from dislocation distribution and crystal mismatch respectively. The difference between these two forces $|F(\alpha_i, \zeta_i, x_i, x) - F'(\alpha_i, \zeta_i, x_i, x)|$ is then minimized by varying the parameters α_i, ζ_i and x_i so that Peierls-Nabarro equation is satisfied numerically. As in Ref. [19], a total of three terms has been used for the expansion of $f(x)$.

4.6 Dislocation Properties

We display in Figure 18 the GSF energies and forces along the relevant directions obtained from our first-principles calculations with the full relaxation of the atomic

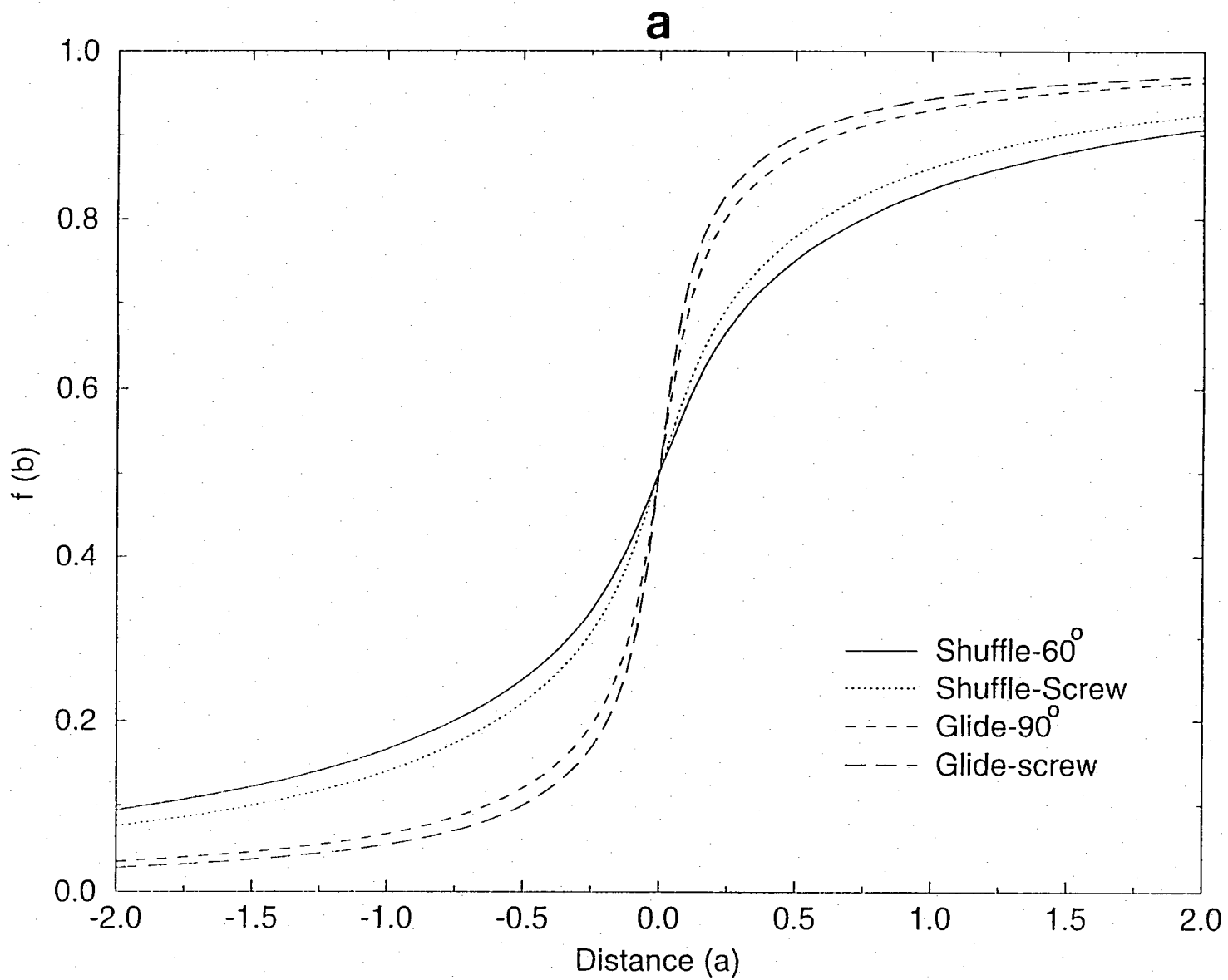
Figure 18: The fully relaxed GSF energy curve and the corresponding stress for the relevant directions on the (111) and (100) planes of silicon. The solid line is for the <110> direction on the shuffle set. The dotted line is for the <110> direction on the glide set. The dashed line is for the <211> direction on the glide set. The long dashed line is for the case of (100) plane.

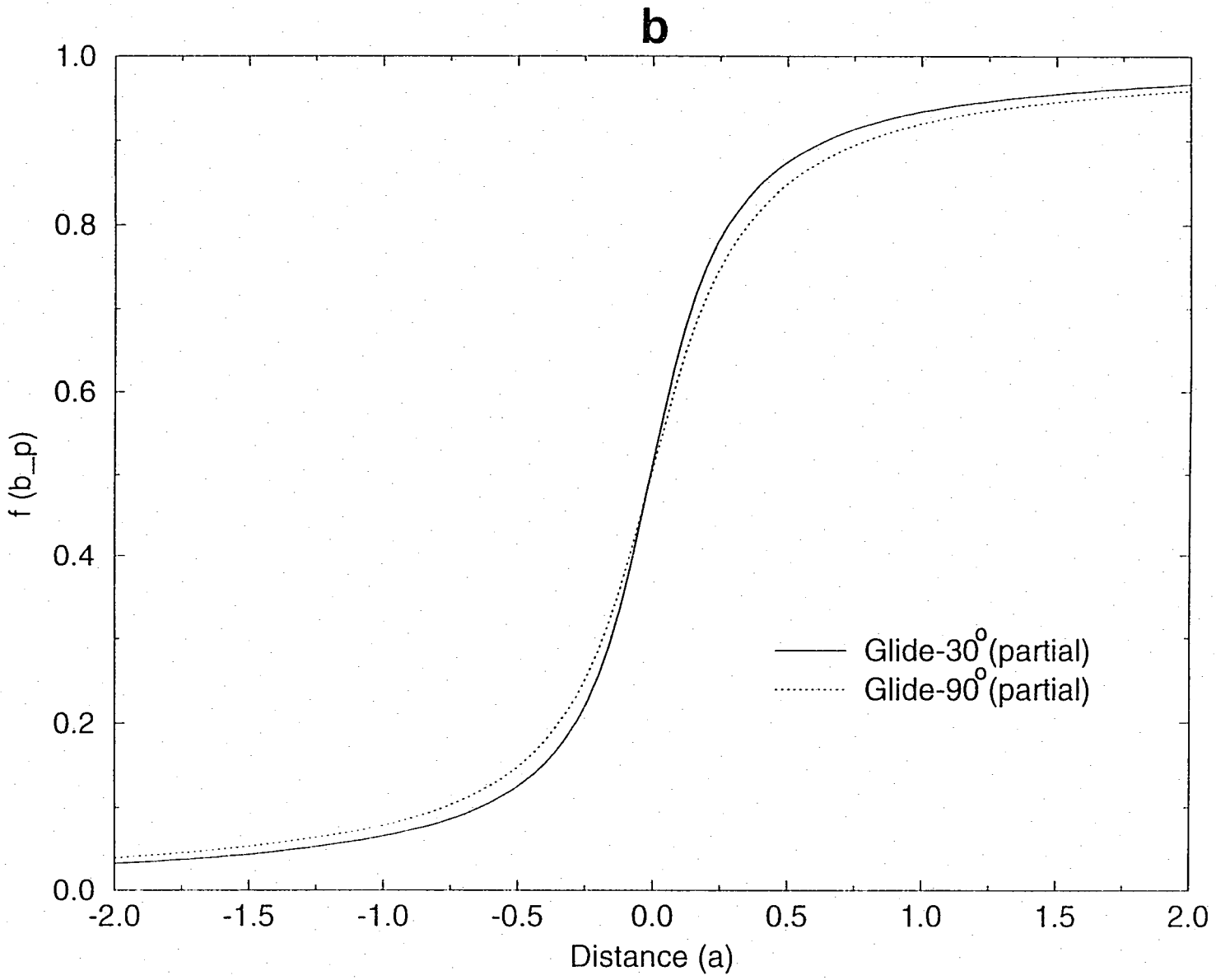

shuffle
glide (110)
glide (112)
(100) plane
GSF energy (eV/A^2)
Restoring force (eV/A^3)
Displacement (b or b_p)
0.4
0.2
0.0
0.1
-0.1
-0.3
0.0
0.5
1.0

coordinates. For the case of <211> direction on the glide plane, only the portion ranging from ideal structure to the stacking fault configuration is shown, since this is the part relevant to the formation of a partial dislocation. One thing that immediately comes to our attention is that not all the restoring forces can be described well by the sinusoidal from. Especially in the case of the (100) plane, the force curve is very flat when the displacement is approximately one half of the full Burgers vector. The significance of this deviation will become obvious when we discuss the profiles of the various dislocations.

The disregistry vector obtained for the full dislocation on the (111) plane, the partial dislocations on the glide set of the (111) plane, and the dislocations on the (100) plane are displayed in Figure 19 (a), (b), and (c) respectively. We can see that the dislocations with Burgers vector along the (110) direction on the glide plane (glide-30^0 and glide-90^0) are more concentrated compared to the other dislocations on the (111) plane, which will become even more obvious when we compare the dislocation density. This is expected since the magnitude of the crystal restoring force is considerably larger along this particular direction which forces the structural distortion to be more localized. A qualitative estimate can be obtained by simply using the formula from the classical model where the half width of the dislocation ζ has been found to be inversely proportional to the magnitude of the crystal restoring force. In Figure 20 we show the representative dislocation densities for the planes we have considered, including (a) the 60^0 dislocation on the shuffle set, (b) the 60^0 dislocation on the glide set, (c) the 90^0 partial dislocation

Figure 19: (a) The disregistry vector obtained for the full dislocation for both the shuffle and the glide set on the (111) plane; (b) the same quantity for the partial dislocations on the glide set of the (111) plane; (c) the disregistry vector for the (100) plane.


a
1.0
0.8
0.6
0.4
0.2
0.0
f (b)
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Distance (a)
Shuffle-60°
Shuffle-Screw
Glide-90°
Glide-screw


b
1.0
0.8
0.6
0.4
0.2
0.0
f (b_p)
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Distance (a)
Glide-30°(partial)
Glide-90°(partial)

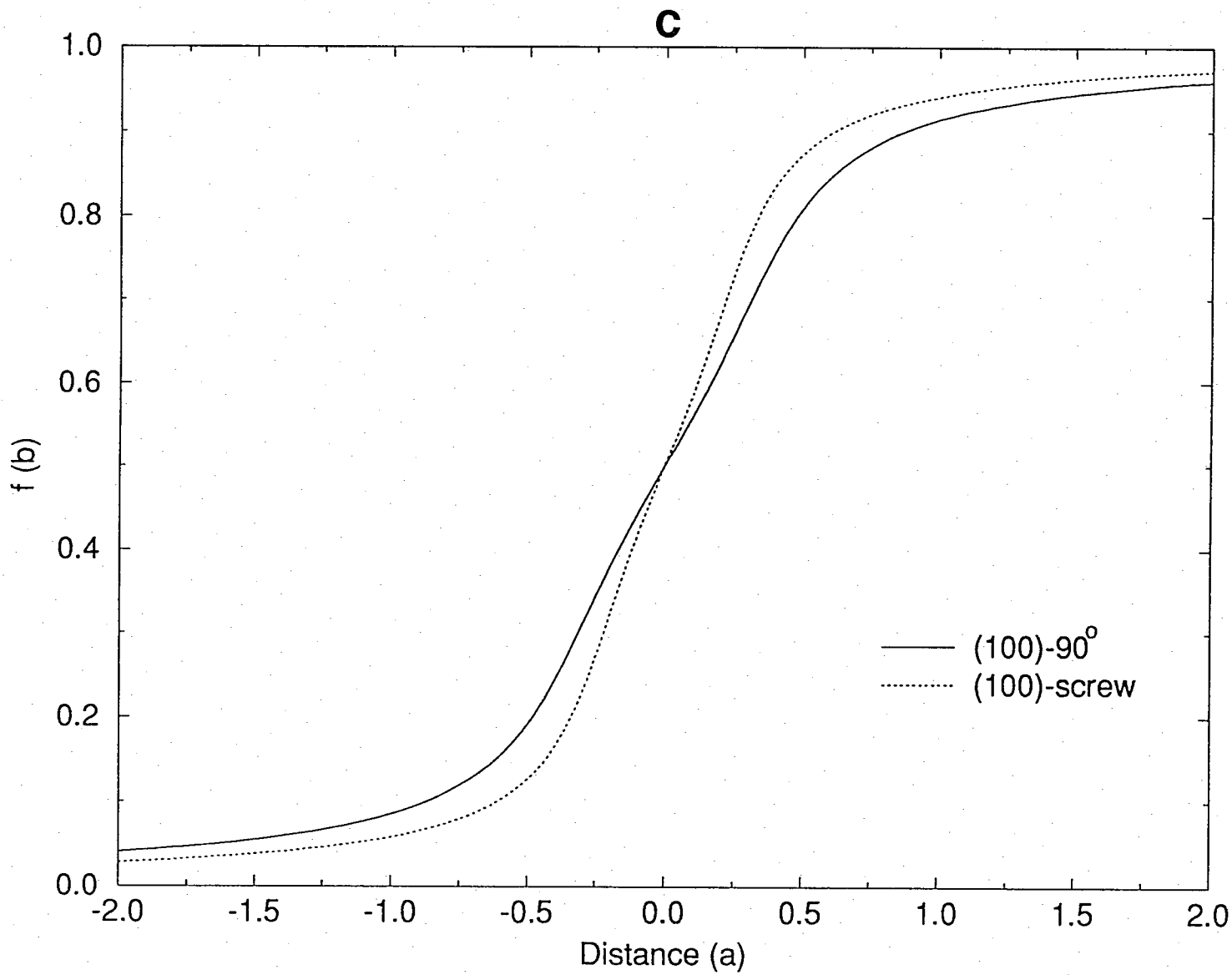
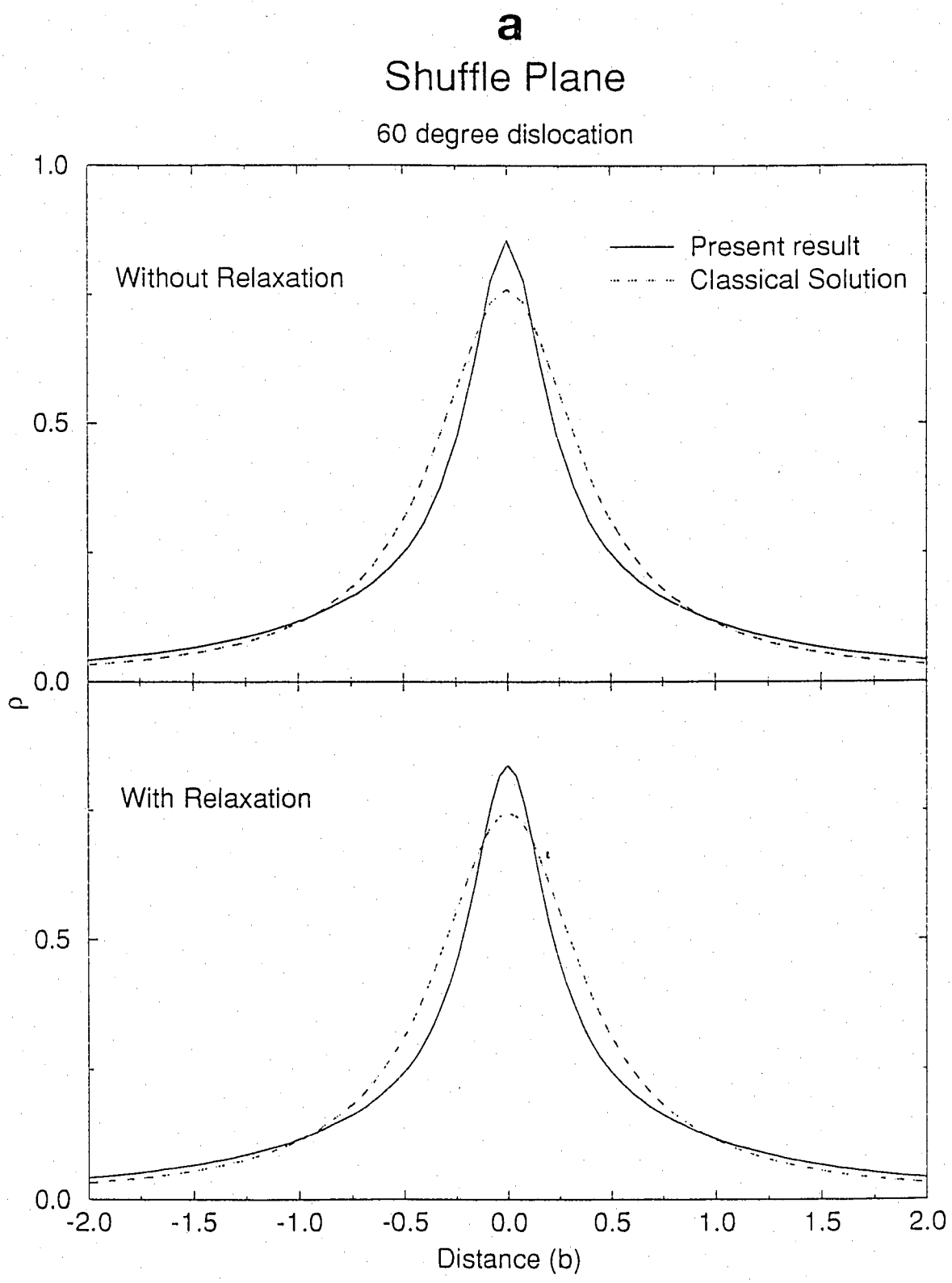
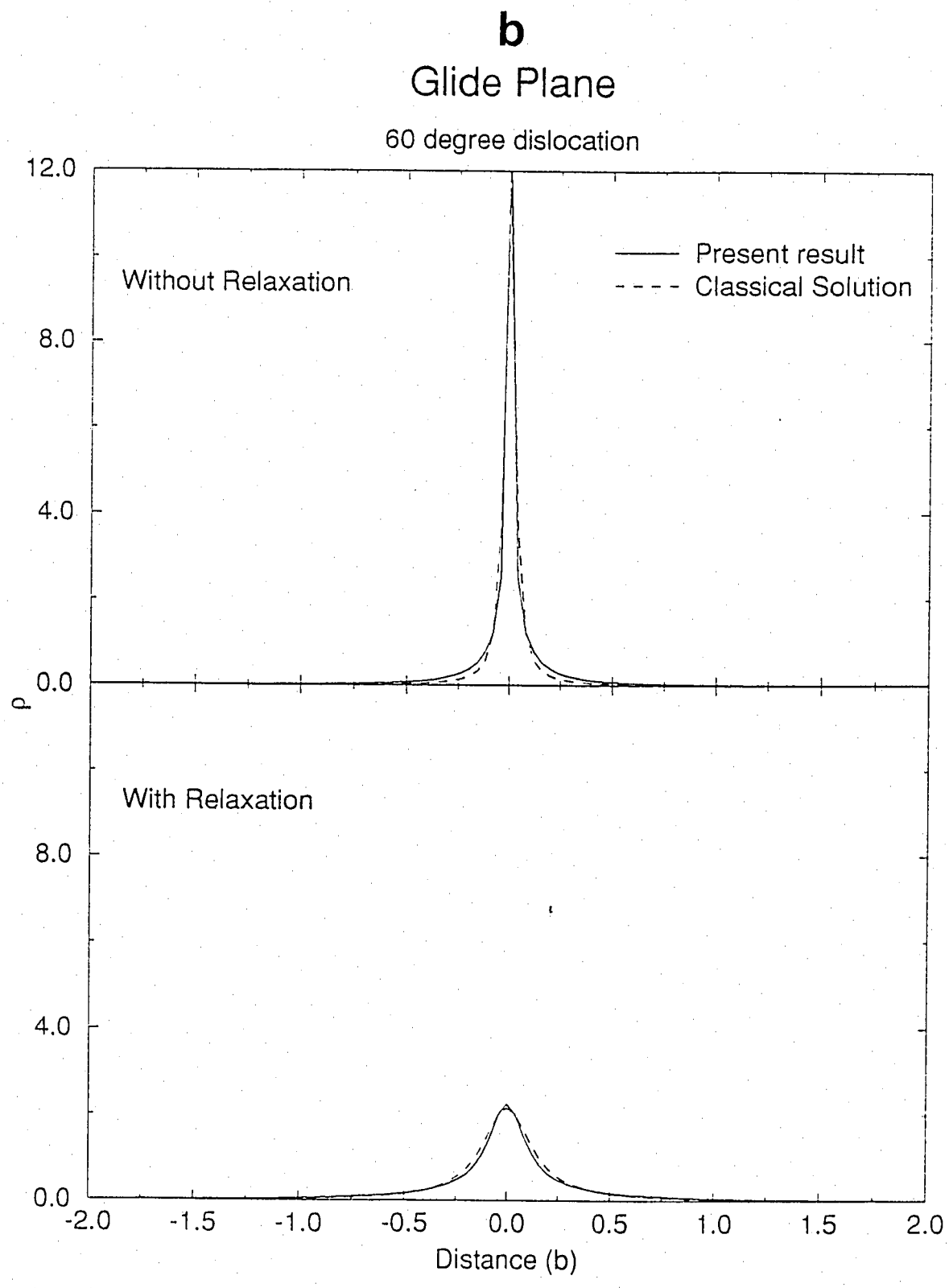
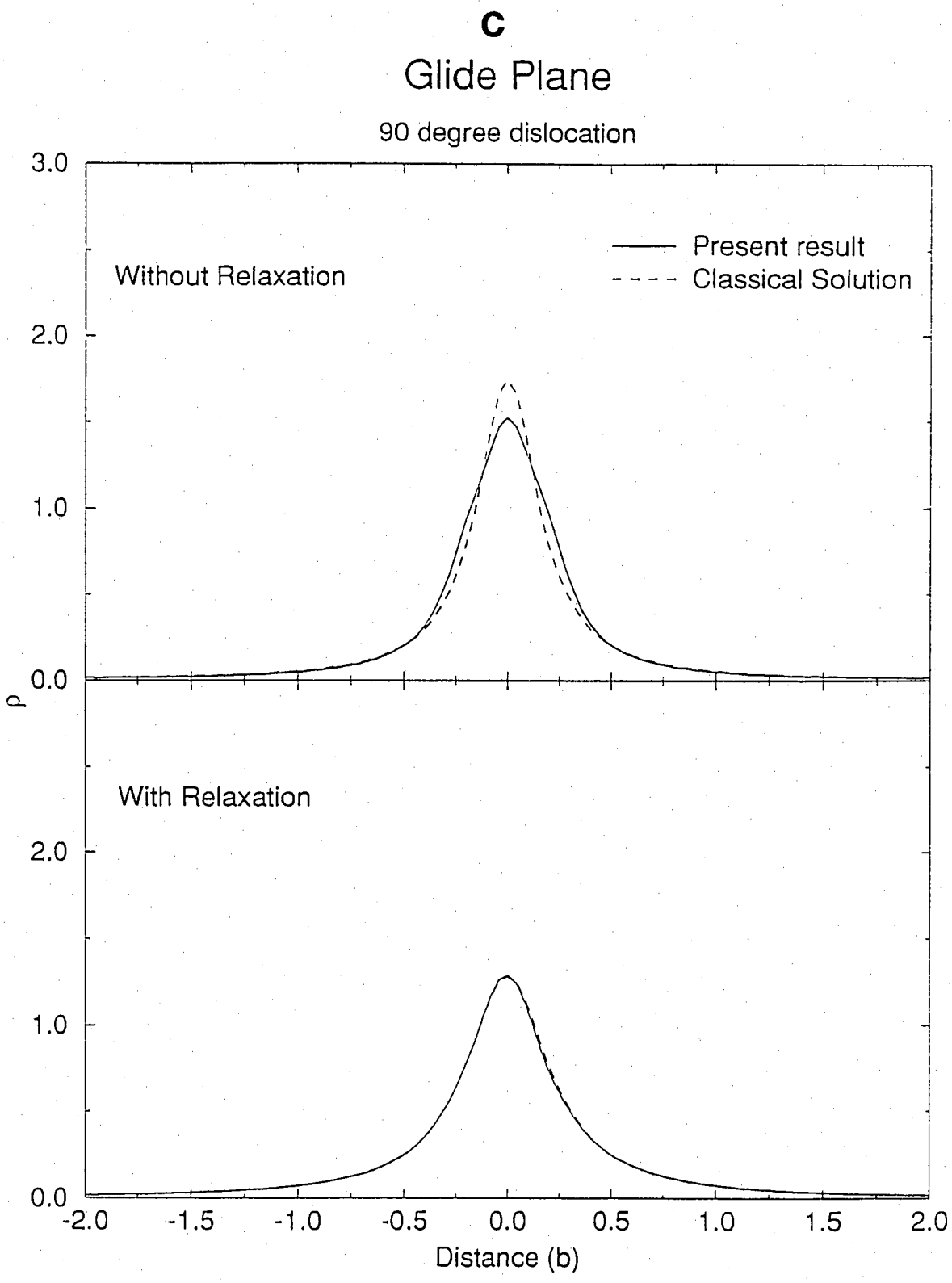
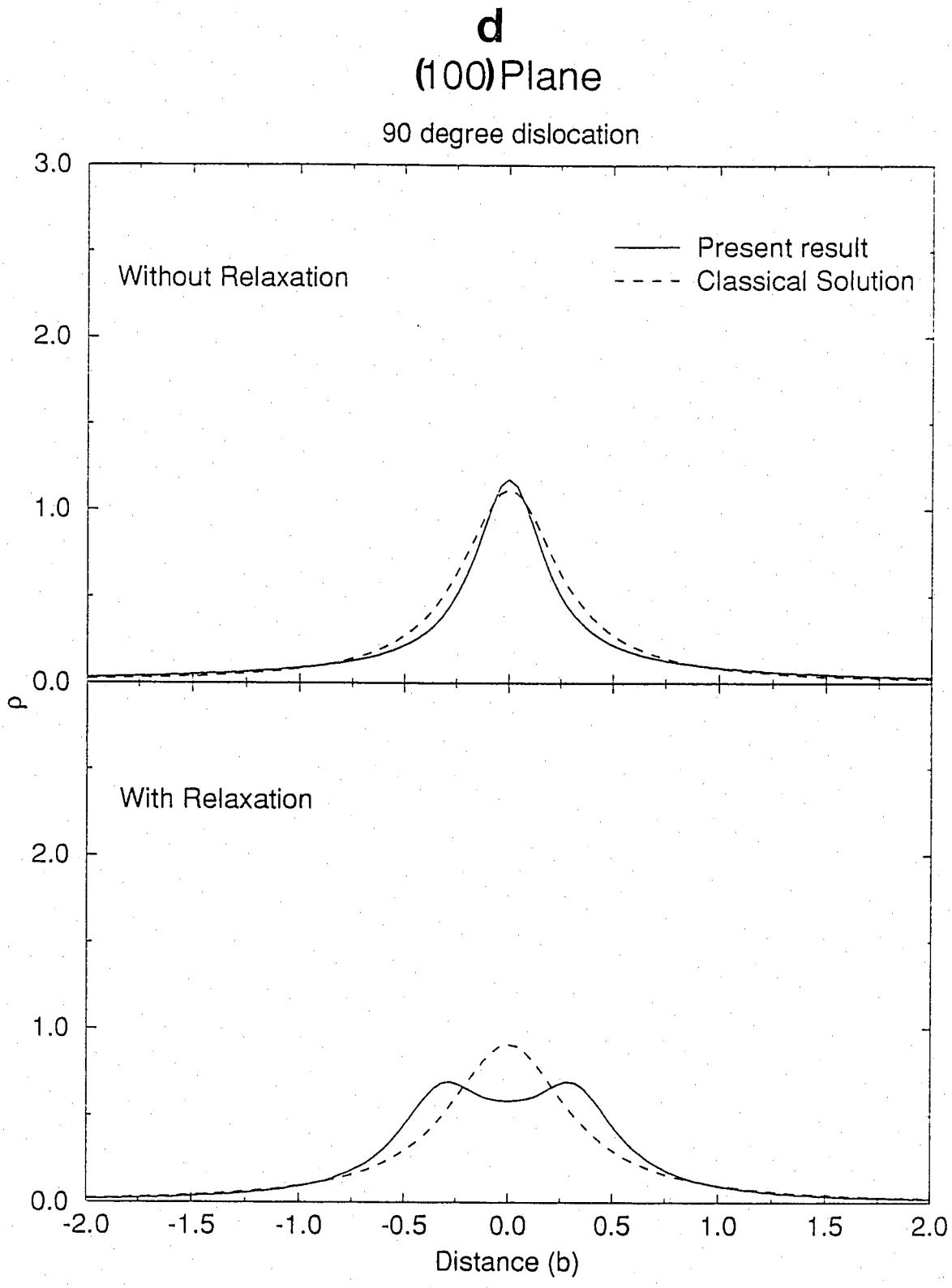

c
1.0
0.8
0.6
0.4
0.2
0.0
f (b)
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Distance (a)
(100)-90°
(100)-screw

Figure 20: The dislocation density obtained from our calculations for (a) the 60^0 dislocation on the shuffle set, (b) the 60^0 dislocation on the glide set, (c) the 90^0 partial dislocation on the glide set, and (d) the 90^0 dislocation on the (100) plane. Both the relaxed (lower panel) and the unrelaxed (upper panel) results are displayed. The classical solution is also shown (dashed curve) for comparison.


a
Shuffle Plane
60 degree dislocation
1.0
0.5
0.0
ρ
Without Relaxation
Present result
Classical Solution
With Relaxation
0.5
0.0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Distance (b)


b
Glide Plane
60 degree dislocation
12.0
8.0
4.0
0.0
Without Relaxation
Present result
Classical Solution
ρ
With Relaxation
8.0
4.0
0.0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Distance (b)


c
Glide Plane
90 degree dislocation
Present result
Classical Solution
Without Relaxation
With Relaxation
ρ
3.0
2.0
1.0
0.0
2.0
1.0
0.0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Distance (b)


d
(100) Plane
90 degree dislocation
Present result
Classical Solution
Without Relaxation
With Relaxation
ρ
3.0
2.0
1.0
0.0
2.0
1.0
0.0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
Distance (b)

on the glide set, and (d) the 90^0 dislocation on the (100) plane. We have displayed the results obtained from both the relaxed (lower panel) and the unrelaxed GSF energy surfaces (upper panel) to examine the effects due to the relaxation of the atomic coordinates. For comparison we have also displayed the results from the classical solution (dashed curves). The classical solution is obtained by assuming a sinusoidal form for the crystal restoring force with its maximum equal to the value obtained by first-principles calculations. For the case of the 60^0 dislocation on the shuffle set and the partial dislocation on the glide set, it is apparent that the use of the relaxed GSF energy surface does not cause significant change on the dislocation profile both qualitatively and quantitatively. This ia a consequence of the fact that the relaxation does not change the GSF energies along these directions in any significant way. For the 60^0 dislocation on the glide set, the use of the relaxed energy surface makes the dislocation profile much smoother and wider even though the general shape does not change very much. This is due to the fact that the relaxation reduces the maximum of the crystal restoring force but does not change its functional form significantly. For the case of the 90^0 dislocation on the (100) plane, the dislocation profile is changed even qualitatively with the use of the relaxed energy surface. This reflects the fact that the crystal restoring force for this plane is significantly different from the sinusoidal function as we mentioned at the beginning of this section. A brief summary of the obtained dislocation properties are given in Table 6.

To study the motion of a dislocation, the useful quantity to consider is the

Table 6: The quantities related to the properties of dislocations for shuffle ang glide sets on (111) plane, as well as the (100) plane of silicon, obtained form our calculations. The meanings of these quantities are explained in details in the text.

dislocation	Shuffle		Glide		Glide partial		(100)	
	60^0	Screw	60^0	Screw	30^0	90^0	90^0	Screw
$K(10^{11}\frac{dyne}{cm^2})$	8.02	6.37	8.02	6.37	6.92	8.58	9.04	6.375
$\tau_{max}(10^{11}\frac{dyne}{cm^2})$	1.49	1.49	4.29	4.29	2.78	2.78	2.06	2.06
ζ (Å)	1.08	0.92	0.46	0.37	0.77	0.92	2.15	1.54
b (Å)	3.84	3.84	3.84	3.84	2.22	2.22	3.84	3.84

energy barrier associated with this process. Within the Peierls-Nabarro model, this energy barrier is called the Peierls energy and is defined to be the amplitude of the variation of the misfit energy on the glide plane as one changes the position of the dislocation. With the obtained dislocation profiles, we can calculate the misfit energy across the glide plane as a function of the position of the dislocation line u following the definition given in Ref. [19] and Ref. [21]. To be consistent with the Peierls-Nabarro model where the discreteness of the crystal is taken into account at the glide plane, the misfit energy is defined to be the sum of the misfit energies between pairs of atom rows, which is the GSF energy at the local disregistry. The misfit energy $W(u)$ is given by the following formula,

$$W(u) = \sum_{m=-\infty}^{+\infty} \gamma(f(ma' - u))a'. \tag{31}$$

a' is the distance between adjacent atomic planes in the direction perpendicular to the dislocation line. One can immediately notice that this formula meets the two known basic requirements. First, it has the correct period a' of the crystal,

$$W(u + a') = W(u) \tag{32}$$

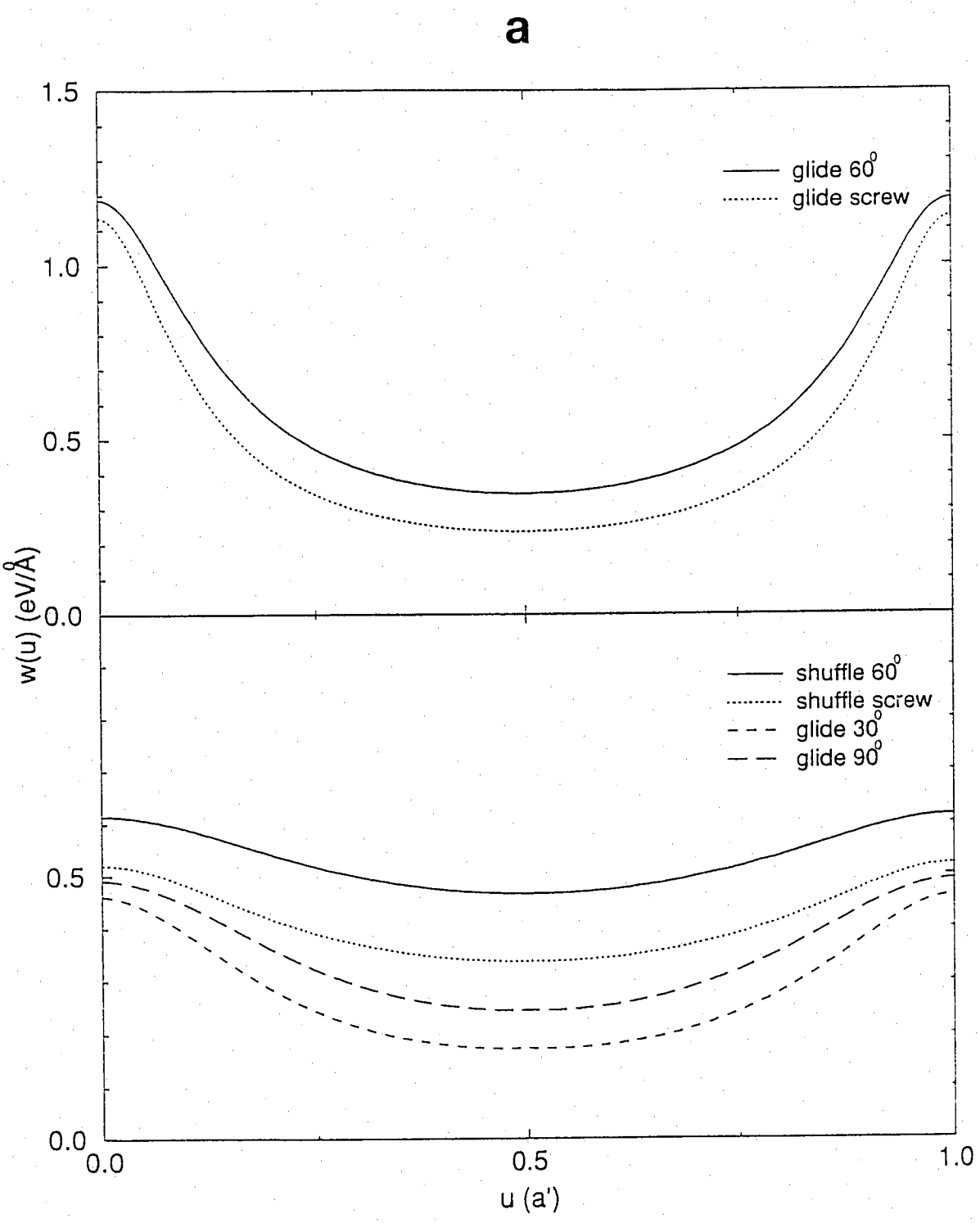
Second, at the limit where one has a very narrow dislocation it reproduces the correct maximum $\gamma_{us}a'$. The other quantity which might be of interest is $W(a'/2)$ where the minimum of the misfit energy function occurs. Since this quantity measures the nonelastic part of the energy of the dislocation, it should provide a reasonable estimate of the core energy. The stress associated with the energy

function $W(u)$ can then be defined as:

$$\sigma(u) = \frac{1}{a'}\frac{dW(u')}{du'}|_u \tag{33}$$

The maximum of this stress function is the Peierls stress which is the stress one has to apply in order to move the dislocation. The values of the energy function $W(u)$ and stress function $\sigma(u)$ for the (111) plane are shown in Figure 21 (a) and (b). It is obvious that the dislocation with Burgers vector along the (110) direction on the glide set has larger Peierls energy and Peierls stress. This is expected since we know from our GSF energy surface calculations that the energy barrier along that direction is higher than those along the other two vectors. Figure 22 is the energy and stress function for the case of (100) plane. The different behavior between the 90^0 and the screw dislocations on the (100) plane is primarily due to the difference of the constant K. The Peierls energy, stress, and core energy of various dislocations are summarized in Table 7. For the (111) plane, it appears that the dislocations belonging to the shuffle set are the easiest to move under zero temperature and zero pressure conditions. However, as we discussed earlier, the effects due to temperature and pressure may change the picture. Finally, we would like to mention that there is experimental evidence that on the glide set the 90^0 partial is more mobile than the 30^0 partial [22, 23, 24, 25], which is consistent with the results of our calculations.

Figure 21: (a) The energy function $W(u)$ with respect to the displacement vector u for the dislocations belonging to (111) plane. (b) The corresponding stress $\sigma(u)$ as a function of u.


a
1.5
1.0
0.5
0.0
w(u) (eV/Å)
glide 60°
glide screw
shuffle 60°
shuffle screw
glide 30°
glide 90°
0.5
0.0
0.0
0.5
1.0
u (a')

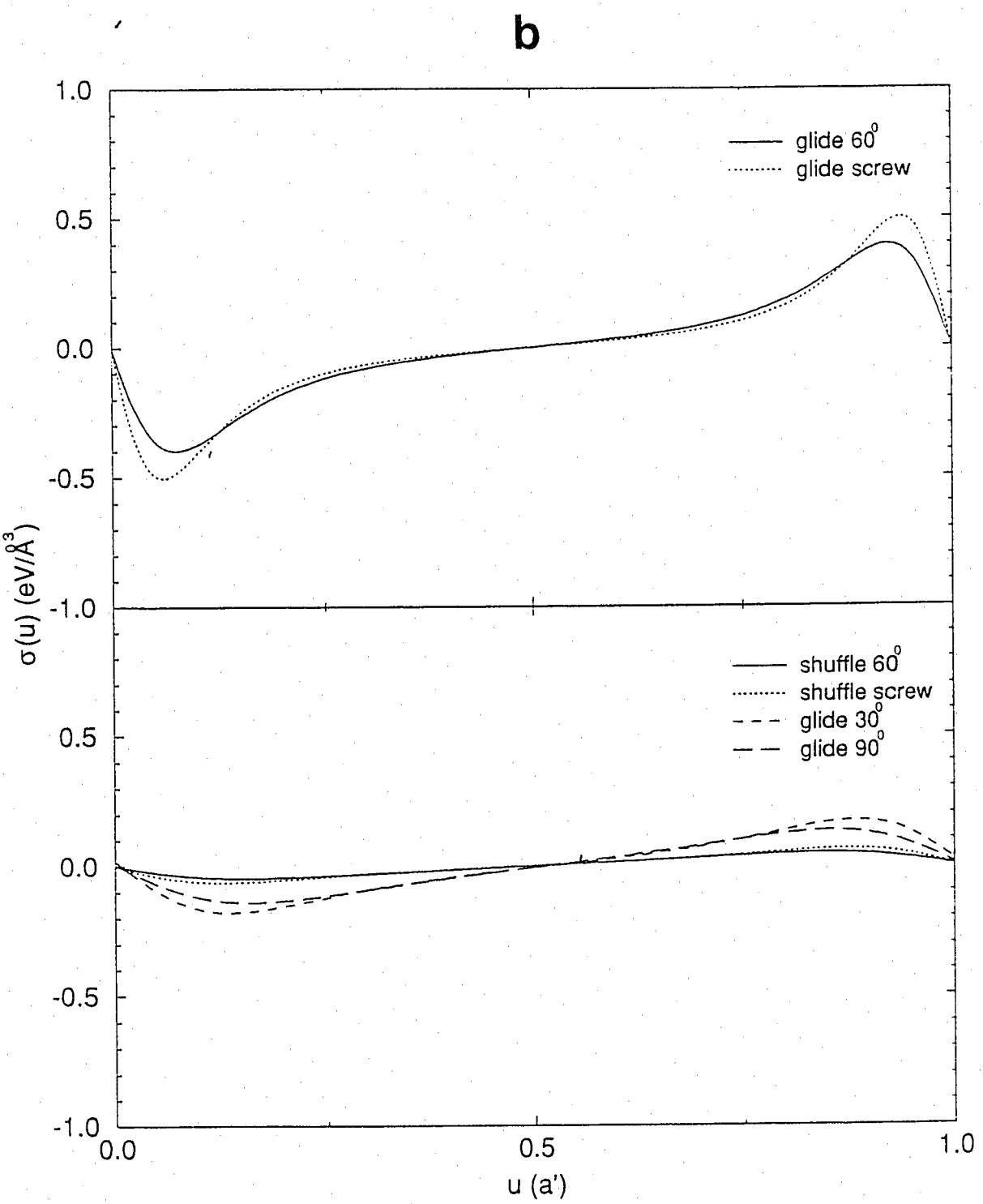

b
1.0
0.5
0.0
-0.5
-1.0
σ(u) (eV/Å^3)
glide 60^0
glide screw
shuffle 60^0
shuffle screw
glide 30^0
glide 90^0
0.0
0.5
1.0
u (a')

Figure 22: The energy function $W(u)$ (upper part) and the corresponding stress $\sigma(u)$ (lower part) for the dislocations belonging to (100) plane displayed with respect to the displacement vector u.

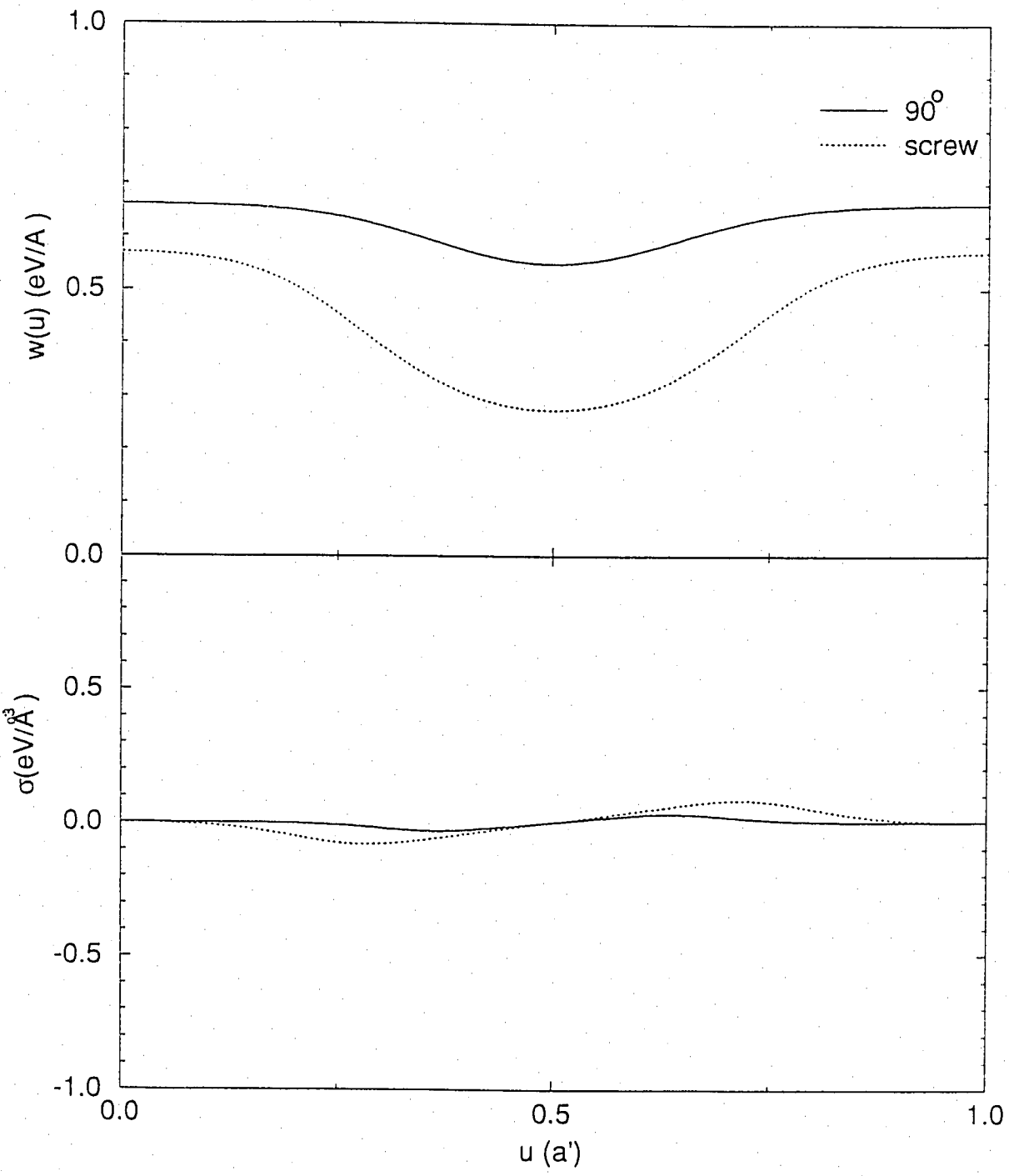

(100) Plane
1.0
0.5
0.0
w(u) (eV/A)
90°
screw
0.5
0.0
-0.5
-1.0
σ(eV/Å³)
0.0
0.5
1.0
u (a')

Table 7: The Peierls energy, Peierls stress, and the estimated core energy obtained from Peierls-Nabarro model for various dislocations on the (111) and (100) plane of silicon.

dislocation	Shuffle		Glide		Glide partial		(100)	
	60^0	Screw	60^0	Screw	30^0	90^0	90^0	Screw
$W_p(\frac{eV}{\AA})$	0.149	0.183	0.842	0.898	0.287	0.246	0.112	0.296
$\sigma_p(\frac{eV}{\AA^3})$	0.046	0.062	0.399	0.504	0.176	0.139	0.032	0.081
W(a'/2)	0.467	0.338	0.345	0.238	0.174	0.246	0.549	0.274

4.7 Summary

In summary, we have performed first-principles calculations to obtain the GSF energy surfaces for both the shuffle and glide sets of the (111) plane, as well as the (100) plane of silicon. We showed that for the glide set and the (100) plane, the effects on the GSF energy surfaces due to the relaxation of the atomic coordinates are significant. The unstable stacking fault energies γ_{us} for these planes are obtained from our calculations. By combining these values with Rice's criterion for dislocation nucleation, the shuffle set appears to be the most favored one for dislocation emission under zero temperature and zero pressure conditions. The effects due to entropy are taken into account with the use of Vineyard's transition state theory. To compare the two different sets (shuffle and glide) of slipping on the (111) plane in a more complete thermodynamical picture, we wrote down the free energy associated with each set in terms of the two thermodynamical parameters (pressure and temperature). By comparing this free energy, we showed that either set (shuffle or glide) can dominate under different thermodynamical conditions. This might shed some light on the long-standing glide-shuffle controversy [10].

The GSF energy surfaces were then combined with the Peierls-Nabarro model to investigate the dislocation properties of silicon. The crystal restoring forces are obtained directly from our first-principles GSF energy surfaces. We demonstrated the importance of using the relaxed energy surface for the calculation of these properties. The differences between dislocation profiles obtained with relaxed

GSF energy surfaces and unrelaxed GSF energy surface are significant in certain cases. The Peierls energy and Peierls stress have also been calculated within the framework of the Peierls-Nabarro model.

We would like to conclude this chapter with a discussion on the approximations involved in the Peierls-Nabarro model: (1) The dislocation is assumed to be planar within this model. We expect that the errors introduced with this approximation for a stiff material like silicon should be minimal. (2) The response from the crystal is treated within elastic theory, which may not apply for a very narrow core situation, where non-elastic effects are expected to be important. (3) The Peierls stress concept assumes that the only relevant quantity in determining the stress is the local disregistry, which is true only when the dislocation density is very smooth. Therefore, we would like to caution the reader that all the numbers should not be taken literally. Rather, it is the qualitative features of the results which should help us understand the physical mechanisms determining the dislocation properties.

References

[1] A.A. Griffth, Phil. Trans. R. Soc. **A 184**, 181 (1920)

[2] J.R. Rice, J. Mech. Phys. Solids, **40**, 239 (1992).

[3] G.E. Beltz and J. R. Rice, in *Modeling the Deformation of Crystalline Solids: Physical Theory, Applications, and Experimental Comparisons*, edited by T.C. Lowe, A.D. Rollett, P.S. Follansbee and G.S. Dehn, (TMS Minerals, Metals and Materials Society, Warrendale, 1991); G.E. Beltz and J.R. Rice, Acta Metall. **40** S321 (1992).

[4] J.R. Rice and G.E. Beltz, J. Mech. Phys. Solids, **42**, 333 (1994).

[5] J.R. Rice, G.E. Beltz, and Y. Sun, in *Topics in Fracture and Fatigue*, edited by A.S. Argon, (Springer, Berlin, 1992).

[6] G.E. Beltz, Ph.D Thesis, Harvard University, Cambridge, MA.

[7] Y. Sun, G.E. Beltz, and J.R. Rice, Mater. Sci. Engng A (in press).

[8] G.H. Vineyard, J. Phys. Chem. Solids **3**, 121 (1957).

[9] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. Sham, Phys. Rev. **140**, A1133 (1965).

[10] F. Louchet and J. Thibault-Desseaux, Rev. Phys. Appl. **22**, 207 (1987); A brief description of this glide-shuffle controversy can also be found in J.P.

Hirth and J. Lothe, Theory of Dislocations, 2nd ed., p. 382 (Wiley, New York, 1982)

[11] We use the expressions for the exchange-correlation functionals proposed by J. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1984).

[12] G. Bachelet, H. Greenside, G. Barraff, and M. Schlüter, Phys. Rev. **24**, 4745 (1981).

[13] L. Kleinman and D.M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[14] H.J. Monkhorst and J.D. Pack, Phys. Rev. B. **13**, 5188 (1976).

[15] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[16] E. Kaxiras and M.S. Duesbery, Phys. Rev. Lett. **70**, 3752 (1993).

[17] K.C. Pandey and E. Kaxiras, Phys. Rev. Lett. **66**, 915 (1991).

[18] F. Kroupa and Lejček, Czech. J. Phys. Rev. B. **22**, 813 (1972).

[19] B. Jòos, Q. Ren, and M.S. Duesbery, Phys. Rev. B **50**, 5890 (1994).

[20] J.P. Hirth and J. Lothe, Theory of Dislocations, 2nd ed. (Wiley, New York, 1982)

[21] V. Vitek and M. Yamaguchi, J. Phys. F. **3**, 537 (1973); M. Yamaguchi and V. Vitek, J. Phys. F. **5**, 11 (1975).

[22] K. Wessel and H. Alexander, Philos. Mag. **35**, 1523 (1977).

[23] H. Alexander, H. Gottschalk, and C. Kisielowski-Kemmerich, in *Dislocations in Solids*, edited by H. Suzuki, T.Minomoya, K. Sumino, and S. Takeuchi (University of Tokyo Press, Tokyo, 1985)

[24] P. Grosbras, J. L. Demenet, H. Garem, and J.C. Desoyer, Phys. Status Solidi A **84**, 481 (1984).

[25] J. L. Demenet, P. Grosbras, H. Garem, and J.C. Desoyer, Philos. Mag. A **59**, 501 (1989).

5 Ledge Energy for Dislocation Emission in Silicon

5.1 Introduction

As was discussed in the previous chapter, Rice [1] has recently proposed a new scheme to analyze the criterion for dislocation emission from the crack tip based on the Peierls stress concept. By assuming a periodical relation between the shear traction τ and the slip displacement δ along a crystal plane emanating from the crack tip, the unstable stacking fault energy γ_{us} (the energy barrier encountered when one half of the crystal is sliding with respect to the other half on a slip plane) has been identified to be the solid state parameter controlling the critical loading condition for the dislocation nucleation from the crack tip. However, the effects due to the surface which is created during this dislocation emission process are neglected in this approach, by assuming that the created surface is traction free. There have been efforts to try to extend the above model by including these surface effects [2, 3]. The results derived are not completely satisfactory due to the fact that the analysis is based on phenomenological atomistic calculations which do not necessarily represent real physical systems. In this chapter, we will show that, using silicon as an example, an accurate description of the surface energy associated with the newly created ledge during dislocation emission can be obtained through first-principles quantum mechanical calculations. The ledge

energy can then be incorporated into Rice's theory by assuming an evanescent force law for the surface effects. The critical loading for the dislocation emission is determined within continuum mechanics.

5.2 Methods

In order to calculate the energy associated with surface creation during dislocation emission, it is necessary to perform the calculation for a finite size block shearing process, as shown schematically in Figure 23. The meanings of the symbols used in the figure are as follows: I indicates the interfacial area corresponding to the generalized stacking fault energy, which has been determined by performing calculations for the infinite size block shearing, as was discussed in Chapter 4; S represents the surface created through the shearing process; and C indicates the coupling between the corner and newly created surface. It is the existence of this coupling term C which causes the energy associated with surface creation during dislocation emission to be different from the free surface energy. In order to perform first-principles calculations, a supercell approximation to this geometry has to be adopted. For the present calculation for silicon, we have used a supercell composed of 108 atoms. The geometry of the cell is displayed in Figure 24. a_1, a_2, and a_3 are the periodical lattice vectors for the supercell. The advantages of this geometry are obvious: First, by simply changing the x-component of the vector a_1, one cell is sheared with respect to the other and a surface is created; Secondly, all the surfaces which are exposed through this process are along the (111) crystallo-

Figure 23: A schematic drawing for the configuration needed to model finite size block shearing. The symbol S, I, and C represent the surface area, interfacial area and coupling term respectively. (See text for details).

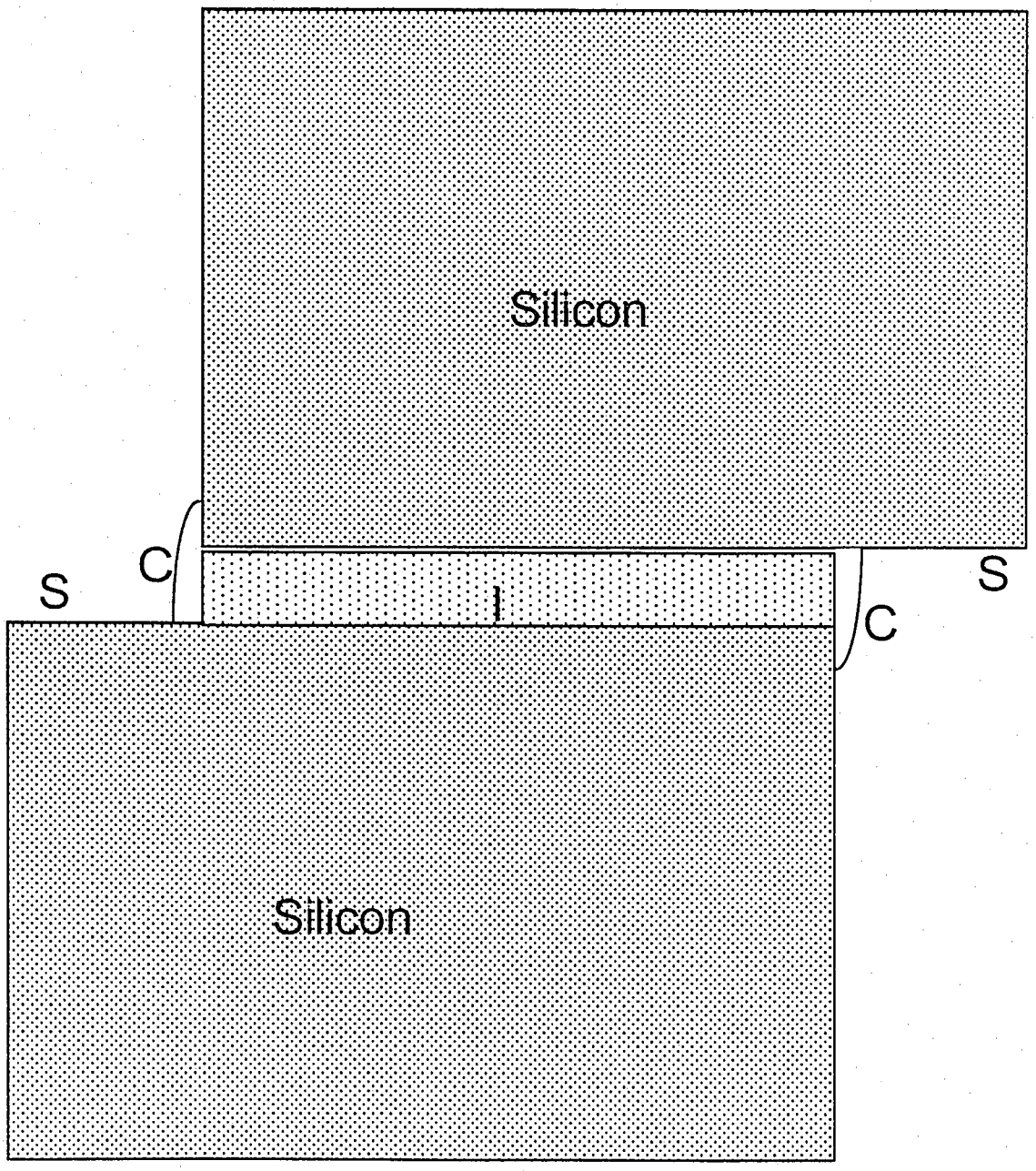
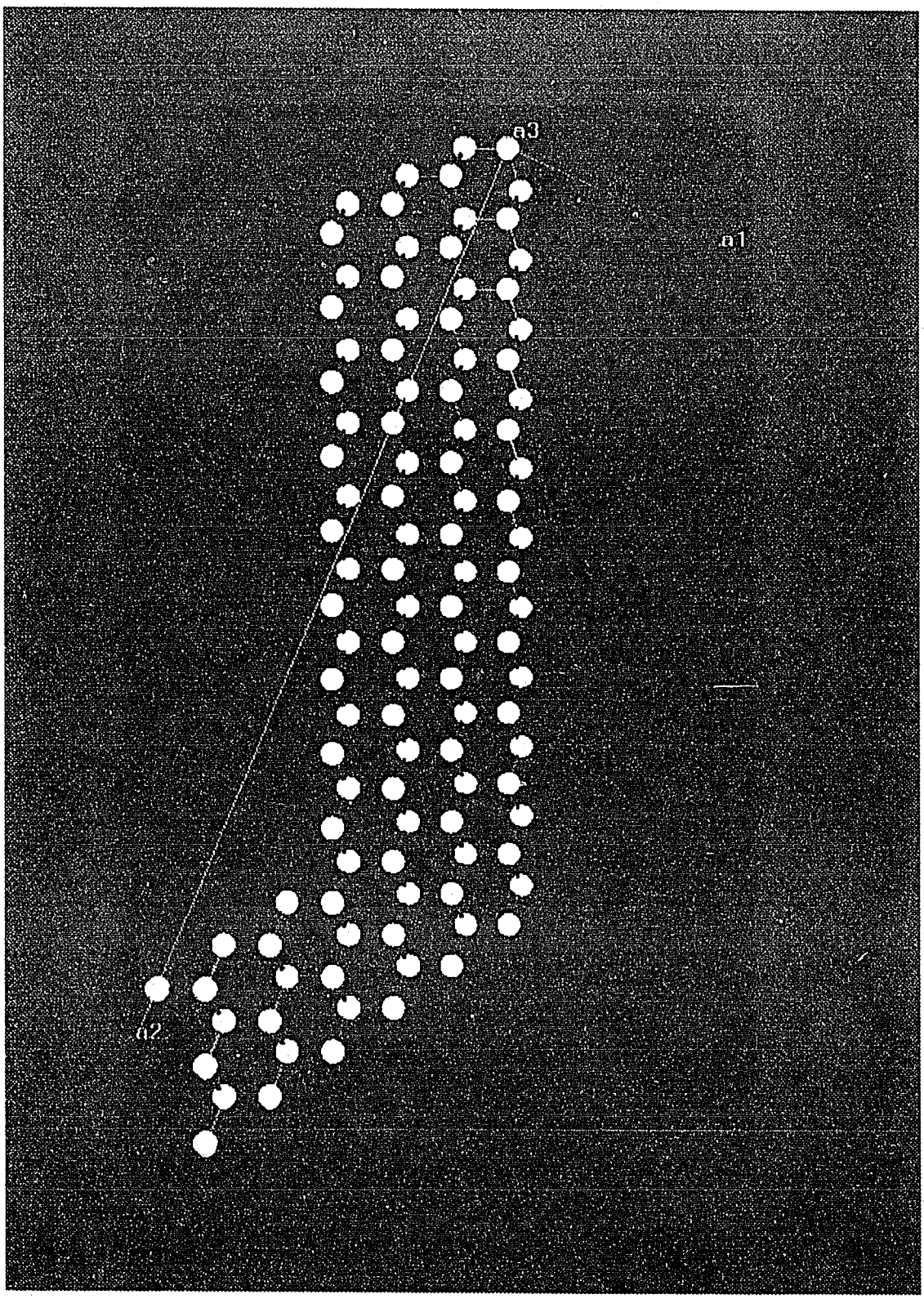

Silicon
S
C
I
C
S
Silicon

Figure 24: The geometry of the silicon supercell used in the present calculations. a_1, a_2 and a_3 are the three primitive lattice vectors of this unit cell. The third vector is in a direction perpendicular to the plane of the figure.


a3
a1
a2

graphic direction of the diamond lattice, which are the natural cleavage planes for this crystal system.

In terms of computational details, we have adopted the local density approximation to the density functional theory [4, 5], which is a first-principles quantum-mechanical approach proven to provide reliable energetic information for both bulk and surface properties of silicon. To obtain the total energy of the system, the iterative scheme due to Car and Parrinello [6] has been employed to solve the Kohn-Sham equations for the electronic wave functions. To make the computation efficient in terms of the plane wave basis, the ionic potential including the screening effects due to the core electrons are approximated by using the norm-conserving pseudopotentials in the Kleinman-Bylander form [7, 8], where d component is treated as the local part with the s and p components contain the non-local contributions. The cutoff of the plane wave basis for the expansion of the valence electron wave functions is 8 Ry. The Γ point has been used for the sampling of the Brillouin zone.

5.3 Results and Discussion

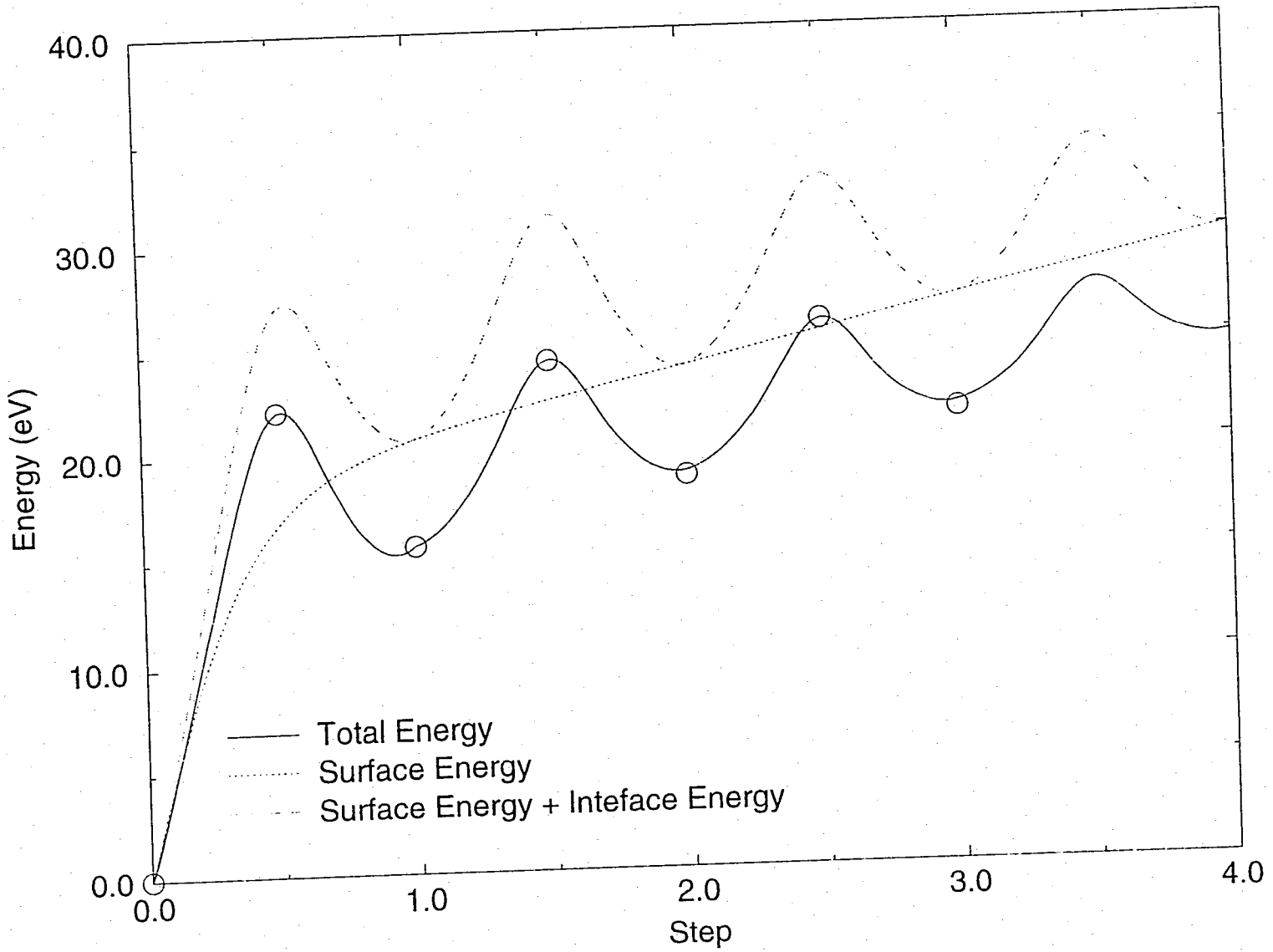
We have calculated the total energy of the system as a function the x-component (the $[1\bar{1}0]$ crystallographic direction) of a_1 from 0 a_0 to 3 a_0, where a_0 is the primitive lattice vector of silicon along that direction. At the largest displacement, the unit cell corresponds to 144 time the atomic volume of bulk Si (the equivalent of 144 Si atoms). In order to simplify the analysis, we separate the change in total

energy into the following three terms: the interfacial energy U_I, the surface energy U_S and the coupling energy U_C. The meanings of the symbols S, I, and C are discussed in the previous paragraph. We now discuss these three energy terms in detail. U_I is the energy associated with the shearing of two solid blocks which can be calculated by multipling the generalized unstable stacking fault energy $\gamma_{GSF}(f)$ with the interfacial area A_I. Because of the way we construct the supercell, the surface energy U_S is composed of two different terms: U_{S_I} and $U_{S_{II}}$. U_{S_I} is the energy associated with ledge surface, while $U_{S_{II}}$ is the energy due to the creation of the side surface. We have performed independent calculations for the block separation process to determine the value of the term $U_{S_{II}}$. The value of the surface term U_{S_I} is determined by taking the simplest linear approximation

$$U_{S_I} = S_I \times \gamma_S \tag{34}$$

where S_I is the exposed surface area and γ_S is the free surface energy. U_C is the difference between the total energy and the sum of the above two terms. The results from our calculations are displayed in Figure 25. The dotted curve is the surface energy U_S. The dashed curve is the sum of the surface energy and the interface energy, $U_S + U_I$. The solid curve is the results we obtained for the finite size block shearing process, where the calculations have been performed at the integral and half-integral steps of unit a_0. The curve is obtained by fitting the first-principles results to an analytic functional form. The coupling energy is simply the difference between the solid curve and the dashed curve. We immediately observe that the

Figure 25: The energy curves obtained from present calculations. Total energy, surface energy and the sum of surface energy and interfacial energy are denoted by the solid line, the dotted line and the dashed line respectively. The physical meaning of these terms are described in the text. The circles are the results from our first-principles calculations.


40.0
30.0
20.0
10.0
0.0
Energy (eV)
0.0
1.0
2.0
3.0
4.0
Step
Total Energy
Surface Energy
Surface Energy + Inteface Energy

the amplitude of the variation of U_C becomes very small after the relative displacement is greater than a_0. This means that the corner-corner interactions are relative small for the present case and can be neglected at this level of approximation. Therefore, as a first order approximation we only need to consider the value of this coupling energy at the integral steps. This number is essentially the difference between the energy for the creation of one area of free surface and one area of surface associated with the dislocation emission from the crack tip. In other words, the effective surface energy $\tilde{\gamma}_s$ for creation of one area of ledge surface is equal to $\gamma_s + U_C$. The number we have obtained for $\tilde{\gamma}_s$ is approximately $0.58\gamma_s$ The effects due to the creation of ledge surface during dislocation emission can then be taken into account within continuum mechanics by assuming a evanescent force law [9, 10]:

$$\tau_s(r) = (\tilde{\gamma}_s/\Lambda)e^{-r/\Lambda} \tag{35}$$

where τ is the stress associated with surface creation, λ is a phenomenological decay length scale and r is the distance between the crack tip and the point where the stress is evaluated. This expression comes from the fact that by doing the integration with respect to the variable r, it will reproduce the corresponding ledge surface energy $\tilde{\gamma}_s$. This should be a reasonable assumption for the case of silicon since surface effects are expected to decay very fast into the bulk. By treating the incipient dislocation as a continuous distribution of infinitesimal dislocations, one can then write down the equation from continuum mechanics based on the force balancing concept. Taking the case where the crack and slip occur on the same

plane with purely shear loading, one can get

$$\tau_I(r) + \tau_s(r) = \sigma(r) - \int_0^\infty g(r,s)\frac{d\delta_s}{ds}ds \tag{36}$$

where $\tau_s(r)$ is the surface stress we defined above, r is the distance from the crack tip, and $\tau_I(r)$ is the stress from the misfit across the slip plane and is defined as

$$\tau_I(r) = \frac{d\gamma_{GSF}(\delta(r))}{d\delta(r)}. \tag{37}$$

$g(r,s)$ describes the stress exerted at one point r on the slip plane by the infinitesimal dislocation at the point s, and has been obtained from the calculation for a line dislocation and a crack tip in the elastic medium [11, 12]:

$$g(r,s) = [\frac{\mu}{2\pi(1-\nu)}]\sqrt{\frac{s}{r}}\frac{1}{r-s} \tag{38}$$

for our present case. $\sigma(r)$ is the stress due to the external loading and is equal to $\frac{K_{II}}{2\pi r}$ for mode II loading. This equation was solved numerically [9]. By increasing the loading continuously, the critical loading condition is determined, which corresponds to the point where a stable solution no longer exists. The advantage of this approach is that it treats the effects from surface and interface on equal footing based on a simple extension of the original Peierls stress concept. The critical loading conditions have been determined for a wide range of $\tilde{\gamma}_s$ and λ. We would like to emphasize two general trends which we have observed: First, the critical loading is almost independent of the decay length scale λ. This can be understood by observing that there are two places where λ enters the expression and it basically only changes the distribution of the surface stress. Furthermore,

the effects due to these two factors appear to cancel each other. Accordingly, the effect due to surface creation should be similar for different values of λ under this approximation. Second, the loading seems to increase linearly with the surface energy. For the case of silicon where $\tilde{\gamma}_s = 0.58\gamma_s$, we found that the energy release rate G corresponding to the critical loading condition is increased by a factor of 1.45 compared to the value obtained without taking into account surface effects.

5.4 Summary

In conclusion, we have used silicon as an example to show that surface effects due to the ledge when a dislocation is emitted from a crack tip can be obtained from first-principles calculations. For the case of silicon, the energy associated with ledge surface creation is approximately one half of the free surface energy. By assuming an evanescent force law for the surface effects, the ledge effects are incorporated into Rice's original theory within continuum mechanics. The inclusion of surface effects increases the necessary loading for the dislocation emission. This can be expected from a simple energetic point of view since surface creation will increase the energy consumption in the process.

References

[1] J.R. Rice, J. Mech. Phys. Solids, **40**, 239 (1992).

[2] S.J. Zhou, A.E. Carlsson and Robb Thomson, Phys. Rev. Lett. **72**, 852 (1994).

[3] Robb Thomson and A.E. Carlsson, Phil. Mag. A **70**, 893 (1994).

[4] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964); W. Kohn and L. Sham, Phys. Rev. bf 140 A1133 (1965).

[5] The expression for local density approximation used in the present calculation is the one proposed by J. Perdew and A. Zunger, Phys. Rev. B bf 23, 5048 (1984).

[6] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[7] G. Bachelet, H. Greenside, G. Barraff, and M. Schlüter, Phys. Rev. **24**, 4745 (1981).

[8] L. Kleinman and D.M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[9] Y. Sun (private communication).

[10] G. Xu, A.S. Argon, and M. Ortiz (preprint).

[11] I.-H. Lin and R. Thomson, Acta Metall. **35**, 185 (1987).

[12] G.E. Beltz, PhD Thesis, Division of Applied Sciences, Harvard University, Cambridge, MA, 1992.

6 Conclusions

We have presented a detailed analysis of the application of recently proposed GGA functionals to density functional calculations. We found that these functionals provide substantial improvements for finite systems like atoms. However, for the properties of solids, there is no conclusive evidence that the use of gradient corrections gives more accurate results. In fact, the bulk moduli of semiconductor systems are found to be worse than the results from LDA. We also demonstrated that it is essential to take into account the nonlinearity of the exchange-correlation functional when GGA functional is used. We concluded that LDA remains a reliable choice for energetic comparisons in covalently bonded systems.

We then investigated problems related to microstructural deformations in silicon, using LDA. First, we studied the energetics associated with the high-pressure plastic flow in silicon within the model of a homogeneous structural transformation. The magic strain concept was applied to search for possible low energy paths in the structural phase space. Entropy effects were taken into account with the use of Vineyard's transition state theory. An upper bound for the energy barrier of this process was determined.

As a second application, we studied dislocation properties of silicon, by performing detailed calculations to obtain the generalized stacking fault energy surfaces in the supercell approximation. Both the shuffle set and the glide set of the (111) plane, as well as the (100) plane of silicon have been examined. Rice's criterion for

dislocation nucleation from a crack tip was applied to compare the energy barriers for dislocation emission on different gliding planes of this system. By combining the Peierls-Nabarro model with the calculated energy surfaces, we obtained the profiles for specific dislocations. We also calculated the Peierls energy and Peierls stress within the Peierls-Nabarro model from the first-principles results.

Finally, we attempted to estimate the effects on the criterion for dislocation nucleation due to surface creation. We performed first-principles calculations to obtain the energy associated with the creation of ledge surface for the case of silicon. Based on this energy factor, an evanescent decaying law was assumed for the surface stress, as suggested by Y. Sun. With this formulation, we were able to incorporate the surface effects into Rice's original theory within continuum mechanics.