

## Linear scaling nonorthogonal tight-binding molecular dynamics for nonperiodic systems

N. BERNSTEIN

*Center for Computational Materials Science, Naval Research Lab  
Washington, DC 20375, USA*

(received 14 February 2001; accepted in final form 26 April 2001)

PACS. 71.15.Dx – Computational methodology (Brillouin zone sampling, iterative diagonalization, pseudopotential construction).

PACS. 61.72.Lk – Linear defects: dislocations, disclinations.

PACS. 71.15.Pd – Molecular dynamics calculations (Car-Parrinello) and other numerical simulations.

**Abstract.** – I present a method for efficiently calculating atomic forces from nonorthogonal tight-binding models. This method is applicable to systems that cannot be described by a periodic supercell. I use the method to determine the temperature-dependent structure of a silicon screw dislocation core. The core reconstruction changes with temperature; a broad, disordered core with many coordination defects appears at high temperatures.

*Introduction.* – Simulations using minimal basis tight-binding (TB) models for semiconductors are being applied to a wide range of materials systems, from bulk lattices and point defects to more complex structures such as dislocations, grain boundaries, and amorphous systems [1]. Dynamics and finite-temperature properties of these structures are commonly studied using TB molecular dynamics (MD) simulations of a representative system with periodic boundary conditions, chosen to be far from the region of interest. Often there is no particular symmetry relation between different boundary atoms in the physical system, even though the *local* environment of each boundary atom resembles an ideal lattice. Dislocations, which play a vital role in the mechanical response of crystals, are an important example. The atomic structure of the dislocation core and the energy needed to move it control the response of the dislocation to applied strain, and therefore the ability of the material to deform plastically. Periodic supercell calculations accommodate the topological changes of the crystal lattice caused by dislocations by including multiple dislocations with Burgers vectors that sum to zero [2]. This increases the number of atoms in the simulation and leads to elastic and electronic interactions with periodic images that could change the dislocation structure [3]. Simulations of dislocations using clusters are unaffected by the topological changes, and continuum elasticity can be used to fix cluster boundary atoms at positions that are compatible with the longest range contribution to the dislocation strain field minimizing image forces. However, cluster simulations must use hydrogen passivated surfaces to reduce the interactions between surface states and defect states at the dislocation core [4]. To eliminate the effects of the periodic images and the need for hydrogen passivation, a method is required that provides a quantum-mechanical description of the interatomic forces in a finite region where each boundary atom has a local environment resembling an ideal-lattice geometry. Such a method would also enable atomic resolution in small regions of multiscale simulations [5, 6].

I present a new method for computing the forces from a nonorthogonal basis TB model that can be applied to a subsection of a large atomic configuration with a stationary boundary. Since the rest of the system acts as a reservoir of electrons, the forces are calculated at fixed electronic chemical potential. The computational effort for the calculation scales linearly with the number of atoms. I demonstrate the accuracy of the method by a direct comparison with an exact diagonalization supercell calculation. I use this method to simulate a single shuffle-set screw dislocation in silicon. This configuration was also studied by Arias and Joannopoulos who noted that while screw dislocations are observed only as dissociated partials on the glide set, shuffle-set dislocations are expected to be stable against dissociation. Their results indeed confirmed this for zero temperature. [7]. Results of my simulations indicate that at low temperatures the core is narrow and has an ordered reconstruction as expected. In contrast, at high temperature the core becomes broad and disordered, and the dislocation cannot be described in terms of straight line segments lying in the Peierls valley connected by kinks and jogs. I argue that this has a significant effect on the mobility of the dislocation.

While all of the approaches suggested for linear scaling methods [8,9] have been generalized for nonorthogonal models [9–15], only those based on integration of the Green’s function [15] or Fermi operator rational expansion [12] are equally efficient for both. Other methods require either an explicit computation of the overlap matrix inverse [13,14], or additional multiplications by the overlap matrix [10,11] which reduce the sparsity of matrices used during intermediate steps of the algorithm, or both [9]. The method I present for simulating a finite subsection of an infinite system is based on an iterative inversion calculation of a Green’s function combined with the Fermi function approximation of Nicholson and Zhang to minimize the computational effort [16]. Coupling to the infinite system is achieved through a shared electronic chemical potential and a constraint on the Green’s function at the boundary.

*Method.* – In TB MD nuclear positions are treated as classical degrees of freedom, interacting through an effective potential created by electrons described by the quantum-mechanical TB model. Since they move much more quickly than atoms, the electrons are assumed to remain in equilibrium [17]. In a finite TB region the number of electrons treated in the grand canonical ensemble is set by the local chemical potential. Forces are computed from derivatives of the grand thermodynamic potential for the electrons:

$$\Omega(\{\epsilon_i\}, T, \mu) = 2 \sum_i \epsilon_i f(\epsilon_i) - T \Sigma(\epsilon_i) - \mu f(\epsilon_i), \quad (1)$$

with respect to atomic position. Here  $f(\epsilon)$  is the Fermi function,  $T$  is the electronic temperature,  $\mu$  is the electronic chemical potential, and

$$\Sigma(\epsilon) = -f(\epsilon) \ln f(\epsilon) - (1 - f(\epsilon)) \ln (1 - f(\epsilon)) \quad (2)$$

is the electronic entropy. The eigenvalues  $\epsilon_i$  are the solutions of the generalized eigensystem

$$H_{lm} c_m^i = \epsilon_i S_{lm} c_m^i, \quad (3)$$

where  $\mathbf{H}$ ,  $\mathbf{S}$ , and  $\mathbf{c}^i$  are the TB model Hamiltonian, overlap matrix, and eigenvectors. The force on atom  $\alpha$  is

$$F_\alpha = -\frac{\partial \Omega}{\partial r_\alpha} = -2 \sum_i f(\epsilon_i) c_i^i \left( \frac{\partial H_{lm}}{\partial r_\alpha} - \epsilon_i \frac{\partial S_{lm}}{\partial r_\alpha} \right) c_m^i. \quad (4)$$

Evaluation of eq. (4) as written would require the explicit calculation of the eigenvalues and eigenvectors, a procedure that scales as the number of atoms cubed. Instead, it is useful

to introduce a Green's function matrix in the TB basis,

$$\mathbf{G}(z) = (z\mathbf{S} - \mathbf{H})^{-1}. \quad (5)$$

The sum in eq. (4) can then be rewritten as

$$F_\alpha = \frac{2}{\pi} \lim_{\epsilon \rightarrow 0^+} \text{Im} \int_{-\infty}^{\infty} dE f(E + i\epsilon) \text{Tr} \left[ \mathbf{G}(E + i\epsilon) \left( \frac{\partial \mathbf{H}}{\partial r_\alpha} - (E + i\epsilon) \frac{\partial \mathbf{S}}{\partial r_\alpha} \right) \right]. \quad (6)$$

This integral <sup>(1)</sup> can be evaluated by including a contour at infinity in the upper half plane. Since  $\mathbf{G}(z)$  is analytic in the upper half plane, the value of the integral is determined entirely by the poles of  $f$ . Evaluating eq. (6) in terms of the residues of the exact Fermi function leads to a slowly converging infinite sum. Instead I use an approximation with  $N$  poles [16]:

$$f^{-1}(\epsilon) \approx f_N^{-1}(\epsilon) = 1 + \left[ 1 + \frac{\epsilon}{kT} \left( \frac{4 - \sqrt{8}}{2N} \right) \right]^N \left[ 1 - \frac{\epsilon}{kT} \left( \frac{-2 + \sqrt{8}}{N} \right) \right]^{-N/2} \quad (7)$$

to evaluate  $F_\alpha$  as a sum over the poles  $N_{\text{poles}} = N/2$  of  $f_N$  in the upper half plane,

$$F_\alpha = 4 \text{Re} \sum_{i=1}^{N_{\text{poles}}} \text{Tr} \left[ a_i \mathbf{G}(z_i) \left( \frac{\partial \mathbf{H}}{\partial r_\alpha} - z_i \frac{\partial \mathbf{S}}{\partial r_\alpha} \right) \right]. \quad (8)$$

Here  $z_i$  is the  $i$ -th pole of  $f_N$  and  $a_i$  is its residue. Errors due to the approximate nature of  $f_N$  are minimized when the energy or free energy used is stationary with respect to the electronic occupation function [16]. Defining the forces in terms of derivatives of the grand thermodynamic potential satisfies this criterion.

The density matrix at finite electronic temperature is localized in real space for both semiconductors and metals [18,19], in accord with the principle of nearsightedness formulated by Kohn [20]. While a short-range density matrix could be the result of cancellation between long-range parts of  $\mathbf{G}(z)$  in the residue sum, in silicon  $\mathbf{G}(z)$  at each pole of  $f_N$  is localized, as I discuss below. The localization enables the storage and calculation of  $\mathbf{G}(z)$  in sparse form by including only matrix elements corresponding to pairs of atoms within a cutoff distance. Using the biconjugate gradient algorithm [21] to iteratively invert  $z\mathbf{S} - \mathbf{H}$  by solving

$$(z\mathbf{S} - \mathbf{H})\mathbf{x} = \mathbf{1} \quad (9)$$

is efficient, and enables the application of fixed boundary conditions on  $\mathbf{G}(z)$ , as I describe below. While the short range of the TB model makes  $z\mathbf{S} - \mathbf{H}$  exactly sparse by construction,  $\mathbf{G}(z)$  is only approximately sparse, with a localization range that may differ from that of  $\mathbf{H}$  or  $\mathbf{S}$ . In practice, the range of  $\mathbf{G}(z)$  is known, and its sparsity is ensured by setting to zero any components of the gradient that fall outside the sparsity pattern.

To improve the convergence rate, I precondition the linear solver by explicitly multiplying the left- and right-hand sides of eq. (9) by an approximate matrix inverse. During MD simulation, the preconditioner is  $\mathbf{G}(z)$  from the previous time step. To start the simulation, I use an initial guess for  $\mathbf{G}(z)$ , where matrix elements between a pair of orbitals are set to  $\mathbf{G}(z)$  matrix elements between two orbitals in an ideal lattice with a similar relative distance and orientation. When applied to a sparse matrix the iterative solver converges to a nonsymmetric matrix, and the solution must be symmetrized.

---

<sup>(1)</sup>An expression, valid only for systems with no fractional occupation, that is similar to eq. (6) was proposed by Jayanthi *et al.* for the derivative of the zero-temperature electronic energy at fixed electron number [15].

To use the iterative inversion method (IIM) for simulating a subsection of an infinite system, I perform the inversion procedure with a constraint on  $\mathbf{G}(z)$  matrix elements in the boundary. Atoms outside the simulated region are neglected. Atoms within 4 Å of the outer surface of the region define the boundary region. All  $\mathbf{G}(z)$  matrix elements involving boundary atoms are held fixed during the iterative matrix inversion procedure to the values they would have if they were part of an ideal lattice. This constraint is applied by splitting the unknown  $\mathbf{x}$  of eq. (9) into the sum of two parts, a constrained part  $\mathbf{x}_c$  and a free part  $\mathbf{x}_f$ , and rewriting the equation as

$$(z\mathbf{S} - \mathbf{H})\mathbf{x}_f = \mathbf{1} - (z\mathbf{S} - \mathbf{H})\mathbf{x}_c. \quad (10)$$

Any elements of  $\mathbf{x}_f$  that are constrained and therefore accounted for in  $\mathbf{x}_c$  are set to zero. Without fixing  $\mathbf{G}(z)$  in the boundary region at the ideal-lattice value, states at the fictitious surface of the TB region would appear in mid-gap and modify the forces computed within the region. Computation of the density of states confirms that such gap states are nearly eliminated by the constraints. One possible refinement would be to compute the local strain at each atom in the boundary, and fix  $\mathbf{G}(z)$  matrix elements to values from a uniformly strained periodic system.

*Tests.* – In the following I use a nonorthogonal TB model for silicon with a range of  $\sim 5.05$  Å [22]. The function  $f_N$  has eight poles in the upper half plane, scaled to occupy a valence band width of 15 eV, leading to an effective electronic  $T$  of about 0.25 eV. The exact  $\mathbf{G}(z)$  evaluated at the poles of  $f_N$  is localized in real space, decaying exponentially with range  $r_G < 2.86$  Å, with the least localized  $\mathbf{G}(z_i)$  corresponding to the pole closest to the Fermi level. Comparison with an exact calculation shows that imposing a range cutoff on  $\mathbf{G}(z)$  of 8.0 Å at all poles leads to a good approximation of the forces. All computations are carried out at fixed  $\mu$ , set in the middle of the ideal-lattice band gap. The solver is implemented on an IBM SP parallel computer using MPI [23] for interprocess communications. The work is divided among the processors by splitting  $\mathbf{x}$  and the right-hand side of eq. (9) by column, and duplicating  $z\mathbf{S} - \mathbf{H}$  and the preconditioner on each processor. The linear scaling of the computational time for systems with 128 to 1728 atoms is shown in fig. 1. Parallel efficiency for up to 50 processors is greater than 0.7.

As a test, I calculate the forces around an ideal silicon vacancy to compare periodic supercell exact diagonalization with the IIM for a nonperiodic system. In fig. 2, I plot the magnitude of the force on each atom as a function of its distance from the vacancy for a 12 Å radius TB region, compared with an exact diagonalization calculation of a 16.29 Å cubic supercell. Both calculations are performed at constant  $\mu$  set in the middle of the ideal-lattice band gap, with the same electronic temperature. The forces on atoms near the vacancy are very accurately computed by the IIM, with an RMS deviation of 0.025 eV/Å. As expected, the disturbance caused by the fictitious surface of the TB region is only significant in the constrained boundary region.

*Application to dislocation core structures.* – The simulation of a single dislocation is an application that takes advantage of the IIM's ability to simulate nonperiodic systems. The TB model used [22] was fit to the energies of strained bonds and coordination defects in silicon, and should accurately predict the energies of geometries likely to appear in the dislocation core. The simulations follow a three-step process. The initial configuration is created by applying the solution for the displacement field around an isolated screw dislocation in an infinite isotropic elastic medium to a diamond structure lattice (see ref. [24], p. 60). The dislocation line is aligned along a [110] direction in the shuffle planes. The system comprises a cylinder parallel to the [110] direction, 22.0 Å in radius and 15.36 Å long, with periodic

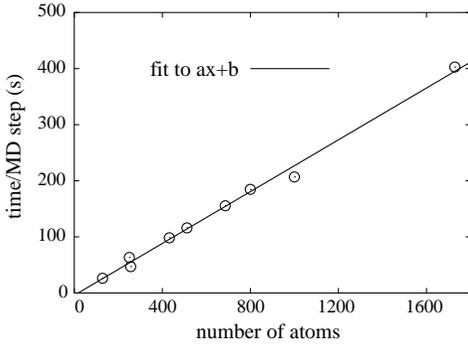


Fig. 1

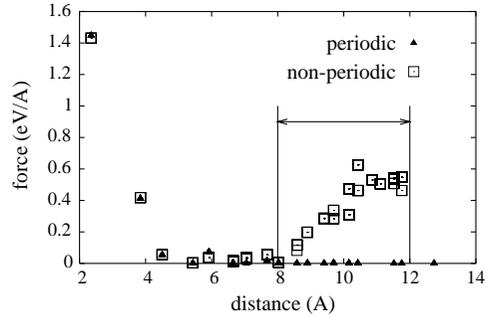


Fig. 2

Fig. 1 – Computational time for one MD time step as a function of number of atoms in the system, measured on 32 nodes of an IBM SP/P3.

Fig. 2 – Force on each atom in a system with a vacancy as a function of distance from the vacancy computed in a nonperiodic system with the IIM, and compared with a periodic system with exact diagonalization. Vertical lines indicate extent of electronic constraint region.

boundary conditions along the axis. The configuration is relaxed with a simple pair potential, quadratic in the interatomic distance with a minimum at the ideal nearest-neighbor distance of 2.35 Å. Atoms farther than 18.0 Å from the axis of the cylinder are held fixed. The TB MD code is then used to evolve the atomic configurations at constant temperature using the velocity Verlet algorithm [25, 26] with a 0.5 fs time step. Atoms that were held fixed during the initial pair potential relaxation are considered part of the boundary region and their  $\mathbf{G}(z)$  matrix elements are constrained.

The left panel of fig. 3 shows the TB MD relaxed core structure, generated by evolving the configuration for 50 fs while rescaling the atomic velocities after every time step to keep the temperature fixed at 50 K. The core shows substantial bond bending, but no coordination

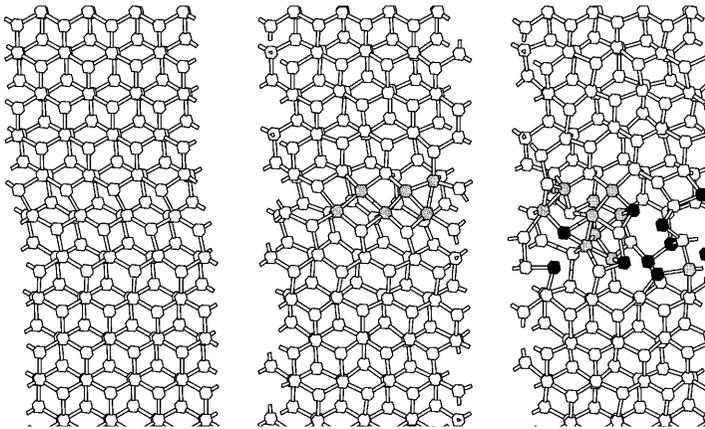


Fig. 3 – Core structures of a screw dislocation in the shuffle plane at 50 K (left), 500 K (middle), and 1000 K (right). A 9 Å thick slice is shown, viewed along the  $[1\bar{1}1]$  direction, with the  $[110]$  dislocation line direction pointing to the right. Bonds are drawn between atoms that are less than 2.8 Å apart. Fourfold coordinated atoms are in white, undercoordinated atoms are black, and overcoordinated atoms are grey.

defects. Since silicon dislocations are believed to be active only at high temperatures (above 850 K) [27], I plot the structures of the screw dislocation after a 500 fs anneal at 500 K (fig. 3 middle panel) or 1000 K (fig. 3 right panel). The melting point of this model in MD simulation is between 1250 K and 1500 K, lower than the experimental value of 1685 K [28]. At 500 K the dislocation core has distorted, creating a partially ordered reconstruction with overcoordination defects. The 50 K and 500 K structures are conventional ordered structures, reminiscent of the asymmetric and symmetric reconstructions proposed for the  $90^\circ$  partial dislocation [29]. Between 500 K and 1000 K, the structure undergoes a dramatic change: it becomes broader and disordered, with many coordination defects. This high-temperature structure is unlike the conventional description, which characterizes thermally activated defects in the dislocation core in terms of geometrically simple features such as dislocation line kinks, jogs, and reconstruction solitons [24, 30, 31]. Annealing the 1000 K structure at 500 K does not restore the 500 K structure in fig. 3. Since the constraints on  $\mathbf{G}$  make the energies ill defined, it is not possible to determine whether this is because the 500 K structure in fig. 3 is not the lowest energy structure, or whether it is because of the limited length of the anneal.

*Discussion.* – The appearance of disorder is driven by a combination of increased entropy and strain relief that lower the free energy. It is perhaps analogous to surface premelting [32], with high strain at the core, rather than the free surface, weakening the bonds. If this observation of disorder is also applicable to the partial dislocations that are experimentally observed in silicon [31, 33], it could affect their mobility through several competing mechanisms. As in diffusion in diamond structure materials [34], coordination defects could reduce the energy barriers to core motion by providing the mobility necessary for rearrangement of interatomic bonds. The broadening of the core could lead to increased mobility by reducing the Peierls stress, as would be predicted by the Frenkel-Kontorova model (see ref. [24], p. 240). However, disorder in the core could hinder motion by creating a configurational entropy barrier to dislocation motion, which must leave behind an ordered lattice. The net effect of core disorder on dislocation mobility is therefore controlled by a subtle balance that will have to be studied by direct simulation of the motion of a dislocation under an applied stress.

The iterative inversion method enables the use of nonorthogonal TB models for the simulation of large numbers of atoms. It is well suited for simulating a subsection of an infinite system with a complex boundary, such as the nonperiodic configurations that arise in a wide range of systems including dislocations and multiscale phenomena. The application of this method to the structure of an isolated dislocation in silicon has revealed a new level of complexity in this system. The role of disorder on the properties of dislocations at high temperatures will be studied in future work on this subject.

\* \* \*

I thank the NRL for financial support through the NRC Associateship Program. This work was supported in part by a Challenge Project grant of computer time from the DoD HPCMO at the ASC MSRC and the MHPCC DSRC. I thank D. W. HESS, R. E. RUDD, N. A. MODINE and G. M. STOCKS for valuable discussions.

## REFERENCES

- [1] GORINGE C. M., BOWLER D.R. and HÉRNANDEZ E., *Rep. Prog. Phys.*, **60** (1997) 1447; LEWIS L. J. and N. MOUSSEAU, *Comp. Mat. Sci.*, **12** (1998) 210.
- [2] MARKLUND S., *Phys. Status Solidi B*, **85** (1978) 673.
- [3] LIU F., MOSTOLLER M., MILMAN V., CHISHOLM M. F. and KAPLAN T., *Phys. Rev. B*, **51** (1995) 17192.

- [4] LEHTO N. and ÖBERG S., *Phys. Rev. B*, **56** (1997) R12706.
- [5] ABRAHAM F. F., BROUGHTON J. Q., BERNSTEIN N. and KAXIRAS N., *Europhys. Lett.*, **44** (1998) 783; BROUGHTON J. Q., ABRAHAM F. F., BERNSTEIN N. and KAXIRAS E., *Phys. Rev. B*, **60** (1999) 2391.
- [6] KOHLHOFF S., GUMBSCH P. and FISCHMEISTER H. F., *Philos. Mag. A*, **64** (1991) 851; GUMBSCH P., *J. Mater. Res.*, **10** (1995) 2897.
- [7] ARIAS T. A. and JOANNOPOULOS J. D., *Phys. Rev. Lett.*, **73** (1994) 680;
- [8] GOEDECKER S., *Rev. Mod. Phys.*, **71** (1999) 1085.
- [9] PALSER A. H. R. and MANOLOPOULOS D. E., *Phys. Rev. B*, **58** (1998) 12704.
- [10] NUNES R. W. and VANDERBILT D., *Phys. Rev. B*, **50** (1994) 17611.
- [11] ORDEJÓN P., DRABOLD D. A., MARTIN R. M. and GRUMBACH M. P., *Phys. Rev. B*, **51** (1995) 1456.
- [12] GOEDECKER S., *Phys. Rev. B*, **48** (1993) 17573.
- [13] MILLAM J. M. and SCUSERIA G. E., *J. Chem. Phys.*, **106** (1997) 5569.
- [14] STEPHAN W. and DRABOLD D. A., *Phys. Rev. B*, **57** (1998) 6391.
- [15] JAYANTHI C. S., WU S. Y., COCKS J., LUO N. S., XIE Z. L., MENON M. and YANG G., *Phys. Rev. B*, **57** (1998) 3799.
- [16] NICHOLSON D. M. C. and ZHANG X.-G., *Phys. Rev. B*, **56** (1997) 12805.
- [17] BORN M. and OPPENHEIMER R., *Ann. Phys.*, **84** (1927) 457.
- [18] GOEDECKER S., *Phys. Rev. B*, **58** (1998) 3501.
- [19] ISMAIL-BEIGI S. and ARIAS T. A., *Phys. Rev. Lett.*, **82** (1999) 2127.
- [20] KOHN W., *Phys. Rev. Lett.*, **76** (1996) 3168.
- [21] PRESS W. H., TEUKOLSKY S. A., VETTERLING W. T. and FLANNERY B. P., *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edition (Cambridge University Press, Cambridge, UK) 1992, p. 83.
- [22] BERNSTEIN N. and KAXIRAS E., *Phys. Rev. B*, **56** (1997) 10488.
- [23] GROPP W., LUSK E. and SKJELLUM A., *Using MPI: Portable Parallel Programming with the Message-Passing Interface* (The MIT Press, Cambridge) 1994.
- [24] HIRTH J. P. and LOTHE J., *Theory of Dislocations* (McGraw-Hill, New York) 1982.
- [25] SWOPE W. C., ANDERSEN H. C., BERENS P. H. and WILSON K. R., *J. Chem. Phys.*, **76** (1982) 637.
- [26] TUCKERMAN M., BERNE B. J. and MARTYNA G. J., *J. Chem. Phys.*, **97** (1992) 1990.
- [27] SAMUELS J. and ROBERTS S. G., *Proc. R. Soc. London, Ser. A*, **421** (1989) 1.
- [28] WEAST R. C. (Editor), *CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton) 1986, p. E-102.
- [29] BIGGER J. R. K., MCINNES D. A., SUTTON A. P., PAYNE M. C., STICH I., KING-SMITH R. D., BIRD D. M. and CLARKE L. J., *Phys. Rev. Lett.*, **69** (1992) 2224.
- [30] HUANG Y. M., SPENCE J. C. H. and SANKEY O. F., *Phys. Rev. Lett.*, **74** (1995) 3392.
- [31] NUNES R. W., BENNETTO J. and VANDERBILT D., *Phys. Rev. Lett.*, **77** (1996) 1516.
- [32] MODESTI S., DHANAK V. R., SANCROTTI M., SANTONI A., PERSSON B. N. J. and TOSATTI E., *Phys. Rev. Lett.*, **73** (1994) 1951.
- [33] FARBER B. YA., IUNIN YU. L. and NIKITENKO V. I., *Phys. Status Solidi A*, **97** (1986) 469; NIKITENKO V. I., FARBER B. YA. and IUNIN YU. L., *Sov. Phys. JETP*, **66** (1987) 738; FARBER B. YA., IUNIN YU. L., NIKITENKO V. I., ORLOV V. I., ALEXANDER H., GOTTSCHALK H. and SPECHT P., *Phys. Status Solidi A*, **138** (1993) 557.
- [34] S. M. HU, *Mater. Sci. Eng. R: Reports*, **13** (1994) 105.