

phys. stat. sol. (b) **217**, 685 (2000)

Subject classification: 71.15.Mb; 71.15.Pd; 73.25.+i; S5; S5.11; S7.14

Large-Scale Applications of Real-Space Multigrid Methods to Surfaces, Nanotubes, and Quantum Transport

J. BERNHOLC, E. L. BRIGGS, C. BUNGARO, M. BUONGIORNO NARDELLI,
J.-L. FATTEBERT, K. RAPCEWICZ, C. ROLAND, W. G. SCHMIDT, and Q. ZHAO

North Carolina State University, Raleigh, NC 27695-8202, USA

(Received August 10, 1999)

The development and applications of real-space multigrid methods are discussed. Multigrid techniques provide preconditioning and convergence acceleration at all length scales, and therefore lead to particularly efficient algorithms. When using localization regions and optimized, non-orthogonal orbitals, calculations involving over 1000 atoms become practical on massively parallel computers. The applications discussed in this chapter include: (i) dopant incorporation and ordering effects during surface incorporation of boron, which lead to the formation of ordered domains at half-monolayer coverage; (ii) incorporation of Mg into GaN during growth, and in particular the conditions that would lead to maximum p-type doping; (iii) optical fingerprints of surface structures for use in real-time feedback control of growth; and (iv) mechanisms of stress release and quantum transport properties of carbon nanotubes.

1. Introduction

Electronic structure calculations based on density functional theory have been enormously successful in predicting and explaining the properties of materials. Pseudopotentials provide a further simplification, since they eliminate the need to explicitly consider the core electrons. Nevertheless, the computational demands of calculations necessary to study complex or artificially structured materials are still very large. This has stimulated the interest in new “real-space” methods, which can enable even larger calculations and/or reduce the computational cost.

The real-space methods avoid the use of plane waves, which extend throughout the entire system, so that the vast majority of operations are “local” in real-space. This has several advantages. Advanced mathematical techniques can be used, which automatically separate the various length scales present in the problem, substantially accelerating convergence (see below). Parallelization becomes much easier, because each processor can be assigned a given region of space. Finally, a real-space formulation is needed to develop and use $O(N)$ techniques, which overcome the $O(N^3)$ scaling of the computational cost of traditional electronic structure techniques.

Real-space methods that employ atom-centered functions, e.g., gaussians or atomic orbitals, have long been used in electronic structure calculations. However, large numerical bases are also very interesting since they allow one to attain any desired numerical accuracy by systematically increasing the number of degrees of freedom. Furthermore, the multiscale aspects of real-space calculations are easiest to exploit with methods based on techniques of applied mathematics: multigrids, wavelets and finite elements. The numerical bases used in recent calculations include finite elements [1, 2, 4, 5], grids

[6 to 14] and wavelets [15 to 17]. These bases allow for natural implementation of mesh refinement [6, 8, 10, 13], cluster boundary conditions, and efficient domain decomposition approaches on massively parallel computers [9, 10].

The computational effort in traditional electronic structure calculations must ultimately scale as $O(N^3)$, where N is the number of atoms. This is because the wavefunction of each electron can in general extend over the whole solid, and therefore computing one wavefunction will take at least $O(N)$ operations. Since the number of electrons grows linearly with the number of atoms, the computational effort must grow at least as $O(N^2)$. Furthermore, the individual wavefunctions must be orthogonal to each other and the orthogonalization or diagonalization effort will ultimately dominate, since they scale as $O(N^3)$. Recently, a number of ingenious methods have been proposed for evaluating the total energy in $O(N)$ operations, see Refs. [18, 19] for recent reviews. The main idea in most methods is to rewrite the total energy expression with help of optimized *localized* functions. Each function is confined to a given region of space, but the various confinement regions overlap so that there is little loss of generality. Clearly, overlaps between functions localized in regions sufficiently far away will vanish, which results in the number of non-zero overlaps being $O(N)$. By exploring this sparsity while avoiding the calculations of the individual eigenfunctions it is possible to evaluate the total energy in $O(N)$ operations. The $O(N)$ methods have already been successful for tight-binding Hamiltonians, where interactions between the various atoms are described with short-range parameters. For true density-functional calculations, however, the advancement has been much slower, due to convergence problems in optimizing the many degrees of freedom.

The present chapter reviews the development and applications of multigrid-based methods for electronic structure calculations, focusing mainly on the work performed at North Carolina State University. In the current implementation, a real-space grid is used as a basis. The grid is fine enough to accurately describe the wavefunctions, the density, and the first-principles ionic pseudopotentials that represent the atoms. An important ingredient of the method is a compact, accurate and efficient discretization scheme, which also aids in parallelization. Multigrid methods [20], which provide automatic preconditioning on all length scales, are used to greatly reduce the number of iterations needed to converge the electronic wavefunctions. In particular, the advantage of multigrids became apparent for systems that are difficult to converge, due to, e.g., charge sloshing or multiple length scales. Quantum molecular dynamics, which must conserve total energy to very high accuracy, works very well and allows for large time steps. Parallelization has also been very effective, and the computational speed on the massively parallel Cray T3E supercomputer has scaled linearly with the number of processors, up to the maximum number we had access to (1024). Our method was already tested and applied to a large number of systems. These include the C_{60} molecule, diamond, Si and GaN supercells, and ab initio MD simulations. Further applications include simulations of surface melting of Si and of strain-induced deformations in carbon nanotubes, properties of Si, GaN, and InP surfaces, $In_xGa_{1-x}N$ quantum wells, and initial studies of a large amyloid β peptide ($C_{146}O_{45}N_{42}H_{210}$) implicated in Alzheimer's disease. The accuracy of the method was also tested by direct comparison with plane-wave calculations, when possible, and the results were in excellent agreement in all cases.

Very recently, we have investigated the performance of $O(N)$ techniques in the context of grid-based DFT calculations. By developing a generalization of the Galli-Parri-

nello method [21], we are able to use non-orthogonal, grid-based localized orbitals, while accelerating convergence through multigrid techniques and maintaining a high accuracy discretization. The localization of the orbitals makes the computation of the wavefunctions and the electron density scale as $O(N)$. These are the most expensive parts of the calculations. However, in order to minimize the number of iterations, we found it advantageous to also include partially occupied and unoccupied orbitals in the calculations, which requires an explicit diagonalization. The unoccupied orbitals accelerate the convergence even further, leading to the overall number of iterations that is not much greater than when orthonormal, fully delocalized orbitals are used. Since the codes are fully parallel, efficient SCALAPACK routines are used for the diagonalization and matrix operation, so that the $O(N^3)$ part is less than 20% even for calculations involving more than a thousand atoms.

The optimized localized orbitals are also very useful apart from facilitating exceedingly large calculations. They enable efficient calculations of quantum transport properties, which rely on the expansion of Hamiltonian and Green's function matrices in terms of localized functions [22]. Since one can use very few localized functions per atom, e.g., four for the case of carbon atoms, the dimensions of the relevant matrices are minimized, which has enabled ab initio calculations of quantum conductance for large nanotube structures [23].

The rest of this paper is organized as follows. In Section 2 the methodology of real-space multigrid calculations [9] is briefly summarized, together with its generalization to the case of localized, non-orthogonal orbitals [24]. This is followed by several examples of recent applications of multigrid methods. Impurity and dopant incorporation at Si(100) surfaces is discussed in Section 3. Section 4 focuses on wide gap nitride semiconductors, and in particular on the incorporation modes of Mg, the dominant p-type dopant, into GaN(0001) surfaces. Section 5 describes the calculations of optical signatures of surface reconstructions, for in situ monitoring and control of growth. Section 6 focuses on defect formation mechanisms in carbon nanotubes under severe tensile stress and the electronic effects of such defects on quantum conductance.

2. Summary of Real-Space Multigrid Methodology

2.1 Grid calculations with multigrid acceleration

In the method described in Ref. [9], a real-space mesh is used to represent the wavefunctions, the charge density, and the ionic pseudopotentials. The density functional equations are discretized using a generalized eigenvalue form:

$$\mathbf{H}_{\text{mehr}}[\psi_n] = \frac{1}{2} \mathbf{A}_{\text{mehr}}[\psi_n] + \mathbf{B}_{\text{mehr}}[\mathbf{V}_{\text{eff}}\psi_n] = \epsilon_n \mathbf{B}_{\text{mehr}}[\psi_n],$$

where \mathbf{A}_{mehr} and \mathbf{B}_{mehr} are the components of the *Mehrstellen* procedure [25], which is based on Hermite's generalization of Taylor's theorem. It uses a weighted sum of the wavefunction and potential values to improve the accuracy of the discretization of the *entire* differential equation, not just the kinetic energy operator. Only nearest and next-nearest neighbor points are used in the discretization, and the short-ranged representation of the Hamiltonian leads to an efficient domain-decomposition-based implementation on massively parallel computers.

To efficiently solve the discretized equations, we use multigrid iteration techniques that accelerate convergence by employing a sequence of grids of varying resolutions.

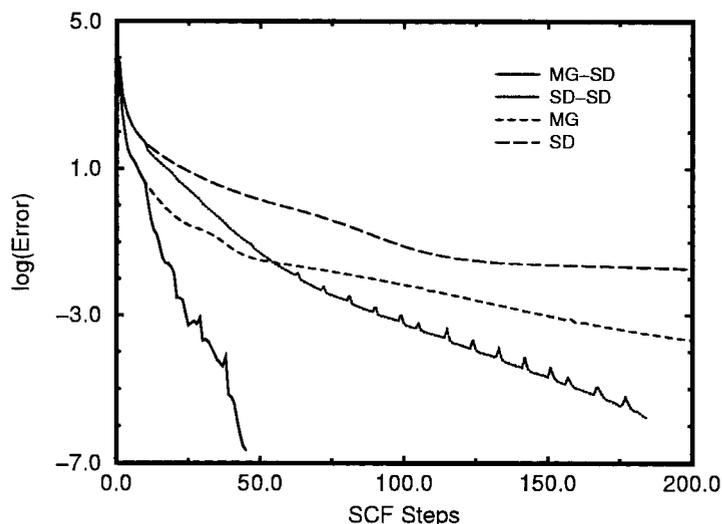


Fig. 1. Convergence rates from a random start for a 64-atom diamond cell with a N impurity: SD steepest descents, MG multigrid, SD-SD steepest descents with subspace diagonalizations, and MG-SD multigrid with subspace diagonalizations

The solution is obtained on a grid fine enough to accurately represent the pseudopotentials and the electronic wavefunctions. However, the iterations are accelerated by solving both the Poisson and the density functional equations on coarser grids and transferring the resulting corrections to the fine grid. This procedure provides excellent

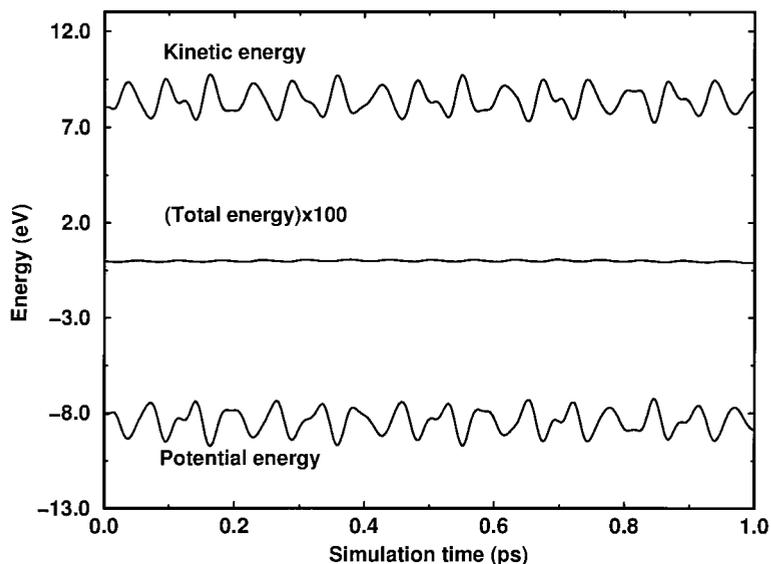


Fig. 2. The potential, kinetic, and total energies of a MD simulation of a 64-atom Si cell at 1100 K with an 80 a.u. time step. The total energy curve is multiplied by a factor of 100. The potential and total energies have been shifted by 251.171 a.u.

preconditioning for all length scales present in a system and leads to very fast convergence rates, see Fig. 1. The operation count to converge one wavefunction with a fixed potential is $O(N_{\text{grid}})$.

The real space methodology is also very suitable for high quality quantum molecular dynamics simulations. For example, a 1 ps test simulation of bulk Si at 1100 K conserved the total energy to within 27 μeV per atom, see Fig. 2. The details of the current multigrid implementation, which include new compact discretization schemes in real space for systems with cubic, orthorhombic, and hexagonal symmetries, and multilevel algorithms for the iterative solution of density functional equations are described in Ref. [9].

2.2 Towards linear scaling in *ab initio* electronic structure calculations

Real-space methods are inherently local, and therefore appropriate for imposing localization constraints on the basis representing the orbitals or for local mesh refinements [6, 8 to 10, 13]. In the Galli-Parrinello method [21], the electron density is expanded in terms of localized, non-orthogonal orbitals ϕ_i . The conjugate orbitals, defined by $\bar{\phi}_i = \sum_j S_{ji}^{-1} \phi_j$, where S is the overlap matrix, satisfy the relation $\langle \bar{\phi}_i | \phi_j \rangle = \delta_{ij}$. The total energy can then be written in a form where the relevant quantities depend on the product $\bar{\phi}_i^* \phi_i$ and the corresponding Euler equations are derived. If the orbitals ϕ_i are constrained to be zero outside of fixed localization radii, one obtains a variational $O(N)$ expression for the total energy whose accuracy, compared to the full density-functional solution, depends on the size of the localization radii. One should stress that the shapes of the orbitals inside their respective localization regions are not predetermined, but are optimized variationally in an iterative fashion. Our initial tests of this approach on 64- and 216-atom cells of Si have demonstrated that while good accuracy can be obtained for sufficiently large localization radii, the convergence rate in optimizing the localized functions on a grid was unacceptably slow. However, by including unoccupied orbitals and explicit multigrid preconditioning, the number of iterations was reduced to a level not greatly different from that of the standard multigrid method without any localization approximation. An outline of the essential steps of our approach is given below. A full discussion will be published elsewhere [24].

Consider the Mehrstellen finite difference discretization of the Kohn-Sham equations, which leads to a generalized eigenvalue problem

$$H\psi = \epsilon B\psi,$$

where H and B are real $M \times M$ sparse matrices. It is convenient to write the eigenvectors ψ as the columns of a $M \times N$ matrix denoted Ψ . In a non-orthogonal basis given by

$$\chi = (\varphi_1, \dots, \varphi_N), \quad (1)$$

the matrix steepest-descent direction, D , is

$$D = B\chi\Theta - H\chi, \quad (2)$$

where

$$\Theta = (\chi^\dagger B\chi)^{-1} (\chi^\dagger H\chi). \quad (3)$$

In the particular case of B being the identity matrix, the result (2) is equivalent to that given in [21]. However, eq. (2) provides an invariant expression for the true steepest descent (SD) direction for any non-orthogonal basis. Without preconditioning, the step along the SD direction must be very small for numerical stability reasons and the convergence of the process can be very slow, especially if $M \gg N$. To prevent this slow-down, we use an iterative multigrid operator as a preconditioner that reduces the high frequency components of the steepest-descent vectors.

When working with non-orthogonal functions χ instead of the eigenfunctions Ψ , localization constraints can easily be imposed on χ to reduce the cost of the calculation. At this time we are using spherical localization regions around each atom, forcing the functions to be zero outside. In particular, this truncation linearizes the cost of computing updates to the orbitals and forming the electron density – the most expensive parts of the algorithm.

As mentioned above, inclusion of some unoccupied and/or partially occupied orbitals in the N computed orbitals substantially improves the convergence rate and the range of applicability of the method. For a chemical potential μ , we define the matrix Υ by its matrix elements:

$$\Upsilon_{ij} = \delta_{ij} f((\epsilon_i - \mu)/kT),$$

where f is the Fermi-Dirac distribution. The density matrix is then given by:

$$\rho = \Psi \Upsilon \Psi^*.$$

For $T = 0$, the density matrix ρ is a projector onto the states with eigenvalues lower than μ . Its dimension is given by the number of degrees of freedom in the discretization, i.e., the number of nodes in the real-space grid. In the finite differences approach, this number is far too large to apply iterative methods that require $\rho(r, r')$. However, one can express ρ in the basis χ :

$$\rho_\chi = \chi^* \rho \chi.$$

Introducing the matrix $\bar{\rho}_\chi = S^{-1} \rho_\chi S^{-1}$ as in [26], the expectation value \bar{A} of an operator A is then given by

$$\bar{A} = \text{tr}(\bar{\rho}_\chi A_\chi).$$

The matrix $\bar{\rho}_\chi$ can be computed by operations on submatrices of dimension N . The knowledge of $\bar{\rho}_\chi$ allows for the computation of the electronic density ρ and the total energy from the non-orthogonal wavefunctions χ .

The above method still contains parts that scale as $O(N^3)$, although with a much reduced prefactor. These include the inversion of the overlap matrix and a submatrix diagonalization. However, by using standard parallel linear algebra packages [27], even without taking advantage of the sparsity of either the overlap or the Hamiltonian matrices, the cost of $O(N^3)$ operations remained below 20% for calculations involving over 1100 atoms.

The localized functions obtained in the process of minimizing the total energy are also quite useful in their own right. Fig. 3 shows one such function for a carbon nanotube. The optimization, which was started from random numbers, led to a chemically very intuitive result: a function strongly localized on a carbon-carbon bond. The quantum-mechanical basis that automatically localizes in bonding regions should lead to ac-

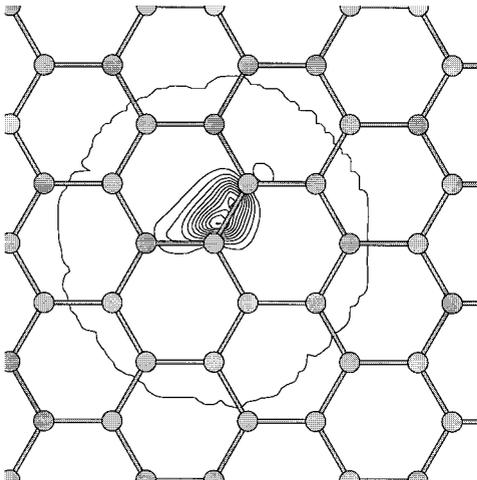


Fig. 3. An optimized electron-density function for a carbon nanotube. Note that although the allowed localization region extends over 6 Å, this function is largely confined to one carbon-carbon bond. The plotting plane is along the surface of the nanotube

curate results while keeping the number of basis functions to a minimum. The latter fact has already been very useful in performing efficient *ab initio* calculations of quantum transport, see Section 6.

The (2×1) reconstructed Si(100) surface, which is preferred for growth of Si-based devices, has complex structural and

electronic properties. It is covered by rows of dimerized three-fold-coordinated Si atoms, with a high density of dangling bonds and anisotropic stress-fields. Correspondingly, the adsorption, incorporation and segregation of impurities at this surface is very complex. In an extensive set of studies, we have investigated the adsorption, incorporation and segregation of a number of common impurities at this surface. Some of our results are summarized below, while a full account is given in Refs. [28 to 30].

Among the considered impurities, two qualitatively distinct classes of behavior were uncovered. Al, Ga, In and Sn exhibit similar behavior. They prefer to adsorb in the trenches between surface dimer rows and their incorporation into the surface is highly unfavorable. Impurities in the other class (which include C, B, P, As and Sb) prefer to adsorb on top of surface dimer rows. Their incorporation into the surface is either energetically favorable or only marginally unfavorable. Within each group, atomic size differences result in quantitative shifts of the incorporation energies. The majority of these impurities, which include donors, acceptors and isoelectronic impurities, are surfactants, in that they prefer the adlayer or surface layer. Boron and carbon alone, which are first-row elements, prefer subsurface incorporation. Furthermore, we found that kinetic effects play an important role in determining the distribution of incorporated configurations.

The above results explain a variety of experimental investigations. Indeed, Al, Ga and In of group III are found in the adlayer. In group V, P and As readily incorporate into the surface [31, 32], while Sb does not incorporate to any significant extent [33]. For B, different models of subsurface B incorporation into the second [34] or the third layer [35] of the surface have been proposed in the literature. Our calculations support the former model over the latter. The results for C and Sn are also consistent with published theory and experiment [33, 36, 37].

Two of the impurities, B and As, have been selected for in-depth investigations [29, 30]. Due to space limitations, we only discuss B here, which is known to form ordered structures at 0.5 monolayer coverage. This effect is important, because an ordered dopant layer could substantially reduce electron scattering phenomena.

We found that the segregation and ordering of boron impurities at high concentration levels near the Si(100) surface is due to a complex interplay between chemical and

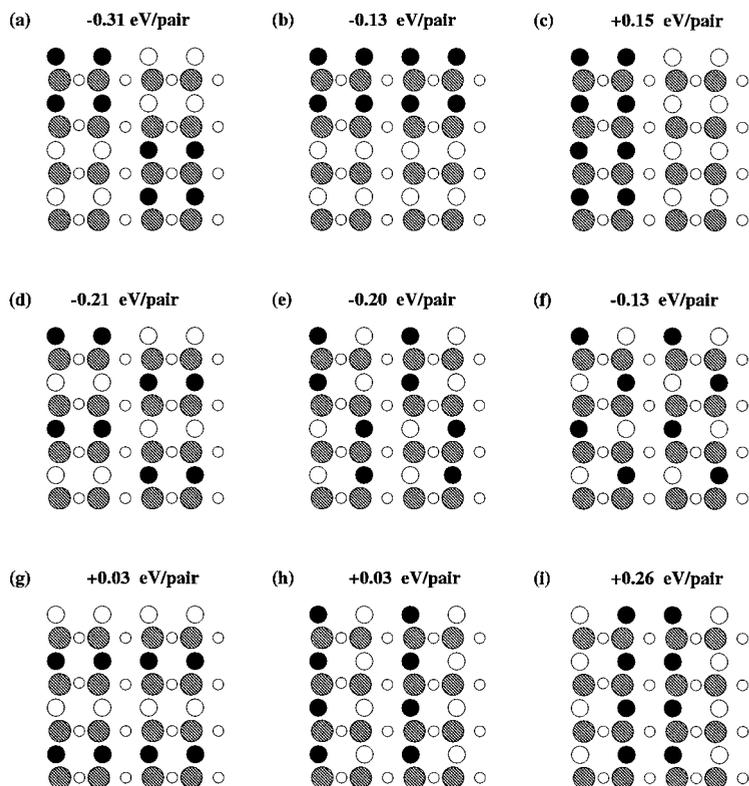


Fig. 4. Schematic structures and energies for 0.5 ML surface doping

strain effects. Like the other acceptors in group III (Al, Ga, In), boron is energetically favored to segregate to the surface, in order to saturate surface dangling bonds and to relieve the strain arising from its large atomic size mismatch with host Si atoms. However, unlike other group III acceptors, it prefers substitutional configurations in the first and second layers of the surface. With increasing doping, aggregation of boron impurities in the second layer of the surface becomes more and more favorable, up to a critical level corresponding to 0.5 ML of the impurities. The low-energy modes of ordering correspond to zigzag patterns of subsurface pairs, see Fig. 4. Above this doping level, the binding energies of ordered impurity configurations decrease sharply, and they become highly repulsive for a complete surface or subsurface impurity monolayer.

A number of experimental observations can be explained on the basis of our results. The equilibrium structure found at 0.5 ML doping has $c(4 \times 4)$ symmetry. However, both at low and high doping levels, several distinct modes of ordering are found to be essentially degenerate. This favors the formation of domain structures over long-range ordering of any one structure. Therefore, even at 0.5 ML coverage only (2×1) LEED patterns are observed [38 to 40], while domains with $c(4 \times 4)$ and (4×4) periodicities are observed in STM [34]. If some of the metastable ordered structures can be stabilized, there would be a potential for engineering complex impurity patterns at the nanoscale.

One common feature of all the surfaces with subsurface boron atoms is the pronounced flattening and lowering of surface Si dimers with B neighbors, of up to 40...50% of the interlayer spacing. The lowered Si dimers are not visible in simulated STM images. However, the removal of such boron-bonded Si dimers to create surface divacancies is energetically very expensive, much more so than on the clean Si(100) surface [28]. Therefore, the strain fields induced by heavy boron doping are not relieved by the creation of ordered patterns of surface Si divacancies. These results provide an alternative and more consistent explanation of the dark features extensively observed on B-doped Si(100) in STM investigations [34, 35, 41].

3. Mg Incorporation at GaN Surfaces

Controlled incorporation of impurities is critical to semiconductor science and technology. Specific dopant concentrations are necessary to achieve desired conductivities, recombination rates in light emitting devices or other electrical characteristics. In general, doping of bulk materials can be performed in three different ways: (i) by in-diffusion following growth, (ii) by ion implantation, and (iii) during growth. The in-diffusion of impurities is limited by the maximum solubility, which is a thermodynamic quantity that can be determined from either experiments or calculations. Ion implantation has no such limit, but in many materials the implantation damage is severe and cannot be annealed out, limiting the usefulness of this technique. During growth, the incorporation of an impurity occurs at the surface and depends sensitively upon the surface and its environment. For example, impurity energetics and consequently its solubility in the near-surface region can be significantly different from that in the bulk. This is because under typical growth conditions diffusion at the surface is much faster than in the bulk and the impurity density of the near-surface region can be frozen in as the film grows, leading to a concentration in the grown film different from that expected on the basis of its bulk solubility. Indeed, under these conditions it is the surface properties, instead of the bulk properties, that most strongly influence the impurity concentration. This is particularly true in a compound semiconductor, where many variables affect impurity incorporation during growth. Surface reconstruction patterns are often complex; in a given growth direction there are at least two possible surface terminations, and the surface structure varies as a function of the chemical potentials of the atomic constituents. The interplay of these effects leads to a complicated doping behavior, which, however, can be used to achieve a desired impurity concentration [42].

We have examined the incorporation of Mg in wurtzite GaN, which is an important paradigmatic case. There is a growing awareness that the two polarities of the (0001) GaN surface, nominally corresponding to the Ga and N faces, exhibit very different behavior. Indeed, striking differences have been observed in the morphology of the (0001) and (000 $\bar{1}$) surfaces [43 to 49], and the strongly ionic nature of the GaN bond together with the low symmetry of the surface results in unusual reconstruction patterns that differ significantly from those observed in other III-V semiconductors. Since as-grown GaN exhibits unintentional n-type conductivity and controlled doping of GaN is necessary for optoelectronic devices, the achievement of good p-type conductivity has been a priority. However, Mg, which is the preferred p-type dopant, has a relatively large ionization energy. Consequently, high Mg concentrations are required to achieve the desired hole densities. In practice, the refined control of doping required to obtain

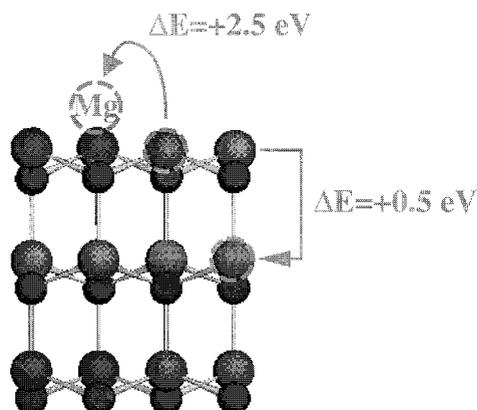


Fig. 5. The energetics of Mg incorporation at the Ga-terminated surface

reproducibly the particularly high Mg concentrations needed has been difficult to accomplish. Our calculations show that the interplay of surface orientation, reconstruction patterns and the availability of the species involved (as measured by their chemical potentials) determines the incorporation characteristics, and that intuition derived from bulk solubility considerations may be misleading [42]. For example, superior incorporation of Mg at the Ga-substitutional site (Mg_{Ga}) is expected under N-rich conditions, because of the decreased competition between Mg and Ga for the Ga site. However, this reasoning is a poor guide for determining the conditions for efficient incorporation. Indeed, we found that the growth conditions for optimal incorporation are reversed when the Mg is incorporated through a surface of different orientation (in this case, the film has different polarity). For the (0001) Ga-surface (see Fig. 5), we find that best incorporation occurs at N-rich and moderate Ga-rich conditions. On the contrary, for the (000 $\bar{1}$) N-surface, in N-rich conditions Mg displays a strong tendency to segregate and superior incorporation occurs in a Ga-rich environment, when the surface is gallium terminated. Growth procedures may therefore need to be tailored in a nontrivial fashion to the properties of the particular growth surface(s) in order to achieve effective doping with specific impurities.

Our other work in this area examined the reconstruction patterns at GaN surfaces [48], band offsets at heterointerfaces [50], the effects of pyroelectric fields on recombination and lasing [51], and defect formation in wide gap nitride alloys [52].

4. Reflectance Anisotropy Spectra for Large Systems

The techniques of optical spectroscopy are evolving from fundamental semiconductor-surface studies to the control of semiconductor processing in situ and with real-time feedback. In particular, Reflectance Difference Spectroscopy (RDS), also known as Reflectance Anisotropy Spectroscopy (RAS), is a reliable in situ monitoring tool, applicable to ultrahigh vacuum as well as to gas-phase environments [53 to 56]. It is particularly useful for the wide range of materials that are grown by metal-organic vapor phase epitaxy (MOVPE), where the well-developed electron- or ion-beam techniques cannot be used. However, first-principles calculations of RDS spectra are very expensive, because thick slabs and a large number of excited states must be used in computing the surface optical response. We have incorporated the formalism [57, 58] for calculating RDS into our parallel multigrid code. This has enabled us to perform, to our knowledge for the first time, a well-converged study of the optical response of realistic III-V(001) surface structures [59]. In particular, we were able to resolve the controversies regarding the $\text{InP}(001)(2 \times 4)$ surface structure [60, 61].

The comparison between theory and experiment, although hindered by the computational shortcomings of the DFT-LDA (see below), allowed for clear identification of the dominant surface features, or structural fingerprints, present under various growth conditions. This characterization should be very useful in real-time feedback and control of InP growth by MOVPE. However, the theoretical energy scale had to be “stretched” in order to match the experimental data. This is due to the well-known deficiencies of DFT-LDA in computing excited state properties.

In contrast to InP, the precise GaP(001) surface structure was essentially unknown prior to our study [62]. On the basis of comprehensive total energy results we established the surface phase diagram. The cation-rich surface turned out to contain mixed dimers, in analogy to the corresponding InP surface. Our theoretical finding is corroborated by recent core-level results [63].

The RA spectra calculated for the energetically favored structures are strongly dependent on the surface geometry: The structures featuring cation–cation surface bonds show a pronounced negative anisotropy in the low energy region, with minima between 2.0 and 2.3 eV. The strength of that anisotropy is directly correlated to the number of Ga–Ga bonds along the [110] direction. On the other hand, we find that P–P dimers give rise to a relatively broad positive anisotropy between about 2.4 and 4.4 eV. These relations between the calculated RA and the surface stoichiometry agree very well with the experimental observations [62]. Furthermore, the results obtained for GaP are in line with the earlier findings for InP, supporting the existence of structural “fingerprints” in the optical spectra. This is an important outcome with respect to the technological application of optical spectroscopy for semiconductor growth.

While a good qualitative description of the measured optical anisotropy was achieved, in particular with respect to the chemical trends, the calculated spectra still deviate significantly from experiment, in particular concerning the energetic positions of the peaks and the line shape. In Fig. 6 we show the DFT-LDA calculation for the Ga-rich GaP surface in comparison with the measured curve. Highlighted in the figure is the pronounced negative anisotropy at low energy (denoted by S). This feature is mainly related to optical transitions between filled Ga–Ga bonding states and empty dangling

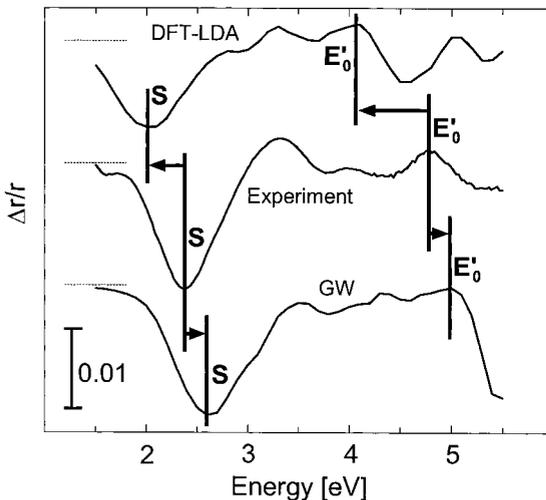


Fig. 6. Calculated and measured RA spectra for the Ga-rich GaP surface (see text)

bond states located at surface Ga atoms. Its energetic position is underestimated in the DFT-LDA calculation by about 0.3 eV with respect to the measured spectrum [62]. A much stronger underestimation of about 0.7 eV occurs for the essentially bulk-related features at higher photon energies. This non-uniform shift of the calculated spectra cannot be remedied by a simple scissor-operator approach. We applied an approximate GW-approach, based on a model dielectric function in conjunction with an approximate treatment of local-field effects and the dynamical screening [64, 65]. The resulting RA spectrum is also presented in Fig. 6. Compared to the DFT-LDA results we observe considerable improvements: The most important effect of the quasiparticle corrections is a non-uniform shift of the spectrum towards higher photon energies. We also observe a stronger shift for the bulk-related features than of the surface signatures.

5. Nanotubes

5.1 Axial tension and plastic transformations

Nanotubes are extremely unusual in their ability to withstand huge deformations, and are thus particularly suitable to mechanical applications. We have previously shown that nanotubes can be reversibly bent to very large angles [66] and this prediction has already been confirmed experimentally. Furthermore, our early classical molecular dynamics simulations indicated exceptional strength, which needed to be investigated further. However, strength measurements on nanotubes are extremely difficult, due to their very short lengths. In lack of any experimental data we performed extensive ab initio simulations to determine the dominant mechanism of strain release. Since the longest ab initio simulations correspond to real times of the order of only tens of picoseconds, simulations at 10% tensile strain and 1800 K were performed, using multiple Nose-Hoover thermostats [67]. Under those conditions the dynamical processes are greatly accelerated and a lengthy simulation has identified the key transformation that releases tension in highly strained carbon nanotubes. This transformation rotates a carbon-carbon bond, converting four hexagons into two pentagons and two hexagons coupled in pairs, the so-called (5-7-7-5) defect. Depending on tube symmetry, temperature and strain rate, it can lead to either breakage or plastic behavior in carbon nanotubes. Once the defect transformation is identified, the thermodynamic stability of nanotubes under strain can be determined by computing defect formation energies as a function of strain. Fig. 7 shows that beyond about 5% tension, a nanotube of “arm-chair” symmetry, in which some carbon-carbon bonds are oriented perpendicular to its axis, can release its excess strain via spontaneous formation of such defects. The predicted strength is indeed the highest by far of any material to date, and there should be no shortage of exciting applications once nanotubes are grown in sufficient quantity and quality. Furthermore, the activation energy for the bond rotation is quite high (see Fig. 8) so that the observed nanotube strengths at ambient temperatures could be substantially higher.

Detailed ab initio simulations for many tube geometries are too expensive at present, but much qualitative and semiquantitative information has been obtained through classical and tight-binding molecular dynamics [67, 68]. It turns out that the appearance of a (5-7-7-5) defect can be interpreted as a nucleation of a degenerate dislocation loop in the hexagonal network of the graphite sheet. The configuration of this primary dipole is a (5-7) core attached to an inverted (7-5) core. The (5-7) defect behaves thus as a

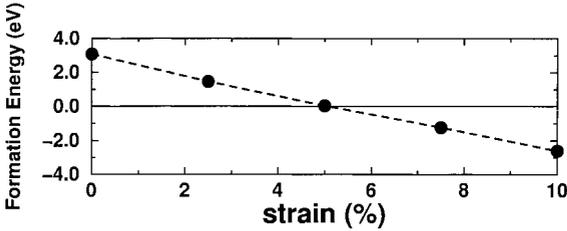


Fig. 7. Formation energy of the (5-7-7-5) defect for the (5,5) tube at different strains, obtained in ab initio calculations

single edge dislocation in the graphitic plane. Once nucleated, the (5-7-7-5) dislocation loop can ease further relaxation by separating the two dislocation cores, which glide through successive bond rotations [67]. This corresponds to a plastic flow of dislocations and gives rise to ductile behavior. In contrast, in the case of nanotubes of the so-called zig-zag symmetry, the same C–C bond is parallel to the applied tension, which is already the minimum energy configuration for the strained bond. The formation of a defect is then limited to rotation of the bonds oriented 120° with respect to the tube axis. Our calculations show that the formation energies of these defects are strongly dependent on curvature, i.e., on the diameter of the tube. Simulations show that nanotubes with a diameter smaller 1.1 nm can deform plastically regardless of their symmetry, while wider tubes must eventually break in a brittle fashion.

As mentioned above, activation energies for the formation defects in perfect nanotubes are very high (cf. Fig. 8). However, due to the high temperature at which the growth of nanotubes occurs, defects will form for thermodynamic reasons and then remain frozen in. For a material at thermal equilibrium, the number of defects of a particular type is given by $N_{\text{sites}}e^{(-G_F/kT)}$, where N_{sites} is the number of potential sites and G_F is the Gibbs free energy of formation of the defect. The entropic contributions are usually a small part of G_F ; one can thus use the energy of formation to obtain a lower bound. For single-walled tubes, which are grown at ~ 1500 K, the above formula suggests that even at zero strain there might be point defects in nanotubes every few tenths of a mm. The presence of the frozen-in defects will certainly limit the ultimate strength of nanotubes and thus some of their proposed uses.

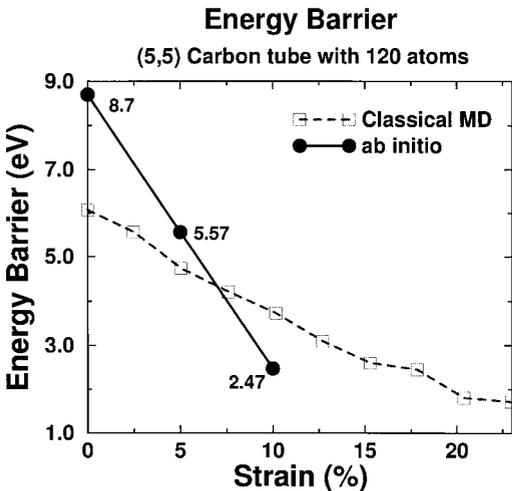


Fig. 8. Preliminary results for the activation energy for bond rotation for the (5,5) tube computed by the multigrid method. The corresponding values obtained by molecular statics using the Tersoff-Brenner potential are also shown. Note that while the qualitative trends are similar between the two approaches, there are substantial quantitative differences

5.2 Quantum transport

Nanotubes can be either metallic or semiconducting, depending on their symmetry (or index). The plastic transformations described above change nanotube index between the two (5-7) pairs. This raises the prospect of easily fabricating metal–semiconductor junctions [67], and thus all-carbon electronic devices [69]. This possibility has stimulated our interest in quantum transport in such structures.

The electronic and transmission properties of carbon nanotubes have already been extensively studied both experimentally and theoretically [70 to 80]. The general problem of calculating the conductance in carbon nanotubes has been addressed with a variety of techniques that reflect the various approaches in the theory of quantum transport in ballistic systems. Most of the existing calculations derive the electronic structure of carbon nanotubes from a simple π -orbital tight-binding Hamiltonian that describes the bands of the graphitic network via a single nearest-neighbor hopping parameter. Since the electronic properties of carbon nanotubes are basically determined by the sp^2 π -orbitals, the model gives a reasonably good qualitative description of their behavior and, given its simplicity, it has become the model of choice in a number of theoretical investigations. However, although qualitatively useful to interpret experimental results, this simple Hamiltonian lacks the accuracy that more sophisticated methods are able to provide. In particular, the π -orbital model is limited to studies of topological changes, and cannot address the effects of topology-preserving distortions, such as bending, on conductance.

Recently, we have developed a general scheme for calculating quantum conductance that is particularly suitable for realistic calculations of electronic transport properties in extended systems [22]. This method is based on the Surface Green's Function Matching formalism and efficiently combines the iterative calculation of transfer matrices with the Landauer formula for coherent conductance. It is very flexible and applicable to any system described by a Hamiltonian with a localized orbital basis. The only quantities that enter are the matrix elements of the Hamiltonian operator, with no need for the explicit knowledge of the electron wavefunctions for the multichannel expansion. The last fact makes the numerical calculations efficient also for systems described by multi-orbital Hamiltonians. In order to obtain the highly-localized *ab initio* orbitals and potentials necessary for the efficient application of this scheme, we use the localized-orbi-

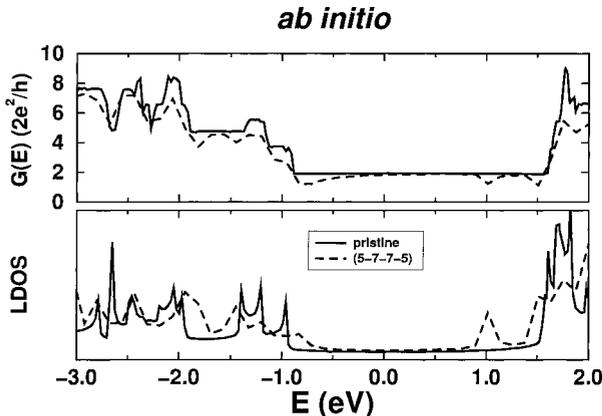


Fig. 9. *Ab initio* conductance and local density of states for a pristine (5,5) nanotube and for one with a (5-7-7-5) defect. The Fermi energy is used as reference

tal method described in Section 2.2. This allows for DFT calculations of quantum conductance to be carried out for systems with complicated topology, contacts, etc.

As the first application, we evaluated the conductance of a (5,5) tube, pristine and with a (5-7-7-5) defect, see Fig. 9. It is clear from the results that a single defect does not affect the conductance near the Fermi level, but there are significant effects near the valence bands maximum and in the upper part of the gap. Our results for the (5,5) tube with a single (5-7-7-5) defect compare very well with a recent ab initio calculation [79] for a larger (10,10) tube, and with tight-binding calculations for the same system. Further applications of this formalism to multi-component nanotube systems are in progress.

Acknowledgement This work was supported in part by ONR, NSF, NASA, and DoE.

References

- [1] B. HERMANSSON and D. YEVICK, *Phys. Rev. B* **33**, 7241 (1986).
- [2] S. R. WHITE, J. W. WILKINS, and M. P. TETER, *Phys. Rev. B* **39**, 5819 (1989).
- [3] H. MURAKAMI, V. SONNAD, and E. CLEMENTI, *Internat. J. Quant. Chem.* **42**, 785 (1992).
- [4] E. TSUCHIDA and M. TSUKADA, *Phys. Rev. B* **52**, 5573 (1995).
- [5] J. E. PASK, B. M. KLEIN, C. Y. FONG, and P. A. STERNE, *Phys. Rev. B* **59**, 12352 (1999).
- [6] J. BERNHOLC, J.-Y. YI, and D. J. SULLIVAN, *Faraday Disc. Chem. Soc.* **92**, 217 (1991).
- [7] J. R. CHELIKOWSKY, N. TROULLIER, and Y. SAAD, *Phys. Rev. Lett.* **72**, 1240 (1994).
- [8] F. GYGI and G. GALLI, *Phys. Rev. B* **52**, R2229 (1995).
- [9] E. L. BRIGGS, D. J. SULLIVAN, and J. BERNHOLC, *Phys. Rev. B* **52**, R5471 (1995); **54**, 14362 (1996).
- [10] G. ZUMBACH, N. A. MODINE, and E. KAXIRAS, *Solid State Commun.* **99**, 57 (1996).
- [11] K. A. IYER, M. P. MERRICK, and T. L. BECK, *J. Chem. Phys.* **103**, 227 (1995).
- [12] A. P. SEITSONEN, M. J. PUSKA, and R. M. NIEMINEN, *Phys. Rev. B* **51**, 14057 (1995).
- [13] J.-L. FATTEBERT, *J. Comput. Phys.* **149**, 75 (1999).
- [14] F. ANCILOTTO, P. BLANDIN, and F. TOIGO, *Phys. Rev. B* **12**, 7868 (1999).
- [15] K. CHO, T. A. ARIAS, J. D. JOANNPOULOS, and P. K. LAM, *Phys. Rev. Lett.* **71**, 1808 (1993).
- [16] S. WEI and M. Y. CHOU, *Phys. Rev. Lett.* **76**, 2650 (1996).
- [17] C. J. TYMCZAK and X.-Q. WANG, *Phys. Rev. Lett.* **78**, 3654 (1997).
- [18] G. GALLI, *Curr. Opin. Solid State Mater. Sci.* **1**, 864 (1996).
- [19] S. GOEDECKER, *Rev. Mod. Phys.*, in press (1999).
- [20] A. BRANDT, *Math. Comp.* **31**, 333 (1977); *GMD Studien* **85**, 1 (1984).
- [21] G. GALLI and M. PARRINELLO, *Phys. Rev. Lett.* **69**, 3547 (1992).
- [22] M. BUONGIORNO NARDELLI, *Phys. Rev. B* **60**, 7828 (1999).
- [23] M. BUONGIORNO NARDELLI, J.-L. FATTEBERT, and J. BERNHOLC, to be published (1999).
- [24] J.-L. FATTEBERT and J. BERNHOLC, to be published (1999).
- [25] L. COLLATZ, *The Numerical Treatment of Differential Equations*, Springer-Verlag, Berlin 1966.
- [26] R. NUNES and D. VANDERBILT, *Phys. Rev. B* **50**, 17611 (1994).
- [27] L. S. BACKFORD, J. CHOI, A. CLEARY, E. D'AZEVEDO, J. DEMMEL, I. DHILLON, J. DONGARRA, S. HAMMARLING, G. HENRY, A. PETTIT, K. STANLEY, D. WALKER, and R. C. WHALEY, *SCALAPACK User's Guide*, SIAM, Philadelphia 1997.
- [28] M. RAMAMOORTHY, E. L. BRIGGS, and J. BERNHOLC, *Phys. Rev. Lett.* **81**, 1642 (1998).
- [29] M. RAMAMOORTHY, E. L. BRIGGS, and J. BERNHOLC, *Phys. Rev. B* **59**, 4813 (1999).
- [30] M. RAMAMOORTHY, E. L. BRIGGS, and J. BERNHOLC, *Phys. Rev. B* **60**, 8178 (1999).
- [31] R. D. BRINGANS, D. K. BIEGELSEN, and L. E. SWARTZ, *Phys. Rev. B* **44**, 3054 (1991).
- [32] Y. WANG, X. CHEN, and R. J. HAMERS, *Phys. Rev. B* **50**, 4534 (1994).
- [33] J. NOGAMI, A. A. BASKI, and C. F. QUATE, *J. Vac. Sci. Technol. A* **8**, 245 (1990); *Phys. Rev. B* **43**, 9316 (1991); **44**, 1415 (1991); **44**, 11167 (1991); *Appl. Phys. Lett.* **53**, 2086 (1991); **58**, 475 (1991).
- [34] Y. WANG, R. J. HAMERS, and E. KAXIRAS, *Phys. Rev. Lett.* **74**, 403 (1995).
- [35] Z. ZHANG et al., *J. Vac. Sci. Technol. B* **14**, 2684 (1996).

- [36] J. TERSOFF, *Phys. Rev. Lett.* **74**, 5080 (1995).
- [37] P. C. KELIRES and E. KAXIRAS, *Phys. Rev. Lett.* **78**, 3479 (1997).
- [38] R. L. HEADRICK, B. E. WEIR, A. F. J. LEVI, D. J. EAGLESHAM, and L. C. FELDMAN, *Phys. Rev. Lett.* **57**, 2779 (1990).
- [39] B. E. WEIR, R. L. HEADRICK, Q. SHEN, L. C. FELDMAN, M. S. HYBERTSEN, M. NEEDELS, M. SCHLUTER, and T. R. HART, *Phys. Rev. B* **46**, 12861 (1992).
- [40] R. CAO, X. YANG, and P. PIANETTA, *J. Vac. Sci. Technol. B* **11**, 1455 (1993).
- [41] T. KOMEDA and Y. NISHIOKA, *Appl. Phys. Lett.* **71**, 2277 (1997).
- [42] C. BUNGARO, K. RAPCEWICZ, and J. BERNHOLC, *Phys. Rev. B* **59**, 9771 (1999).
- [43] F. A. PONCE, D. P. BOUR, W. T. YOUNG, M. SAUNDERS, and J. W. STEEDS, *Appl. Phys. Lett.* **69**, 337 (1996).
- [44] B. DAUDIN, J. L. ROVIÉRE, and M. ARLERY, *Appl. Phys. Lett.* **69**, 2480 (1996).
- [45] Z. LILIENTAL-WEBER, Y. CHEN, S. RUVIMOV, and J. WASHBURN, *Phys. Rev. Lett.* **79**, 2835 (1997).
- [46] M. M. SUNG, J. AHN, V. BYKOV, J. W. RABALAIS, D. D. KOLESKE, and A. E. WICKENDEN, *Phys. Rev. B* **54**, 14652 (1996).
- [47] E. S. HELLMAN, D. N. E. BUCHANAN, D. WIESMANN, and I. BRENER, *MRS Internet J. Nitride Semicond. Res.* **1**, 16 (1996).
- [48] K. RAPCEWICZ, M. BUONGIORNO NARDELLI, and J. BERNHOLC, *Phys. Rev. B* **56**, R12725 (1997).
- [49] A. R. SMITH, R. M. FEENSTRA, D. W. GREVE, J. NEUGEBAUER, and J. E. NORTHRUP, *Phys. Rev. Lett.* **79**, 3934 (1997).
- A. R. SMITH, R. M. FEENSTRA, D. W. GREVE, M.-S. SHIN, M. SKOWRONSKI, J. NEUGEBAUER, and J. E. NORTHRUP, *Appl. Phys. Lett.* **72**, 2114 (1998).
- [50] M. BUONGIORNO NARDELLI, K. RAPCEWICZ, and J. BERNHOLC, *Phys. Rev. B* **55**, R7323 (1997).
- [51] M. BUONGIORNO NARDELLI, K. RAPCEWICZ, and J. BERNHOLC, *Appl. Phys. Lett.* **71**, 3135 (1997).
- [52] P. BOGUSLAWSKI and J. BERNHOLC, *Phys. Rev. B* **59**, 1567 (1999).
- [53] D. E. ASPNES and A. A. STUDNA, *Phys. Rev. Lett.* **54**, 1956 (1985).
- [54] D. E. ASPNES, *Surf. Sci.* **307/309**, 1017 (1994).
- [55] W. RICHTER and J. T. ZETTLER, *Appl. Surf. Sci.* **101**, 465 (1996).
- [56] J. F. MCGILP, D. WEAIRE, and C. H. PATTERSON, *Epioptics. Linear and Nonlinear Optical Spectroscopy of Surfaces and Interfaces*, Springer-Verlag, Berlin 1995.
- [57] R. DEL SOLE, *Solid State Commun.* **37**, 537 (1981).
- [58] F. MANGHI, R. DEL SOLE, A. SELLONI, and E. MOLINARI, *Phys. Rev. B* **41**, 9935 (1990).
- [59] W. G. SCHMIDT, E. L. BRIGGS, J. BERNHOLC, and F. BECHSTEDT, *Phys. Rev. B* **59**, 2234 (1999).
- [60] D. PAHLKE, J. KINSKY, C. SCHULTZ, M. PRISTOVSEK, M. ZORN, N. ESSER, and W. RICHTER, *Phys. Rev. B* **56**, R1661 (1997).
- K. B. OZANYAN, P. J. PARBROOK, M. HOPKINSON, C. R. WHITEHOUSE, Z. SOBIESIERSKI, and D.I. WESTWOOD, *J. Appl. Phys.* **82**, 474 (1997).
- M. ZORN, T. TREPKE, J. T. ZETTLER, B. JUNNO, C. MEYNE, K. KNORR, T. WETHKAMP, M. KLEIN, M. MILLER, W. RICHTER, and L. SAMUELSON, *Appl. Phys. A* **65**, 333 (1997).
- J. KINSKY, C. SCHULTZ, D. PAHLKE, A. M. FRISCH, T. HERRMANN, N. ESSER, and W. RICHTER, *Appl. Surf. Sci.* **123**, 228 (1998).
- [61] C. D. MACPHERSON, R. A. WOLKOW, C. E. J. MITCHELL, and A. B. MCLEAN, *Phys. Rev. Lett.* **77**, 691 (1996).
- M. SHIMOMURA, N. SANADA, Y. FUKUDA, and P. J. MOLLER, *Surf. Sci.* **359**, L451 (1996).
- [62] A. M. FRISCH, W. G. SCHMIDT, J. BERNHOLC, M. PRISTOVSEK, N. ESSER, and W. RICHTER, *Phys. Rev. B* **60**, 2488 (1999).
- [63] N. SANADA et al., *Surf. Sci.* **419**, 120 (1999).
- [64] F. BECHSTEDT et al., *Solid State Commun.* **84**, 765 (1992).
- [65] W. G. SCHMIDT et al., in preparation.
- [66] S. IJIMA, C. BRABEC, A. MAITI, and J. BERNHOLC, *J. Chem. Phys.* **104**, 2089 (1996).
- [67] M. BUONGIORNO NARDELLI, B. I. YAKOBSON, and J. BERNHOLC, *Phys. Rev. B* **57**, R4277 (1998).
- [68] M. BUONGIORNO NARDELLI, B. I. YAKOBSON, and J. BERNHOLC, *Phys. Rev. Lett.* **81**, 4656 (1998).
- [69] B. I. DUNLAP, *Phys. Rev. B* **46**, 1933 (1992).
- L. CHICO, V. H. CRESPI, L. X. BENEDICT, S. G. LOUIE, and M. L. COHEN, *Phys. Rev. Lett.* **76**, 971 (1996).
- [70] P. G. COLLINS, A. ZETTL, H. BANDO, A. TRESS, and R. SMALLEY, *Science* **278**, 100 (1996).
- S. N. SONG, X. K. WANG, R. P. H. CHANG, and J. B. KETTERSON, *Phys. Rev. Lett.* **72**, 697 (1994).

- L. LANGER, L. STOCKMAN, J. P. HEREMANS, V. BAYOT, C. H. OLK, C. VAN HAESENDONCK, Y. BRUYNSERAEDE, and J.-P. ISSI, *J. Mater. Res.* **9**, 927 (1994).
- L. LANGER, V. BAYOT, E. GRIVEI, J.-P. ISSI, J. P. HEREMANS, C. H. OLK, L. STOCKMAN, C. VAN HAESENDONCK, and Y. BRUYNSERAEDE, *Phys. Rev. Lett.* **76**, 479 (1996).
- S. J. TANS, M. H. DEVORET, H. DAI, A. THESS, R. E. SMALLEY, L. J. GEORLIGA, and C. DEKKER, *Nature* **386**, 474 (1997).
- A. BACHTOLD, C. STRUNK, J.-P. SALVETAT, J.-M. BONNARD, L. FORRÓ, T. NUSSBAUMER, and C. SCHÖNENBERGER, *Nature* **397**, 673 (1999).
- [71] A. BEZRYADIN, A. R. M. VERSCHUEREN, S. J. TANS, and C. DEKKER, *Phys. Rev. Lett.* **80**, 4036 (1998).
- [72] S. PAULSON, M. R. FALVO, N. SNIDER, A. HELSER, T. HUDSON, A. SEEGER, R. M. TAYLOR II, R. SUPERFINE, and S. WASHBURN, <http://xxx.lanl.gov/abs/cond-mat/9905304>, preprint (1999).
- [73] W. TIAN and S. DATTA, *Phys. Rev. B* **49**, 5097 (1994).
- [74] R. SAITO, G. DRESSELHAUS, and M. S. DRESSELHAUS, *Phys. Rev. B* **53**, 2044 (1996).
- [75] L. CHICO, L. X. BENEDICT, S. G. LOUIE, and M. L. COHEN, *Phys. Rev. B* **54**, 2600 (1996).
- [76] R. TAMURA and M. TSUKADA, *Phys. Rev. B* **55**, 4991 (1997); **58**, 8120 (1998).
- [77] M. P. ANANTRAN and T. R. GOVINDAN, *Phys. Rev. B* **58**, 4882 (1998).
- [78] A. A. FARAJIAN, K. ESFARJANI, and Y. KAWAZOE, *Phys. Rev. Lett.* **82**, 5084 (1999).
- [79] H. J. CHOI and J. IHM, *Phys. Rev. B* **59**, 2267 (1999).
- [80] A. ROCHEFORT, F. LESAGE, D. R. SALAHUB, and P. AVOURIS, <http://xxx.lanl.gov/abs/cond-mat/9904083>, preprint (1999).