Linear combination of bulk bands method for large-scale electronic structure calculations on strained nanostructures

Lin-Wang Wang and Alex Zunger
National Renewable Energy Laboratory, Golden, Colorado 80401
(Received 3 February 1999)

A ''strained linear combination of bulk bands'' method is introduced for calculating the single-particle electronic states of strained, million-atom nanostructure systems, within an empirical pseudopotential Hamiltonian. This method expands the wave functions of a nanostructure (superlattice, wire, and dot) as linear combinations of bulk Bloch states of the constituent materials, over band indices n and wave vectors k. This allows one to use physical intuition in selecting the n and k that are most relevant for a given problem. This constitutes a useful approximation over the ''direct diagonalization'' approach where the basis is complete (individual plane waves) but unintuitive. It also constitutes a dramatic improvement upon the $\mathbf{k}\cdot\mathbf{p}$ approach, where the continuum model Hamiltonian is used, losing the atomistic details of the system. For a pyramidal InAs quantum dot embedded in GaAs, we find electronic eigenenergies that are within 20 meV of the exact direct diagonalization calculation, while the speed of the current method is 100–1000 times faster. The sublinear scaling of the current method with the size of the system enables one to calculate the atomistic electronic states of a million-atom system on a personal computer in about 10 h. Sufficient detail is provided in the formalism, so that the method can be promptly implemented. [S0163-1829(99)01024-3]

I. INTRODUCTION

Heteronanostructures such as superlattices,[1] embedded quantum wires,[2,3] or embedded quantum dots[4] today constitute the main platforms for electronic structure engineering.[5] This calls for accurate predictions of their electronic properties. The standard approaches to the problem are the effective-mass envelope-function approximation (EMA) and its multiband $\mathbf{k}\cdot\mathbf{p}$ generalization,[6,7] where the wave function $\psi(\mathbf{x})$ of the nanostructure is expanded in terms of zone-center ($\mathbf{k}=0$ or ''Γ point'') Bloch bands (indexed n) of the underlying periodic solid

$$\psi(\mathbf{x})=\frac{1}{\sqrt{N}}\sum_{n}^{N_B}\sum_{k}^{N_k} C'_{k,n}[u_{\Gamma,n}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}]\equiv\sum_{n}^{N_B} f_n(\mathbf{x})u_{\Gamma,n}(\mathbf{x}). \tag{1}$$

Here N_B is the number of bands included in the expansion, $u_{\Gamma,n}(\mathbf{x})$ is the periodic Bloch function, $\mathbf{k}$ is the wave vector, $f_n(\mathbf{x})$ is the envelope function, and N is the number of primary cells in the system. This method is exact if all ($N_B=\infty$) zone-center bulk bands are included. In practice, this approach has been implemented in the spirit of a model Hamiltonian, with a heavy truncation of the basis set (N_B), mitigated by extensive empirical parametrization of the Hamiltonian matrix. Retension of a single band ($N_B=1$) leads to the ''particle in a box'' description, whereas use of the highest occupied bulk valence band Γ_{15v} leads to the ''6×6 $\mathbf{k}\cdot\mathbf{p}$'' generalization, and addition of the lowest unoccupied bulk conduction band Γ_{1c} leads to the ''8×8 $\mathbf{k}\cdot\mathbf{p}$'' generalization, etc.

The simplicity of the model comes, however, at a cost that has been recognized only recently, when large basis set calculations of the *same* Hamiltonian became available for comparison.[8–10] These include the following.

(1) *The actual physical symmetry of the nanostructure is replaced in the EMA by an artificially higher-symmetry group*. This distorts qualitative features of the electronic structure. Examples include the following: (a) The omission of odd vs even oscillation in quantum films,[1,8] leading to an artificially monotonic energy-thickness curve. (b) The replacement of the true zincblende T_d symmetry of spherical dots by spherical symmetry, which artificially disallows s-p envelope function mixing.[11] (c) The replacement of the true C_{2v} symmetry of a square-based pyramidal dot by a $C_{4\nu}$ symmetry which misses the splitting of p states and the in-plane polarization anisotropy.[12]

(2) *The coupling of nanostructure states to X or L bulk bands is disallowed*. Thus Γ-to-X conversion of a quantum dot electron state as a function of pressure[13,14] or quantum dot size[15] cannot be described by the 8×8 $\mathbf{k}\cdot\mathbf{p}$ model, nor can the L character of some of the lowest dot electron states, or the X mixing in (001) GaAs/AlAs superlattices.[16]

(3) *The coupling of strain to the electron state is linearized*. This could lead to errors as large as a few hundred meV.

(4) *Interfaces are nonexistent or idealized*. For example, in superlattices lacking a common atom, such as $(\mathrm{InAs})_n/(\mathrm{GaSb})_m$ with (n,m) integers. The interface can either be Ga-As bonded or In-Sb bonded. They are treated as the same in $\mathbf{k}\cdot\mathbf{p}$ models. In addition, interdiffused and intermixed interfaces cannot be easily described by the $\mathbf{k}\cdot\mathbf{p}$ models.

(5) *Electon-hole interaction energies evaluated via envelope functions can be significantly inaccurate*. Within the $\mathbf{k}\cdot\mathbf{p}$ approach, Coulomb energies can be overestimated by up to 40% in a quantum dot,[17] while exchange energies can only be evaluated using a short-range analytical model which has been shown recently to be incorrect.[18]

However, there are ways of improving the conventional effective mass (or $\mathbf{k}\cdot\mathbf{p}$) models,[19] e.g., by adding a nonpara-

bolic kinetic energy term; adding a state mixing term at the interface,[20] introducing nonlinear strain effects, etc. The formalism can get very complicated, and the rapid increase in the number of parameters makes it quite intractable.

An alternative approach[21] to this problem, which is more along the lines of *ab initio* methods, is to avoid the decomposition of the wave function in Eq. (1) into envelope functions. Instead, the wave function is described by a variationally flexible plane wave basis, and the potential is described as a superposition of atomic potentials[22] $v_\alpha(\mathbf{r})$. This guarantees that the physical symmetry of the system is preserved, and affords an atomistic description of surfaces, interfaces, and strain. Since the basis set is not drawn from Γ-like Bloch states, an off-Γ character in the wave functions can be readily described. In this approach, the single-particle Schrödinger equation is constructed as

$$\left\{-\frac{1}{2}\nabla^2+\sum_{R,\alpha} W_\alpha(\mathbf{R})v_\alpha[\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha(R)]\right\}\psi(\mathbf{x}) \equiv H\psi(\mathbf{x})=\epsilon\psi(\mathbf{x}), \tag{2}$$

where $\psi(\mathbf{x})$ is the single-particle wave function, $\mathbf{R}$ denotes the position of the primary cell, and $\mathbf{d}_\alpha(R)$ is the displacement of atom of type α inside the primary cell $\mathbf{R}$. $v_\alpha(\mathbf{x})$ is a screened, spherical atomic pseudopotential that can be extracted from local density-approximation (LDA) calculations on bulk systems, and then adjusted empirically so as to fit the measured bulk band properties while keeping LDA quality wave functions.[22] In this work, we will deal with local pseudopotentials $v_\alpha(\mathbf{x})$ only. The weight function $W_\alpha(\mathbf{R})$ is used to denote whether an atom of type α occupies site $\mathbf{R}+\mathbf{d}_\alpha(R)$. The atomic positions $\{\mathbf{R}+\mathbf{d}_\alpha(R)\}$ are relaxed with respect to their ''ideal'' positions (e.g., perfect zinc-blende crystal structure) to minimize the strain energy. For strain-dependent pseudopotentials,[23,24] $W_\alpha(\mathbf{R})$ could deviate from 0 or 1, being instead a function of local strain. $\psi(\mathbf{x})$ is expanded in a fixed plane-wave basis set

$$\psi(\mathbf{x})=\sum_\kappa A(\kappa)e^{i\kappa\cdot\mathbf{x}}. \tag{3}$$

This approach (1) includes, via $\{\mathbf{R}+\mathbf{d}_\alpha(R)\}$, the true atomistic symmetry of the heterostructure; (2) is able to describe the multivalley (e.g., Γ-X) mixing *via* the ability to describe the whole Brillouin zone (BZ) Bloch states; (3) could be used to study very small nanostructures, beyond the region of parabolic approximation of the band structure, and beyond the region of linear strain dependence; (4) retains interfacial properties pertaining to the atomic structures; and (5) given the atomistic (i.e., nonenvelope) nature of the wave functions $\psi(\mathbf{x})$, this method can also be used to calculate explicitly the Coulomb interaction and the exchange interactions.[18,17]

This approach has been used to study quantum wells, superlattices, disordered superlattices, quantum wires, colloidal quantum dots, embedded pyramidal quantum dots, and composition modulations in alloys. Excellent agreements with the experiment have been obtained for single-particle energy levels,[11] exchange splittings,[18] optical-absorption spectra[25] and the magnitudes of Γ-X coupling.[13]

The disadvantage of the expansion of Eq. (3) is that it does not lend itself to systematic approximations: The plane-wave basis $\{e^{i\kappa\cdot\mathbf{x}}\}$ of Eq. (3) is classified only in terms of the kinetic energy $\frac{1}{2}\kappa^2$. In contrast, the EMA expansion of Eq. (1) is ''intuitive,'' in that one can classify the basis set in terms of physical Bloch functions $\{u_{\Gamma,n}e^{i\mathbf{k}\cdot\mathbf{x}}\}$ and include only those that are ''low energy.'' The consequence of this distinction is that one uses a large basis set of plane waves in the pseudopotential approach. Even though we employ a linear-in-size ''$O(N)$'' method,[26] treating dots made of 500 000 atoms requires more than 20 h of CPU time on a 128 nodes Cray T3E machine. In this paper, we refer to this direct solution of Eq. (2) as ''exact.''

Here we will present an approximate method to solve Eq. (2) by changing the basis of Eq. (3) to the basis of full zone *bulk Bloch states* [Eq. (4) below]. This approximate method can solve the same problem of Eq. (2) within 10 h on a personal computer. The accuracy of this method is typically 10–20 meV of the exact results.

This work is an extension to the ''linear combination of Bloch bands'' (LCBB) method, which is described briefly in Ref. 13. In the LCBB method, the wave function $\psi(\mathbf{x})$ in Eq. (2) is expanded in terms of full-zone Bloch states of the constituent bulk solids:

$$\phi^0_{k,n}(\mathbf{x})=\frac{1}{\sqrt{N}}u_{k,n}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \tag{4}$$

and

$$\psi(\mathbf{x})=\sum_n^{N_B}\sum_k^{N_k} C_{k,n}\phi^0_{k,n}(\mathbf{x}). \tag{5}$$

The system being calculated is a supercell (large box) periodic in all three dimensions. Again, N is the number of primary cells in the supercell, n is the band index, and k is the supercell reciprocal-lattice vector defined within the first BZ. The Hamiltonian matrix elements are evaluated within the basis set $\{\phi^0_{k,n}(\mathbf{x})\}$, and the resulting Hamiltonian matrix is diagonalized to yield $\{C_{k,n}\}$. There are two advantages to this approach.

(i) *The advantage relative to a plane-wave method* [Eq. (3)]: Unlike the plane wave expansion of $\psi(\mathbf{x})$ in Eq. (3), expansion in Eq. (5) allows one to select the physically important bands n and k points. As a result, the number of basis functions ($N_B\times N_k$) in Eq. (5) can be reduced significantly compared to the plane-wave basis. As a matter of fact (see Secs. III and IV), it is possible to use a fixed number of basis functions in Eq. (5) to achieve the same degree of accuracy for different system sizes. This is in direct contrast to the plane-wave expansions, where the number of basis functions scales linearly with the size of the system. This sublinear scaling makes the LCBB method much faster than the $O(N)$ methods (e.g., the folded spectrum method[26]) for very large systems.

(ii) *The advantage relative to the* $\mathbf{k}\cdot\mathbf{p}$ *method* [Eq. (1)]: In the conventional $\mathbf{k}\cdot\mathbf{p}$ method of Eq. (1), the Bloch function $u_{k,n}(\mathbf{x})$ of Eq. (4) is replaced by the zone center (Γ) part $u_{\Gamma,n}(\mathbf{x})$. As a result, one can describe the band structure only near the Γ point.[8,10] Furthermore, the Hamiltonian has to be changed from Eq. (2) to a parametrized form so as to yield the correct effective masses. As a result, the $\mathbf{k}\cdot\mathbf{p}$ method is detached from the atomistic origins of Eq. (2), and serves

only as a continuum model for the envelope functions. In contrast, the LCBB method (1) retains the full atomistic structure of the wave functions, enabling the use of the Hamiltonian of Eq. (2) without parametrization; Eq. (2) describes the bulk band structure correctly throughout the BZ, *even if only a single band* ($N_B=1$) *is used in Eq. (5)*. Thus the method can be used to describe Γ-X couplings, nonparabolic effects around the Γ point, interfacial anisotropy, and interface-induced state couplings.

The formulation of the LCBB method described in Ref. 13 does not address strain, and thus pertains to unstrained heterostructures, such as GaAs/AlAs. In reality, most semiconductor heterostructures involve some degree of strain, e.g., InAs/GaAs and InP/GaP. Strain can have a significant effect (up to a few hundred meV) on the electronic structure of the system. In this paper, we will extend the previous LCBB method to describe strained systems; thus we call it the ''strained LCBB'' or SLCBB method. This extension expands the usability of this approach from very limited cases (e.g., GaAs/AlAs) to all possible semiconductor heterostructures, and thus constitutes a major improvement on this approach.

Recall that the description of strain effects in the $\mathbf{k}\cdot\mathbf{p}$ model requires as input the continuum elasticity strain profile of the system, and various deformation potential parameters. They are used to construct an explicit linearly strain dependent Hamiltonian. On the other hand, in our atomistic approach, all we need to describe strain effects is to solve the Schrödinger equation (2) when $\{\mathbf{R}+\mathbf{d}_\alpha(R)\}$ are not at the ideal zinc-blende positions. Thus $\{\mathbf{R}+\mathbf{d}_\alpha(R)\}$ and v_α are the only inputs of our method. Explicit deformation-potential parameters and continuum elasticity strain profiles are not needed.

II. FORMALISM

A. Strain-free LCBB formalism

Since the LCBB method is only briefly discussed in Ref. 13, we will review its formalism here, and will provide more discussions on its selection of basis sets and computational details in Secs. III and V, along with the discussions of the SLCBB method. In the LCBB method, the atoms are assumed to be in their unrelaxed ideal zinc-blende positions $\{\mathbf{R}_0+\mathbf{d}_\alpha^0\}$. The nanostructure is placed in a periodic supercell with cell edge vectors $\{\mathbf{L}_1^0,\mathbf{L}_2^0,\mathbf{L}_3^0\}$ and reciprocal-lattice vectors $\{\mathbf{k}\}$. The zinc-blende reciprocal-lattice vectors are denoted as $\{\mathbf{G}\}$. While $\mathbf{k}$ is restricted to the first Brillouin zone of the zinc-blende reciprocal lattice, all $\mathbf{G}$ except $\mathbf{G}=0$ are outside the first BZ. Each spatial point inside the supercell is described by its Cartesian coordinate $\mathbf{x}$.

As described in Eq. (5), the nanostructure wave function is expanded by the constituent bulk Bloch states $\phi_{k,n}^0(\mathbf{x})$. The periodic part $u_{k,n}(\mathbf{x})$ of the Bloch function is described by the plane wave functions as

$$u_{k,n}(\mathbf{x})=\frac{1}{\sqrt{V_0}}\sum_{G}^{N_G} A_{k,n}(\mathbf{G})e^{i\mathbf{G}\cdot\mathbf{x}}, \tag{6}$$

where N_G is the number of zinc-blende reciprocal-lattice vectors $\mathbf{G}$ within an energy cutoff. Using this basis set $\{\phi_{k,n}^0\}$ of Eq. (4), the Hamiltonian matrix elements are

$$\langle\phi_{k',n'}^0|\hat{H}|\phi_{k,n}^0\rangle=\Omega_0\sum_{\mathbf{G},\mathbf{G}'}A_{\mathbf{k}',n'}(\mathbf{G}')\left[\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2\delta_{\mathbf{k},\mathbf{k}'}\delta_{\mathbf{G},\mathbf{G}'}+\sum_\alpha V_\alpha(|\mathbf{k}+\mathbf{G}-\mathbf{k}'-\mathbf{G}'|^2)e^{-i\mathbf{d}_\alpha^0\cdot(\mathbf{k}+\mathbf{G}-\mathbf{k}'-\mathbf{G}')}W_\alpha^0(\mathbf{k}-\mathbf{k}')\right]A_{\mathbf{k},n}(\mathbf{G}), \tag{7}$$

where Ω_0 is the supercell volume and $V_\alpha(q_2)$ is the Fourier transform of $v_\alpha(\mathbf{x})$:

$$V_\alpha(q_2)=\frac{1}{V_0}\int v_\alpha(|\mathbf{x}|)e^{i\mathbf{q}\cdot\mathbf{x}}d^3\mathbf{x}, \tag{8}$$

where $q_2=|\mathbf{q}|^2$ and $V_0=\Omega_0/N$ is the volume of a single primary cell, and $W_\alpha^0(\mathbf{k})$ is a structure factor

$$W_\alpha^0(\mathbf{k})=\frac{1}{\Omega_0}\sum_{\mathbf{R_0}}W_\alpha(\mathbf{R_0})e^{i\mathbf{k}\cdot\mathbf{R_0}}. \tag{9}$$

The unapproximated equation (7) can be evaluated efficiently if $W_\alpha^0(\mathbf{k})$ is available. The calculation of $W_\alpha^0(\mathbf{k})$ is facilitated by the fact that the atoms $\mathbf{R}_0$ are on a regular grid, so fast Fourier transformation (FFT) can be used in Eq. (9). Using this technique, the evaluation of the matrix elements in Eq. (7) requires $\propto(N_B\times N_k\times N_G)^2$ operations. As will be discussed in Sec. III, the number of k points N_k can be kept the same independent of system size; only a few bands N_B are needed in the basis set, and N_G is a constant for a given plane-wave cutoff energy. Consequently, the evaluation of Eq. (7) is fast; the effort is independent of the system's size. A direct diagonalization of the Hamiltonian matrix $\langle\phi_{k',n'}^0|\hat{H}|\phi_{k,n}^0\rangle$ yields the coefficients $C_{k,n}$ in Eq. (5) and the eigenenergy ϵ. This diagonalization step takes $\propto(N_B\times N_k)^3$ operations, which in practice is comparable to the evaluation of the matrix elements. The strain-free LCBB method provides a fast and accurate way to obtain the eigensolutions of unrelaxed nanostructures.

B. SLCBB formalism for strained systems

Reference system: Starting with an unrelaxed system $\{\mathbf{L}_1^0,\mathbf{L}_2^0,\mathbf{L}_3^0\}$ and $\{\mathbf{R}_0+\mathbf{d}_\alpha^0\}$ as a reference system for our SLCBB, we let the atoms and the supercell shape change, minimizing the total elastic energy of the system (using, e.g., an atomistic valence force field model[30]). Let the relaxed

atomic positions be denoted as $\{\mathbf{R}+\mathbf{d}_\alpha(R)\}$, and the relaxed supercell edge vectors be denoted as $\{\mathbf{L}_1,\mathbf{L}_2,\mathbf{L}_3\}$. One can, for example, take the anion positions as $\{\mathbf{R}\}$; then each cation has a displacement $\{\mathbf{d}_\alpha(R)\}$ (the displacements for anions is zero). Notice that, in the unrelaxed Bloch function $\phi^0_{k,n}(\mathbf{x})$ of the reference system, the periodic function $u_{k,n}(\mathbf{x})$ is commensurate with the unrelaxed atomic positions $\{\mathbf{R_0}+\mathbf{d^0_\alpha}\}$, not with the displaced atomic positions $\{\mathbf{R}+\mathbf{d}_\alpha(R)\}$. As a result, the Bloch functions of the unrelaxed system form a poor basis set for the relaxed system. What is needed here is to shift the Bloch wave functions along with the relaxation-induced shift in the atomic positions. This can be done by mapping a point, described by $\mathbf{x}$ in the relaxed system, to a corresponding point $\mathbf{x^0}$ in the unrelaxed reference system: $\boldsymbol{\mu}(\mathbf{x})\equiv\mathbf{x^0}$. One can imagine obtaining this mapping by considering a rubber sheet (with a grid on it) wrapped on the unrelaxed supercell, and letting the rubber sheet deform following the displacements of the anions. Thus, we have the mapping $\boldsymbol{\mu}(\mathbf{R})=\mathbf{R}_0$, and $\boldsymbol{\mu}(\mathbf{L}_j)=\mathbf{L}^0_j$ for $j=1$, 2, and 3. The unrelaxed Bloch function $\phi^0_{k,n}(\mathbf{x})$ will be replaced by $\phi^0_{k,n}[\boldsymbol{\mu}(\mathbf{x})]$ as the basis function.

Strain-deformed Bloch basis set: Let us define a deformed plane wave function as

$$|\mathbf{k}+\mathbf{G}\rangle\equiv\sqrt{\frac{J(\mathbf{x})}{\Omega_0}}e^{i(\mathbf{k}+\mathbf{G})\cdot\boldsymbol{\mu}(\mathbf{x})}, \tag{10}$$

where Ω_0 is the volume of the reference supercell, and $J(\mathbf{x})$ is the Jacobian of the map $\boldsymbol{\mu}(\mathbf{x})$, so that

$$d^3\boldsymbol{\mu}=J(\mathbf{x})d^3\mathbf{x}. \tag{11}$$

Using Eq. (10), the orthogonality condition is

$$\langle\mathbf{k}+\mathbf{G}|\mathbf{k}'+\mathbf{G}'\rangle=\delta_{\mathbf{k},\mathbf{k}'}\delta_{\mathbf{G},\mathbf{G}'}, \tag{12}$$

so the deformed plane waves form an orthonormal set. Suppose that the unrelaxed Bloch basis $\phi^0_{k,n}(\mathbf{x})$ of Eq. (4) has the periodic part $u_{k,n}(\mathbf{x})$ defined by Eq. (6). Then the ''deformed Bloch function'' in the relaxed system can be defined as

$$|\phi_{k,n}\rangle=\sqrt{J(\mathbf{x})}\phi^0_{k,n}[\boldsymbol{\mu}(\mathbf{x})]=\sum_{\mathbf{G}}^{N_G}A_{k,n}(\mathbf{G})|\mathbf{k}+\mathbf{G}\rangle, \tag{13}$$

where $A_{k,n}(\mathbf{G})$ is defined in Eq. (6) for a bulk system. The index n now denotes collectively (i) the band index, (ii) the identity of the constituent bulk materials (e.g., GaAs and InAs), and (iii) the bulk strains used to calculate $A_{k,n}(\mathbf{G})$ in Eq. (6) (e.g., unstrained InAs, hydrostatically compressed InAs, or uniaxially strained InAs). The Bloch functions obtained directly from Eqs. (6) and (13) for different bands, materials, and strains are not mutually orthogonal at a given $\mathbf{k}$. We will thus explicitly orthogonalize these basis functions at each $\mathbf{k}$ point (using the Gram-Schmidt orthogonalization scheme), so that the resulting $|\phi_{k,n}\rangle$'s form an orthonormal basis set:

$$\langle\phi_{k,n}|\phi_{k',n'}\rangle=\sum_{G}^{N_G}A^*_{k,n}(\mathbf{G})A_{k,n'}(\mathbf{G})\delta_{\mathbf{k},\mathbf{k}'}=\delta_{n,n'}\delta_{\mathbf{k},\mathbf{k}'}. \tag{14}$$

As in the case of the strain-free LCBB of Eq. (5), the eigenstate $\psi(\mathbf{x})$ of the deformed system will be expanded by the deformed Bloch functions $|\phi_{k,n}\rangle$

$$\psi(\mathbf{x})=\sum_{n}^{N_B}\sum_{k}^{N_k}C_{k,n}|\phi_{k,n}\rangle. \tag{15}$$

Hamiltonian matrix in a deformed Bloch basis: To solve for the coefficients $\{C_{k,n}\}$, we need to evaluate the matrix elements of the Hamiltonian H of Eq. (2) using the basis set $|\phi_{k,n}\rangle$, then diagonalize the ensuing Hamiltonian matrix.

We will break the Hamiltonian H into the kinetic energy part $H_1=-\frac{1}{2}\nabla^2$ and the potential energy part $H_2=\Sigma_{R,\alpha}W_\alpha(R)v_\alpha[\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha(R)]$, and evaluate their matrix elements separately. For the kinetic-energy part, we have

$$\begin{aligned}\langle\phi_{k,n}|H_1|\phi_{k',n'}\rangle&=\frac{1}{2}\sum_{G,G'}A^*_{k,n}(G)A_{k',n'}(G')\int\frac{J(\mathbf{x})}{\Omega_0}[\nabla e^{-i(\mathbf{k}+\mathbf{G})\cdot\boldsymbol{\mu}(\mathbf{x})}]\cdot[\nabla e^{i(\mathbf{k}'+\mathbf{G}')\cdot\boldsymbol{\mu}(\mathbf{x})}]d^3\mathbf{x}\\&=\frac{1}{2}\sum_{G,G'}A^*_{k,n}(G)A_{k',n'}(G')\frac{1}{\Omega_0}\int[(\mathbf{k}+\mathbf{G})\cdot\hat{\boldsymbol{\epsilon}}(\mathbf{x})]\cdot[(\mathbf{k}'+\mathbf{G}')\cdot\hat{\boldsymbol{\epsilon}}(\mathbf{x})]e^{i(\mathbf{k}'-\mathbf{k}+\mathbf{G}-\mathbf{G}')\cdot\boldsymbol{\mu}(\mathbf{x})}d^3\boldsymbol{\mu}\\&=\frac{1}{2}\sum_{G,G'}A^*_{k,n}(G)A_{k',n'}(G')\delta_{G,G'}(\mathbf{k}+\mathbf{G})\cdot\left[\frac{1}{N}\sum_{R_0}\hat{\boldsymbol{\epsilon}}(R_0)\cdot\hat{\boldsymbol{\epsilon}}(R_0)e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{R}_0}\right]\cdot(\mathbf{k}'+\mathbf{G}').\end{aligned} \tag{16}$$

Here, the strain tensor $\hat{\boldsymbol{\epsilon}}(\mathbf{x})$ is defined as $\partial\mu/\partial\mathbf{x}$. In Eq. (16), we have made two approximations. First, we have ignored the gradient operating on the Jacobi function $J(\mathbf{x})$. This ''smooth $\boldsymbol{\epsilon}(\mathbf{x})$ approximation'' is valid when the change in the strain is smooth in space. Second, we have broken down the full space integral $d^3\boldsymbol{\mu}$ into the integral inside one primary cell and the summation Σ_{R_0} over the primary cells. For the integral inside one primary cell, we have only considered the variation of $e^{i(\mathbf{G}-\mathbf{G}')\cdot\boldsymbol{\mu}}$ in the integrand, and thus obtained $\delta_{G,G'}$. Now, if we define

$$\hat{\sigma}(\mathbf{k})\equiv\frac{1}{N}\sum_{R_0}\hat{\boldsymbol{\epsilon}}(R_0)\cdot\hat{\boldsymbol{\epsilon}}(R_0)e^{i\mathbf{k}\cdot\mathbf{R_0}}, \tag{17}$$

we have

$$\langle \phi_{k,n}|H_1|\phi_{k',n'}\rangle = \frac{1}{2}\sum_G A^*_{k,n}(G)A_{k',n'}(G)(\mathbf{k}+\mathbf{G})\cdot\hat{\sigma}(\mathbf{k}'-\mathbf{k})\cdot(\mathbf{k}'+\mathbf{G}). \tag{18}$$

Next we will evaluate the matrix element for the potential part H_2. Let us first derive the formula for the matrix element in the deformed plane-wave basis function of Eq. (10):

$$\begin{aligned}\langle \mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle &= \frac{1}{\Omega_0}\int J(\mathbf{x})\sum_{R,\alpha} v_\alpha(\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha)W_\alpha(R)e^{i(\mathbf{k}'+\mathbf{G}'-\mathbf{k}-\mathbf{G})\cdot\mu(\mathbf{x})}d^3\mathbf{x}\\ &= \frac{1}{N}\sum_{R_0,\alpha} J(R_0)W_\alpha(R_0)e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_0}\frac{1}{V_0}\int v_\alpha(\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha)e^{i\mathbf{q}\cdot[\mu(\mathbf{x})-\mathbf{R}_0]}d^3\mathbf{x}\\ &= \frac{1}{N}\sum_\alpha e^{i\mathbf{q}\cdot\mathbf{d}^0_\alpha}\sum_{R_0} J(R_0)W_\alpha(R_0)e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_0}e^{i\mathbf{q}\cdot[\hat{\epsilon}(R_0)\cdot\mathbf{d}_\alpha(R_0)-\mathbf{d}^0_\alpha]}\\ &\quad\times\frac{1}{V_0}\int v_\alpha(\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha)e^{i\mathbf{q}\cdot\hat{\epsilon}(R_0)\cdot(\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha)}d^3(\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha).\end{aligned} \tag{19}$$

Here, $V_0=\Omega_0/N$ is the volume of one primary cell. In Eq. (19), we have used $\mathbf{q}\equiv\mathbf{k}'+\mathbf{G}'-\mathbf{k}-\mathbf{G}$ and $\hat{\epsilon}=\partial\mu/\partial\mathbf{x}$. Again, we have ignored the change of $J(\mathbf{x})$ within one primary cell. We have also used $\mu(\mathbf{x})-\mathbf{R}_0\doteq\hat{\epsilon}(R_0)\cdot(\mathbf{x}-\mathbf{R})$. This is valid for small $\mathbf{x}-\mathbf{R}$, which is the case in the integral of $d^3(\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha)$, where the pseudopotential $v_\alpha(\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha)$ decays rapidly with increasing $\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha$. Also notice that, we have changed $W_\alpha(R)$ to $W_\alpha(R_0)$, since, as indices, R and R_0 refer to the same primary cell. Now, let us define

$$\Delta\mathbf{d}_\alpha(R_0)=\hat{\epsilon}(R_0)\cdot\mathbf{d}_\alpha(R_0)-\mathbf{d}^0_\alpha. \tag{20}$$

Using Eqs. (20) and (8), Eq. (19) can be rewritten as

$$\langle \mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle = \frac{1}{N}\sum_\alpha e^{i\mathbf{q}\cdot\mathbf{d}^0_\alpha}\sum_{R_0} J(R_0)W_\alpha(R_0)e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_0}e^{i\mathbf{q}\cdot\Delta\mathbf{d}_\alpha(R_0)}V_\alpha(|\mathbf{q}\cdot\hat{\epsilon}(R_0)|^2). \tag{21}$$

A direct evaluation of the matrix elements formed by the sum in Eq. (21) requires a large number of operations $\propto N\times(N_k\times N_G)^2$, where N is the number of the primary cells in the system. The number of plane waves N_G is typically $\sim$60 when a 5–7 Ry plane-wave cutoff energy is used. The number of k points N_k is typically around 1000. But the number of primary cells N is of the order of millions. Thus a direct evaluation of Eq. (21) is impractical. We thus prefer to simplify Eq. (21). Notice that $\Delta\mathbf{d}_\alpha(R_0)$ denotes the internal displacement of the cation atom from its ''ideal'' position ($\hat{\epsilon}^{-1}\cdot\mathbf{d}^0_\alpha$), thus $\mathbf{q}\cdot\Delta\mathbf{d}_\alpha(R_0)$ is usually much smaller than 1. As a result, we can expand $e^{i\mathbf{q}\cdot\Delta\mathbf{d}_\alpha(R_0)}$ into $1+i\mathbf{q}\cdot\Delta\mathbf{d}_\alpha(R_0)$. Let us also define $\hat{\chi}(R_0)=\hat{\epsilon}(R_0)\cdot\hat{\epsilon}(R_0)-\hat{I}$, then $V_\alpha(|\mathbf{q}\cdot\hat{\epsilon}(R_0)|^2)=V_\alpha(q^2+\mathbf{q}\cdot\hat{\chi}(R_0)\cdot\mathbf{q})$. Assuming the strain is small, then $\hat{\chi}(R_0)$ is small; thus we can expand V_α to second order in $\mathbf{q}\cdot\hat{\chi}(R_0)\cdot\mathbf{q}$. Substituting these expansions into Eq. (21), we have

$$\begin{aligned}\langle \mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle = \frac{1}{N}\sum_\alpha e^{i\mathbf{q}\cdot\mathbf{d}^0_\alpha}\sum_{R_0} J(R_0)W_\alpha(R_0)e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_0}\Big[V_\alpha(q^2)+iV_\alpha(q^2)\mathbf{q}\cdot\Delta\mathbf{d}_\alpha(R_0)\\ +V'_\alpha(q^2)\mathbf{q}\cdot\hat{\chi}(R_0)\cdot\mathbf{q}+\frac{1}{2}V''_\alpha(q^2)\sum_{i1,\ldots,i4} q_{i1}q_{i2}q_{i3}q_{i4}\chi_{i1,i2}(R_0)\chi_{i3,i4}(R_0)\Big],\end{aligned} \tag{22}$$

where $V'_\alpha(q_2)=dV_\alpha(q_2)/dq_2$ and $V''_\alpha(q_2)=d^2V_\alpha(q_2)/dq_2^2$. Each index of $i1,\ldots,i4$ runs through the three dimensions of vector $\mathbf{q}$.

Now let us define a few structure factors:

$$\begin{aligned}W^\alpha(\mathbf{k}) &= \frac{1}{N}\sum_{R_0} J(R_0)W_\alpha(R_0)e^{-i\mathbf{k}\cdot\mathbf{R}_0},\\ \mathbf{W}^\alpha_d(\mathbf{k}) &= \frac{1}{N}\sum_{R_0} J(R_0)W_\alpha(R_0)\Delta\mathbf{d}_\alpha(R_0)e^{-i\mathbf{k}\cdot\mathbf{R}_0},\\ \hat{W}^\alpha_\chi(\mathbf{k}) &= \frac{1}{N}\sum_{R_0} J(R_0)W_\alpha(R_0)\hat{\chi}(R_0)e^{-i\mathbf{k}\cdot\mathbf{R}_0},\\ W^\alpha_{i1,i2,i3,i4}(\mathbf{k}) &= \frac{1}{N}\sum_{R_0} J(R_0)W_\alpha(R_0)\chi_{i1,i2}(R_0)\chi_{i3,i4}(R_0)e^{-i\mathbf{k}\cdot\mathbf{R}_0}.\end{aligned} \tag{23}$$

Substituting these structure factors into Eq. (22), we obtain

$$\langle \mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle = \sum_{\alpha} e^{i\mathbf{q}\cdot\mathbf{d}_{\alpha}^{0}}\Big[V_{\alpha}(q^2)W^{\alpha}(\mathbf{k}-\mathbf{k}') + iV_{\alpha}(q^2)\mathbf{q}\cdot\mathbf{W}_{d}^{\alpha}(\mathbf{k}-\mathbf{k}') + V'_{\alpha}(q^2)\mathbf{q}\cdot\hat{W}_{\chi}^{\alpha}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{q} + \frac{1}{2}V''_{\alpha}(q^2)\sum_{i1,\ldots,i4} q_{i1}q_{i2}q_{i3}q_{i4}W_{i1,i2,i3,i4}^{\alpha}(\mathbf{k}-\mathbf{k}')\Big]. \tag{24}$$

The replacement of Eq. (21) by Eq. (24) leads to a speedup by 10^4, as will be discussed in Sec. IV. Finally, the matrix elements in the deformed Bloch function basis can be expressed in terms of $\langle \mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle$,

$$\langle \phi_{k,n}|H_2|\phi_{k',n'}\rangle = \sum_{G,G'} A_{k,n}^{*}(G)A_{k',n'}(G')\langle \mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle. \tag{25}$$

In our pseudopotential calculation, we have used[22] a smooth plane-wave cutoff function $\omega(|\mathbf{k}+\mathbf{G}|^2)$. This is necessary because we have a small number of plane waves ($N_G \sim 60$) for each primary cell. If $\omega(|\mathbf{k}+\mathbf{G}|^2)$ is not used, the bulk band structure as a function of $\mathbf{k}$ could be discontinuous due to sudden changes of the number of plane waves as the k point changes. The use of $\omega(|\mathbf{k}+\mathbf{G}|^2)$ smooths out this discontinuity. However, the use of $\omega(|\mathbf{k}+\mathbf{G}|^2)$ also introduces additional terms in our formulas above. For a detailed description of these additional terms, see the Appendix. Notice that, althought we started the derivation of our formalism by introducing the maping $\mu(\mathbf{x})$, at the end, in Eq. (24) [or Eq. (A6)], we only need $\hat{\epsilon}(R_0)$ and $J(R_0)$, not the detailed maping $\mu(\mathbf{x})$.

III. DISCUSSIONS OF APPROXIMATIONS

In contrast to the exact solutions of Eq. (2), we have used several approximations in our SLCBB approach. The first type of approximations concerns the evaluation of the matrix elements in Eqs. (18), (24), and (25). Because of these approximations, for a given basis set [Eq. (13)], the calculated matrix elements from Eqs. (18), (24), and (25) are not exact. The ''smooth $\epsilon(\mathbf{x})$ approximation'' and ''Taylor expansion approximation'' belong to this category. The second type of approximations concerns the truncation of the basis sets. Instead of using a complete basis set, we have taken advantage of the fact that, in our Eq. (15), the nanostructure wave functions are expanded in the physically meaningful basis set of bands n and wave vectors k. We have thus truncated the basis set to include only physically relevant $\{n,k\}$.

A. Smooth $\epsilon(\mathbf{x})$ approximation

A major approximation we made in deriving Eqs. (18) and (24) is that we have ignored the change of $\hat{\epsilon}(\mathbf{x})$ [hence $J(\mathbf{x})$] within each primary cell. This is the ''smooth $\hat{\epsilon}(\mathbf{x})$ approximation.'' It is because of this useful approximation that we do not need to know the detailed function $\mu(\mathbf{x})$, but instead we need only to know $J(\mathbf{R}_0)$ and $\hat{\epsilon}(\mathbf{R}_0)$ at each primary cell $\mathbf{R}_0$. This ''smooth $\hat{\epsilon}(\mathbf{x})$ approximation'' is expected to be accurate in the region away from the interface, but near the interface, this approximation could in principle introduce some errors. However, our numerical tests of the SLCBB results against the exact results in Sec. V indicate that such errors are small. One possible reason for this is that the wave functions are not localized around the interface.

B. Taylor expansion approximation

In deriving Eq. (24) from Eq. (21), we have used a Taylor expansion of $e^{i\mathbf{q}\cdot\Delta\mathbf{d}_{\alpha}(R_0)}$ in powers of $\mathbf{q}\cdot\Delta\mathbf{d}_{\alpha}(R_0)$, and a Taylor expansion of $V_{\alpha}(q^2+\mathbf{q}\cdot\hat{\chi}(R_0)\cdot\mathbf{q})$ around q^2. These truncated expansions can lead to errors. These approximations exist even when the strain is uniform as long as $\Delta\mathbf{d}_{\alpha}$ and $\chi(R_0)$ are not zero. Notice that we have expanded $V_{\alpha}[q^2+\mathbf{q}\cdot\hat{\chi}(R_0)\cdot\mathbf{q}]$ to the second order of $\hat{\chi}(R_0)$. That means, for a major part, we have retained the second-order effects of the strain. To see how large the error resulting from the ''Taylor expansion approximation'' is, Fig. 1 shows the SLCBB eigenenergies vs strain for bulk InAs compared with the exact result. The SLCBB was calculated with an unrelaxed reference system $\{\mathbf{R}_0,\mathbf{k}\}$ defined as the $x=0$ point in the figure. We consider three types of strains (A, B, and C) taken from an elastic calculation (Sec. V) at three locations (see the inset to Fig. 1) inside an InAs pyramidal dot. We see that, with our (semi-) second-order expansion, the bulk electronic states of the different strain regions in the quantum dot have been reproduced within an error of ~ 10 meV.

C. Choice of basis sets

1. Unstrained system

We first discuss the convergence of the LCBB method with regard to the number of bulk bands n and wave vectors k for the unstrained AlAs/GaAs system. The issue is how many bands and k points should be included in the basis set to obtain accurate results. Notice that, in this unstrained heterostructure, the approximations of matrix element evaluation discussed in Secs. III A and III B do not exist. Indeed, the matrix elements calculated via Eq. (7) are exact. Thus the only error comes from the limited basis set. As shown in Fig. 2, when six GaAs bands and six AlAs bands are included in the basis set, the electron eigenenergies for $(\mathrm{GaAs})_m/(\mathrm{AlAs})_m$ superlattices are within 2 meV of the exact calculations for m down to 1 ML. When the number of bands is reduced to two for each k point, the eigenenergy

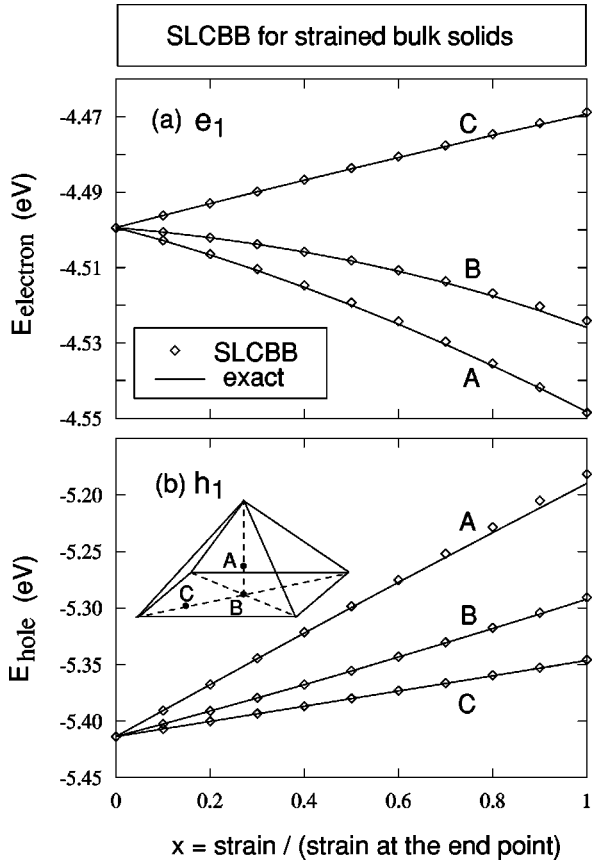

FIG. 1. SLCBB calculations for bulk InAs under different strains. The end point ($x=1$) strains for *A*, *B*, and *C* are taken from the interior of the pyramidal quantum dot studied in Sec. V. *A* is at the center of the pyramid, *B* is at the center of the base of the pyramid, and *C* is shifted from point *B* along the (110) direction, one-third of the distance away from the corner of the pyramid. The unrelaxed reference system ($x=0$) has a lattice constant of $a=5.879$ Å. The basis functions used in the SLCBB calculation for the lowest electron state e_1 and the highest hole state h_1 are described in Sec. V. The strains at *A*,*B*,*C* points, 3.4% relative to the reference system of $x=0$ (not to the nature bulk InAs), are (*A*) $\mathrm{Tr}(\epsilon)-3\simeq 3.4\%$, $\epsilon_{xx}\simeq\epsilon_{yy}\simeq\epsilon_{zz}$, $\epsilon_{xy}\simeq\epsilon_{yz}\simeq\epsilon_{xz}\simeq 0$; (*B*) $\mathrm{Tr}(\epsilon)-3\simeq 0$, $\epsilon_{zz}-\epsilon_{xx}(=\epsilon_{yy})\simeq 7\%$, $\epsilon_{xy}\simeq\epsilon_{yz}\simeq\epsilon_{xz}\simeq 0$; (*C*) $\mathrm{Tr}(\epsilon)-3\simeq 0.3\%$, $\epsilon_{zz}-\epsilon_{xx}(=\epsilon_{yy})\simeq 8\%$, $\epsilon_{xz}\simeq\epsilon_{yz}\simeq 2\%$, $\epsilon_{xy}\simeq 1\%$.

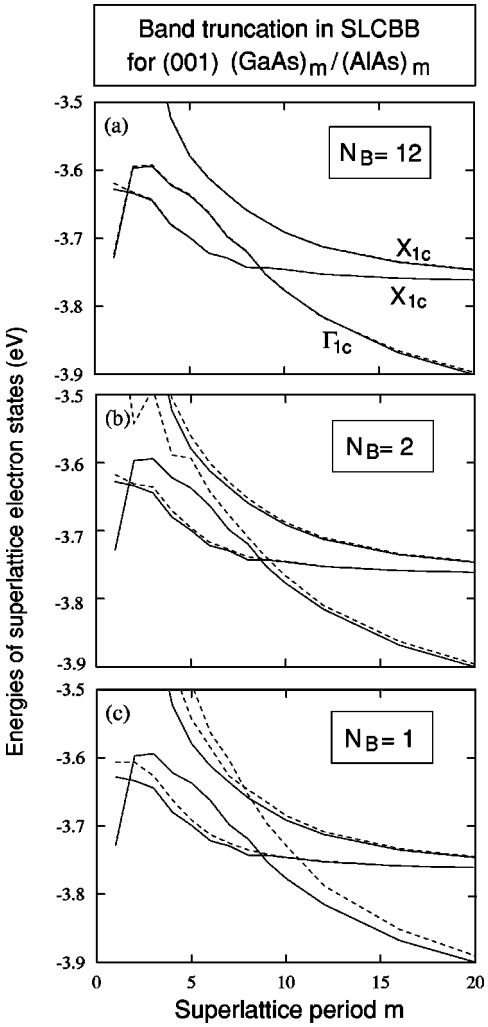

FIG. 2. The effects of the number of bulk band basis functions used in the SLCBB calculations of (001) $(GaAs)_m/(AlAs)_m$ superlattices. The solid lines are the exact results, and the dashed lines are the SLCBB results. (a) The number of bands N_B used in the basis set equals 12: six from GaAs and six from AlAs. (b) $N_B=2$, from the fifth and sixth bands of GaAs for *k* near the Γ point, and from the fifth and sixth bands of AlAs for *k* near the *X* points. (c) $N_B=1$, from the fifth band of GaAs for *k* near the Γ point, and the fifth band of AlAs for *k* near the *X* points. 24 *k* points are used in the basis set, 13 near the Γ point, and 11 near the *X* points.

error increases to ~ 5 meV for $m>10$ ML. For large superlattices, e.g., $m=20$ ML, including only one band in the basis set is enough to obtain the electron energies within 10 meV. Thus, for the large nanostructures for which this method is designed, just a few (all-zone) bands should be enough to describe the electron states. Usually, because of the degeneracy of the bulk bands near the Γ point, more bulk band basis functions are needed to converge the nanostructure hole states than the electron states. Note further that the hole states and electron states of a given nanostructure can be calculated separately using different basis sets.

As for the choice of *k* points in the basis, one usually places them in regions near special *k* points, e.g., the Γ point, the *X* point, and the *L* point. Here one can exercise physical judgment to achieve economy. For example, when studying a Γ-*X* coupling, as in Ref. 13, one needs to include regions both near the Γ and *X* points, while in studying a state primarily derived from the bulk Γ point, as in Sec. V, we only need to include *k* points near the Γ point. An important feature of the LCBB and SLCBB methods is that for simple shaped nanostructures, when the size of the system increases, the number N_k of the *k* points needs not to be increased. The basic idea is that, when the size of the system increases, the envelope function of the electronic state becomes smoother. As a result, the maximum value of the *k* vector needed to describe the envelope function becomes smaller. This results in the same number of *k* points in the basis set, and is thus independent of the system size. This is demonstrated in Fig. 3. The exact lowest electron state of the (001) $(GaAs)_m/(AlAs)_m$ superlattice is analyzed using Eq. (5). The sum of coefficients $\Sigma_n|C_{k,n}|^2$ is plotted as a function of $\mathbf{k}$ and the superlattice period *m*. As *m* increases, we have more and more $\mathbf{k}$ points within the BZ. However, if we choose $\mathbf{k}$ according to a magnitude criterion of $\Sigma_n|C_{k,n}|^2$,

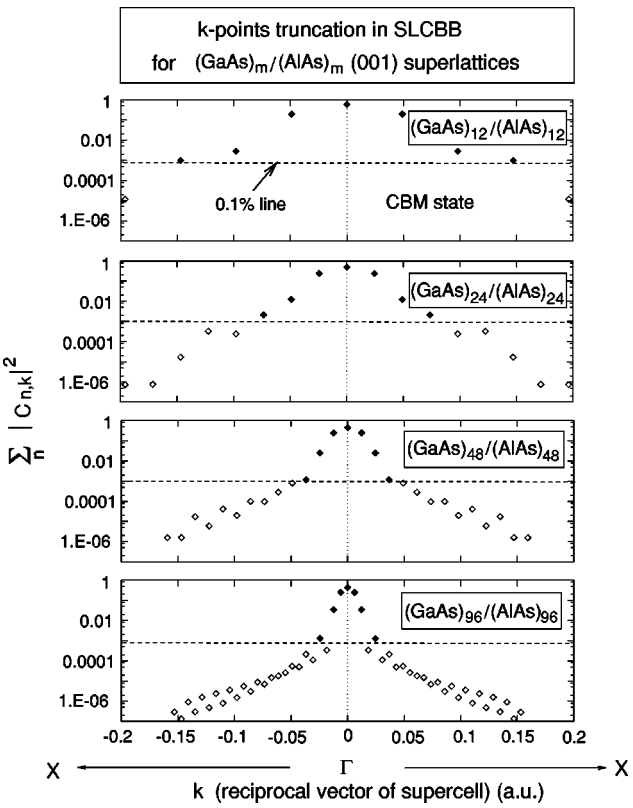

FIG. 3. The selection of k points in the basis set and its size dependence for $(GaAs)_m/(AlAs)_m$ superlattices. Exact electron wave functions $\psi(\mathbf{x})$ of the superlattices are analyzed using Eq. (15) [equivalent to Eq. (5) in this case]. The basis function contribution to the wave function from each k point $[\Sigma_n|C_{n,k}|^2]$ is plotted here as a function of k and the superlattice period m. If only the k points with $\Sigma_n|C_{n,k}|^2>0.001$ are chosen, the number of k points N_k is independent of superlattice size m. Thus one can use the same number of basis functions in SLCBB for different size systems.

e.g., $\Sigma_n|C_{k,n}|^2>1\times10^{-3}$ (above the dashed line in Fig. 3), the number of $\mathbf{k}$ points is almost independent of m. This important feature of the LCBB and SLCBB methods makes their computational efforts scale almost as $O(N^0)$ (see Sec. IV). Thus these methods are suited ideally to studying very large systems.

2. Strained systems

To calculate the effects of strain on the eigenenergy to second order requires that (i) the Hamiltonian matrix elements should be evaluated correctly to second order [as we did in Eq. (24) and discussed in Sec. III B], and (ii) the wave function should be correct to first order in strain. This means that the basis set should have sufficient variational flexibility to allow the wave function to change with strain. We introduce the strain dependence into the basis set by including bulk Bloch functions corresponding to a few types of stains. For example, to describe a bulk solid under various degrees of hydrostatic strains, one needs to include in the basis one hydrostatically strained system in addition to the unstrained system. To describe the (001) uniaxially strained systems, an additional (001) uniaxially strained Bloch function is needed in the basis set. The idea here is to have enough variational degrees of freedom for the wave function to cover the whole strain profile of the system. As will be discussed in Sec. IV, to calculate a pyramidal InAs quantum dot, we need to include 3–4 differently strained bulk systems in the basis set, such as in the calculation of Fig. 1. If we include only a single bulk system in the basis set, the eigenenergies can often be in error of up to 100 meV. Note that the unrelaxed reference system defined in the begining of Sec. II B is used to generate $\{\mathbf{R}_0,\mathbf{k}\}$, but not to generate the basis set. The basis function [coefficients $A_{k,n}(\mathbf{G})$ in Eq. (6)] can be generated from strained bulk systems, as discussed here.

IV. COMPUTATIONAL CONSIDERATIONS

In this section, we discuss some numerical details of this method, which are important for its implementation. There are two computational aspects to the implementation of the SLCBB method: (i) evaluation of the matrix elements using the deformed Bloch function basis [Eqs. (18), (24) and (25)], and (ii) diagonalization of the resulting matrix to get the eigenenergies and eigenstates.

A. Matrix element evaluation

The key to the current implementation is to have a fast evaluation of the matrix elements. This is made possible by the use of Eq. (24) instead of Eq. (21). As mentioned in Sec. II, the derivation of Eq. (24) from Eq. (21) reduces the computational effort of the matrix elements [Eq. (25)] from $\propto N\times(N_B\times N_k\times N_G)^2$ operations to $\propto 30(N_B\times N_k\times N_G)^2$ operations, given that all the structure factors in Eqs. (17) and (23) are known. Here a prefactor 30 is used to represents the evaluation of the many terms in Eq. (24) [or Eq. (A6)]. Since the number of primary cells N is typically about a million, the use of Eq. (24) instead of Eq. (21) represents a $\sim$10 000-fold speedup.

Since all the $\mathbf{R}_0$ points in Eqs. (17) and (23) reside on a regular fcc grid, fast Fourier transformation can be used to calculate the structure factors in Eqs. (17) and (23), in a fraction ($<5\%$) of the total matrix evaluation time. There are only six components to be calculated for the tensor $\hat{W}^{\alpha}_{\chi}(\mathbf{k})$, and 15 grouped components to be calculated for $W^{\alpha}_{i1,i2,i3,i4}(\mathbf{k})$. Furthermore, to save memory, not all $\mathbf{k}$ points within the first BZ need to be stored. This is because, for a given basis set, only N_k k points inside the first BZ are used. Then $\mathbf{k}-\mathbf{k}'$ in Eq. (24) occupies only a small fraction of the total k points in the first BZ and only on those k points need we store $\hat{\sigma}(\mathbf{k})$, $W^{\alpha}(\mathbf{k})$, $\mathbf{W}^{\alpha}_{d}(\mathbf{k})$, $\hat{W}^{\alpha}_{\chi}(\mathbf{k})$, and $W^{\alpha}_{i1,i2,i3,i4}(\mathbf{k})$.

In Eq. (23), each primary cell has one $\hat{\epsilon}(\mathbf{R}_0)$ and $J(R_0)$. In practice, we have assigned one $\hat{\epsilon}(\mathbf{R}_0)$ and $J(R_0)$ for each atom in the zinc-blende primary cell. In other words, $\hat{\chi}(R_0)$ and $J(R_0)$ in Eq. (23) should be changed to $\hat{\chi}_{\alpha}(R_0)$ and $J_{\alpha}(R_0)$. To evaluate $\hat{\epsilon}$ for a given atom, we have considered the four atoms bonded to that atom. If we use $d\mathbf{x}_k$ ($k=1$, 2, and 3) to denote the three edges of the four-atom-cornered tetrahedron, and $d\mu_k$ are the corresponding three edges in the unrelaxed reference system, then $\hat{\epsilon}_{\alpha}$ for the atom at the center of the tetrahedron is calculated as $d\mu_k=\hat{\epsilon}_{\alpha}\cdot d\mathbf{x}_k$. Thereafter, $J_{\alpha}(R_0)$ is calculated as $det|\hat{\epsilon}_{\alpha}(R_0)|$. Notice that $\Delta\mathbf{d}_{\alpha}(R_0)$ in Eq. (23) is nonzero only for cation atoms since

the anion atoms are used to represent the position of the primary cell. In Eq. (17), we have used the averaged $\hat{\epsilon}\cdot\hat{\epsilon}$ from the cation and anion atoms.

The calculation of Eqs. (24) and (25) can be further speeded up by carefully rearranging the ''do loops'' in the program over $\mathbf{k}$, $\mathbf{k}'$, $\mathbf{G}$, $\mathbf{G}'$ and n, n'. The idea is that Eq. (24) [and the major part of Eq. (A6)] depends only on $\mathbf{q}=\mathbf{k}-\mathbf{k}'+\mathbf{G}-\mathbf{G}'$. Thus, for the same $\mathbf{q}$, the time-consuming multiplications need not be repeated. In practice, instead of doing a double loop over $\mathbf{k}$ and $\mathbf{k}'$, one can have an outer loop over all possible $\mathbf{k}-\mathbf{k}'$ values, and an inner loop over the possible $\mathbf{k}$ for each $\mathbf{k}-\mathbf{k}'$. Within each loop for the same $\mathbf{k}-\mathbf{k}'$ value, we can then store the results of Eq. (24) for each $\mathbf{G}-\mathbf{G}'$, and use them for different $\mathbf{G}$ and $\mathbf{k}$ values. In doing so, we have effectively reduced a major part of the $(N_B\times N_k\times N_G)^2$ scaling to $N_B^2\times N_k\times N_G$ scaling.

To carry out Eq. (25), one needs the coefficients $A_{k,n}(\mathbf{G})$ of the bulk Bloch function. This is obtained by direct diagonalization of the bulk Hamiltonian at each individual k point, and for each constituent material and bulk strain. Again, the band index, material index, and bulk strain index are combined into ''n'' in $A_{k,n}(\mathbf{G})$. Furthermore, at each k point, $A_{k,n}(\mathbf{G})$'s for different ''n'' are orthogonalized, as shown in Eq. (14). The generation of $A_{k,n}(\mathbf{G})$ for all of the needed basis functions takes $<5\%$ of the total matrix evaluation time.

Finally, since the Hamiltonian H is real, one can exploit symmetries to reduce the number of matrix elements one needs to evaluate. First, due to the Hermitian symmetry, $\langle\phi_{k,n}|H|\phi_{k',n'}\rangle=\langle\phi_{k',n'}|H|\phi_{k,n}\rangle^*$. Second, due to the time-reversal symmetry, $\langle\phi_{k,n}|H|\phi_{k',n'}\rangle=\langle\phi_{-k,n}|H|\phi_{-k',n'}\rangle^*$. As a result, only a quarter of the $(N_B\times N_k)^2$ matrix elements need to be explicitly calculated.

B. Diagonalization of the Hamiltonian matrix

After the whole matrix $\langle\phi_{k,n}|H|\phi_{k',n'}\rangle$ is obtained, it is diagonalized to get the eigenenergies and eigenstates. This is done using standard numerical packages, e.g., LAPACK. Usually, only the band-edge states are needed for nanostructure calculations. We found it necessary in some cases, to first obtain $A=(H-E_{ref})^2$ by a matrix multiplication, then solve the eigenstates of A. Here, E_{ref} is a reference energy placed inside the band gap. This folded spectrum method[26] changes the band-edge states from the middle of the energy spectrum to the lowest-energy states of the folded spectrum. Thus we solve for only a few states, instead of the whole spectrum. The diagonalization step scales as $(N_B\times N_k)^3$. If the number $N_B\times N_k$ of the basis set is larger than 4000–5000, the time for the diagonalization part becomes larger than the time for the Hamiltonian evaluation part. For a system with a few million atoms and using a few thousand deformed Bloch functions as the basis set, it typically takes 1–2 h to evaluate the matrix elements on a workstation and an additional few hours to diagonalize the resulting matrix. However, up to ~400-MB memory might be needed to store the large Hamiltonian matrix.

V. RESULTS FOR A PYRAMIDAL QUANTUM DOT

In this section, we will apply the SLCBB method to calculate the single-electron states of a pyramidal quantum dot, and compare the results with the exact solutions of Eq. (2). Nanometer-sized semiconductor quantum dots can be grown using the Stranski-Kranstanow growth mode in molecular beam epitaxy (MBE)[27] or metal-organic chemical vapor deposition.[28] When a 1.7-ML InAs is deposited on a GaAs substrate during a 550-570 °C MBE growth, Ruvimov *et al.*[29] reported {101}-faceted InAs pyramids formed on the surface. These surface InAs pyramidal quantum dots are subsequently covered by GaAs capping layers.

We have studied such a {101}-faceted InAs pyramidal quantum dot embedded in a GaAs matrix. The crystal structure is zinc blende. The square base of the pyramid is $b=110$ Å (i.e., $20a$, where a is the lattice constant of GaAs). The height of the pyramid is then $h=b/2=55$ Å. The pyramid is placed in a $40a\times40a\times20a$ supercell, filled with GaAs. The whole system contains a quarter of a million atoms.

There is a 7% lattice mismatch between InAs and GaAs. The atomic positions are relaxed using a Keating valence force field (VFF) model.[30] The strain profile of this system is reported in Ref. 31. The hydrostatic strain and uniaxial strain can be as large as 10% inside the quantum dot. The strain effects on the local confinement potential can be larger than[31] 500 meV. Thus it is essential to treat these strain effects accurately.

The electronic structure of this system using Hamiltonian (2) can be calculated exactly using the parallelized folded spectrum method (FSM).[12,24] A 5-Ry cutoff energy is used in the plane-wave expansion. A smooth cutoff function of $\beta=0.8$ is used, as described in the Appendix. The FSM results were reported in Ref. 24. That study revealed relationships between the shape of the quantum dot with the state splitting, polarization, and number of confined states in the quantum dot. The time to calculate the conduction and valence-band states of this system is about 20 h on 128 nodes of the Cray T3E parallel machine. Here we are interested in using this dot system as a test for our approximated SLCBB method.

First we calculate the electron states of the system. We choose the lattice constant of our unrelaxed reference system to be 5.879 Å, which is between the GaAs lattice constant (5.653 Å) and the InAs lattice constant (6.058 Å). We have used a sphere around the Γ point to select the k points in the SLCBB expansion of Eq. (15). About 800 k points are used. This corresponds to a grid of ten discrete k point in each direction. Four strained bulk systems were used to generate the Bloch wave-function basis. The first one is GaAs in its natural lattice constant of 5.653 Å. The second one is InAs in a compressed lattice constant of 5.879 Å. The third one is InAs in an average lattice constant of 5.879 Å but under $\epsilon_{zz}-1=3\%$ uniaxial strain (z is the pyramid height direction). The fourth system is InAs, as in the third case, but with an additonal 3% ϵ_{xy}, ϵ_{yx} strain. The choice of these bulk systems is designed to cover the strain profile of the system, as shown in Ref. 31. Note that for a simpler system like quantum wells and superlattices, the strain profile would be much simpler. Thus a smaller number of strained bulk systems would be needed in the basis set. For each of the above four systems, we have chosen only one full-zone band (the first conduction band) as the basis set. As a result, there are a total of 3018 basis functions. It takes about 1.4 min on a IBM 595 workstation to generate all the basis functions $A_{k,n}(\mathbf{G})$; 1.2 min to obtain the $\hat{\sigma}(\mathbf{k})$ in Eq. (17) and $W(\mathbf{k})$'s

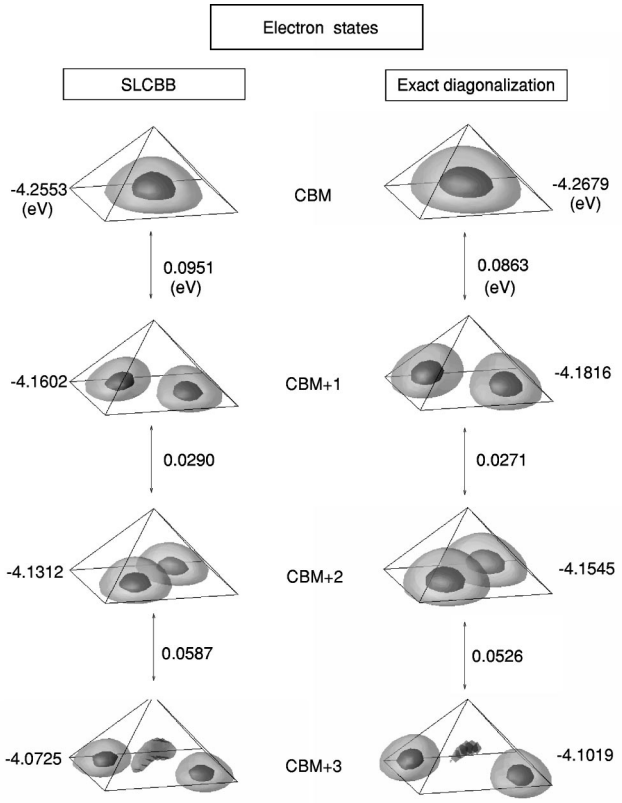

FIG. 4. The SLCBB calculated electron states of an InAs pyramidal quantum dot embedded in a GaAs matrix, with a base length equal 110 Å and a height equal to 55 Å. The results are compared with the exact results of the folded-spectrum method. The numbers are the corresponding eigenenergies (in eV) and energy splittings. The isosurfaces are plotted from the wave-function squares. The level values of the green and blue isosurfaces equal 0.25 and 0.75 of the maximum wave-function square values, respectively.

in Eq. (23); 21 min to evaluate the matrix elements in Eq. (25) [or Eq. (A6)]; and 10 min to diagonalize the 3018 $\times$3018 matrix. Thus, in total it takes about 34 min to calculate the electron states of the quantum dot. This is about 1000 times faster than the direct calculation using the FSM. Should the system be larger, this ratio could also be. This is because the time for the current approach will not change, while the time of FSM will increase linearly with the size of the system.

Figure 4 shows the electron wave function and energies of the pyramidal InAs dot. We see that the SLCBB CBM energy has an error of 12 meV compared with the exact FSM results. The energy of SLCBB CBM+3 has a larger error of 29 meV. Notice that the SLCBB energy is slightly higher than the exact energy. This is an indication that the energy error comes from a limited SLCBB basis set [rather than from the matrix element evaluation error in Eq. (24)]. A strong restriction on the basis set is the number of k points. With the limited k points, it is difficult to describe the more complicated wave functions of CBM+3, causing a larger energy error for this state. Notice also that the energy splitting between states has been reproduced accurately by the SLCBB method. The largest error for the energy splitting is 9 meV. So, in this regard, the result of SLCBB is quite satisfactory. The wave functions of the SLCBB and the direct calculation are quite similar. There are some small differences, like the size of the isosurface. However, the SLCBB wave function plotted in Fig. 4 is only an approximate one. That is, to plot Fig. 4, we have projected the wave function $\psi(\mathbf{x})$ in Eq. (15) to $u_{\Gamma,n}(\mathbf{x})$ [with a multiplication envelope function $F_n(\mathbf{x})$]. The use of $u_{\Gamma,n}(\mathbf{x})$ instead of $\phi_{k,n}(\mathbf{x})$ introduces some small errors in the plotted wave-function amplitude $\Sigma_n|F_n(\mathbf{x})|^2$. Thus the small isosurface difference between SLCBB and direct calculation in Fig. 3 should not be taken too literally.

Next we calculate the hole states. Here more basis functions are needed to describe the degenerated top of the valence band. Like the conduction state calculation, an unrelaxed reference system with a lattice constant of 5.879 Å is used. The same four strained bulk systems and the same k-point selections are used as for the electron state calculation. However, more bands are taken in the basis set for each strained system. For the first bulk system of natural GaAs, four bands are taken, from the first to the fourth (Γ_{1v} and Γ_{15v}). For the other three systems, the three top valence bands (Γ_{15v}) are taken in the basis set. This results in 9909 basis functions. On the same IBM 595 workstation, it takes the same time to generate the basis functions $A_{k,n}(\mathbf{G})$ and to obtain the $\hat{\sigma}(\mathbf{k})$ in Eq. (17) and $W(\mathbf{k})$'s in Eq. (23), as in the case of electron state calculations. It takes however 150 min to evaluate the matrix elements in Eq. (25) [or Eq. (A6)]; 109 min to carry out a matrix multiplication to obtain $(H-E_{ref})^2$; and another 316 min to diagonalize the resulting $(H-E_{ref})^2$ matrix. In total, the calculation of hole states takes about 9.6 h. This is more than ten times slower than the calculation of the electron state, but is still about 100 times faster than the direct calculation *via* the folded spectrum method. Notice that about three quarters of the computer time is spent in calculating $(H-E_{ref})^2$ and diagonalizing the resulting matrix. A standard exact diagonalization routine is used to diagonalize the extrema states of the $(H-E_{ref})^2$ matrix. There may be room to improve the situation by using an iterative diagonalization scheme, like the Lanczos method. To store a 9909$\times$9909 complex Hermitian matrix, about 400-MB memory is needed.

Figure 5 shows the wave function of the dot's hole states and the eigenenergies. The SLCBB hole state energies are typically 11–22 meV lower than the exact energy. Again, this is mainly due to the limited variational degree of freedom in the basis set, especially due to the limited k points. This is a more acute problem for hole state, since they tend to have finer features in their wave function, as shown in Fig. 5. The energy splittings have a maximum error of 10 meV, with the average error being 5.6 meV. The order of the states are the same for the SLCBB and FSM. The wave functions of the SLCBB and FSM look similar. However, for VBM and VBM-1, the two wave-function lumps are more separated in the FSM than in the SLCBB. This may simply indicate that there are not enough k points in the SLCBB basis to describe the small wave-function lumps. Again, some fine structure differences between the SLCBB and FSM in Fig. 5 should not be taken too literally, since we have introduced small errors in plotting the SLCBB wave functions.

It is nontrivial to obtain the correct hole wave functions. Unlike the case for the electron states, where the nodal structure follows s, p, and d classifications, the hole states are more complicated, e.g., they are so intermixed as to have no

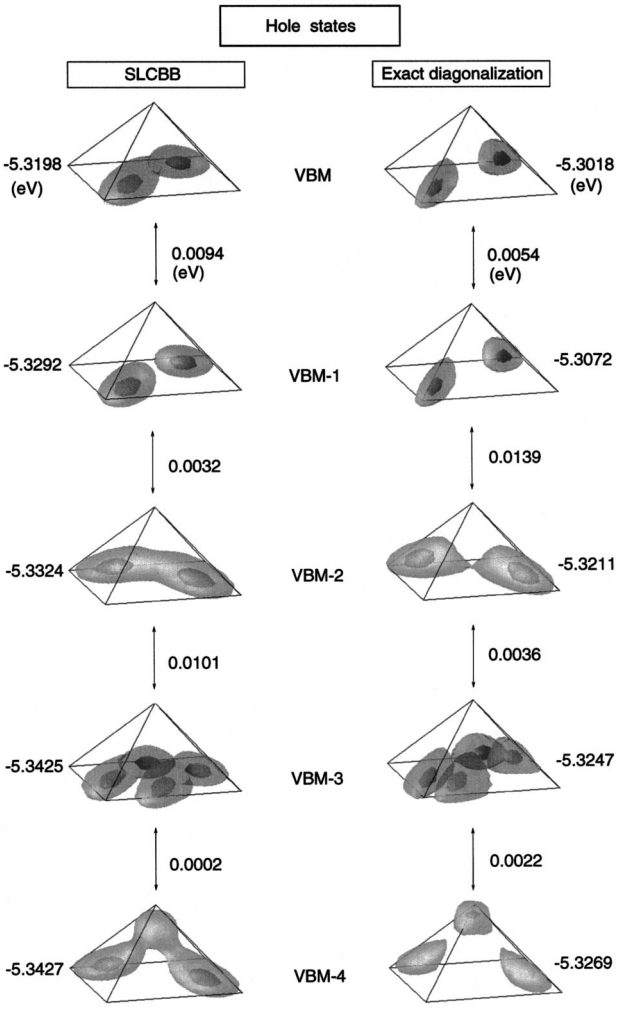

FIG. 5. The SLCBB calculated hole states of the same InAs pyramidal quantum dot as in Fig. 4. The results are compared with the exact results of the folded-spectrum method. The same as in Fig. 4 for the isosurface values.

nodes. These wave functions cannot be estimated without explicit calculation. If the error of our approximation were on the order of 100 meV, instead of the current 10-meV value [for example, by removing the second-order term $W^{\alpha}_{i1,i2,i3,i4}$ in Eq. (24) or by using only a single strained bulk system in the basis set], completely different hole wave functions will result, bearing no resemblance to the states shown in Fig. 5.

VI. CONCLUSIONS

We have presented a strained, linear combination of bulk band (SLCBB) method for calculating single-particle wavefunctions and energies of nanosystems. This method represents an extension of the LCBB method to the strained heterostructure systems. This method correctly treats the effects of strain on the electronic structure up to a major part of the second order. Unlike the conventional $\mathbf{k}\cdot\mathbf{p}$ method, no explicit strain deformation potential parameters are used to represent this strain effect. What is needed are the atomic positions and the atomic pseudopotentials. Sufficient details are presented in this paper, so that a workable program can be developed based on the current formalism. The SLCBB results differ from the exact diagonalization results only by $\lesssim 20$ meV in absolute energy levels, and $\lesssim 10$ meV in the energy splittings. The wave functions of the SLCBB are very similar to those obtained in direct diagonalization. However, for a system containing a quarter of a million atoms, the SLCBB is 100–1000 times faster than the direct FSM calculation. For larger systems, this speedup ratio will be even larger. With the SLCBB method, the atomistic (not the envelope function) electronic structure of a million-atom system can be calculated within 10 h on a personal computer.

Another use of the SLCBB method is to study the contribution of different bulk states to the eigenstates of a nanostructure. This can be done by monitoring the change of the eigenstates when some of the bulk bands in the basis set are removed. This can also be done by plotting $\Sigma_k C_{k,n}|\phi_{k,n}\rangle$ in Eq. (15) for different band index n. A simple summation of $\Sigma_k|C_{k,n}|^2$ will reveal the magnitudes of different bulk band contributions in quantum dot states. For example, the dot's hole states shown in Fig. 5 are consisted of typically 95% of the bulk heavy hole and 5% of the bulk light hole.

Finally, throughout this paper, we have ignored the spin-orbit interaction in our calculation.[12] To include this interaction is not difficult. One approach is to add a spin-orbit interaction term in Eq. (25). One can double the basis set of Eq. (13), so that there are spin-up and spin-down components. Under this basis set, the terms in Eq. (25) are the diagonal terms within the spin-up and spin-down components. The spin-orbit coupling can be introduced as additional diagonal and off-diagonal terms between spin-up and spin-down components. The same technique of FFT can be used as in Eq. (23). A nonlocal potential $V^{so}_{\alpha}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}')$ can be used to replace $V_{\alpha}(q_2)$ for the spin-orbit interaction. However, the formalism of these additional terms can be much simplified comparing to the terms in Eq. (24), since energy involved in these additional terms is small and low order approximations are acceptable. A recent implementation of the spin-orbit interaction version of the SLCBB program shows that (1) the memory of the program will only be doubled due to symmetry; (2) the time to evaluate the matrix elements increases by about 25%.

ACKNOWLEDGMENTS

We would like to thank J. Kim for testing the SLCBB program for different systems, and A. J. Williamson and A. Franceschetti for useful discussions. This work was supported by the U.S. Department of Energy, OER-BES, under Grant No. DE-AC36-83CH10093. The calculation of the pyramidal dot using the parallelized folded spectrum method was carried out on the Cray T3E machine in the National Energy Research Scientific Computer Center (NERSC).

APPENDIX: ADDITIONAL TERMS CAUSED BY THE SMOOTH CUTOFF FUNCTION

In a plane-wave pseudopotential calculation, the plane wave $\mathbf{k}+\mathbf{G}$ is included in the basis set if

$$E(\mathbf{k}+\mathbf{G})=\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2\leqslant E_{cut}, \tag{A1}$$

where E_{cut} is a cutoff energy, usually about 5–8 Ry in our calculations.

A smooth cutoff function $\omega(|\mathbf{k}+\mathbf{G}|^2)$ can be defined as[22]

$$\omega(|\mathbf{k}+\mathbf{G}|^2)=\begin{cases} 1 & \text{if } E(\mathbf{k}+\mathbf{G})\leqslant\beta E_{cut} \\ \sin^2\left[\dfrac{\pi[E_{cut}-E(\mathbf{k}+\mathbf{G})]}{2(1-\beta)E_{cut}}\right] & \text{if } \beta E_{cut}\leqslant E(\mathbf{k}+\mathbf{G})\leqslant E_{cut}, \end{cases} \tag{A2}$$

where $\beta\leqslant 1$ is a control factor ($\beta=0.8$ is used in the current calculation).

To use this smooth cutoff function, the potential part H_2 of the Hamiltonian will be changed to a nonlocal operator

$$H_2(\mathbf{x}_1,\mathbf{x}_2)=\int \omega(\mathbf{x}-\mathbf{x}_1)\left\{\sum_{R,\alpha} v_\alpha[\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha(R)]W_\alpha(R)\right\}\omega(\mathbf{x}-\mathbf{x}_2)d^3\mathbf{x}, \tag{A3}$$

where $\omega(\mathbf{x})$ is the Fourier transformation of $\omega(|\mathbf{k}+\mathbf{G}|^2)$ in real space, and is a short-range function. Then the matrix element $\langle\mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle$ can be written as

$$\begin{aligned}\langle\mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle=\frac{1}{\Omega_0}\int \sqrt{J(\mathbf{x}_1)}e^{-(\mathbf{k}+\mathbf{G})\cdot\mu(\mathbf{x}_1)}\omega(\mathbf{x}-\mathbf{x}_1)\left\{\sum_{i,\alpha} v_\alpha[\mathbf{x}-\mathbf{R}-\mathbf{d}_\alpha(R)]W_\alpha(R)\right\} \\ \times\omega(\mathbf{x}-\mathbf{x}_2)\sqrt{J(\mathbf{x}_2)}e^{(\mathbf{k}'+\mathbf{G}')\cdot\mu(\mathbf{x}_2)}d^3\mathbf{x}_1 d^3\mathbf{x}_2 d^3\mathbf{x}.\end{aligned} \tag{A4}$$

Using the same approximation as we did in Eq. (19), especially making the expansion $\mu(\mathbf{x}_l)-\mu(\mathbf{x})\doteq\hat{\epsilon}(R_0)\cdot(\mathbf{x}_l-\mathbf{x})$, for $l=1$ and 2, we have

$$\begin{aligned}\langle\mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle=\frac{1}{N}\sum_\alpha e^{i\mathbf{q}\cdot\mathbf{d}^0_\alpha}\sum_{R_0} J(R_0)W_\alpha(R_0)e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_0}e^{i\mathbf{q}\cdot\Delta\mathbf{d}_\alpha(R_0)} \\ \times V_\alpha(|\mathbf{q}\cdot\hat{\epsilon}(R_0)|^2)\omega(|(\mathbf{k}+\mathbf{G})\cdot\hat{\epsilon}(R_0)|^2)\omega(|(\mathbf{k}'+\mathbf{G}')\cdot\hat{\epsilon}(R_0)|^2).\end{aligned} \tag{A5}$$

This equation is the counterpart of Eq. (21). Following the same approximation after Eq. (21), and expand $\omega[|\mathbf{k}+\mathbf{G}|^2+(\mathbf{k}+\mathbf{G})\cdot\hat{\chi}\cdot(\mathbf{k}+\mathbf{G})]$ to the first order of $(\mathbf{k}+\mathbf{G})\cdot\hat{\chi}\cdot(\mathbf{k}+\mathbf{G})$, and we have our final result as

$$\begin{aligned}\langle\mathbf{k}+\mathbf{G}|H_2|\mathbf{k}'+\mathbf{G}'\rangle=\sum_\alpha e^{i\mathbf{q}\cdot\mathbf{d}^0_\alpha}\Bigg\{\omega(|\mathbf{k}+\mathbf{G}|^2)\omega(|\mathbf{k}'+\mathbf{G}'|^2)\Bigg[V_\alpha(q^2)W^\alpha(\mathbf{k}-\mathbf{k}')+iV_\alpha(q^2)\mathbf{q}\cdot\mathbf{W}^\alpha_d(\mathbf{k}-\mathbf{k}') \\ +V'_\alpha(q^2)\mathbf{q}\cdot\hat{W}^\alpha_\chi(\mathbf{k}-\mathbf{k}')\cdot\mathbf{q}+\frac{1}{2}V''_\alpha(q^2)\sum_{i1,\ldots,i4} q_{i1}q_{i2}q_{i3}q_{i4}W^\alpha_{i1,i2,i3,i4}(\mathbf{k}-\mathbf{k}')\Bigg] \\ +\omega(|\mathbf{k}+\mathbf{G}|^2)\omega'(|\mathbf{k}'+\mathbf{G}'|^2)V_\alpha(q^2)(\mathbf{k}'+\mathbf{G}')\cdot\hat{W}^\alpha_\chi(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{k}'+\mathbf{G}') \\ +\omega'(|\mathbf{k}+\mathbf{G}|^2)\omega(|\mathbf{k}'+\mathbf{G}'|^2)V_\alpha(q^2)(\mathbf{k}+\mathbf{G})\cdot\hat{W}^\alpha_\chi(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{k}+\mathbf{G})\Bigg\},\end{aligned} \tag{A6}$$

where $\omega'(q_2)\equiv d\omega(q_2)/dq_2$. Equation (A6) should be used to replace Eq. (24) in cases where a smooth cutoff function ω is used in the calculation. The kinetic-energy part [Eq. (18)] remains the same.

[1] W. Ge, W. D. Schmidt, M. D. Sturge, L. N. Pfeiffer, and K. W. West, J. Lumin. **59**, 163 (1994).

[2] H. Akiyama, J. Phys.: Condens. Matter **10**, 3095 (1998).

[3] M. Lomascolo, P. Ciccarese, R. Cingolani, R. Rinaldi, and F. K. Reinhart, J. Appl. Phys. **83**, 302 (1998).

[4] R. J. Warburton, C. S. Durr, K. Karrai, J. P. Kotthaus, G. Medeiros-Ribeiro, and P. M. Petroff, Phys. Rev. Lett. **79**, 5282 (1997).

[5] MRS Bull. **23**, 15 (1998), special issue on semiconductor quantum dot, edited by A. Zunger.

[6] C. Pryor, Phys. Rev. B **57**, 7190 (1998).

[7] H. Jiang and J. Singh, Phys. Rev. B **56**, 4696 (1997).

[8] D. M. Wood and A. Zunger, Phys. Rev. B **53**, 7949 (1996); D. M. Wood, A. Zunger, and D. Gershoni, Europhys. Lett. **33**, 383 (1996).

[9] H. Fu, L. W. Wang, and A. Zunger, Phys. Rev. B **57**, 9971 (1998); Appl. Phys. Lett. **71**, 3433 (1997).

[10] L. W. Wang and A. Zunger, Phys. Rev. B **54**, 11 417 (1996).

[11] L. W. Wang and A. Zunger, J. Phys. Chem. B **102**, 6449 (1998).

[12] L. W. Wang, J. Kim, and A. Zunger, Phys. Rev. B **59**, 5678 (1999).

[13] L. W. Wang, A. Franceschetti, and A. Zunger, Phys. Rev. Lett. **78**, 2819 (1997).

[14] A. Williamson and A. Zunger, Phys. Rev. B **58**, 6724 (1998).

[15] A. Franceschetti and A. Zunger, Phys. Rev. B **52**, 14 664 (1995).

[16] L. W. Wang and A. Zunger, Phys. Rev. B **56**, 12 395 (1997).

[17] A. Franceschetti and A. Zunger, Phys. Rev. Lett. **78**, 915 (1997).
[18] A. Franceschetti, L. W. Wang, H. Fu, and A. Zunger, Phys. Rev. B **58**, R13 367 (1998).
[19] M. G. Burt, J. Phys.: Condens. Matter **4**, 6651 (1992).
[20] B. A. Foreman, Phys. Rev. Lett. **81**, 425 (1998).
[21] L. W. Wang and A. Zunger, in *Semiconductor Nanoclusters*, edited by P. V. Kamat and D. Meisel (Elsevier, Amsterdam, 1996).
[22] L. W. Wang and A. Zunger, Phys. Rev. B **51**, 17 398 (1995).
[23] A. J. Williamson, J. Kim, L. W. Wang, S. H. Wei, and A. Zunger (unpublished).
[24] J. Kim, L. W. Wang, and A. Zunger, Phys. Rev. B **57**, R9408 (1998).
[25] L. W. Wang and A. Zunger, Phys. Rev. B **53**, 9579 (1996).
[26] L. W. Wang and A. Zunger, J. Chem. Phys. **100**, 2394 (1994).
[27] B. G. Levi, Phys. Today **49**(6), 22 (1996).
[28] W. Seifert *et al.*, Prog. Cryst. Growth Charact. Mater. **33**, 423 (1996).
[29] S. Ruvinmov *et al.*, Phys. Rev. B **51**, 14 766 (1995).
[30] P. Keating, Phys. Rev. **145**, 637 (1966).
[31] C. Pryor, J. Kim, L. W. Wang, A. J. Williamson, and A. Zunger, J. Appl. Phys. **83**, 2548 (1998).