Large-scale atomic effective pseudopotential program including an efficient spin-orbit coupling treatment in real space

F. Zirkelbach,[1] P.-Y. Prodhomme,[1] Peng Han,[1] R. Cherian,[1] and G. Bester[1,2,3,*]

[1]*Max-Planck-Institut für Festkörperforschung, Heisenbergstraße 1, D-70569 Stuttgart, Germany*

[2]*Institut für Physikalische Chemie, Universität Hamburg, Grindelallee 117, D-20146 Hamburg, Germany*

[3]*The Hamburg Centre for Ultrafast Imaging, Luruper Chaussee 149, D-22761 Hamburg, Germany*

(Received 1 April 2014; revised manuscript received 21 January 2015; published 20 February 2015)

Within the scheme of the *large-scale atomic effective pseudopotential program* (LATEPP), the Schrödinger equation of an electronic system is solved within an effective single-particle approach. Although not limited to, it focuses on the recently introduced atomic effective pseudopotentials derived from screened local effective crystal potentials as obtained from self-consistent density functional theory calculations. The problem can be solved in both real (real-space grid) and reciprocal space (plane-wave basis functions). Following the idea of atomic effective pseudopotentials, the density, and hence a self-consistent cycle, is not required and not implemented. An iterative solver is implemented to deliver the eigenstates close to a selected reference energy, e.g., around the band gap of a semiconductor. This approach is particularly well suited for theoretical investigations of the electronic structure of semiconductor nanostructures and we demonstrate linear scaling with the system size up to around 100 000 atoms on a single standard compute node. Moreover, an efficient real-space treatment of spin-orbit coupling within the pseudopotential framework is proposed in this work allowing for a fully relativistic description.

DOI: 10.1103/PhysRevB.91.075119 PACS number(s): 71.15.Dx, 31.15.A−

I. INTRODUCTION

Semiconductor nanostructures exhibit highly attractive and technologically relevant electronic and optical properties that depend on their composition, structure, size, and shape. The increasing ability to control these parameters, as well as experimental achievements in the characterization of these nanostructures, enable discoveries of increasingly complex behavior. These include the appearance of different kinds of excitons [1–3], Auger processes [4], as well as quantum entanglement [5], which make these semiconductor nanostructures ideal candidates for functionalized solid-state materials in the field of optoelectronics and quantum information. Moreover, semiconductor quantum wells constitute the first structural class of materials, for which conducting edge states in bulk insulators were predicted [6] to be protected by time-reversal invariance and subsequently observed experimentally [7], establishing the emerging research field of topological insulators [8,9].

While these structures with dimensions of several up to hundreds of nanometers are considered small, they do consist of up to 100 000 atoms. Theoretical methods describing the electronic and optical properties of semiconductor nanostructures must be capable of addressing this large number of atoms including the treatment of excitations at a many-body level, as well as relativistic spin-orbit effects, the latter being, for instance, responsible for the protected metallic surface states of topological insulators or the polarization properties of optical emission.

The currently existing approaches applied for the calculation of the excited state properties of nanostructures range from *ab initio* descriptions to rather empirical methods with limited predictive capabilities. Historically, continuum approaches where the atomic fast oscillating potential is replaced by a smooth (e.g., parabolic) potential, as in the $\mathbf{k}\cdot\mathbf{p}$ method, have been applied to nanostructures revealing interesting trends and delivering an initial basic understanding of confinement effects. Recent and ongoing theoretical efforts are aimed at the development of affordable atomistic descriptions, which naturally capture the correct atomistic symmetry and include all the relevant effects at the onset. Eventually, the atomistic description will make the continuum models obsolete. The mapping of the quantitative results onto simple effective Hamiltonians for the purpose of interpretation is a mere postprocessing of the data.

Existing atomistic methods can be cast into different groups. (i) First-principles methods such as time-dependent density functional theory, GW, and Bethe-Salpeter provide an accurate atomistic description but are limited to up to a few hundreds of atoms. (ii) Empirical tight-binding methods can easily address several millions of atoms [10–13] but lack single-particle wave functions, which are needed in the subsequent treatment of excitations. Worthwhile theoretical effects are presently undertaken to address this issue. (iii) The empirical pseudopotential method [14,15] (EPM) has been generalized [16–18] to be able to address nanostructures with millions of atoms using appropriate basis sets [19] and offered an excellent basis for the calculation of excitonic wave functions via configuration interaction [18,20]. (iv) Recently, a new generation of the EPM was introduced [21], which removes its empirical character by using an analytic connection between results of density functional theory (DFT) calculations on elongated and slightly deformed supercells and the derived atomic effective pseudopotentiall (AEP). This procedure provides access to the long-range interaction, which was lacking in the original EPM and offers an automatic pseudopotential generation leading to unique AEPs.

This work introduces the *large-scale atomic effective pseudopotential program* (LATEPP) developed to efficiently solve

*gabriel.bester@uni-hamburg.de

the electronic eigenvalue problems utilizing AEPs. The formalism underlying the AEPs [21] involves nonlocal angular-momentum-dependent pseudopotentials that are implemented following the Kleinmann-Byland separable formulation in reciprocal and in real space. The eigensolver is distinct from the case of DFT since an inner eigenvalue problem is solved. For the spin-orbit treatment, an efficient formulation was developed in this work and implemented and tested against the implementation used in the established DFT code ABINIT [22]. The kinetic energy operator is implemented in reciprocal space, as in standard plane-wave codes, as well as in real space realized by a finite difference scheme. The latter implementation leads to a fully real-space treatment of the problem. The obtained single-particle wave functions can be used in a subsequent configuration interaction approach [18] to obtain the many-body effects, which, however, is beyond the scope of the present method.

In the first part of this paper, a short overview of the capabilities as well as a summary of the basic concepts of AEPs is presented. This is followed by a more detailed description of the underlying code that solves the Schrödinger equation. Finally, the method is illustrated and the performance is tested by applying it to a GaAs quantum dot embedded in AlAs for a total number of 97 336 atoms. Calculations on bulk GaAs are used to illustrate the capabilities and accuracy of the spin-orbit treatment.

II. METHODOLOGY

A. Capabilities

The LATEPP code solves the Schrödinger equation of an electronic system within an effective single-particle approach. It is aimed at utilizing the recently presented AEPs [21] derived from *ab initio* pseudopotential calculations. Within the AEP scheme, the self-consistent update of the density, as performed in a DFT calculation, is not required. This allows us to focus on a restricted selection of relevant eigenstates close to a reference energy, which must be known or guessed based on previous findings. In contrast to DFT, which computes all occupied states, this method does not provide total energies and forces; equilibrated structures have to be assumed or computed by efficient valence force-field models. Likewise, although exhibiting a high transferability, structures that involve large charge transfer compared to the structure used for the generation of the AEP are not guaranteed to be computed accurately. Keeping these limitations in mind, however, LATEPP is capable of efficiently treating large structures since the eigenstates of the Hamiltonian need to be solved only once. Three different solvers are implemented for this task. The iterative Jacobi-Davidson [23–27] solver is applied to the folded spectrum [28] Hamiltonian and the Arnoldi restart algorithm as implemented in ARPACK [29,30] to deliver eigenstates close to a target reference energy. In addition, a direct diagonalization solver is integrated, which requires the explicit construction and storage of the full Hamiltonian and delivers all the eigenstates, in the case this is required.

The nonlocal contribution of the potential to reconstruct the full crystal potential of the self-consistent *ab initio* pseudopotential is treated in a fully separable form as suggested by Kleinman and Bylander [31] and can be evaluated in reciprocal or in real space.

Not a new idea in the field of electronic-structure calculations [32–35], the kinetic energy, which is diagonal in reciprocal space, can likewise be calculated in real space by a finite difference approach, which leads to a fully real-space treatment scaling linearly with the number of atoms. For transformations between reciprocal and real space, the fast Fourier transformation (FFT) as implemented in FFTW [36,37] or FFTE [38] can be selected.

Spin-orbit interaction can be included within a real- or reciprocal-space treatment using the iterative or direct diagonalization scheme.

B. Atomic effective pseudopotentials

As recently introduced [21], AEPs are constructed by extracting the local part of the self-consistent effective crystal potentials $V^{\mathrm{loc,eff}}(\mathbf{r})$ from DFT calculations, which include the local part of the pseudopotential as well as the Hartree and exchange correlation (XC) contributions. This can be transformed to reciprocal space according to

$$V^{\mathrm{loc,eff}}(\mathbf{G}) = \frac{1}{\Omega_c}\int_{\Omega_c} V^{\mathrm{loc,eff}}(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} d^3\mathbf{r}, \tag{1}$$

with Ω_c being the volume of the supercell. By rewriting the crystal potential in Eq. (1) as a sum over atom-centered pseudopotentials

$$V^{\mathrm{loc,eff}}(\mathbf{r}) = \sum_{\alpha}^{N_{\mathrm{species}}} \sum_{n}^{N_\alpha} v_\alpha(\mathbf{r} - \boldsymbol{\tau}_{\alpha,n}) \tag{2}$$

and substituting $\mathbf{r} = \mathbf{r}' + \boldsymbol{\tau}_{\alpha,n}$, the total reciprocal space potential can be expressed as a Fourier sum

$$V^{\mathrm{loc,eff}}(\mathbf{G}) = \frac{1}{\Omega_c} \sum_{\alpha}^{N_{\mathrm{species}}} \sum_{n}^{N_\alpha} e^{-i\mathbf{G}\cdot\boldsymbol{\tau}_{\alpha,n}} v_\alpha(\mathbf{G}) \tag{3}$$

of atomic reciprocal space potentials

$$v_\alpha(\mathbf{G}) = \int_{\infty} v_\alpha(\mathbf{r}) e^{-i\mathbf{G}\cdot\mathbf{r}} d^3\mathbf{r}. \tag{4}$$

A detailed description of the analytic connection between the DFT results obtained for elongated and slightly deformed supercells and the AEPs $v_\alpha(|\mathbf{G}|)$ is given in Ref. [21], along with benchmark tests of their accuracy. The AEPs are available for over 20 binary semiconductors in a tabulated form [39] on a dense grid in reciprocal space. These files are read by LATEPP.

C. Solving the Schrödinger equation

The LATEPP code solves the Schrödinger equation of the electronic problem (in atomic units)

$$\left(-\frac{\nabla^2}{2} + V_{\mathrm{L}}(\mathbf{r}) + \hat{V}_{\mathrm{NL}} + \hat{V}_{\mathrm{SO}}\right)\psi_{i,\mathbf{k}}(\mathbf{r}) = \epsilon_{i,\mathbf{k}}\psi_{i,\mathbf{k}}(\mathbf{r}). \tag{5}$$

Here, $V_{\mathrm{L}}(\mathbf{r})$ is the local effective pseudopotential, $\hat{V}_{\mathrm{NL}}$ represents the nonlocal part of the pseudopotential, $\hat{V}_{\mathrm{SO}}$ is the spin-orbit contribution, and the first term corresponds to the kinetic energy.

1. Basis sets

The wave functions are represented either in a plane-wave basis

$$\psi_{i,\mathbf{k}}(\mathbf{r}) = \sum_{\substack{\mathbf{G} \\ |\mathbf{k}+\mathbf{G}| < G_{\text{cut}}}} c_{i,\mathbf{k}}\, e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \tag{6}$$

truncated at a certain cutoff energy $E_{\text{cut}} = \frac{1}{2}G_{\text{cut}}^2$, or on a regular real-space grid.

2. Solvers and representations

The evaluation of the different parts of the Hamiltonian depends on the selected solver as well as on the representation chosen for the wave functions. The possibilities provided by LATEPP are shortly summarized in the following.

The wave functions are represented on a regular grid, either in real or reciprocal space, both linked via Fourier transforms. If the full Hamiltonian is diagonalized directly, which is only possible for small systems, the wave functions as well as the nonlocal and spin-orbit potentials are exclusively represented and evaluated in reciprocal space.

For larger systems, the iterative Jacobi-Davidson [23–27] solver or Arnoldi restart [29,30] algorithm can be used. These algorithms iteratively approach the solution based on a repeated operation of the operator on the wave function. The following discussions will be restricted to the Arnoldi method since it turned out to be more efficient than the Jacobi-Davidson solver and has established itself as the standard solver.

Within the iterative treatment, $\hat{V}_{\text{NL}}$ can be implemented in the fully separable formulation proposed by Kleinman and Bylander [31] in reciprocal or in real space, whereas the inclusion of spin-orbit interaction is restricted to the real-space treatment.

3. Kinetic energy

In reciprocal space, the matrix elements of the kinetic energy operator $\hat{T}$ are diagonal and easily obtained by

$$T_{\mathbf{G},\mathbf{G}'}(\mathbf{k}) = \langle \mathbf{k}+\mathbf{G}|\hat{T}|\mathbf{k}+\mathbf{G}'\rangle = \delta_{\mathbf{G},\mathbf{G}'}\tfrac{1}{2}(\mathbf{k}+\mathbf{G})^2. \tag{7}$$

Within an iterative treatment, the kinetic energy is calculated by the sum of the product of these diagonal elements and the coefficients of the wave function.

If the remaining parts of the Hamiltonian are evaluated in real space, Fourier transforms of the wave function into reciprocal space and back are needed. The scaling is, thus, determined by the $N\log(N)$ scaling of the FFT algorithm, if N is the grid size, which is proportional to the number of atoms in the system. The kinetic energy can likewise be calculated directly in real space by a finite difference scheme:

$$\begin{aligned}\frac{d^2}{dx^2}\psi\bigg|_{(i,j,k)} \approx & \sum_{l=-O/2}^{O/2} c_{l,0}^{d=2}\psi(i+l,j,k)/h_x^2 \\ & + \sum_{l=-O/2}^{O/2} c_{l,0}^{d=2}\psi(i,j+l,k)/h_y^2 \\ & + \sum_{l=-O/2}^{O/2} c_{l,0}^{d=2}\psi(i,j,k+l)/h_z^2, \end{aligned} \tag{8}$$

for any even order O with the optimal weights $c_{l,m}^{d}$ being determined by an algorithm given by Fornberg [40,41], and $h_{x,y,z}$ being the grid spacings. This preserves linear scaling and becomes advantageous for large systems, as will be shown subsequently.

4. Local potential

The AEPs are tabulated in reciprocal space [39]. The corresponding matrix elements are given by

$$\begin{aligned} V_{\text{L}}^{\mathbf{G},\mathbf{G}'}(\mathbf{k}) &= \langle \mathbf{k}+\mathbf{G}|V_{\text{L}}|\mathbf{k}+\mathbf{G}'\rangle \\ &= \frac{1}{\Omega_{\text{c}}}\int V_{\text{L}}(\mathbf{r})e^{i(\mathbf{G}-\mathbf{G}')\cdot\mathbf{r}}d^3\mathbf{r} \\ &= V_{\text{L}}(\mathbf{G}-\mathbf{G}') \end{aligned} \tag{9}$$

and computed according to Eq. (3). Thus, for fully converged results with respect to the chosen plane-wave cutoff energy, the grid must be chosen large enough ($2G_{\text{cut}}$ in each direction) to include all possible reciprocal lattice vectors $\mathbf{G}-\mathbf{G}'$. However, in most cases, smaller grid sizes have been found to give accurate results, which is demonstrated in Sec. III B.

For surfaces, interfaces, or for an alloy, weights are introduced to account for the difference in the chemical environment compared to the binary or pure bulk system. Including the weights $w_{\alpha,n}$, the sum in Eq. (3) is adjusted to read as

$$V^{\text{loc,eff}}(\mathbf{G}) = \frac{1}{\Omega_{\text{c}}}\sum_{\alpha}^{N_{\text{species}}} v_{\alpha}(\mathbf{G})\sum_{n}^{N_{\alpha}} e^{-i\mathbf{G}\cdot\boldsymbol{\tau}_{\alpha,n}} w_{\alpha,n}. \tag{10}$$

Here, N_{species} specifies the number of different types of atoms that exist in the system in its various occurring environments, i.e., it corresponds to the number of utilized AEPs. For instance, As atoms surrounded by Ga atoms constitute a different species than As atoms surrounded by Al atoms. In the case of surfaces, interfaces, or alloys, an atom, which has different types of next neighbors is represented by more than one such atom species. The respective contributions to the local potential are modified by weights determined by

$$w_{\alpha,n} = \frac{n_{\alpha,n}}{4} \text{ (for tetrahedrally coordinated materials)}, \tag{11}$$

with $n_{\alpha,n}$ being the number of next neighbors of type α.

For evaluation in real space, the potential is initially transformed and stored on a real-space grid. This is likewise done for the iterative methods in reciprocal space. In this way, the $O(N^2)$ complexity due to the matrix-vector multiplication in reciprocal space is reduced to $O[N\log(N)]$ complexity governed by the Fourier transforms of the wave functions, which are then multiplied with the local potential operator that is diagonal in real space.

5. Nonlocal potential

The nonlocal part of the potential is given by

$$\hat{V}_{\text{NL}} = \sum_{l,m} |l,m\rangle\delta V_l(r)\langle l,m|, \tag{12}$$

with the spherical harmonics $|l,m\rangle$ and $\delta V_l(r)$ being the difference of the l-dependent pseudopotential $V_l(r)$ and the selected local part $V_{\text{loc}}(r)$. The local part of the pseudopotential

contains the long-ranged part of the potential, which leads to short-ranged nonlocal operators. In this way, $\delta V_l(r)$ is only nonzero within a sphere of a certain radius.

LATEPP supports the semilocal form in Eq. (12), which is nonlocal only in the angular part and requires a radial integration

$$\langle \mathbf{q}|\hat{V}_{\mathrm{NL}}|\mathbf{q}'\rangle = \frac{4\pi}{\Omega_{\mathrm{c}}}\sum_l (2l+1)P_l(\widehat{q}\cdot\widehat{q}') \times \int r^2 j_l(qr)\delta V_l(r) j_l(q'r)\,dr \qquad (13)$$

with the spherical Bessel functions j_l, the Legendre polynomials P_l, $\mathbf{q} = \mathbf{k} + \mathbf{G}$ and $\mathbf{q}' = \mathbf{k} + \mathbf{G}'$.

For larger system sizes, the separable formulation of Kleinman and Bylander [31] (KB)

$$\hat{V}_{\mathrm{NL}}^{\mathrm{KB}} = \sum_{l,m} \left|\chi_{lm}^{\mathrm{KB}}\right\rangle E_l^{\mathrm{KB}} \left\langle\chi_{lm}^{\mathrm{KB}}\right| \qquad (14)$$

with the KB eigenvalue

$$E_l^{\mathrm{KB}} = \frac{\langle u_l \delta V_l | \delta V_l u_l\rangle}{\langle u_l|\delta V_l|u_l\rangle} \qquad (15)$$

and the normalized KB projectors

$$\left|\chi_{lm}^{\mathrm{KB}}\right\rangle = \frac{|\delta V_l \phi_{lm}\rangle}{\langle u_l \delta V_l|\delta V_l u_l\rangle^{1/2}} \qquad (16)$$

can be used. Here, ϕ_{lm} is the atomic pseudopotential wave function and u_l its radial part multiplied with the radius. The radial parts as well as the respective potentials δV_l of the pseudopotential wave functions are stored on a real-space grid in a file and are constructed according to the procedures by Hamann, Schlüter, and Chiang [42] or Troullier and Martins [43].

To calculate E_l^{KB}, the integrals of Eq. (15) are evaluated on the one-dimensional real-space grid of the pseudopotential. The same grid is used to obtain the radial part of the KB projectors

$$u_l^{\mathrm{KB}}(q) = \int_r j_l(qr)\delta V_l(r) u_l(r) r\,dr \qquad (17)$$

on a likewise one-dimensional, equally spaced auxiliary grid consisting of 600 points up to the cutoff value in reciprocal space. Interpolated to the reciprocal space grid of LATEPP and multiplied by the respective spherical harmonics of the angular part $\Omega_{\boldsymbol{q}}$ of the reciprocal lattice vector $\mathbf{q} = \mathbf{k} + \mathbf{G}$, the KB projector

$$\left\langle \mathbf{q}\middle|\chi_{lm}^{\mathrm{KB}}\right\rangle = 4\pi(-i)^l Y_{lm}^*(\Omega_{\mathbf{q}}) u_l^{\mathrm{KB}}(\mathbf{q}) \qquad (18)$$

can be used to either construct the matrix elements

$$\langle \mathbf{q}|\hat{V}_{\mathrm{NL}}^{\mathrm{KB}}|\mathbf{q}'\rangle = \sum_{lm} \left\langle \mathbf{q}\middle|\chi_{lm}^{\mathrm{KB}}\right\rangle E_l^{\mathrm{KB}} \left\langle\chi_{lm}^{\mathrm{KB}}\middle|\mathbf{q}'\right\rangle \qquad (19)$$

or to get the new components $\mathbf{q}$ of the matrix multiplication in the iterative solver methods

$$\langle \mathbf{q}|\hat{V}_{\mathrm{NL}}^{\mathrm{KB}}|\psi_n\rangle = \sum_{lm} \left\langle \mathbf{q}\middle|\chi_{lm}^{\mathrm{KB}}\right\rangle E_l^{\mathrm{KB}} \sum_{\mathbf{q}'} \left\langle\chi_{lm}^{\mathrm{KB}}\middle|\mathbf{q}'\right\rangle \langle \mathbf{q}'|\psi_n\rangle. \qquad (20)$$

Again, including atoms and the concepts of weights as in Eqs. (3) and (10) results in a structure factor, which can be stored in memory or recalculated in each iteration. The first approach is faster but is limited by the available memory, whereas the recalculation of structure factors in each iteration is extremely time consuming. The required memory or number of evaluations scales with the square of the system size since the number of structure factors is given by the product of the number of plane waves and the number of atoms.

The real-space representation of the projector is given by

$$\langle \mathbf{r}|\chi_{lm}^{\mathrm{KB}}\rangle = \sum_{\alpha,n} \frac{\delta V_l(|\boldsymbol{r}-\boldsymbol{\tau}_{\alpha,n}|)\frac{u_l(|\boldsymbol{r}-\boldsymbol{\tau}_{\alpha,n}|)}{|\boldsymbol{r}-\boldsymbol{\tau}_{\alpha,n}|}}{\langle u_l \delta V_l|\delta V_l u_l\rangle^{\frac{1}{2}}} Y_{lm}(\Omega_{\boldsymbol{r}-\boldsymbol{\tau}_{\alpha,n}}). \qquad (21)$$

Due to the short-ranged character of the nonlocal projectors, the integration is restricted to a small sphere around the atoms. Thus, assuming a constant density of the grid, the evaluation scales linearly with the number of atoms of the system.

The wave functions in reciprocal space $\langle \mathbf{k}+\mathbf{G}|\psi\rangle = \psi_{\mathbf{k}}(\mathbf{G})$ are stored on the reciprocal lattice sites $\mathbf{G}$ and the real-space representation is obtained by Fourier transformation

$$\begin{aligned}\psi_{\mathbf{k}}(\mathbf{r}) = \langle \mathbf{r}|\psi_{\mathbf{k}}\rangle &= \sum_{\mathbf{G}} \langle \mathbf{r}|\mathbf{k}+\mathbf{G}\rangle\langle \mathbf{k}+\mathbf{G}|\psi\rangle \\ &= \frac{1}{\sqrt{N_{\mathbf{G}}}} \sum_{\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} \psi_{\mathbf{k}}(\mathbf{G}) \\ &= e^{i\mathbf{k}\mathbf{r}} \frac{1}{\sqrt{N_{\mathbf{G}}}} \sum_{\mathbf{G}} e^{i\mathbf{G}\mathbf{r}} \psi_{\mathbf{k}}(\mathbf{G}). \end{aligned} \qquad (22)$$

Thus, in real-space treatments for nonzero $\boldsymbol{k}$ points, the phase appearing in the last line of Eq. (22) must be taken into account.

6. Spin-orbit coupling

The relativistic effect of the coupling of the orbital angular momentum of the electron and its spin [44] can be incorporated within the formulation of norm-conserving pseudopotentials [45,46]

$$\begin{aligned}\hat{V}_{\mathrm{ps}}^{\mathrm{ion}} = &\sum_{l,M} \left|l+\frac{1}{2},M\right\rangle V_{l,l+\frac{1}{2}}(r) \left\langle l+\frac{1}{2},M\right| \\ &+ \sum_{l,M'} \left|l-\frac{1}{2},M'\right\rangle V_{l,l-\frac{1}{2}}(r) \left\langle l-\frac{1}{2},M'\right|, \end{aligned} \qquad (23)$$

with M ranging from $-(l+\frac{1}{2})$ to $l+\frac{1}{2}$, M' ranging from $-(l-\frac{1}{2})$ to $l-\frac{1}{2}$, and

$$\left|l\pm\tfrac{1}{2},M\right\rangle = \frac{1}{\sqrt{2l+1}} \begin{pmatrix} \pm\sqrt{l\pm M+\frac{1}{2}} & \left|l,M-\frac{1}{2}\right\rangle \\ \sqrt{l\mp M+\frac{1}{2}} & \left|l,M+\frac{1}{2}\right\rangle \end{pmatrix} \qquad (24)$$

being the spin angular functions in the two-component spinor formalism. These pseudopotentials now depend on the joint orbital and spin angular momentum $j = l \pm s$. By defining an averaged l-dependent potential weighted by the different j degeneracies of the $|l \pm \frac{1}{2}\rangle$ states

$$\overline{V}_l(r) = \frac{1}{2l+1}\left(l V_{l,l-\frac{1}{2}}(r) + (l+1)V_{l,l+\frac{1}{2}}(r)\right) \qquad (25)$$

and a potential describing the difference in the potential with respect to the spin

$$V_l^{\mathrm{SO}}(r) = \frac{2}{2l+1}\left(V_{l,l+\frac{1}{2}}(r) - V_{l,l-\frac{1}{2}}(r)\right), \tag{26}$$

the total pseudopotential operator reads as

$$\hat{V}_{\mathrm{ps}}^{\mathrm{ion}} = \hat{V}_{\mathrm{NL}} + \hat{V}_{\mathrm{SO}} = \sum_{l,m} |l,m\rangle[\overline{V}_l(r) + V_l^{\mathrm{SO}}(r)\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}]\langle l,m|, \tag{27}$$

with m ranging from $-l$ to l. The first term corresponds to the scalar relativistic mass velocity and Darwin corrections, which are most often included in the available pseudopotentials. The latter term is associated with the spin-orbit coupling, which remains to be evaluated for a fully relativistic description.

In real space, a simple yet efficient treatment is implemented, which takes advantage of an expression encountered along the lines of the derivation of Eq. (27) from Eq. (23), which is demonstrated in the following. Reversing Eqs. (25) and (26),

$$V_{l,l+\frac{1}{2}}(r) = \overline{V}_l(r) + \frac{l}{2}V_l^{\mathrm{SO}}(r), \tag{28}$$

$$V_{l,l-\frac{1}{2}}(r) = \overline{V}_l(r) - \frac{l+1}{2}V_l^{\mathrm{SO}}(r), \tag{29}$$

the ionic pseudopotential can be rewritten to read as

$$\begin{aligned}\hat{V}_{\mathrm{ps}}^{\mathrm{ion}} &= \sum_{l,M}\left(\overline{V}_l(r) + \frac{l}{2}V_l^{\mathrm{SO}}(r)\right)\left|l+\tfrac{1}{2},M\right\rangle\left\langle l+\tfrac{1}{2},M\right| \\ &\quad + \sum_{l,M'}\left(\overline{V}_l(r) - \frac{l+1}{2}V_l^{\mathrm{SO}}(r)\right)\left|l-\tfrac{1}{2},M'\right\rangle \\ &\quad \times\left\langle l-\tfrac{1}{2},M'\right| \\ &= \sum_l \overline{V}_l(r)\Bigg(\sum_M \left|l+\tfrac{1}{2},M\right\rangle\left\langle l+\tfrac{1}{2},M\right| \\ &\quad + \sum_{M'}\left|l-\tfrac{1}{2},M\right\rangle\left\langle l-\tfrac{1}{2},M\right|\Bigg) \\ &\quad + \sum_{l,M}\frac{l}{2}V_l^{\mathrm{SO}}(r)\left|l+\tfrac{1}{2},M\right\rangle\left\langle l+\tfrac{1}{2},M\right| \\ &\quad - \sum_{l,M'}\frac{l+1}{2}V_l^{\mathrm{SO}}(r)\left|l-\tfrac{1}{2},M'\right\rangle\left\langle l-\tfrac{1}{2},M'\right|. \end{aligned} \tag{30}$$

As outlined in more detail in the Appendix, the relation of the operator in the basis of the $|l\pm\frac{1}{2},M\rangle$ in Eq. (30) and in the basis of the $|l,m\rangle\otimes|\pm\frac{1}{2}\rangle$ is given by

$$\begin{aligned}&\sum_M \left|l+\tfrac{1}{2},M\right\rangle\left\langle l+\tfrac{1}{2},M\right| + \sum_{M'}\left|l-\tfrac{1}{2},M'\right\rangle\left\langle l-\tfrac{1}{2},M'\right| \\ &\quad = \sum_m\left(\left|l,m,\tfrac{1}{2}\right\rangle\left\langle l,m,\tfrac{1}{2}\right| + \left|l,m,-\tfrac{1}{2}\right\rangle\left\langle l,m,-\tfrac{1}{2}\right|\right)\end{aligned} \tag{31}$$

with

$$\left|l,m,\tfrac{1}{2}\right\rangle \doteq \begin{pmatrix}|l,m\rangle\\0\end{pmatrix}, \tag{32}$$

$$\left|l,m,-\tfrac{1}{2}\right\rangle \doteq \begin{pmatrix}0\\|l,m\rangle\end{pmatrix}. \tag{33}$$

Here, the $|\pm\frac{1}{2}\rangle$ are the eigenstates of the $\hat{S}_z$ operator. Using the relation in Eq. (31) and the fact that

$$\begin{aligned}&\sum_m\left(\left|l,m,\tfrac{1}{2}\right\rangle\left\langle l,m,\tfrac{1}{2}\right| + \left|l,m,-\tfrac{1}{2}\right\rangle\left\langle l,m,-\tfrac{1}{2}\right|\right) \\ &\quad = \sum_m |l,m\rangle\langle l,m|\left(\left|\tfrac{1}{2}\right\rangle\left\langle\tfrac{1}{2}\right| + \left|-\tfrac{1}{2}\right\rangle\left\langle-\tfrac{1}{2}\right|\right)\end{aligned} \tag{34}$$

$$= \sum_m |l,m\rangle\langle l,m|\cdot\mathbb{1} \equiv \sum_m |l,m\rangle\langle l,m| \tag{35}$$

as well as the eigenvalue relation of the $\hat{\mathbf{L}}\cdot\hat{\mathbf{S}} = 1/2(\hat{\mathbf{J}}^2 - \hat{\mathbf{L}}^2 - \hat{\mathbf{S}}^2)$ operator applied on the spin angular functions

$$\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}|j,M\rangle = \begin{cases}\frac{l}{2}|l+\frac{1}{2},M\rangle & \text{for} \quad j = l+\frac{1}{2},\\ -\frac{l+1}{2}|l-\frac{1}{2},M\rangle & \text{for} \quad j = l-\frac{1}{2},\end{cases} \tag{36}$$

the potential can be further evaluated to read as

$$\begin{aligned}\hat{V}_{\mathrm{ps}}^{\mathrm{ion}} &= \sum_l \overline{V}_l(r)\sum_m\left(\left|l,m,\tfrac{1}{2}\right\rangle\left\langle l,m,\tfrac{1}{2}\right| + \left|l,m,-\tfrac{1}{2}\right\rangle\left\langle l,m,-\tfrac{1}{2}\right|\right) \\ &\quad + \sum_{l,M}\frac{l}{2}V_l^{\mathrm{SO}}(r)\left|l+\tfrac{1}{2},M\right\rangle\left\langle l+\tfrac{1}{2},M\right| - \sum_{l,M'}\frac{l+1}{2}V_l^{\mathrm{SO}}(r)\left|l-\tfrac{1}{2},M'\right\rangle\left\langle l-\tfrac{1}{2},M'\right|\end{aligned} \tag{37}$$

$$= \sum_{l,m}\overline{V}_l(r)|l,m\rangle\langle l,m| + \sum_l V_l^{\mathrm{SO}}(r)\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}\Bigg(\sum_M\left|l+\tfrac{1}{2},M\right\rangle\left\langle l+\tfrac{1}{2},M\right| + \sum_{M'}\left|l-\tfrac{1}{2},M'\right\rangle\left\langle l-\tfrac{1}{2},M'\right|\Bigg) \tag{38}$$

$$= \sum_{l,m}\left(\overline{V}_l(r) + V_l^{\mathrm{SO}}(r)\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}\right)|l,m\rangle\langle l,m|. \tag{39}$$

The summation indices m, M, and M' cover the values as before.

However, within the formalism suggested here and as implemented in LATEPP, the $\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}$ operator is never evaluated.

Instead, Eq. (37) is used. Moreover, the expression is extended to the fully nonlocal formulation of Kleinman and Bylander. As stated above, the scalar relativistic contributions are already contained in the nonlocal treatment and only the two remaining spin-orbit parts need to be considered, which are reformulated to read as

$$\hat{V}_{\mathrm{SO}} = \sum_{l} \left(\frac{l}{2} E_l \sum_{M=-(l+1/2)}^{l+1/2} \left|\chi^{\mathrm{SO}}_{l+1/2,M}\right\rangle\left\langle\chi^{\mathrm{SO}}_{l+1/2,M}\right| - \frac{l+1}{2} E_l \sum_{M=-(l-1/2)}^{l-1/2} \left|\chi^{\mathrm{SO}}_{l-1/2,M}\right\rangle\left\langle\chi^{\mathrm{SO}}_{l-1/2,M}\right| \right) \tag{40}$$

with

$$\left|\chi^{\mathrm{SO}}_{l\pm\frac{1}{2},M}\right\rangle = \frac{\left|R_{l\pm\frac{1}{2}} \quad V_l^{\mathrm{SO}} \quad l\pm\frac{1}{2}, M\right\rangle}{\left\langle u_{l\pm\frac{1}{2}} V_l^{\mathrm{SO}} \middle| u_{\pm\frac{1}{2}} l V_l^{\mathrm{SO}}\right\rangle^{\frac{1}{2}}}, \tag{41}$$

$$E_{l\pm\frac{1}{2}} = \frac{\left\langle u_{l\pm\frac{1}{2}} V_l^{\mathrm{SO}} \middle| u_{l\pm\frac{1}{2}} V_l^{\mathrm{SO}}\right\rangle}{\left\langle u_{l\pm\frac{1}{2}} \middle| V_l^{\mathrm{SO}} \middle| u_{l\pm\frac{1}{2}}\right\rangle}, \tag{42}$$

$$u_{l\pm\frac{1}{2}}(r) = R_{l\pm\frac{1}{2}}(r) r. \tag{43}$$

The present approach is different than the one used in other codes [47,48] that are based on a formulation of Hemstreet *et al.* [49], who apply the transformation according to Kleinman and Bylander already to Eq. (23) (after subtracting a local potential) followed by expressing the $\delta V_{l\pm1/2}(r) R_{l\pm1/2}(r)$ in terms of a scalar relativistic and spin-orbit difference part. This results in an intricate expression for the pseudopotential consisting of projectors constructed by all possible combinations of the scalar relativistic and spin-orbit difference parts appearing in the ket and bra states of the operator. The slim spin-orbit operator in the present approach comes at the small additional expense of a one-time initial evaluation of the angular functions of the second spinor component for the grid points within the atomic spheres [similar to Eq. (21)] for both the $|\chi^{\mathrm{SO}}_{l+1/2,M}\rangle$ and $|\chi^{\mathrm{SO}}_{l-1/2,M}\rangle$ states used in the projector. The big advantage, instead, is the circumvention of evaluating the angular and spin momentum operator. In the case of the implementation in Parsec [47], expressing $\hat{\mathbf{L}}\cdot\hat{\mathbf{S}} = \hat{L}_z\hat{S}_z + \frac{1}{2}(\hat{L}_+\hat{S}_- + \hat{L}_-\hat{S}_+)$, three of the six stencils describing the spin-orbit coupling become more complicated off-diagonal (in terms of the $|l,m\rangle$ basis) stencils due to the creation and annihilation operators. In ABINIT [50], evaluating $\hat{\mathbf{L}} = \hat{\mathbf{r}} \times \hat{\mathbf{p}}$ on the Legendre polynomial introduced by the vector addition theorem of the spherical harmonics, the formalism is no longer separable resulting in an inner loop over all grid points nested within an outer loop over all grid points of an atomic sphere.

Moreover, it is worth to note that the wrong normalization constants

$$C_{l\pm1/2} = \langle R_{l\pm1/2}|\delta V_{l\pm1/2}|R_{l\pm1/2}\rangle^{1/2} \tag{44}$$

are used in some of the previously stated formalisms [47–49], which instead should read as

$$C_{l\pm1/2} = \langle u_{l\pm1/2}|\delta V_{l\pm1/2}|u_{l\pm1/2}\rangle^{1/2} \tag{45}$$

due to an additional factor of r^2 originating from the integration in spherical coordinates. As later illustrated in Fig. 17 in Sec. III F, the introduced error is in the range of a few to 30 meV for the spin splitting in GaAs.

The $V_l^{\mathrm{SO}}(r)$ and $R_{l\pm\frac{1}{2}}$ of fully relativistic pseudopotentials are used. These are imported from files other than the files containing the scalar relativistic nonlocal pseudopotentials, which allows for a more flexible spin-orbit treatment.

Moreover, the size of the wave-function vector has to be doubled. Next to the spin-orbit treatment, the additional subspace is likewise considered in the kinetic energy as well as the local and nonlocal potential evaluation.

III. APPLICATIONS AND TESTS

To present the features of LATEPP and point out its ability to deal with nanoclusters, it is applied to GaAs quantum dots (QDs) embedded in AlAs. Next to some basic illustrative results on the electronic properties of these QDs, the calculations are meant to test the scaling and performance behavior of the code, which are supplemented by calculations of supercells of bulk GaAs of varying size. The capability of describing spin-orbit interactions is demonstrated in calculations of primitive cells of bulk GaAs.

For a comparison of AEPs with results from DFT, the reader is referred to the work introducing the AEPs [21], in which bulk structures as well as different types of quantum wells are studied in detail, demonstrating good accuracy and a high level of transferability of the potentials.

A. Eigensolver

The eigensolver, enabling access to a specific fraction of the eigenspectrum of the Hamiltonian, is one of the most important components of the machinery. As a first test, the number of iterations of the solver requesting 4 complete blocks of degenerate eigenvalues that consist of 14 eigenstates in total with respect to the specified reference energy is investigated on a bulk GaAs structure of 512 atoms. The number of basis vectors of the operating Krylov subspace is chosen to be four times larger than the number of requested states. The results are shown in Fig 1. Data points enclosed by a circle correspond to calculations yielding the requested blocks of solutions. They are all located in the shaded region marking the energy range for which the same solutions are expected. This is only true if the tolerance of the ARPACK solver is set close to machine precision. For larger tolerances, some of the solutions fail to appear, which is attributed to the high degree of degeneracy in the eigenspectrum of the large supercell of bulk GaAs. If the reference energy is shifted outside of this region, new states appear while others disappear from the solved spectrum, resulting in incomplete blocks of degenerate states, which are obviously harder to converge for ARPACK. This supports the finding [30] of optimal performance if complete clusters of eigenvalues are requested.

In a second test, the behavior of the solver with respect to the number of requested states is analyzed. To reduce the influence of clusters of degenerate eigenvalues, a large GaAs bulk structure consisting of 1728 atoms with a single distorted Ga atom is used as a test configuration. A second distortion

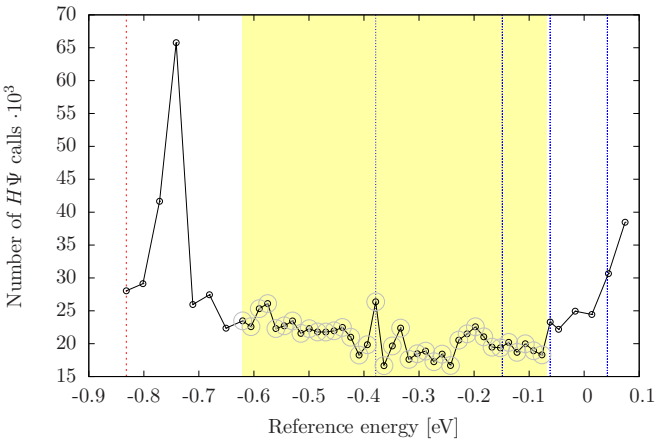

FIG. 1. (Color online) Number of iterations of the ARPACK solver requesting 14 states with respect to the utilized reference energy. The shaded region marks the range of reference energies for which the same solutions are expected. Data points supplied with a circle identify calculations that indeed contain these solutions. The blue lines correspond to the four blocks of eigenvalues of the requested states. The red line corresponds to the middle of the band gap.

different from the first one is investigated to exclude a possibly remaining high-symmetry configuration in the first case. For the second distortion, the behavior is shown for two different reference energies. The reference energies, in both cases, are chosen to be below the lowest eigenvalue of the spectrum to be solved. Figure 2 shows the number of iterations and total time of the ARPACK computation with respect to the number of requested states. No obvious dependency of the number of requested states can be identified. For the first distortion, there is a maximum at 19 states and no solution is found if 16 states are requested. Using a different distortion does not significantly improve the behavior. In fact, ARPACK still does not converge if 16 states are requested and again the maximum appears at 19 states with an almost identical number of required iterations. However, a variation of the reference

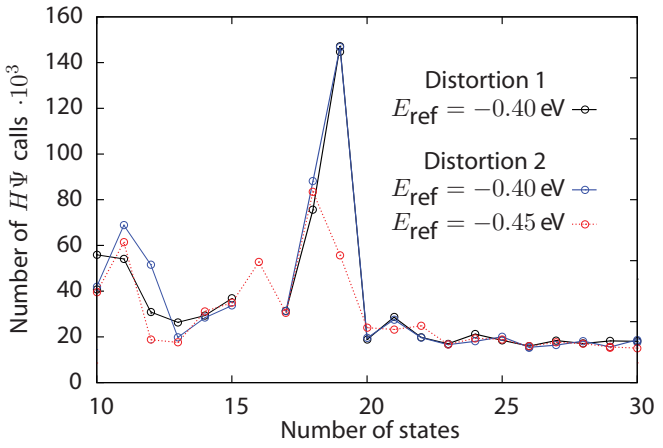

FIG. 2. (Color online) Number of iterations required to achieve convergence with respect to the number of requested states for a GaAs bulk structure consisting of 1728 atoms, for which two different distortions are applied to a single Ga atom of the structure.

energy changes this behavior. A solution is found if 16 states are requested and the former maximum number of iterations at 19 states is drastically reduced. This suggests the effective matrix $H - E_{\text{ref}}\mathbb{1}$ to be ill conditioned for the first reference energy E_{ref} of -0.40 eV in the respective subspace determined by the number of requested states. In the second case, for a reference energy of -0.45 eV, the effective matrix seems to be well behaved for all the respective subspaces. Unfortunately, a numerically cheap *a priori* test of the condition number of the matrix is not available. A pragmatic solution is a slight shift of the reference energy in cases where convergence cannot be achieved.

B. Numerical accuracy tests and optimization of computational parameters

Before moving on to the applications, the influence of the plane-wave energy cutoff, grid size, and polynomial order of the finite difference approximation are investigated.

1. Size of the FFT grid

In the reciprocal-space basis, the minimum grid needs to include all possible plane waves up to the cutoff energy. However, to include all possible reciprocal lattice vectors $\mathbf{G} - \mathbf{G}'$ as they occur in the expression of the local potential in Eq. (9), the size of the full grid must be chosen to range from $-2G_{\text{cut}}$ to $2G_{\text{cut}}$ in each dimension. Next to the local potential, the grid size affects the sampling of the projectors and the wave function of the nonlocal potential if evaluated in real space. More accurate sampling is naturally expected for higher grid densities. Both effects are illustrated in Fig. 3, which shows the convergence of the band gap of a bulk GaAs supercell containing eight atoms for two plane-wave energy cutoffs with respect to the utilized grid size in each dimension

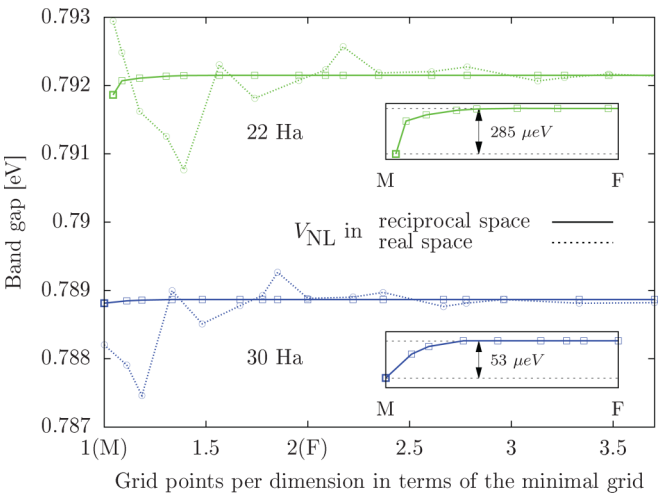

FIG. 3. (Color online) Convergence of the band-gap energy of a bulk GaAs supercell containing eight atoms for two plane-wave energy cutoffs (green: 22 Ha, blue: 30 Ha) with respect to the grid size in one dimension in terms of the minimum grid (M); the full grid is indicated by (F). Results are shown for calculations in the reciprocal-space basis with the nonlocal potential being evaluated in real space (dashed line) and reciprocal space (solid line).

for calculations in a reciprocal-space basis with the nonlocal potential being evaluated in real as well as in reciprocal space. Already the results of the 22 Ha calculations are considered converged employing a convergence tolerance of ±5 meV. In the case of the nonlocal potential evaluation in reciprocal space (solid line), the observed convergence is entirely governed by the local potential contribution. As can be seen from the insets, the contribution turns out to be small and already the minimum grid yields converged results. If evaluated in real space, a more distinct convergence characteristic represented by the dashed lines is observed. This is due to the real-space sampling of the nonlocal projectors and the wave function, which are refined and become more accurate with an increasing number of grid points. Still, the deviations from the respectively converged band gap are always below 2 meV, which suggests a minimum grid ranging from $-G_{\text{cut}}$ to $+G_{\text{cut}}$ to be sufficient for proper sampling.

2. *Comparison of the real and reciprocal space implementations*

To compare calculations that are entirely carried out in real space with calculations performed in the reciprocal-space basis, the previous finding of a sufficient minimum grid is used to relate the number of grid points per dimension N_d with the plane-wave energy cutoff E_{cut}, i.e.,

$$E_{\text{cut}} = \frac{1}{2}\left(\frac{N_d \pi}{a}\right)^2. \tag{46}$$

A comparison using the respective mapping is displayed in Fig. 4, which shows the convergence of the band gap with respect to the actual and equivalent plane-wave cutoff energy. As mentioned before, converged band-gap energies are obtained beyond 22 Ha in the plane-wave basis (red and green lines). The full real-space treatment (blue) converges slightly later at equivalent 25 Ha due to the additional approximation

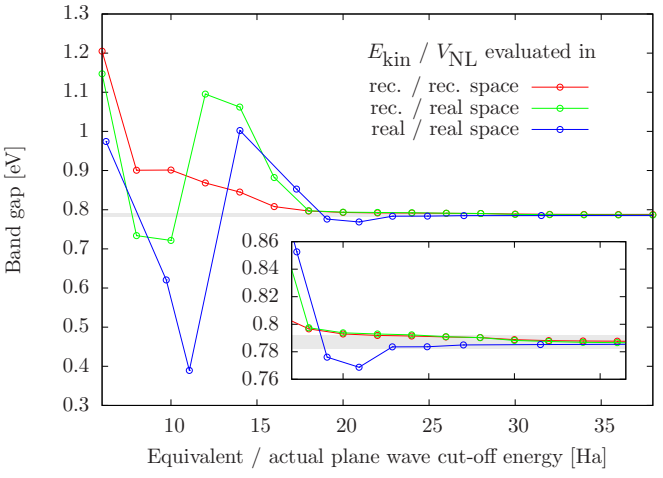

FIG. 4. (Color online) Convergence of the band-gap energy of a bulk GaAs supercell containing eight atoms with respect to the plane-wave cutoff energy. For the full real-space calculations, the equivalent cutoff energy is calculated according to Eq. (46). The kinetic energy is calculated in real (blue) or reciprocal space (red or green) and the nonlocal potential in real (blue or green) or reciprocal (red) space.

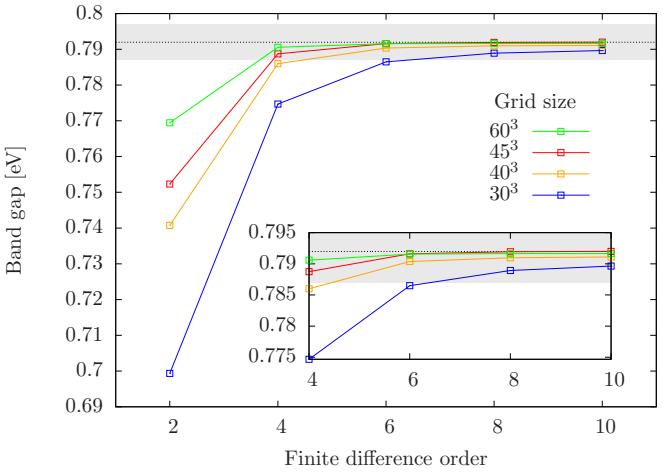

FIG. 5. (Color online) Convergence of the band-gap energy of a bulk GaAs supercell containing eight atoms with respect to the polynomial order of the finite difference scheme in the fully real-space calculations. Results for different grid sizes are shown.

in the kinetic energy evaluation, which depends on the grid density, as discussed in the following.

3. *Polynomial order of the finite difference approximation*

The band gap with respect to the polynomial order according to Eq. (8) is shown in Fig. 5. As expected, the error introduced by the finite difference scheme decreases with increasing polynomial order. Moreover, the error for a fixed value of the polynomial order likewise decreases with increasing grid size. The remainder of the finite difference expansion obviously depends directly on the grid spacing, which becomes smaller for denser grids. Because the number of operations to calculate the kinetic energy per real-space grid point depends linearly on the polynomial order O, but the number of real-space grid points N itself scales cubic with the number of grid points in each direction N_d, it is advised to use big enough polynomial orders in favor of smaller grid sizes and, equivalently, vector sizes. For instance, similar accuracy is obtained for grid sizes $N_d = 40$, $O = 4$ and $N_d = 30$, $O = 6$, which involve $3 \times 4 \times 40^3 = 768\,000$ and $3 \times 6 \times 30^3 = 486\,000$ operations, respectively. Next to the smaller number of required operations, the vector size of $30^3 = 27\,000$ in contrast to $40^3 = 64\,000$ grid points is much smaller in the latter approach.

C. Eigenstates and wave functions of GaAs quantum dots in AlAs

Spherical GaAs QDs are embedded in a cubic supercell of AlAs. The AEPs for Ga are derived from an *ab initio* pseudopotential for Ga not considering the $3d$ electrons as valence electrons. The lattice constant of the GaAs QD as well as the AlAs bulk is set to an averaged experimental value of 10.68 a_0. The equilibrium local density approximation (LDA) lattice constants for GaAs and AlAs are 10.13 a_0 and 10.72 a_0, respectively. Different sizes of the QD ranging from 1 up to 12 nm corresponding to 2 up to 21 lattice constants in diameter are investigated. A minimum distance

TABLE I. QD diameter and size of the cubic supercells including the number of atoms for three different GaAs QDs embedded in AlAs with $a = 10.68\,a_0$ being the averaged experimental lattice constant. The last rows show the utilized LATEPP grid size and the full grid size (in brackets) to contain all possible reciprocal lattice vectors $\mathbf{G} - \mathbf{G}'$ with respect to the cutoff energy as well as the number of plane waves.

Supercell size (a^3)	12^3	20^3	23^3
Dot diameter (a)	10	18	21
Number of Al atoms	4800	19 768	29 156
Number of Ga atoms	2112	12 232	19 512
Number of As atoms	6912	32 000	48 668
Total number of atoms	13 824	64 000	97 336
LATEPP grid size	200^3 (365^3)	240^3 (469^3)	288^3 (541^3)
Plane wave cut-off (Ha)	10	6	6
Number of plane waves	9 545 913	6 847 715	31 243 629

of two lattice constants between the periodically repeated QDs is sufficient to isolate them electronically. The respective dimensions and parameters of interest of the three largest structures are summarized in Table I. The reciprocal-space basis is chosen for all calculations with the nonlocal potential being evaluated in real space. The calculations are performed at the Γ point, utilizing a plane-wave cutoff energy of 10 Ha. For the larger supercell calculations, consisting of more than 10 000 atoms, a reduced cutoff energy of 6 Ha has been used.

In Fig. 6, the band structures of bulk GaAs and AlAs at the averaged experimental lattice constant are shown. The direct band gap of GaAs is 0.90 eV, which is lower than the indirect gap of 1.33 eV for AlAs, which involves a transition from the X point in the conduction to the Γ point in the valence band.

Solutions in the vicinity of the valence band maximum (VBM) and conduction band minimum (CBM) are obtained separately by two independent calculations with a different reference energy. Considering the band gaps and band offsets obtained from the bulk calculations, the wave functions of the highest occupied and lowest unoccupied states in the QD structure are, in a first quick thought, expected to be localized within the GaAs QD. In fact, this is true for the VBM as can be seen in Fig. 7, which displays the wave functions of the largest of the investigated QD structures consisting of $23 \times 23 \times 23$ lattice constants corresponding to 97 336 atoms. The red wave function corresponds to the highest occupied state, which is localized in the GaAs QD. In contrast, the lowest unoccupied state turns out to be localized within the AlAs

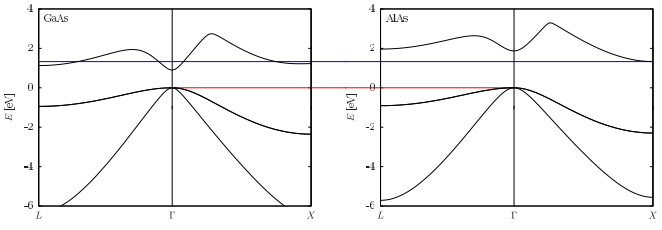

FIG. 6. (Color online) Excerpt of the electronic band structure of the highest occupied p and the lowest unoccupied s state of GaAs (left) and AlAs (right) bulk structures.

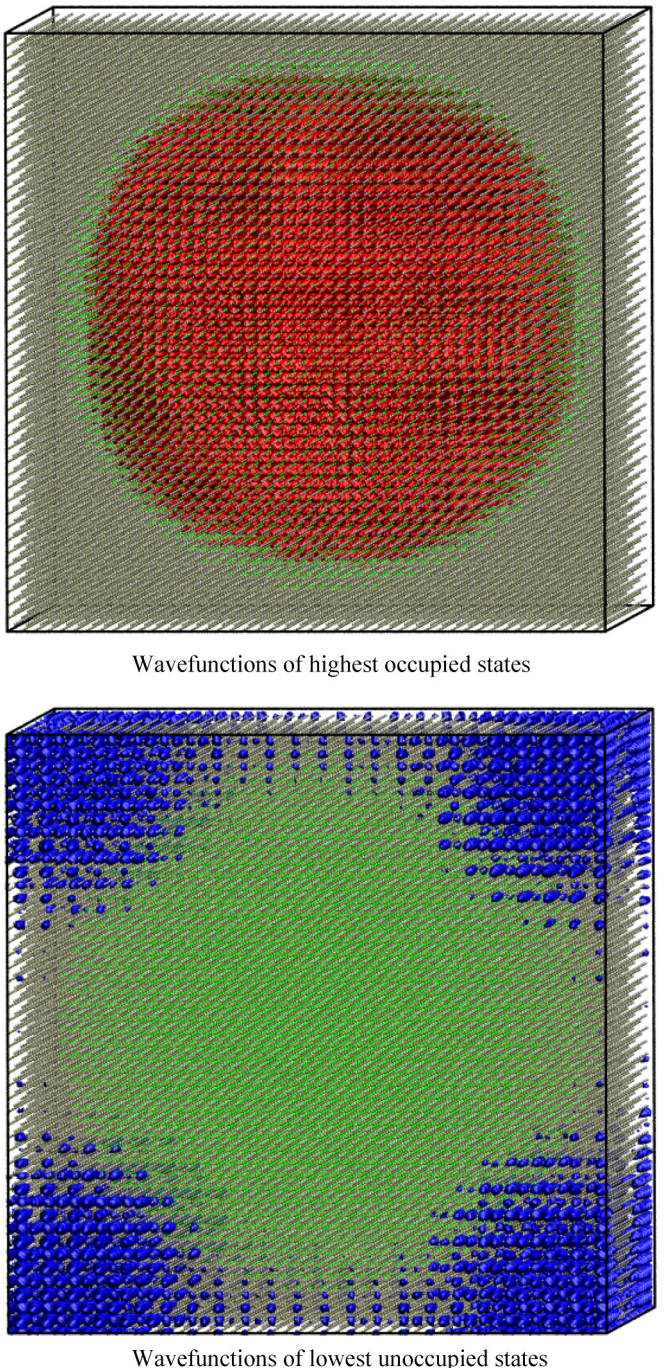

FIG. 7. (Color online) Squared wave function in real space of the highest occupied (top, red) and lowest unoccupied (bottom, blue) states of a GaAs QD embedded in AlAs. Green, tan, and silver spheres correspond to Ga, Al, and As atoms. The structure consists of 97 336 atoms.

barrier as indicated by the blue isosurface. Using an empirical pseudopotential method [51], a direct-to-indirect transition in real space was observed for this material system if the radius of the GaAs QD decreases below 5 nm. Due to confinement, the lowest unoccupied molecular orbital state of the QD shifts above the CBM of the barrier, with the latter originating from the X point in the Brillouin zone. Thus, the band gap is indirect in reciprocal as well as in real space, which is in agreement

with the present band-structure calculation of bulk AlAs in Fig. 6 determining the CBM at the X point.

Although these results are in qualitative agreement to the EPM calculations, they remain quantitatively inaccurate due to the deficiencies of the local density approximation, which severely underestimates band gaps and effective masses. A correction to the band gap is therefore required to compare to experiment. Although possible in LATEPP, this is not the scope of this paper.

D. Parallelization

Parts of the LATEPP algorithms are parallelized using the shared memory OpenMP standard [52]. To test its performance and scalability, QD as well as bulk structures of different sizes are solved on one compute node containing two quad-core CPUs [Intel(R) Xeon(R) E5540 2.53 GHz] equipped with 24-GB of memory up to a maximum number of eight threads. Furthermore, different versions of the numerical libraries, i.e., ARPACK [29,30] for the iterative solver and the Intel Math Kernel Library providing optimized LAPACK [53] and BLAS [54] routines, were tested.

Figure 8 shows the total calculation time of the smallest QD structure consisting of 512 atoms for all possible combinations of employed libraries with respect to the number of threads used in the calculation. The solid line corresponds to calculations using ARPACK routines utilizing multithreaded BLAS routines provided by the Intel MKL, whereas the dashed lines utilize the serial BLAS routines. Circles and triangles represent calculations of LATEPP executables linked against the sequential and multithreaded MKL library, respectively. A linear relation of the calculation time with respect to the inverse of the number of threads with almost identical slope can be identified for all combinations. Clearly, the multithreaded ARPACK setups outperform its sequential variants for runs on more than one thread. The combination of the preferential multithreaded ARPACK linked against the sequential version of the MKL yields shorter calculation times on less or equal

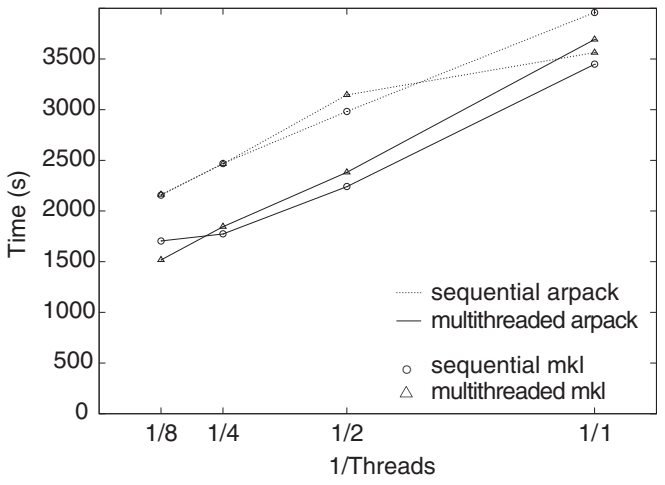

FIG. 8. Total calculation time for a QD structure consisting of $4 \times 4 \times 4$ lattice constants containing 512 atoms with respect to the number of threads. Results of four different combinations of libraries utilized in LATEPP are shown.

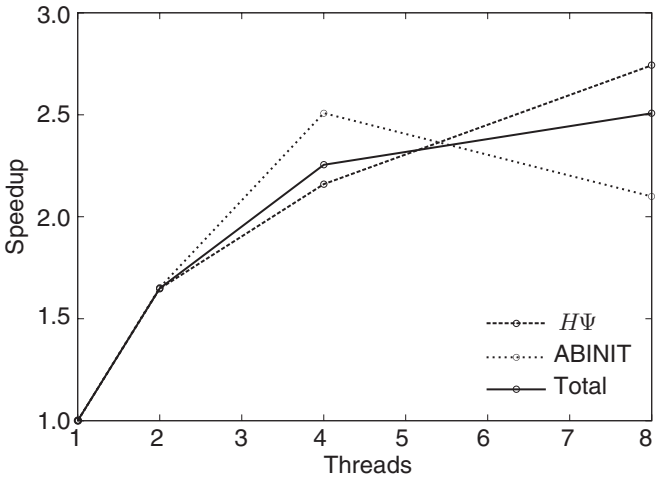

FIG. 9. Speedup of LATEPP gained by parallelization with respect to the number of threads. Here, FFTE [38] is used for the transformations.

than four threads. However, the continuous linear decrease in calculation time for runs on more than four threads suggests to use the multithreaded versions of both libraries.

In Fig. 9, the speedup gained by parallelization with respect to the number of threads is plotted for calculations of the smallest QD structure using FFTE [38] for transformations between real and reciprocal space. Speedups are additionally shown for the two parts of the iterative procedure, i.e., the evaluation of $H\psi$ and the ARPACK algorithms realizing the implicitly restarted Arnoldi method. Although the gain in speed of the ARPACK routines increases up to a number of four threads, the external solver library seems to behave badly and even slows down if a higher number of threads is used. In contrast, the implemented evaluation of $H\psi$ shows continuously increasing speedups. Since the ARPACK part only occupies a relatively small fraction of the iteration time, the total speedup likewise shows an upward trend.

Although these are modest speedups, they are not unusual for an OpenMP parallelization on symmetric multiprocessing systems due to limitations introduced by the memory architecture. The advantage, on the other hand, is the simplicity of developing applications and adding new functionalities. The advantage of LATEPP is the possibility to address large structures on a single node or even on inexpensive desktop hardware.

E. Scaling

The capability of solving large-scale problems is illustrated and supported by the scaling behavior with respect to the system size. The time for one iteration step in the iterative eigensolver as a function of the number of atoms for bulk GaAs structures is displayed in Fig. 10. The times are averaged over 20 iterations and 20 eigenstates are requested. Calculations in reciprocal space with the nonlocal potential being evaluated in real space (solid lines) are compared to fully real-space calculations (dashed lines). A plane-wave cutoff energy of 22 Ha is used for the reciprocal-space basis. The corresponding size of the minimum grid in the real-space basis is determined by Eq. (46). The total time (black lines) is composed of

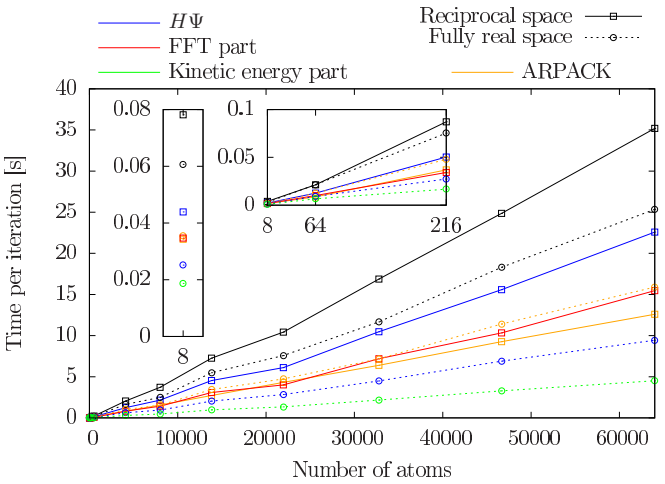

FIG. 10. (Color online) Time for one iteration (black lines) for bulk GaAs structures with respect to the number of atoms. Results are shown for fully real-space calculations (dashed lines) as well as for calculations in the reciprocal-space basis with the nonlocal potential being evaluated in real space (solid lines). The total time is composed of the time of the $H\psi$ evaluation (blue lines) and the ARPACK routine (orange lines). The treatment in reciprocal space requires FFTs (red line) while a more computationally intensive kinetic energy evaluation (green line) is used in the real-space treatment. A minimum grid (see Fig. 3) and 22 Ha cutoff energy for the plane waves are used.

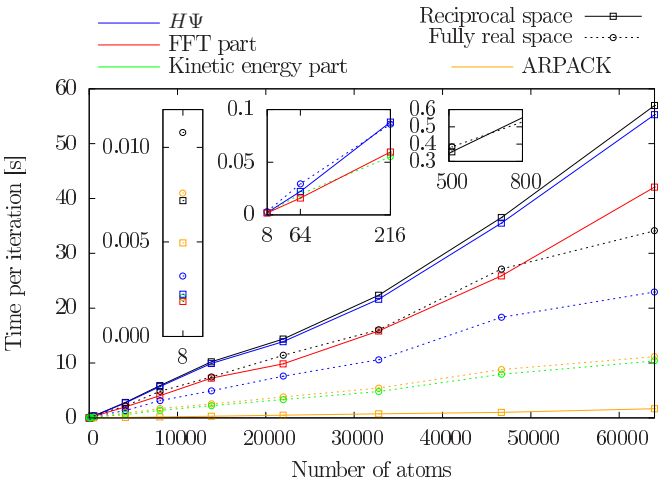

FIG. 11. (Color online) Time for the first iteration of fully real-space calculations (dashed line) and calculations in the reciprocal-space basis with the nonlocal potential being evaluated in real space (solid line) for different sizes of bulk GaAs structures. The insets show magnified regions of the plot. The contribution of the $H\psi$ evaluation and the ARPACK routine is shown by the blue and orange lines, respectively. The contributions of the FFT and kinetic energy calculation within the $H\psi$ evaluation are highlighted by red and green colors.

the $H\Psi$ (blue lines) and ARPACK (orange lines) operations. Contributions to the $H\Psi$ routine in turn are the FFTs (red line) and the kinetic energy evaluation (green line) as well as the local and nonlocal potentials that are evaluated in real space in both methods, which take the same amount of time and, therefore, need not be compared with each other.

The FFTs are only required in calculations in the reciprocal-space basis to transform the wave functions into real space for evaluating the local and nonlocal potential. The kinetic energy evaluated in reciprocal space is negligible and not shown.

In contrast, the fully real-space treatment does not require any FFT but needs a more computationally intensive kinetic energy evaluation and, in addition, exhibits a larger vector size by a factor of $6/\pi$.

As suggested from the graph, the fully real-space treatment is always faster, although more time is spent for the ARPACK routine due to the larger vector size. This can be traced back to the FFTs, that are slower than the real-space kinetic energy evaluation.

In both, the reciprocal space as well as the fully real-space case, the total computational costs as well as the respective contributions scale linearly with the number of atoms N of the system. Unexpectedly, this is likewise true for the FFT, which is expected to obey a complexity of $O[N\log(N)]$. Obviously, within the investigated size regime, the FFT operation is likewise proportional to the number of atoms. In fact, minimum FFT grids according to Eq. (46) corresponding to 480^3 grid points for the structure containing 64 000 atoms are used, which result in an overall linear scaling behavior.

In contrast, Fig. 11 shows the time per iteration for a cutoff energy of 10 Ha but using the full FFT grid, i.e., 625^3 grid points for the structure consisting of 64 000 atoms. Due to the increased grid size, the expected $O[N\log(N)]$ behavior of the FFTs is noticeable.

Linear scaling is preserved in the fully real-space treatment. However, using a full grid, the vector size of the real-space basis is now $48/\pi$ times larger than the reciprocal-space basis, which is penalized by a significantly slower ARPACK routine (compare dashed and solid orange lines). In fact, the fully real-space approach is less favorable for the eight-atom case as can be seen in the left inset of Fig. 11. This is due to slower ARPACK routines that operate on a much longer vector and to slightly slower kinetic energy evaluations in real space (compare green and red FFT times). However, for a critical number of atoms between 64 and 216, the kinetic energy calculation in real space becomes faster than the FFTs. *Finally, the overall $H\Psi$ multiplication in the real-space basis becomes advantageous for system sizes exceeding 512 atoms.*

However, these findings must be taken with care. The increased vector size typically requires a much larger number of iterations to reach convergence and the reciprocal-space basis remains a competitive approach in LATEPP.

To summarize, although the code does not show demonstrative speedups gained by parallelization, the linear scaling behavior is promising and a proper MPI implementation could take full advantage of this fact. As shown, LATEPP is able to describe structures consisting of 100 000 atoms at a level of accuracy very comparable to DFT on a single node, possibly a modern desktop computer.

F. Spin-orbit splitting in GaAs

The new implementation of the spin-orbit interaction is demonstrated on bulk GaAs. The investigations are restricted to the real-space spin-orbit treatment, which does not introduce any empirical character and provides a computational advan-

tage, thus, constituting the preferred method. For the scalar relativistic contribution, the AEPs and nonlocal pseudopotentials for Ga considering $4s$, $4p$, and $3d$ states are used. Likewise, for the spin-orbit coupling, the difference potentials of Eq. (26) of the $4p$ and $3d$ states are employed. For As, an AEP derived from a pseudopotential considering only $4s$ and $4p$ states is used. For the spin-orbit potentials, the unbound $4d$ state must be considered in addition to the $4p$ state. The APE [55] package is used for the derivation of the *ab initio* difference potentials. The radial cutoff in the psuedopotential generation for the $4d$ state was set to 2.3 a_0 to avoid overlapping of the atomic spheres.

For comparison, the ABINIT code is used, which, in its present version, is restricted to the usage of the HGH pseudopotentials [56] if spin-orbit interaction is considered using norm-conserving pseudpotentials. The equilibrium lattice constant of the HGH pseudopotential for GaAs is 10.576 a_0, in rather good agreement with experiment. The determined lattice constant of 10.599 a_0 for the AEP to match the HGH band gap is slightly increased with respect to the equilibrium lattice constant of 10.596 a_0 used in the DFT calculations to generate the GaAs AEP [21].

The band structures generated with ABINIT and LATEPP are shown in Fig. 12. The band structures are aligned with respect to the VBM. Due to the difference of the utilized pseudopotentials, an exact agreement, even of the band structures not considering spin-orbit effects, cannot be expected. However, an excellent agreement of the band structure over the whole path calculated by DFT using HGH pseudopotentials and by LATEPP employing the AEPs and spin-orbit difference potentials obtained from APE is found. Next to the lifting of the degeneracy of the split-off and the light-/heavy-hole bands at the Γ point, the characteristic splitting of the heavy- and light-hole bands approaching the L and X points is very well described. Moreover, the spin splitting of the two spin components represented by dashed and solid lines along the Σ direction between the X and Γ points is excellently captured within the utilized spin-orbit treatment.

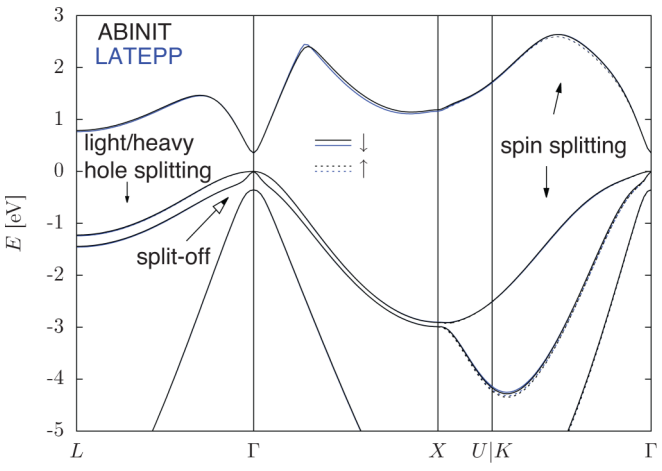

FIG. 12. (Color online) Electronic band structure of GaAs illustrating the lifting of the degeneracy of the split-off and heavy-/light-hole band at the Γ point as well as the characteristic splitting of the heavy- and light-hole bands approaching the L and X points due to spin-orbit interaction. In addition, the spin splitting of the two spin components (dashed and solid lines) between the X and Γ points along the [1 1 0] can be seen. The results of LATEPP (blue lines) are compared to DFT results using ABINIT employing the HGH [56] pseudopotentials (black lines).

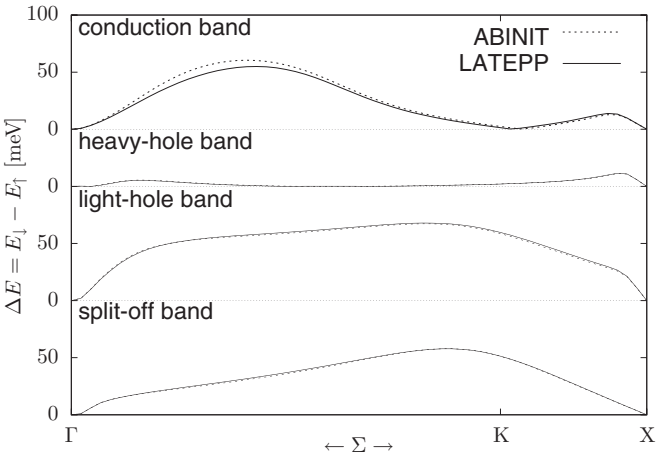

FIG. 13. Lifting of the light-/heavy-hole degeneracy in GaAs due to spin-orbit interaction. The results of LATEPP (solid line) are compared to a DFT result using ABINIT employing the HGH [56] pseudopotentials (dashed line).

The lifting of the light-/heavy-hole degeneracy is shown once more on a larger scale in Fig. 13. The LATEPP and DFT results are in good agreement. The whole path from L to Γ is very accurately reproduced. The character of the splitting, i.e., the maxima as well as the curvature of the splitting of the two bands is very well reproduced.

Spin splitting is pronounced along the Σ direction starting from the Γ point. To a lesser extent, it also also occurs between L and Γ along the Λ ([1 1 1]) direction. The difference in energy of the two spin components between Γ and X is

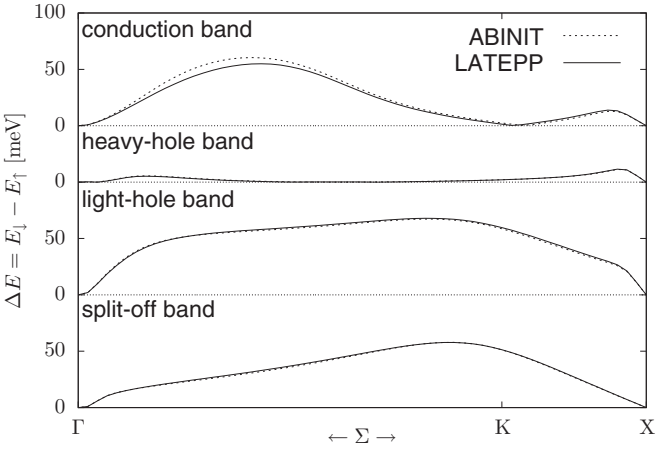

FIG. 14. Spin splitting in the valence and conduction bands of GaAs between the Γ and X points along the [1 1 0] direction due to spin-orbit interaction. Results of LATEPP (solid lines) are compared to DFT results using ABINIT employing the HGH [56] potentials (dashed lines).

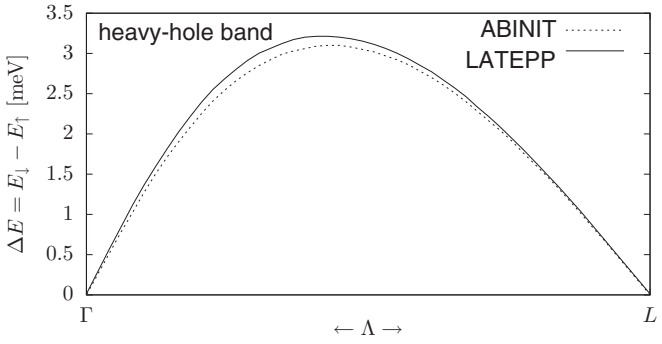

FIG. 15. Spin splitting in the heavy-hole band of GaAs between the L and Γ points along the Λ direction due to spin-orbit interaction. The result of LATEPP (solid line) is compared to the DFT result using ABINIT employing the HGH [56] potentials (dashed line).

displayed in Fig. 14 for the conduction and the three highest valence bands. Figure 15 shows the respective splitting of the heavy-hole band between Γ and L. The spin splitting is very well reproduced by the spin-orbit treatment of LATEPP, although two different pseudopotentials were used in the calculations.

In Fig. 16, the spin splitting of the heavy-hole band from Γ to X along the Σ direction is displayed separately including a magnified region near the Γ point illustrating the capability of describing very accurately the characteristic bump and its respective linear and cubic dependence on the distance from the band extremum [57,58].

Finally, the importance of the correct normalization of the spin-orbit projectors is illustrated. As shown in Fig. 17, the error introduced by the wrong normalization constant as reported in Refs. [47–49] is in the range of a few to 30 meV for the spin splitting in GaAs.

In summary, these results show that the treatment of spin-orbit interaction as implemented in LATEPP can be considered suitable to properly account for spin-orbit coupling.

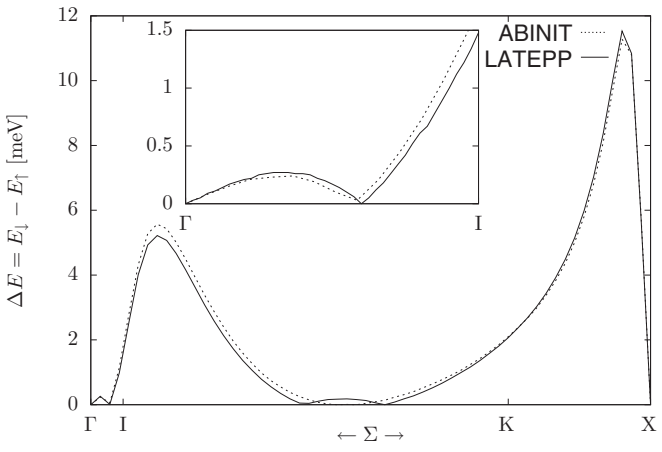

FIG. 16. Spin splitting of the GaAs heavy-hole band from Γ to X along Σ direction. The inset shows a magnified region near the Γ point. Results of LATEPP (solid line) are compared to results of ABINIT employing the HGH [56] potentials (dashed line).

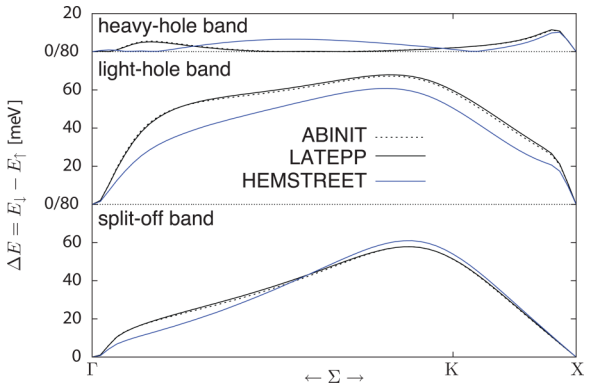

FIG. 17. (Color online) Spin splitting in the valence bands of GaAs between the Γ and X points along the [1 1 0] direction due to spin-orbit interaction. Results of LATEPP (solid lines) are compared to DFT results using ABINIT employing the HGH [56] potentials (dashed lines). In addition, results with the wrong normalization as given in Refs. [47–49] are shown (blue lines).

IV. CONCLUSION AND SUMMARY

The LATEPP code solving the Schrödinger equation of an electronic system has been introduced. It uses atomic effective pseudopotentials derived from self-consistent *ab initio* DFT calculations. A self-consistent cycle is not required, which reduces computational costs and allows for the calculation of only a few eigenstates of interest. Thus, LATEPP is specifically suited for the investigation of the electronic structure in the vicinity of the band gap of semiconductor nanostructures. These are intended to be used in a subsequent calculation of many-body effects using configuration interaction [18]. As has been shown, structures containing up to 100 000 atoms can be treated at an atomistic *ab initio* level comparable to DFT, with the respective approximation for exchange and correlation on a single node.

The possibility to represent the wave function and to evaluate parts of the Hamiltonian either in a plane wave or real-space basis allowed for a coherent analysis and direct comparison of different approaches. Among the different possible combinations, the reciprocal space method with the local and nonlocal pseudopotential being evaluated in real space as well as the fully real-space method constitute the most competitive procedures.

Linear scaling with the system size is achieved in the fully real-space treatment for all investigated system sizes.

The minimum FFT grid, which is just large enough to contain the plane-wave sphere while truncating high-frequency components of the local potential, is shown to be sufficient to reach an accuracy below ±2 meV in the eigenvalues. Using this minimum grid we obtain for calculations up to 64 000 atoms using the reciprocal space basis a nearly linear scaling, despite the $O[N \log(N)]$ scaling of the FFTs.

We showed that results converged to the same accuracy (below ±2 meV) can be obtained in the fully real-space method using a grid that is equivalent to the minimum FFT grid. Moreover, the convergence with respect to the polynomial order of the finite difference approximation for the kinetic energy in real space is discussed and a polynomial order of six is found to be ideal.

Finally, an efficient real-space implementation of spin-orbit interaction based on the formulation within the basis of spin angular functions is presented. This implementation, which does not require the evaluation of the product of angular momentum and spin operator, is different from other codes. Direct comparison with standard DFT codes show that spin-orbit interaction in our formalism is accurately accounted for.

ACKNOWLEDGMENTS

We acknowledge funding by the DFG (Project No. BE 4292/2). We are grateful to L.-W. Wang for interesting and fruitful discussions concerning the treatment of spin-orbit interaction.

APPENDIX: CHANGE OF BASIS

To prove the equivalence of Eqs. (23) and (27), the basis transformation

$$\left|l \pm \tfrac{1}{2}, M\right\rangle \rightarrow |l, m\rangle \otimes \left|\pm\tfrac{1}{2}\right\rangle \tag{A1}$$

to obtain Eq. (31) has yet to be shown. The Clebsch-Gordon coefficients

$$C^{+}_{l+\frac{1}{2},M} = \frac{\sqrt{l+M+\frac{1}{2}}}{\sqrt{2l+1}}, \quad C^{-}_{l+\frac{1}{2},M} = \frac{\sqrt{l-M+\frac{1}{2}}}{\sqrt{2l+1}},$$

$$C^{+}_{l-\frac{1}{2},M} = \frac{-\sqrt{l-M+\frac{1}{2}}}{\sqrt{2l+1}}, \quad C^{-}_{l-\frac{1}{2},M} = \frac{\sqrt{l+M+\frac{1}{2}}}{\sqrt{2l+1}}$$

connecting the spin angular functions as defined in Eq. (24) with the spherical harmonics have the following properties:

$$C^{+}_{l+\frac{1}{2},M} C^{-}_{l+\frac{1}{2},M} = -C^{+}_{l-\frac{1}{2},M} C^{-}_{l-\frac{1}{2},M}, \tag{A2}$$

$$C^{\pm}_{l+\frac{1}{2},M}{}^{2} + C^{\pm}_{l-\frac{1}{2},M}{}^{2} = 1. \tag{A3}$$

Writing short $|\pm\rangle = |\pm\tfrac{1}{2}\rangle$ and assuming spinor notation

$$|\alpha\rangle = |+\rangle\langle+|\alpha\rangle + |-\rangle\langle-|\alpha\rangle = \alpha^{+}|+\rangle + \alpha^{-}|-\rangle$$
$$\doteq \begin{pmatrix}\langle+|\alpha\rangle\\ \langle-|\alpha\rangle\end{pmatrix} = \begin{pmatrix}\alpha^{+}\\ \alpha^{-}\end{pmatrix}, \tag{A4}$$

the projectors constructed out of the spin angular functions for a specific quantum number M are

$$\begin{aligned}\left|l+\tfrac{1}{2},M\right\rangle\!\left\langle l+\tfrac{1}{2},M\right| &= \left(C^{+}_{l+\frac{1}{2},M}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle + C^{-}_{l+\frac{1}{2},M}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\right)\left(\langle+|\!\left\langle l,M-\tfrac{1}{2}\right|C^{+}_{l+\frac{1}{2},M} + \langle-|\!\left\langle l,M+\tfrac{1}{2}\right|C^{-}_{l+\frac{1}{2},M}\right)\\ &= C^{+}_{l+\frac{1}{2},M}{}^{2}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + C^{+}_{l+\frac{1}{2},M}C^{-}_{l+\frac{1}{2},M}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|\\ &\quad + C^{-}_{l+\frac{1}{2},M}C^{+}_{l+\frac{1}{2},M}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + C^{-}_{l+\frac{1}{2},M}{}^{2}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|,\end{aligned} \tag{A5}$$

$$\begin{aligned}\left|l-\tfrac{1}{2},M\right\rangle\!\left\langle l-\tfrac{1}{2},M\right| &= \left(C^{+}_{l-\frac{1}{2},M}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle + C^{-}_{l-\frac{1}{2},M}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\right)\left(\langle+|\!\left\langle l,M-\tfrac{1}{2}\right|C^{+}_{l-\frac{1}{2},M} + \langle-|\!\left\langle l,M+\tfrac{1}{2}\right|C^{-}_{l-\frac{1}{2},M}\right)\\ &= C^{+}_{l-\frac{1}{2},M}{}^{2}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + C^{+}_{l-\frac{1}{2},M}C^{-}_{l-\frac{1}{2},M}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|\\ &\quad + C^{-}_{l-\frac{1}{2},M}C^{+}_{l-\frac{1}{2},M}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + C^{-}_{l-\frac{1}{2},M}{}^{2}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|.\end{aligned} \tag{A6}$$

Due to the properties of the coefficients [Eqs. (A2) and (A3)], the sum of both parts becomes

$$\begin{aligned}&\left|l+\tfrac{1}{2},M\right\rangle\!\left\langle l+\tfrac{1}{2},M\right| + \left|l-\tfrac{1}{2},M\right\rangle\!\left\langle l-\tfrac{1}{2},M\right|\\ &= C^{+}_{l+\frac{1}{2},M}{}^{2}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + C^{-}_{l+\frac{1}{2},M}{}^{2}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|\\ &\quad + C^{+}_{l-\frac{1}{2},M}{}^{2}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + C^{-}_{l-\frac{1}{2},M}{}^{2}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|\\ &= C^{+}_{l+\frac{1}{2},M}{}^{2}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + C^{+}_{l-\frac{1}{2},M}{}^{2}\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right|\\ &\quad + C^{-}_{l+\frac{1}{2},M}{}^{2}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right| + C^{-}_{l-\frac{1}{2},M}{}^{2}\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|\\ &= \left(C^{+}_{l+\frac{1}{2},M}{}^{2} + C^{+}_{l-\frac{1}{2},M}{}^{2}\right)\left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + \left(C^{-}_{l+\frac{1}{2},M}{}^{2} + C^{-}_{l-\frac{1}{2},M}{}^{2}\right)\left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|\\ &= \left|l,M-\tfrac{1}{2}\right\rangle|+\rangle\langle+|\!\left\langle l,M-\tfrac{1}{2}\right| + \left|l,M+\tfrac{1}{2}\right\rangle|-\rangle\langle-|\!\left\langle l,M+\tfrac{1}{2}\right|.\end{aligned} \tag{A7}$$

The off-diagonal contributions $|\ldots\rangle|\pm\rangle\langle\mp|\langle\ldots|$ with respect to spin vanish due to Eq. (A2) and the prefactors of the diagonal contributions $|\ldots\rangle|\pm\rangle\langle\pm|\langle\ldots|$ sum up to one [Eq. (A3)]. Finally, taking the sum over all M populates both spinor components with all possible magnetic quantum numbers m of the spherical harmonic of angular momentum l. Then, writing $|l,m\rangle|\pm\rangle = |l,m,\pm\rangle$

proves the basis transformation as outlined in Eq. (31). The closure relation

$$\mathbb{1} = |+\rangle\langle+| + |-\rangle\langle-| \tag{A8}$$

is used to obtain Eq. (35):

$$\begin{aligned} |l,m,+\rangle\langle l,m,+| + |l,m,-\rangle\langle l,m,-| &= (|l,m\rangle \otimes |+\rangle)(\langle l,m| \otimes \langle+|) + (|l,m\rangle \otimes |-\rangle)(\langle l,m| \otimes \langle-|) \\ &= |l,m\rangle\langle l,m| \otimes |+\rangle\langle+| + |l,m\rangle\langle l,m| \otimes |-\rangle\langle-| \\ &= |l,m\rangle\langle l,m| \otimes (|+\rangle\langle+| + |-\rangle\langle-|) \\ &= |l,m\rangle\langle l,m| \otimes \mathbb{1} \equiv |l,m\rangle\langle l,m| \end{aligned} \tag{A9}$$

and illustrates how Eq. (27) of the original literature [45,46] is to be understood.

[1] E. Dekel, D. Gershoni, E. Ehrenfreund, D. Spektor, J. M. Garcia, and P. M. Petroff, Phys. Rev. Lett. **80**, 4991 (1998).
[2] M. Bayer, G. Ortner, O. Stern, A. Kuther, A. A. Gorbunov, A. Forchel, P. Hawrylak, S. Fafard, K. Hinzer, T. L. Reinecke *et al.*, Phys. Rev. B **65**, 195315 (2002).
[3] B. Urbaszek, R. J. Warburton, K. Karrai, B. D. Gerardot, P. M. Petroff, and J. M. Garcia, Phys. Rev. Lett. **90**, 247403 (2003).
[4] R. D. Schaller and V. I. Klimov, Phys. Rev. Lett. **92**, 186601 (2004).
[5] D. Loss and D. P. DiVincenzo, Phys. Rev. A **57**, 120 (1998).
[6] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, Science **314**, 1757 (2006).
[7] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).
[8] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[9] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306 (2007).
[10] G. Klimeck, F. Oyafuso, T. B. Boykin, R. Chris Bowen, and P. von Allmen, Comput. Model. Eng. Sci. **3**, 601 (2002).
[11] G. W. Bryant and W. Jaskólski, Phys. Rev. B **67**, 205320 (2003).
[12] G. Klimeck, S. S. Ahmed, H. Bae, N. Kharche, S. Clark, B. Haley, S. Lee, M. Naumov, H. Ryu, F. Saied *et al.*, IEEE Trans. Electron Devices **54**, 2079 (2007).
[13] C. Delerue and M. Lannoo, *Nanostructures: Theory and Modelling* (Springer, Berlin, 2004).
[14] M. L. Cohen and V. Heine, *Solid State Physics* (Academic, New York, 1970), Vol. 24, pp. 37–248.
[15] M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).
[16] L. W. Wang and A. Zunger, J. Phys. Chem. **98**, 2158 (1994).
[17] K. A. Mäder and A. Zunger, Phys. Rev. B **50**, 17393 (1994).
[18] G. Bester, J. Phys.: Condens. Matter **21**, 023202 (2009).
[19] L.-W. Wang and A. Zunger, Phys. Rev. B **59**, 15806 (1999).
[20] A. Franceschetti, H. Fu, L. W. Wang, and A. Zunger, Phys. Rev. B **60**, 1819 (1999).
[21] J. R. Cárdenas and G. Bester, Phys. Rev. B **86**, 115332 (2012).
[22] X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Côté *et al.*, Comput. Phys. Commun. **180**, 2582 (2009).
[23] C. G. J. Jacobi, Astronom. Nachr. **22**, 297 (1845).
[24] C. G. J. Jacobi, J. Reine und Angew. Math. **1846**, 51 (1846).
[25] E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[26] G. G. Sleijpen and H. Van der Vorst, SIAM J. Matrix Anal. Appl. **17**, 401 (1996).
[27] G. Sleijpen and H. Van der Vorst, SIAM Rev. **42**, 267 (2000).
[28] L.-W. Wang and A. Zunger, J. Chem. Phys. **100**, 2394 (1994).
[29] D. Sorensen, SIAM J. Matrix Anal. Appl. **13**, 357 (1992).
[30] ARPACK library developed by D. C. Sorensen, R. B. Lehoucq, C. Yang, and K. Maschhoff, Rice University, 1996.
[31] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
[32] J. R. Chelikowsky, N. Troullier, and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994).
[33] E. L. Briggs, D. J. Sullivan, and J. Bernholc, Phys. Rev. B **54**, 14362 (1996).
[34] T. L. Beck, Rev. Mod. Phys. **72**, 1041 (2000).
[35] L. Kronik, A. Makmal, M. L. Tiago, M. M. G. Alemany, M. Jain, X. Huang, Y. Saad, and J. R. Chelikowsky, Phys. Status Solidi B **243**, 1063 (2006).
[36] M. Frigo and S. G. Johnson, in *Acoustics, Speech and Signal Processing*, Proceedings of the 1998 IEEE International Conference (IEEE, Piscataway, NJ, 1998), Vol. 3, pp. 1381–1384.
[37] M. Frigo and S. G. Johnson, Proc. IEEE **93**, 216 (2005).
[38] D. Takahashi, in *Parallel Processing and Applied Mathematics*, edited by R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Wasniewski, Vol. 6067 of Lecture Notes in Computer Science (Springer, Berlin, 2010), pp. 606–614.
[39] J. R. Cárdenas and G. Bester, *Atomic effective pseudopotentials for semiconductors:* http://www.fkf.mpg.de/bester/downloads/downloads.html.
[40] B. Fornberg, Math. Comput. **51**, 699 (1988).
[41] B. Fornberg and D. M. Sloan, Acta Numerica **3**, 203 (1994).
[42] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).
[43] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[44] P. A. M. Dirac, Proc. R. Soc. London, Ser. A **117**, 610 (1928).
[45] L. Kleinman, Phys. Rev. B **21**, 2630 (1980).
[46] G. B. Bachelet and M. Schlüter, Phys. Rev. B **25**, 2103 (1982).
[47] D. Naveh, L. Kronik, M. L. Tiago, and J. R. Chelikowsky, Phys. Rev. B **76**, 153407 (2007).
[48] R. Cuadrado and J. I. Cerdä, J. Phys.: Condens. Matter **24**, 086005 (2012).
[49] L. A. Hemstreet, C. Y. Fong, and J. S. Nelson, Phys. Rev. B **47**, 4238 (1993).

[50] M. J. Verstraete, M. Torrent, F. Jollet, G. Zérah, and X. Gonze, Phys. Rev. B **78**, 045119 (2008).

[51] J.-W. Luo, A. Franceschetti, and A. Zunger, Phys. Rev. B **78**, 035306 (2008).

[52] OpenMP Architecture Review Board, OpenMP application program interface version 3.0, 2008.

[53] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. Dongarra, J. D. Croz, A. Greenbaum, S. Hammarling, A. McKenney *et al.*, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, 1999).

[54] C. L. Lawson, R. J. Hanson, D. R. Kincaid, and F. T. Krogh, ACM Trans. Math. Software **5**, 324 (1979).

[55] M. J. T. Oliveira and F. Nogueira, Comput. Phys. Commun. **178**, 524 (2008).

[56] C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).

[57] M. Cardona, N. E. Christensen, and G. Fasol, Phys. Rev. B **38**, 1806 (1988).

[58] J.-W. Luo, A. N. Chantis, M. van Schilfgaarde, G. Bester, and A. Zunger, Phys. Rev. Lett. **104**, 066405 (2010).