PHYSICAL REVIEW B **94**, 045440 (2016)

Linear scaling approach for atomistic calculation of excitonic properties of 10-million-atom nanostructures

Piotr T. Różański[*] and Michał Zieliński[†]

Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100 Torun, Poland

(Received 8 March 2016; revised manuscript received 27 May 2016; published 29 July 2016)

Numerical calculations of excitonic properties of novel nanostructures, such as nanowire and crystal phase quantum dots, must combine atomistic accuracy with an approachable computational complexity. The key difficulty comes from the fact that excitonic spectra details arise from atomic-scale contributions that must be integrated over a large spatial domain containing a million and more atoms. In this work we present a step-by-step solution to this problem: a combined empirical tight-binding and configuration interaction scheme that unites linearly scaling computational time with the essentials of the atomistic modeling. We benchmark our method on the example of well-studied self-assembled InAs/GaAs quantum dots. Next, we apply our atomistic approach to crystal phase quantum dots containing more than 10 million atoms.

DOI: 10.1103/PhysRevB.94.045440

I. INTRODUCTION

Accurate calculations of excitonic properties of semiconductor quantum dots [1] must fulfill two apparently contradictory conditions. The first constraint is the necessity of handling computational domains containing millions of atoms [2,3]. The second requirement is the atomistic accuracy necessary for the accurate description of the excitonic spectra [4–7]. First-principles modeling such as the density functional theory [8] or GW-Bethe-Salpeter-equation approach [9] for self-assembled [10] or nanowire [11] quantum dots is still beyond the reach of modern computers. Continuous matter approaches like the effective mass approximation or $k \cdot p$ method [12,13] have demonstrated the capability of describing main features in QDs spectra [10,14]; however, these methods are restricted by the resolution on the scale of a unit cell [15]. One of the key examples is the bright exciton splitting [16–18], where sophisticated multiband approaches are needed to accommodate for the correct symmetry of the underlying crystal lattice [19–22]. Apart from symmetry issues, even the most elaborate continuous matter tactics cannot fully account for effects of alloying [3], lattice randomness [23], interface effects [24,25], or crystal phase symmetry [26,27] for which a truly atomistic approach is needed.

Practical methods, capturing atomistic structure of quantum dots and their surrounding matrix, include two semiempirical approaches: the empirical tight-binding [28–35] and the empirical pseudopotential method [36–38]. Typical calculations using these approximate approaches involve three subsequent stages [39–42]: (a) calculation of equilibrium position of constituent atoms, (b) calculation of quasielectron and quasihole states, and (c) inclusion of excited quasiparticles interactions via screened Coulomb potential, solved by some sort of configuration interaction (exact diagonalization) method.

Although the stages of finding positions (strains) and quasiparticle calculations are often far from trivial, the key computational issue for approaches aiming to quantitatively describe nanosystem spectra is the ultimate step: the many-body calculation. This part of the calculation is of particular importance as in a typical quantum dot photoluminescence experiment one does not observe the single-particle spectra, but rather many-body spectra of charged and neutral (multi)excitonic complexes [1]. Therefore in a realistic calculation the single-particle part of the computation must be followed by a many-body calculation. Similarly to post-Hartree-Fock methods of quantum chemistry [43], the many-body calculation following the single-particle part is often the most time demanding stage of the entire empirical computations sequence. In particular, as a necessary prerequisite for the many-body calculation typically numerous two-particle integrals must be first calculated using quasiparticle functions. In the tight-binding (TB) approach, for N atoms, A basis functions per atom, and M quasiparticle states there are $O(N^4A^4M^4)$ two-particle integrals, and the computational effort will thus formally scale as the fourth power in number of atoms, what leads to a practically untractable problem for typical quantum dots with numbers of atoms exceeding a million.

In this paper we present an approximated method for efficient order-N calculation of screened Coulomb and exchange integrals within the tight-binding framework. Apart from linear scaling our method accounts for terms that are customarily neglected in a typical tight-binding calculation of Coulomb matrix elements. We test our approach on the example of million-atom InAs/GaAs lens-shape self-assembled quantum dots and finally apply our approach to multimillion-atom crystal phase quantum dots.

II. COULOMB MATRIX ELEMENTS

At the moment one of the most successful empirical computational procedures is the empirical pseudopotential method [36,39]. This approach uses "realistic" empirical pseudopotentials that reproduce correct band gaps and effective masses. Those potentials can be accommodated with a relatively small plane-wave basis set and are used to obtain the single-particle eigenvectors as solution of the Schrödinger equation with an auxiliary basis set of

---

[*] Present address: College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences (MISMaP), University of Warsaw, ul. Banacha 2c, 02-097 Warszawa, Poland.

[†] mzielin@fizyka.umk.pl

strained Bloch function of the underlying bulk [44]. Whereas conceptually straightforward, the practical implementation of the empirical pseudopotential method for quantum dots has been so far limited only to Zunger and co-workers.

On the contrary, numerous theoretical groups utilize different flavors of the empirical tight-binding method. The latter approach origins from the Slater-Koster scheme of orthogonal tight binding [45], whereas Hamiltonian matrix elements are given in terms of several empirical constants [46] determined to reproduce bulk properties such as effective masses, bulk deformation potentials, and gaps at different points of the Brillouin zone [47]. The fitting process is usually far from trivial [48]; however, for a wide family of materials, there are reliable and accurate tight-binding parameter sets available in the literature [47–52].

The tight-binding method applied to a typical nanowire or self-assembled quantum dot problem produces Hamiltonian matrices of dimensions extending 10^6; however, due to the nearest-neighbors approximation the Hamiltonian matrix is sparse. Then, for selected single particle states (close to energy band gap) the eigenproblem can be solved efficiently using, e.g., Krylov iterative methods, such as the Lanczos [53] or Arnoldi [54] algorithm, therefore resulting in a linear scaling of computational time with respect to the number of atoms in a computational domain. Like many other numerical approaches the empirical tight binding benefits considerably from the parallelization that further reduces the overall computation time [3].

The tight-binding or linear combination of atomic orbitals (LCAO) wave function is given as

$$\psi(\vec{r}) = \sum_{\vec{R}}^{N}\sum_{\alpha}^{A} b_{\vec{R}\alpha}|\vec{R}\alpha\rangle = \sum_{\vec{R}}^{N}\sum_{\alpha}^{A} b_{\vec{R}\alpha}\phi_{\vec{R}\alpha}(\vec{r}), \tag{1}$$

where the summation goes over all (N) atoms $\vec{R}$ and all (A) over atomic orbitals α centered on a given atom, whereas $b_{\vec{R}\alpha}$ are basis expansion coefficients. Assuming a statically screened [9] Coulomb interaction, the Coulomb matrix elements V_{ijkl} are given by [33]

$$V_{ijkl} = \iint \psi_i^*(\vec{r}_1)\psi_j^*(\vec{r}_2)\frac{e^2}{\epsilon(\vec{r}_1,\vec{r}_2)|\vec{r}_1-\vec{r}_2|}\psi_k(\vec{r}_2)\psi_l(\vec{r}_1), \tag{2}$$

where $\epsilon(\vec{r}_1,\vec{r}_2)$ is the position-dependent dielectric function and ψ's are single-particle, electron or hole, wave functions. By substituting single-particle wave functions in the form of Eq. (1) into Eq. (2) one obtains

$$\begin{aligned} V_{ijkl} = &\sum_{\vec{R}_1\alpha_1}\sum_{\vec{R}_2\alpha_2}\sum_{\vec{R}_3\alpha_3}\sum_{\vec{R}_4\alpha_4} b^{i*}_{\vec{R}_1\alpha_1} b^{j*}_{\vec{R}_2\alpha_2} b^{l}_{\vec{R}_3\alpha_3} b^{k}_{\vec{R}_4\alpha_4} \\ &\times \omega(\vec{R}_1\alpha_1,\vec{R}_2\alpha_2,\vec{R}_3\alpha_3,\vec{R}_4\alpha_4), \end{aligned} \tag{3}$$

where fourfold summation goes over all atomic positions and orbitals and

$$\begin{aligned} &\omega(\vec{R}_1\alpha_1,\vec{R}_2\alpha_2,\vec{R}_3\alpha_3,\vec{R}_4\alpha_4) \\ &\quad\equiv \iint \phi_{\vec{R}_{1\alpha_1}}(\vec{r}_1)\phi_{\vec{R}_{2\alpha_2}}(\vec{r}_2)\frac{e^2}{\epsilon(\vec{r}_1,\vec{r}_2)|\vec{r}_1-\vec{r}_2|}\phi_{\vec{R}_{3\alpha_3}}(\vec{r}_2)\phi_{\vec{R}_{4\alpha_4}}(\vec{r}_1) \end{aligned} \tag{4}$$

is an integral calculated in a basis of tight-binding ("atomic") orbitals.

If treated directly, this procedure would result in $O(N^4A^4)$ terms (atomic integrals) constituting one quasiparticle Coulomb matrix element, where N is the number of atoms in the domain (typically $N \sim 10^6$ for self-assembled and nanowire quantum dots) and A is the number of (spin) orbitals associated with each of atoms (e.g., 20 for the $sp^3d^5s^\star$ TB model [34]). To further complicate matters, practical calculation demands computation not of one, but numerous electron-electron, hole-hole, and electron-hole Coulomb matrix elements calculated using Eq. (3). For example, in a typical quantum dot calculation involving 12 electron states and 12 hole states (including spin), the total number of these quasiparticle Coulomb matrix elements reaches 10^5, whereas in certain situations [22] the number of quasiparticle states (M) used for the calculation must be further extended. Either way, the overall computation time scales as $O(N^4A^4M^4)$.

Typically, this formidable problem is resolved by utilizing a series of approximations [32,33], including neglect of three- and four-center integrals ω [29]. Further approximations involve multipole expansion of single integral ω and retaining monopole-monopole contributions only [32,33].

Finally one gets an approximate form of Coulomb matrix elements [33]:

$$\begin{aligned} V_{ijkl} = &\sum_{\vec{R}_1}\sum_{\vec{R}_2\neq\vec{R}_1}\left[\sum_{\alpha_1} b^{i*}_{\vec{R}_1\alpha_1} b^{l}_{\vec{R}_1\alpha_1}\right]\left[\sum_{\alpha_2} b^{j*}_{\vec{R}_2\alpha_2} b^{k}_{\vec{R}_2\alpha_2}\right] \\ &\times \frac{e^2}{\epsilon|\vec{R}_1-\vec{R}_2|} + \sum_{\vec{R}_1}\sum_{\alpha_1\alpha_2\alpha_3\alpha_4} b^{i*}_{\vec{R}_1\alpha_1} b^{j*}_{\vec{R}_1\alpha_2} b^{k}_{\vec{R}_1\alpha_3} b^{l}_{\vec{R}_1\alpha_4} \\ &\times \omega(\vec{R}_1\alpha_1,\vec{R}_1\alpha_2,\vec{R}_1\alpha_3,\vec{R}_1\alpha_4), \end{aligned} \tag{5}$$

where the first term is the long-range, bulk-screened contribution to the two-center integral built from the monopole-monopole interaction [55,56] of two charge densities localized at different atomic sites. The second term is the on-site unscreened part, calculated by direct integration using atomic orbitals [28,29]. This approach is justified by the fact that the screening (Thomas-Fermi) radius (≈2–4 Å) is on the order of a bond length [15,29] resulting in nearly bulk screening of off-site (long-range) terms and limited screening of on-site (short-range) terms contribution. A potential problem arising here is the choice of atomic basis used for calculation of ω on-site integrals. The generally important matter of basis dependence will be discussed later in more detail.

The second summation in Eq. (5) over on-site terms can be further simplified by neglecting exchange terms and multiple ω's replaced by a single on-site contribution [32]:

$$\begin{aligned} &V_{ijkl} = \sum_{\vec{R}_1}\sum_{\vec{R}_2}\left[\sum_{\alpha_1} b^{i*}_{\vec{R}_1\alpha_1} b^{l}_{\vec{R}_1\alpha_1}\right]\left[\sum_{\alpha_2} b^{j*}_{\vec{R}_2\alpha_2} b^{k}_{\vec{R}_2\alpha_2}\right] V_{\vec{R}_1\vec{R}_2}, \\ &V_{\vec{R}_1\vec{R}_2} = \begin{cases} \frac{e^2}{\epsilon|\vec{R}_1-\vec{R}_2|}, & \vec{R} \neq \vec{R}', \\ U_{\vec{R}}, & \vec{R} = \vec{R}', \end{cases} \end{aligned} \tag{6}$$

where the on-site atomic contribution $U_{\vec{R}}$ can be calculated (estimated) using different approaches [15,32]. Whereas relatively uncomplicated, Eqs. (5) and (6) used for excitonic calculations give the excitonic energy and optical spectra in the reasonable agreement with the experiment and other computational approaches [32,34,35]. The above procedures [Eqs. (5) and (6)] give also far more approachable $O(N^2)$ scaling of a single Coulomb matrix element calculation rather than impractical $O(N^4)$ of the straightforward method [Eq. (3)]. Further reduction of the computational time (but not the scaling factor) can be achieved by utilizing multiscale [57] approaches (i.e., using smaller domains at different stages of the computation). While not affecting the scaling properties, the usage of numerical libraries such as BLAS can further reduce the computation time by avoiding time-consuming recalculation [40] of intermediate terms.

Apart from the advantages, the above methods [Eqs. (5) and (6)] reveal two apparent issues. The first is a nonlinear $O(N^2)$ scaling of the computational time leading to exploding computational complexity for domains containing millions of atoms. The second problem is a possible inaccuracy related to the two-center and the monopole-monopole approximations.

III. WAVE-FUNCTION RECONSTRUCTION

In what follows we present a numerical method that effectively addresses the above issues. Let us start by reiterating that the LCAO representation of the tight-binding wave function leads to a nonlinear scaling of the two-particle Coulomb matrix elements calculation with respect to the number of atoms in the domain. Therefore, after TB stage of calculation, we convert TB eigenstates from a typical LCAO wave function to a real-space tight-binding wave-function representation. This stage is achieved by an introduction of a three dimensional, uniform real-space grid with complex values. At each point of the spatial grid we use Eq. (1) (assuming a particular basis set) to calculate the wave-function values. We repeat this procedure and process in this way each of several lowest electron and hole functions obtained by the tight-binding procedure, obtaining the real-space grid representation for each of the considered functions. The resolution of the grid, common for all functions, is determined by convergence studies and will be discussed later. We denote this stage of calculation as the "wave-function grid reconstruction" or simply "wave-function reconstruction." The benefits of this transformation will become apparent soon. Later we will discuss in detail the effect of the basis choice used in Eq. (1).

The computational cost of the straightforward single wave-function reconstruction using Eq. (1) is proportional to the number of atoms N and the number of grid points P resulting in $O(NP)$ computation time for the single-quasiparticle function reconstruction. For accurate computations and high grid resolution, P should clearly be proportional (even exceeding) to N leading to the unfavorable $O(N^2)$ scaling at this stage of calculation. However, let us notice that in an orthogonal tight-binding method it is legitimate to assume that for any point in space $\vec{r}$ the value of the tight-binding wave function ψ can be replaced by a contribution from several atomic neighbors only lying within a certain "cutoff" radius R_{cut}:

$$\psi(\vec{r}) = \sum_{\vec{R}}^{N}\sum_{\alpha}^{A} b_{\vec{R}\alpha}\phi_{\vec{R}\alpha}(\vec{r}) \approx \sum_{|\vec{r}|<R_{\text{cut}}}\sum_{\alpha}^{A} b_{\vec{R}\alpha}\phi_{\vec{R}\alpha}(\vec{r}). \tag{7}$$

More strictly, rather than due to orthogonality, this assumption originates from atomic orbitals' spatial locality, i.e., relatively small spatial extent of atomic orbitals compared to dimensions of the entire system. Similar assumptions are typically utilized in the context of other linear-scaling approaches [58], that rely heavily on the use of strictly confined basis orbitals, i.e., orbitals zeroing beyond a certain radius. One way of achieving this goal is the method of Sankey and Niklewski [59] where the (pseudo)atom is embedded within a spherical box of finite radius.

Should R_{cut} be much smaller than the system size then the number of atoms (N_{cut}) within the cutoff radius is $N_{\text{cut}} \ll N$. More importantly, for quantum dots embedded in a bulklike matrix, N_{cut} should be relatively constant and independent from N, leading to $O(N)$ scaling of the tight-binding wave-function reconstruction time.

A straightforward implementation of Eq. (7) would involve iteration over grid points and therefore implicitly perform costly calculations of distances between all grid points and all atoms. Instead, in our implementation, we iterate over all atoms and update only grid points within a neighborhood defined as a cube of edge length equal to $2R_{\text{cut}}$. This procedure can also be efficiently parallelized—in our approach we use a hybrid OpenMP/MPI parallelization scheme allowing for utilization of multicore computer clusters.

In the case of semiconductor quantum dots the above assumptions and exact value of R_{cut} will be verified later by numerical tests. The size of the reconstruction grid should be equal to spatial system dimensions increased in all directions by necessary margins equal to R_{cut}. We also note that we separately compute and store spin-up and spin-down wave-function components at each grid point. Once the reconstruction is finished, the wave function is finally renormalized on the grid.

With the TB wave function given in real space we can follow with the calculation of Coulomb matrix elements using Eq. (2). Straightforward substitution of real-space wave functions into Eq. (2) and direct integration would however lead to $O(P^2) \propto O(N^2)$ complexity. Fortunately, this equation can be conveniently evaluated in the reciprocal space. Equation (2) can be cast in a general form [60].

$$V_{ijkl} = \iint \rho_{il}(\vec{r})\, G(\vec{r}-\vec{r}')\, \rho_{jk}(\vec{r}')\, dV\, dV', \tag{8}$$

where $\rho_{il} = \psi^{i*}\psi^{l}$, $\rho_{jk} = \psi^{j*}\psi^{k}$, and $G(\vec{r}-\vec{r}')$ is the screened Coulomb interaction. Then, potential V_{jk} calculated from quasidensity ρ_{jk} is given as

$$V_{jk}(\vec{r}) = \int G(\vec{r}-\vec{r}')\, \rho_{jk}(\vec{r}')\, dV'. \tag{9}$$

For a discrete case, assuming a regular grid with the grid step h, we obtain

$$V_{jk}[\vec{r}] = \sum_{\vec{r}'} G[\vec{r}-\vec{r}']\,\rho_{jk}[\vec{r}']\,h^3 = (G*\rho_{jk})[\vec{r}]\,h^3, \qquad (10)$$

where $\vec{r}$ and $\vec{r}'$ are points on a discrete grid and $(G*\rho_{jk})$ is a full convolution between quasidensity ρ_{jk} and the screened Coulomb interaction $G(\vec{r}-\vec{r}')$, which is defined on a three-dimensional domain twice as large as ρ_{jk} (as it includes both positive and negative shifts) in every dimension. Therefore, the density grid must also be padded with zeros, resulting in $8P$ grid points instead of P. Due to this padding, the full convolution $(G*\rho_{jk})$ is equivalent to the circular convolution, which in turn can be effectively computed using fast Fourier transform (FFT) algorithm [61] in $O(P\log P)$ time.

Since we use FFT purely as a computational tool to calculate full (not circular) convolution, we do not introduce any undesired periodicity to the problem. Details of this approach are given, e.g., in Ref. [62]. This technique is a standard in digital signal processing for a discrete Fourier transform treatment of nonperiodic signals convoluted with a response function of a finite duration. Otherwise, Eq. (10) would typically converge slowly with respect to the grid (supercell) volume. Different techniques including multipole expansion [60,63,64] or truncated Coulomb interaction [64] would be typically necessary to speed up this convergence.

The method we utilize seems not to be often used in the electronic structure calculation, most likely due to the increased memory demand. In our case the size of the FFT domain is effectively doubled in all three directions. As a practical benefit, we do not need to perform convergence tests with respect to the domain size as Eq. (11) gives exactly the same results as the direct integration of Eq. (2). In the latter part of the text we will additionally verify that statement by performing numerical tests of the effective range of the Coulomb interaction. We also note that in other approaches the size of the supercell is often much larger than the actual system size [64]. A potential solution to the problem without the increase of the supercell size was proposed by Genovese *et al.* [65], where a set of interpolation functions was used to represent the charge density. The method is inherently free from false long-range interactions between cells, and could be applied to our problem. However, extension of the FFT domain in our approach is not affecting the overall time needed for the wave-function reconstruction, but only the (relatively short) time of the FFT calculation. In other words, the FFT part is not the most time-consuming part of the approach and, as will be shown in the latter part of the text, the computational time is dominated by the wave-function reconstruction stage.

Last but not least, we note that the kernel singularity at the origin is overcome by choosing the value of $1/r$ averaged over the unit cell, as calculated in work by Mura *et al.* [66].

V_{ijkl} is finally given as a straightforward $O(P)\propto O(N)$ summation over all grid points:

$$V_{ijkl} = \sum_{\vec{r}} \rho_{il}[\vec{r}]\,V_{jk}[\vec{r}]h^3. \qquad (11)$$

For a given jk pair of tight-binding (grid reconstructed) functions, V_{jk} and the FFT transform are calculated only once and are used to calculate all resulting Coulomb matrix elements V_{ijkl}, therefore avoiding costly recalculation of V_{jk}.

Our approach accounts for dielectric effects at different levels of approximation. For example, we can use any form of distance dependent dielectric function, e.g., a Thomas-Fermi model of Resta [67] customarily utilized in the empirical-pseudopotential method calculations [64]. Apart from the distance dependence, we can account for the spatial dependence of the dielectric medium. Details of this latter approach will be discussed elsewhere. In the current paper G is either taken as a Fourier transform of Coulomb bulk-screened interaction or is given by a Thomas-Fermi model of Resta [67]. Finally, we note that due to the introduction of a large spatial grid our method demands substantially larger computer memory than the simple model [Eq. (5)].

IV. LATTICE AND THE SINGLE-PARTICLE SPECTRA

As discussed in the Introduction, the calculation consists of several major steps: first atomic positions are calculated. For a lattice mismatched system (such as InAs/GaAs quantum dots) to calculate strain relaxed positions we use the atomistic valence force field (VFF) approach of Keating [68,69]. This method is described in more detail in Refs. [70,71] and in our previous papers [4,33–35]. We note here only that the VFF approach is a $O(N)$ method and can be efficiently parallelized allowing for treatment of domains containing 10^8 atoms [35]. For InP crystal phase quantum dots, we neglect strain effects [26,27].

Once the atomic positions are given, we use them to calculate single-particle energies with the empirical nearest-neighbor tight-binding model that accounts for strain, spin-orbit interactions, crystal lattice symmetry, and wurtzite crystal-field splitting in the case of zinc-blende/wurtzite mixed crystal phase quantum dots [34,35].

The single-particle tight-binding Hamiltonian for the system of N atoms and m orbitals per atom can be written, in the language of the second quantization, in the following form:

$$\hat{H}_{TB} = \sum_{i=1}^{N}\sum_{\alpha=1}^{m} E_{i\alpha}c^{+}_{i\alpha}c_{i\alpha} + \sum_{i=1}^{N}\sum_{\alpha=1,\beta=1}^{m} \lambda_{i\alpha,\beta}c^{+}_{i\alpha}c_{i\beta} + \sum_{i=1}^{N}\sum_{j=1}^{4}\sum_{\alpha,\beta=1}^{m} t_{i\alpha,j\beta}c^{+}_{i\alpha}c_{j\beta}, \qquad (12)$$

where $c^{+}_{i\alpha}$ $(c_{i\alpha})$ is the creation (annihilation) operator of a carrier on the orbital α localized on the site i, $E_{i\alpha}$ is the corresponding on-site (diagonal) energy, and $t_{i\alpha,j\beta}$ describes the hopping (off-site, off-diagonal) of the particle between the orbitals on (four) nearest-neighboring sites. Coupling to further neighbors is neglected, whereas $\lambda_{i\alpha,\beta}$ (on-site, off-diagonal) accounts for the spin-orbit interaction following the description given by Chadi [72].

For the InAs/GaAs system we use tight-binding parameters set from Ref. [47] in $sp^3d^5s^\star$ parametrization. More details of the $sp^3d^5s^\star$ tight-binding calculation were discussed thoroughly in our earlier papers [34,35].

For InP crystal phase quantum dots we have used Vogl *et al.* [49] $sp^3s^\star$ tight-binding parameters augmented to

TABLE I. Tight-binding $sp^3s^\star$ parameters (eV) for InP zinc-blende and wurtzite phases. Right-hand (11) parameters (eV) are identical for both phases.

	E_{sa}	E_{pa}	E_{p_za}	E_{sc}	E_{pc}	E_{p_zc}	$E_{s^\star a}$	$E_{s^\star c}$	V_{ss}	V_{xx}	V_{xy}	$V_{sa,pc}$	$V_{sc,pa}$	$V_{s^\star a,pc}$	$V_{s^\star c,pa}$	λ_a	λ_c
ZB	−8.5274	0.7677	0.7677	−1.4826	3.9407	3.9407	8.2635	7.0665	−5.3614	1.8801	4.2324	2.2265	5.5825	3.4623	4.4814	0.067	0.392
W	−8.4634	0.8323	0.8063	−1.4186	4.0053	3.9793						=					

account for the spin-orbit splitting (126 meV [72]). We have additionally modified these parameters to account for the increased wurtzite's band gap (1.474 eV [73]), the valence-band offset (64.6 meV [73]) between the wurtzite and zinc-blende segments, and the wurtzite crystal-field splitting (26 meV [74]). The tight-binding parameters used in the calculation for both phases are summarized in Table I.

For million-atom systems the size of the tight-binding Hamiltonian [34,47] typically exceeds 10^7, reaching 10^8 for the largest crystal phase quantum dot considered in this work. However, due to the nearest-neighbor approximation, the Hamiltonian matrix is sparse and the number of nonzero matrix elements scales as $O(N)$. Hamiltonian elements are calculated on demand without occupying the computer's memory. Several lowest electron and hole states are found by means of the Lanczos algorithm with matrix-vector multiplication parallelized using the *MPI* library. We reiterate that thanks to the application of Lanczos algorithm, and the sparse form of Hamiltonian, the process of partial Hamiltonian diagonalization scales linearly with the domain size. More details will be shown in the following section.

V. BENCHMARKS

Once single-particle states are calculated, we move to the efficient calculation of Coulomb matrix elements. We first illustrate our method on the example of the "standard" [34,39] lens-shaped InAs/GaAs quantum dot. The quantum dot has a diameter of 25 nm and a height of 3.5 nm and is located on a 0.6-nm-thick wetting layer. The InAs quantum dot and the wetting layer are embedded into the GaAs matrix, with the total number of atoms in the tight-binding computational domain reaching 0.6×10^6 atoms. This "standard" quantum dot will be used to benchmark our approach, while later in the paper we will present results for InP crystal phase quantum dots.

In the tight-binding method the Hamiltonian matrix elements are treated as empirical parameters and the basis is not explicitly specified. Due to this freedom there are several practical choices possible for the basis functions. These typically include Slater orbitals [29,75] and Hermann-Skillman orbitals [76,77]. Slater-type orbitals are defined by simple rules giving approximate analytic atomic wave functions, in the form of $\phi_\mu = N_\mu Y_\mu(\theta,\phi) r^{n-1} e^{-\alpha r}$, where n is the principal quantum number, N_μ is the normalization constant, μ denotes orbital symmetry, and $Y_\mu(\theta,\phi)$ denotes the corresponding spherical harmonic; finally α is a screening constant obtained by a set of semiempirical principles. On the other hand, Hermann-Skillman orbitals are obtained by self-consistent numerical calculations for free atoms and ions. Importantly, we note here that there is ongoing research for tight-binding schemes [78,79] that would remove the basis ambiguity and directly relate tight-binding parameters with a well defined orbital set.

Slater orbitals have analytical formulation, however Hermann-Skillman orbitals should correspond more closely to the actual atomic states. None of these basis sets is orthogonal (due to nonzero overlaps between orbitals on neighboring atoms), as assumed by the Slater-Koster tight-binding approach, however Hermann-Skillman orbitals are also better localized in space (Fig. 1) and have a smaller overlap between neighboring sites, whereas Slater orbitals have tails extending over many lattice sites.

Our implementation accepts any form of orbitals defined as a product of radial and angular parts, whereas the radial part can be given by an analytical or numerical form. Therefore our method can, in principle, utilize orbitals such as those generated by the TB-density functional theory (DFT) approach [80], or recent schemes of empirical tight-binding parametrization [78,79].

Figure 2 shows electron-electron J_{ee}, electron-hole J_{eh}, and hole-hole J_{hh} Coulomb integrals for electrons and holes occupying their ground s states [35], calculated as a function of the wave-function reconstruction cutoff radius using the Herman-Skillman basis, grid step $h = 0.8$ Å, and Thomas-Fermi screening model for a lens-shaped InAs/GaAs quantum dot. Here we use the notation where, e.g., $J_{eh} \equiv$

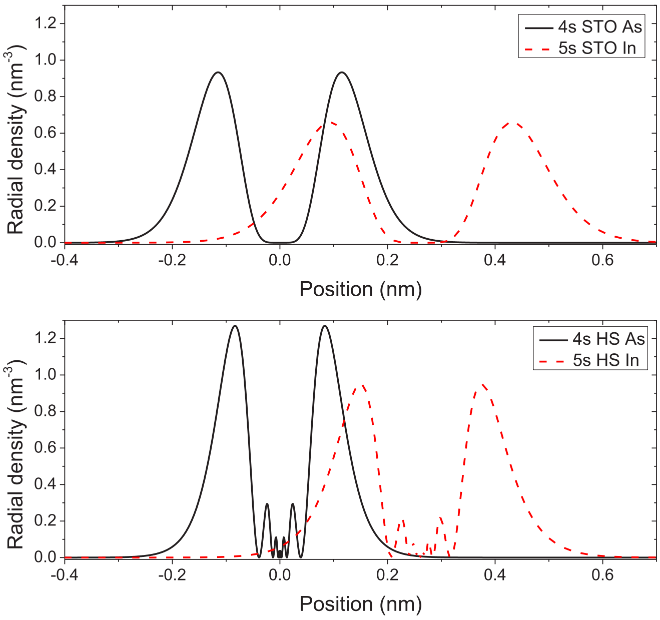

FIG. 1. Comparison of Slater-type (upper row) orbitals and Herman-Skillman (lower row) orbitals charge density distribution for indium and arsenic atomic s valence orbitals. The distance between indium and arsenic corresponds to InAs bond length.

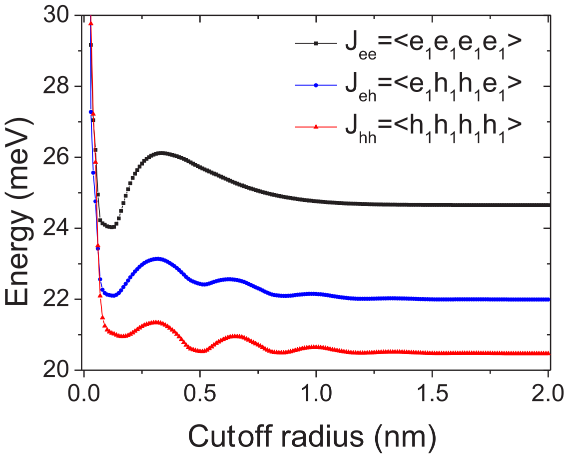

FIG. 2. Electron-electron J^{ee}, electron-hole J^{eh}, and hole-hole J^{hh} Coulomb integrals for electron and hole occupying their ground states calculated for the lens-shaped (see the text) InAs/GaAs quantum dot as a function of the wave-function cutoff radius.

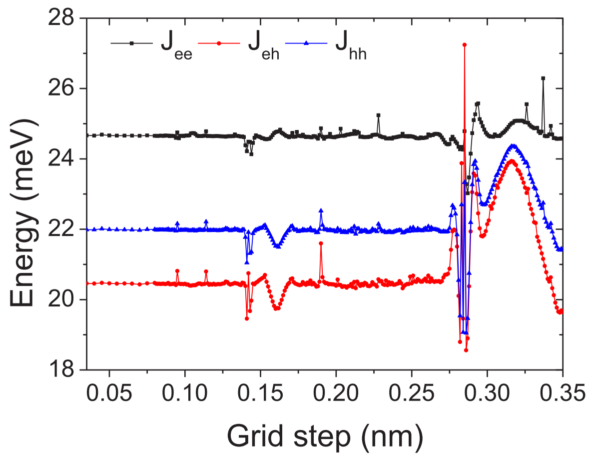

FIG. 3. Electron-electron J^{ee}, electron-hole J^{eh}, and hole-hole J^{hh} Coulomb integrals for electron and hole occupying their ground states calculated for the the lens-shaped (see the text) InAs/GaAs quantum dot as a function of the wave-function reconstruction grid step; calculated using Herman-Skillman orbitals basis set.

$V_{e_1h_1h_1e_1} \equiv \langle e_1h_1h_1e_1 \rangle$. These three integrals are of particular importance as they allow us to estimate the single exciton and excitonic complexes binding energies at the level of Hartree-Fock approximation [35]. Figure 2 shows that the cutoff radius of ≈1.5 nm is sufficient for the converged (within 0.1% accuracy) calculation of these Coulomb integrals. We performed numerical tests for numerous quantum dot systems (including nanocrystals, self-assembled and nanowire quantum dots) and generally found that the cutoff radius of 1.5–2.0 nm for Herman-Skillman and 2.0–2.5 nm for Slater orbitals is sufficient to achieve this level of accuracy. We also found that the cutoff radius is, to a large degree, independent from the size and shape of the investigated nanosystem.

We note as well that there are ≈500 atoms within a 1.5-nm cutoff radius, potentially contributing to a single grid point. As discussed earlier this is a much smaller number than the total $\approx 10^6$ number of atoms in the computational domain. Therefore without the R_{cut} optimization [Eq. (7)] the reconstruction stage would take 10^3 times longer, rendering the entire calculation impractical.

Figure 3 shows the evolution of the same Coulomb integrals for the same quantum dot system as discussed earlier (using Herman-Skillman orbitals as well), however, as a function of the grid spatial resolution and assuming a fixed cutoff radius equal to 1.5 nm. For large grid steps there are noticeable oscillations related to the overlap of the regular computational grid and the underlaying zinc-blende crystal lattice. These oscillations are particularly pronounced at the grid spacings 0.2–0.3 nm comparable to typical InAs and GaAs bond lengths, however for a grid step h lower than 0.09-nm integral values stabilize. Importantly, already for $h \approx 0.1$ nm the relative uncertainty due to grid spacing errors is <1%. We checked that these conclusions can be generalized to all other integrals entering the excitonic calculation. We note as well that we performed a similar study (Fig. 4) for Slater orbitals and found analogous convergence dependence. Unless specified otherwise, the results presented below refer to the $h = 0.08$ nm $= 0.8$ Å grid step. Further increase of the grid resolution (decrease of h) is typically unfeasible due to the computational time limits. Additionally it seem unnecessary as the excitonic properties of quantum dots are determined by contributions from valence orbitals rather than subatomic details of atomic cores. We note, however, that in the case of Herman-Skillman orbitals the rapid oscillations near the origin could certainly pester the convergence below a certain h value. Therefore, should one aim for a small h and use Herman-Skillman orbitals (or similar) for highly accurate (artifact-free) calculations, then a smoothing of the charge density would seem necessary before further processing in the FFT solver [81]. Another systematic approach to convergence would involve utilization of basis sets deliberately created [78,79] for tight-binding calculations: such functions neglect core oscillations inherently. Finally, we note that Slater orbitals are also by construction free from near-core oscillations and

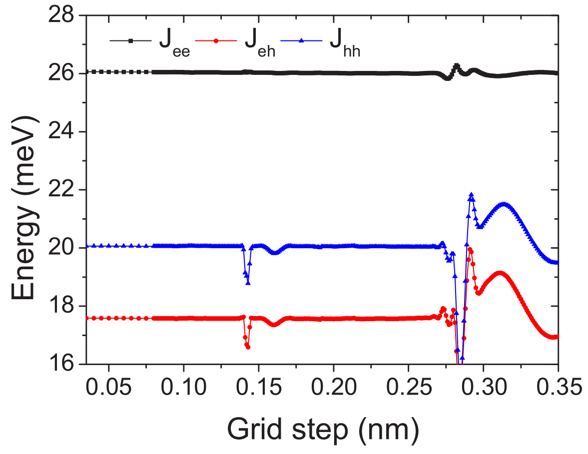

FIG. 4. Electron-electron J^{ee}, electron-hole J^{eh}, and hole-hole J^{hh} Coulomb integrals for electron and hole occupying their ground states calculated for the the lens-shaped (see the text) InAs/GaAs quantum dot as a function of the wave-function reconstruction grid step; calculated using Slater orbitals basis set.

TABLE II. Selected Coulomb and exchange integrals calculated for lens-shaped (see the text) InAs/GaAs quantum dot for different bases and two different models of dielectric screening: bulklike ϵ_{InAs} screening and Thomas-Fermi (TF) model of Resta (see the text). Results obtained using simplified atomistic model [TB-LCAO; Eq. (6)] are shown for comparison.

	STO		HS orbitals		TB-LCAO	Opt. STO
	ϵ_{InAs}	TF	ϵ_{InAs}	TF	ϵ_{InAs}	TF
$J_{ee} = \langle e_1 e_1 e_1 e_1 \rangle$ (meV)	26.00	26.05	24.61	24.65	24.75	25.82
$J_{eh} = \langle e_1 h_1 h_1 e_1 \rangle$ (meV)	20.03	20.06	21.96	21.99	22.54	22.66
$J_{hh} = \langle h_1 h_1 h_1 h_1 \rangle$ (meV)	17.54	17.58	20.41	20.47	21.25	20.81
$\lvert\langle e_\uparrow h_\Uparrow e_\uparrow h_\Uparrow \rangle\rvert$ (μeV)	708.5	722.8	233.5	254.9	238.6	154.2
$\lvert\langle e_\uparrow h_\Downarrow e_\downarrow h_\Uparrow \rangle\rvert$ (μeV)	553.8	555.9	50.1	49.7	22.5	20.8
$\lvert\langle e_\uparrow h_\Uparrow e_\downarrow h_\Downarrow \rangle\rvert$ (μeV)	13.4	13.5	0.6	0.6	0.2	0.6

generally reveal smoother grid step dependence (Fig. 4) than Herman-Skillman orbitals.

For $h \approx 0.1$ nm there is about 100 grid points per atom, each point storing a spin-up and spin-down part of the wave function. For comparison, in the LCAO form and $sp^3d^5s^\star$ parametrization there are 20 spin orbitals per atom and therefore 20 (complex) expansion coefficients per single atom. Therefore we note that the memory needed to store the wave function in the real-space representation is increased by a factor $\approx$10 (20 for $sp^3s^\star$) when compared to the conventional TB-LCAO form. For a million-atom system this corresponds to $\approx$3 GB per single TB wave function and can be efficiently handled by modern day computers. To summarize, the cutoff radius of about 2 nm and grid step of about 0.1 nm were found satisfactory for the reasonable convergence of selected Coulomb matrix elements.

Table II shows various Coulomb direct and exchange integrals calculated for a InAs/GaAs lens-shaped quantum dot ($D = 25$ nm, $h = 3.5$ nm) using different approaches. For this particular nanosystem the electron-electron repulsion has larger magnitude than electron-hole attraction or hole-hole repulsion. In all considered cases there is only a minor difference due to the dielectric screening model used in a calculation. This confirms our general conclusion that Coulomb interactions in typical self-assembled quantum dots are nearly bulk screened as discussed earlier in the text. For the case of J_{ee} all approaches give a similar value of 25–26 meV. There is, however, a substantial discrepancy between different approaches for direct integrals involving hole states, i.e., J_{eh} and J_{hh}. This is particularly noticeable for $J_{hh} \approx 17.5$ meV obtained using Slater-type orbitals, whereas both Herman-Skillman (HS) orbitals and the simple model of Eq. (6) predict J_{hh} to vary between 20.5 and 21.25 meV.

The above differences can be understood in terms of atomic-orbital contributions to single-particle states. Whereas the ground electron and hole states have quite similar envelopes [34], they have much different (s- and p-type correspondingly) dominant orbital contributions. In particular, p-type (and d-type) Slater-type orbitals are typically substantially more delocalized in space than corresponding Herman-Skillman orbitals. Therefore the difference between both basis sets is due to the spatial extent of basis orbitals, which will be verified later in the text.

Apart from Coulomb direct integrals, Table II shows three selected exchange integrals. Those integrals play an important role in the control of quantum dot excitonic fine structure [16]. The dark-bright exciton exchange splitting is determined predominantly by a (real) exchange matrix element, which also conserves spin: $\langle e_\uparrow h_\Uparrow e_\uparrow h_\Uparrow \rangle$, whereas $\langle e_\uparrow h_\Downarrow e_\downarrow h_\Uparrow \rangle$ is responsible for the mixing of two bright-excitonic states ($e_\uparrow h_\Downarrow$ and $e_\downarrow h_\Uparrow$) and therefore leads to the bright-exciton splitting. Finally $\langle e_\uparrow h_\Uparrow e_\downarrow h_\Downarrow \rangle$ mixes two dark states ($e_\uparrow h_\Uparrow$ and $e_\downarrow h_\Downarrow$) and leads to the dark-exciton splitting. In all the above cases we used notation where $e_\uparrow$ and $e_\downarrow$ is a Kramers degenerate pair of states corresponding to the electron ground-state energy. Analogous notation has been used for hole states as well.

For the case of exchange integrals the effect of the basis has very noticeable consequences. Again, whereas results obtained using Herman-Skillman orbitals and Eq. (6) agree reasonably, especially for the electron-hole exchange ($\approx$0.25 meV), the value calculated by using Slater-type orbitals is about three time larger and close to 0.8 meV. Even larger difference is reported for integrals related to the bright-exciton splitting (anisotropic electron-hole exchange) and the dark-exciton splitting. In this case the Slater-type orbitals basis overestimates the other two approaches by more than an order of magnitude. It should be noted that the bright-exciton splitting (fine-structure splitting) for cylindrical (or close to cylindrical) quantum dots observed in the experiment is typically on the order of 10–100 μeV, therefore in clear disagreement with the approach using Slater-type orbitals.

To study this effect further, Fig. 5 shows the same integrals as presented in Table II, however calculated as a function of the Coulomb interaction radius which can be artificially limited to a certain radius $r_{\text{max}} \equiv \max |\mathbf{R} - \mathbf{R}'|$. In this case for a simple model, Eq. (6) is simply replaced with the following formula:

$$V_{nm} = \begin{cases} U_n & : \quad n = m\,, \\ \frac{1}{\epsilon|\vec{R}_n - \vec{R}_m|} & : \quad n \neq m \quad \wedge \quad |\vec{R}_n - \vec{R}_m| \leqslant r_{\text{max}}\,, \\ 0 & : \quad n \neq m \quad \wedge \quad |\vec{R}_n - \vec{R}_m| > r_{\text{max}}\,, \end{cases} \tag{13}$$

whereas for grid calculations, in order to avoid granularity artifacts we use the sigmoid-type function:

$$G(\vec{r}) = \frac{e^2}{4\pi\epsilon}\frac{1}{|\vec{r}|}\left(1 + e^{\frac{1}{\lambda}(|\vec{r}| - r_{\text{max}})}\right)^{-1}, \tag{14}$$

where $\lambda > 0$ plays a role of a smoothing parameter. For clarity we emphasize that the Coulomb interaction cutoff radius is a

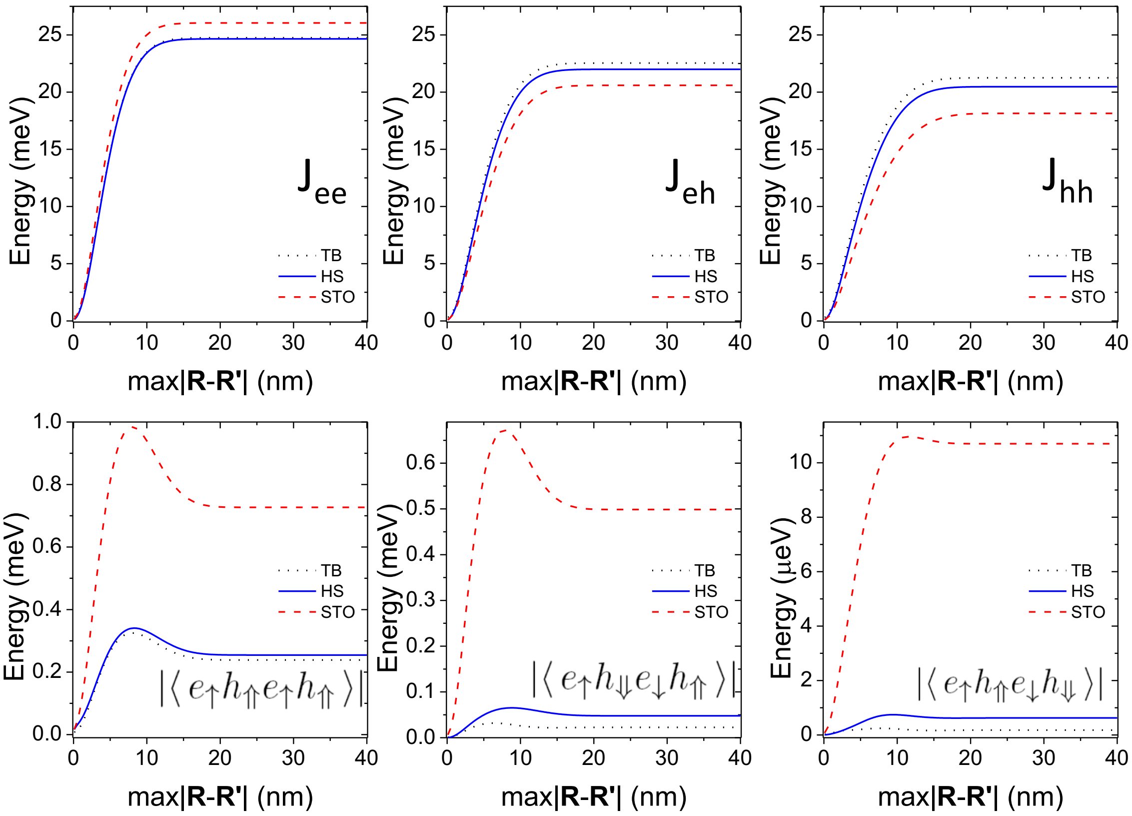

FIG. 5. Several Coulomb and exchange integrals calculated for lens-shaped (see the text) InAs/GaAs quantum dot as a function of a cutoff radius of Coulomb interaction.

different quantity from the wave-function reconstruction cutoff radius studied in Fig. 2.

For all considered integrals their values stabilize for the Coulomb interaction cutoff radius ≈20 nm comparable to the quantum dot diameter. This further proves that the calculation is free from any effects of the image charges. Figure 5 (upper row) demonstrates rather small contributions to the direct matrix elements for this nanosystem that originate from on-site and nearest-neighbor contributions, and that these integrals are dominated by the long-range contributions. Interestingly, for the case of exchange integrals there are noticeable maxima in their modulus at about 8–10 nm, corresponding to approximately one-third of the quantum dot diameter (or three times the quantum dot height). With further increase of the Coulomb interaction cutoff radius modules of exchange integrals are reduced and their values stabilize at 20 nm. As this effect is visible for the simple model [Eq. (6)] as well as for calculations involving basis, this suggests that there are two long-range, monopole-monopole contributions to exchange integrals of opposite character. On the other hand, for the case of exchange integrals calculated using Slater-type orbitals the short-range (<3 nm) contribution is apparently dominant and is likely responsible for the overestimation of integral values as compared to other approaches.

As discussed earlier, in the case of semiconductor quantum dots, one could assume that for sites which are far enough apart from each other the exact structure of the localized orbitals is not important [32] and that the long-range contributions are dominated by the monopole-monopole interaction of two charge densities localized at different sites. This assumption is clearly not fulfilled for Slater-type orbitals which extend over many lattice constants from the site center. This is especially important for the bright-exciton splitting that has been shown to be strongly related to the (local) electron-hole nonorthogonality on the scale of a unit cell [55,56]. Slater-type orbitals localized on neighboring sites, i.e., within the same unit cell, are clearly (Fig. 1) far from orthogonal, resulting in a strongly overestimated value of anisotropic exchange integral.

In order to study these effects further, we note that bulk on-site atomic energies are shifted with respect to their free atom counterparts [47]. By the same token, the spatial extent of basis orbitals in bulk should be reduced as compared with free atoms. This claim has been recently supported by work by Benchamekh *et al.* [78] where microscopic (Bloch) functions for the tight-binding model were obtained by a process in which screening constants α of Slater-type orbitals were optimized by the fitting. Reference [78] shows that the fitting procedure has a relatively small effect on well localized $4s$ and $4p$ orbitals (sp^3 in tight binding), e.g., altering $4s$ screening constant from 1.7 to 1.94. On the other hand, the screening constant of Slater-type $4d$ arsenic orbital is increased significantly from 0.27 in the free atom case to 0.96 for arsenic site in GaAs bulk crystal. The screening constant of $5s$ (i.e., $s^\star$) orbital is affected even more, being increased from 0.4 to over 1.74 for the bulk case. This corresponds to effective "compression" of atomic orbitals in bulk, and is especially pronounced for higher orbitals of the largest spatial extent. These results indicate strong (exponential) dumping of

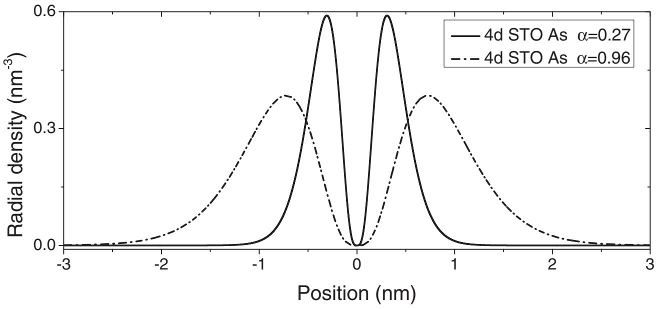

FIG. 6. Comparison of radial densities of Slater-type orbitals for arsenic 4*d* orbital, with original ($\alpha = 0.27$) and modified ($\alpha = 0.96$; see the text) screening constants.

atomic orbital tails at long distance as compared to free atoms counterparts, which is the apparent radial density plot as seen in Fig. 6. This procedure reduces significantly basis function tails and the contribution from a given site is on average effectively limited to the radius of about 1 nm.

Whereas modified screening constants are currently not available for indium arsenide, we repeated our calculations using an approach in which we increased screening constants on s, p, d, and $s^\star$ orbitals for all atomic species in the nanosystem by the following values: 0.36, 0.25, 0.7, 1.3. These values are close to modifications reported by Ref. [78]. Results of our calculations are shown in the right-hand column (Opt. STO) in Table II.

In the optimized case Coulomb and exchange integral values are similar to those given by a more confined Herman-Skillman basis or the asymptotic model of Eq. (6) (TB-LCAO). This is particularly pronounced for exchange integrals, e.g., for $|\langle e_\uparrow h_\Downarrow e_\downarrow h_\Uparrow \rangle|$ (related to the bright-exciton splitting) where the "compressed" Slater-type orbitals give values an order of magnitude smaller than unmodified orbitals and very close to that of Eq. (6). Otherwise, we note that the choice of unmodified Slater-type basis results in severe overestimation of the above integrals due to the overlap of orbitals between far (>1.0 nm) neighbors.

VI. CRYSTAL PHASE QUANTUM DOTS

Crystal phase quantum dots [26,27,74] gained recently a lot of attraction due to nearly perfect interface between crystal phases constituting the system, well defined (with monolayer accuracy) height, and the lack of alloying effects pestering spectral reproducibility [3,82,83] of typical self-assembled quantum dots. In this section we apply our method to a single InP crystal phase quantum dot. We model this nanostructure by a zinc-blende InP segment of 1 nm height (three monolayers (MLs), single ABC stacking sequence) along the [111] direction, embedded between two 30-nm (100 MLs)-long wurtzite InP segments grown along the [0001] direction. Figure 7 shows corresponding energy band alignment of the studied system, plotted along the growth axis, and schematics of the nanosystem. The total height of the system exceeds 60 nm, whereas we vary nanowire diameter from 12 to 70 nm, therefore a number of atoms in the system varies from close to 0.3 to over 10 million atoms.

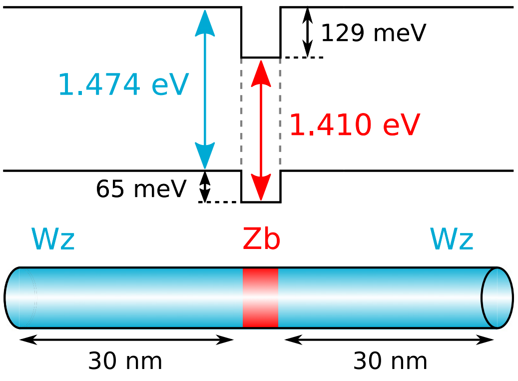

FIG. 7. Band alignment of InP zinc-blende and wurtzite used in our calculations, and schematics of the system.

Figure 8 shows convergence of eigenvalues of the single-particle tight-binding Hamiltonian matrix obtained as a result of Lanczos (tridiagonalization) process [84] for crystal phase quantum dot of 35 nm diameter. By its nature Lanczos algorithm converges to extreme eigenvalues [84] first, i.e., eigenvalues most separated from the rest of the spectra. Since there is a $\approx$1.5 eV energy gap (conduction band–valence band) in the system, the ground electron state is effectively separated from other eigenvalues. As a result, the ground electron state (1.464 62 eV) converges after a couple hundred iterations only. The higher excited electron states follow systematically and for example at about 10 000 iterations there are already eight converged electron states. The convergence rate for hole states is somewhat smaller since there are much smaller spacings between hole states in the vicinity of valence-band edge and therefore smaller spectral separation. Typically we use 35 000 Lanczos iterations to obtain about ten lowest hole eigenvalues. As an additional benefit at this number of iterations we obtain as well typically more than 40 converged electron states. The Lanczos algorithm is somewhat affected by the finite numerical precision of computations. This is manifested by the presence of spurious states seen as random points in Fig. 8. Another class

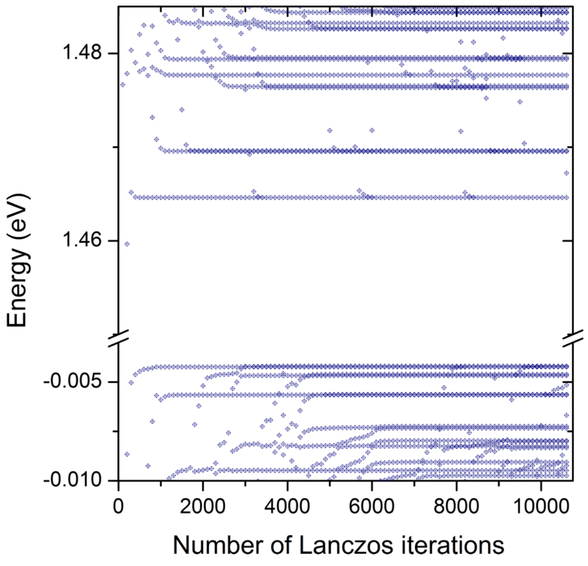

FIG. 8. Spectra of the tight-binding matrix produced by Lanczos algorithm for the crystal phase quantum dot of diameter 35 nm.

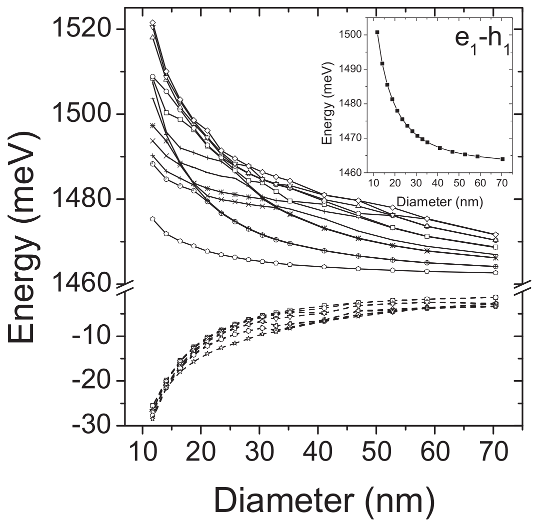

FIG. 9. Several lowest electron and hole eigenenergies calculated for crystal phase quantum dots of different diameters (see the text). Please note different energy scales and the presence of the energy gap. Inset shows single-particle energy gap ($e_1 - h_1$) evolution with the diameter.

of spurious states due to finite numerical precision is formed by so-called ghost states, energies of which are very close to the real eigenvalues [84]. Both classes of spurious states are effectively removed by the analysis and postprocessing of eigenvalues obtained at different Lanczos iterations [84].

Figure 9 shows energy levels corresponding to several lowest electron and hole states for crystal phase quantum dots as a function of diameter. For both electron and hole we observed clear trends due to reduced lateral confinement: the electron's energies are reduced, whereas hole energies are increased with the increasing diameters. Additionally, for both carriers, the energy spacings between single-particle states are reduced with the diameter. This is visible particularly well for $e_2 - e_1$ spacing between ground electron state (e_1) and first excited electron state (e_2), which drops from nearly 13 meV for 11 nm diameter to 1.5 meV for the 70-nm system. The spacings between lowest hole states are much smaller, typically well below 0.1 meV, and as low as several μeV's only for largest considered diameters.

The inset in Fig. 9 shows the effective single-particle gap (defined as the energy difference between ground electron e_1 and hole h_1 state) calculated for the crystal phase quantum dot as a function of a diameter. Importantly, as a result of e_1 and h_1 evolution with diameter, the effective gap drops substantially from over 1.5 eV for the lowest considered diameter of 12 nm to 1.46 eV for the largest considered diameter of 70.4 nm. These results show that it is in principle possible to tailor the effective gap of these nanostructures by control of a nanowire diameter. On the other hand, Fig. 9 reveals the significant role of lateral confinement, therefore without precise control of diameter crystal phase quantum dots are subject to a potential spectral inhomogeneity.

Next, in Fig. 10 we show probability charge densities for several lowest electron and hole states. The electron states are predominantly localized in the zinc-blende region (marked in Fig. 10 as dotted/red lines), yet with significant tails in the wurtzite area of the nanowire. The pronounced density oscillations seen in the plot are related to the presence of atomic monolayers constituting the nanowire. Looking at the general shape of the charge distribution we judge that the ground electron state is generally of s-like character, whereas two excited electron states are of p-like character with nodal points visible in the perpendicular projection. The higher electron states (not shown here) have more complicated in-plane (lateral) nodal structure and the entire electron spectra in general resembles that of the disk-shaped self-assembled quantum dot [85]. On the other hand, hole states are delocalized over the wurtzite wire sections, with nodes occurring at the zinc-blende part, and therefore limited overlap with electron states: a hallmark of the type-II system. Interestingly h_1 and h_2 states are of moleculariike character, being nanostructure analogs of bonding and antibonding orbitals, whereas the artificial molecule is effectively formed by two wurtzite (left and right) nanowire sections surrounding the zinc-blende barrier. Higher excited hole states (h_3 and h_4) form a similar pair with additional nodal structure. Interestingly, the lowest-lying hole states have no nodal structure in the perpendicular (lateral) view and appear to be more localized along the nanowire central axis. We note that the peculiar structure of hole states varies with diameter as well. Finally, we note that three (with spin, six) lowest electron states and lowest four (with spin, eight) hole states will be used as a basis set for configuration interaction (exact diagonalization) calculations in the following part of this paper. Such a limited basis is chosen here for computational time reasons mainly, yet it should be large enough to demonstrate the scaling properties of our approach and, at the same time, capture the most relevant physical properties of the system. We note, however, that more accurate calculations for crystal phase quantum dots should involve a larger configuration interaction basis set. Alternatively, the exact diagonalization step of the calculation could be preceded by a self-consistent calculation of the electron-hole interaction. Such an approach is actively developed by authors of this paper and will be the subject of our future work.

Figure 11 shows the time of the single-particle (tight-binding) part of the calculation as a function of the number of atoms (which is growing quadratically with the nanowire diameter). Each of the computations was performed on the same 48-core computer system and for the same number (35 000) of Lanczos iterations. The time of the computation scales practically linearly as a function of number of atoms. The small steps in the plot are related to parallel computation and load-balancing issues, i.e., the problem of uniform division of a discrete atomic grid to a discrete number of processors.

As mentioned in the Introduction, experimental spectra are obtained for the interacting electron-hole pair, i.e., exciton. Therefore we follow our single-particle calculation with a many-body calculation. The Hamiltonian for the interacting electrons and holes can be written in second quantization as [1]

$$\begin{aligned}\hat{H}_{ex} = {} & \sum_i E_i^e c_i^\dagger c_i + \sum_i E_i^h h_i^\dagger h_i \\ & + \frac{1}{2}\sum_{ijkl} V_{ijkl}^{ee} c_i^\dagger c_j^\dagger c_k c_l + \frac{1}{2}\sum_{ijkl} V_{ijkl}^{hh} h_i^\dagger h_j^\dagger h_k h_l \\ & - \sum_{ijkl} V_{ijkl}^{eh,\mathrm{dir}} c_i^\dagger h_j^\dagger h_k c_l + \sum_{ijkl} V_{ijkl}^{eh,\mathrm{exchg}} c_i^\dagger h_j^\dagger c_k h_l, \quad (15)\end{aligned}$$

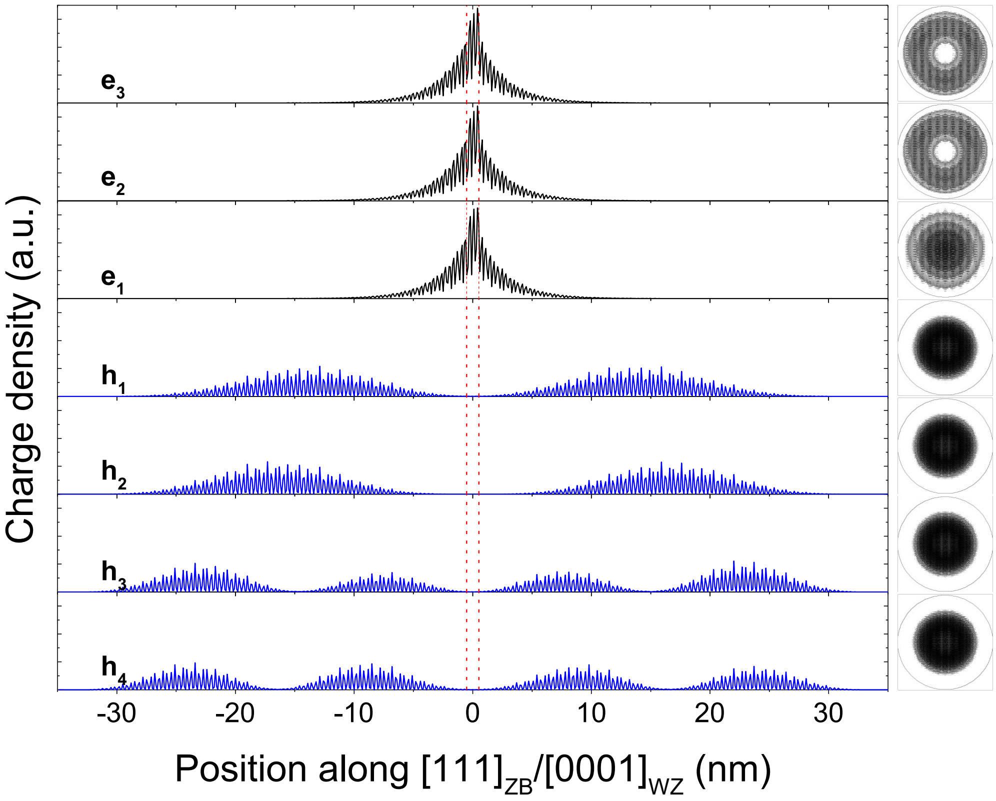

FIG. 10. Probability density for several three lowest electron (e) and four lowest hole (h) states calculated along the growth direction of the nanowire for crystal phase quantum dot of 30 nm diameter. Right-hand side of the plot shows corresponding probability densities summed along the growth direction and projected on the nanowire base.

where E_i^e and E_i^h are the single-particle electron and hole energies, obtained at the single-particle stage of calculations, and V_{ijkl} are Coulomb matrix elements (Coulomb direct and exchange integrals) defined as in Eq. (3) and calculated using tight-binding electron and hole eigenstates. The many-body Hamiltonian for the electron-hole pair (exciton) is solved using the configuration interaction (CI) approach [33,40].

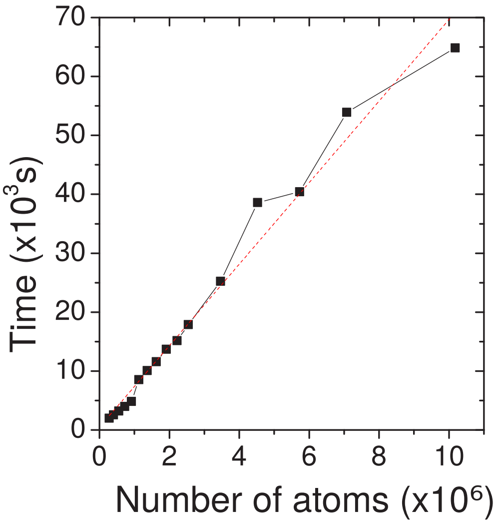

FIG. 11. Time of the tight-binding (singe-particle) calculations for crystal phase quantum dots (see the text) of different number of atoms (quadratically dependent on the nanostructure diameter); 10.2 millon atoms corresponds to 70.4-nm nanometers in diameter. The red (dashed) line is a linear fit that is a guide to the eye.

We performed our calculations using two approaches for calculation of Coulomb matrix elements: first is the traditional TB-LCAO method of Eq. (6) that involved $O(N^2)$ summations, and the second is our linear-scaling approach. In order to execute this comparison we have utilized a limited CI basis involving lowest three (with spin, six) electron states and lowest four (with spin, eight) hole states. For each diameter, we have calculated the total of 1296 electron-electron integrals, $2 \times 2304 = 4608$ electron-hole Coulomb and exchange integral, and 4096 hole-hole integrals. (Note that the number of integrals to be calculated is effectively reduced by a factor of 4 due to the symmetries of Coulomb matrix elements.) Computation of these integrals is a prerequisite not only for the single exciton calculations (which involve only electron-hole interactions), but also for other excitonic complexes such as a biexciton (which additionally needs electron-electron and hole-hole integrals). In particular, the above basis results in total 48 configurations for the single exciton and 420 configurations for the biexciton. Such number of configurations presents little computational effort for the CI Hamiltonian diagonalization; however, due to rapidly (M^4) increasing number of Coulomb matrix elements, further increase of the CI basis would results in prohibitive computational times for the method of Eq. (6), and would render the comparison between methods impractical.

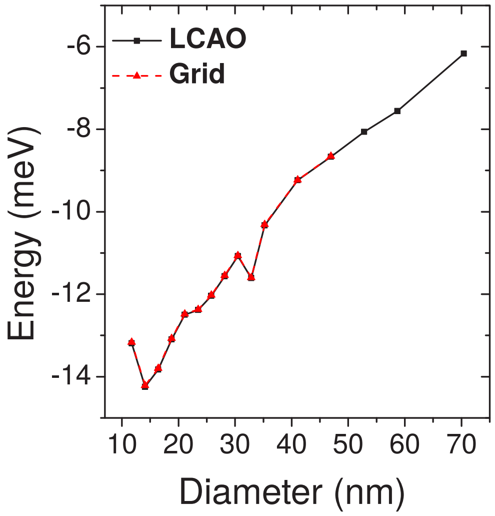

FIG. 12. Excitonic binding energy (see the text) calculated for crystal phase quantum dots of different diameters (see the text).

Figure 12 shows the excitonic binding energy for the discussed crystal phase quantum dot as a function of diameter. The exciton binding energy is defined as an energy difference between the ground state of the interacting electron-hole pair and the single-particle gap studied earlier. With increasing diameter, the magnitude (absolute value) of the binding energy decreases from a high-lateral confinement regime (≈14 meV) for small diameters, to nearly bulklike binding energy of ≈6 meV for the largest diameter. This figure reveals a transition from a quantum dotlike to a quantum welllike confinement for large diameter systems. The unexpected steps (e.g., 33 nm) in otherwise monotonous function are an artefact of a limited (not converged) configuration interaction basis set. Most importantly, we note that in the range where both methods can be compared with each other ($d < 50$ nm) they produce practically identical output. We note as well the results shown here were obtained for the grid step of 1.0 Å, yet they differ only by a tiny fraction of a meV (typically 0.05 meV) from results obtained on the 0.8-Å grid. Additionally, we point out that due to C_{3v} symmetry crystal phase quantum dots have exactly vanishing excitonic fine structure [6] as confirmed by our studies, therefore these systems are to a large degree free from basis dependence artifacts mentioned earlier.

Whereas the excitonic results reported by both methods are nearly identical, the cost of the computation differs dramatically as shown on Fig. 13. With the domain size growing from 0.28 to 4.5 million atoms, the time spent in the calculation of Eq. (5) is increased by more than ≈460 times, showing strong nonlinear scaling. This scaling is even more unfavorable than $O(N^2)$, most likely due to the increased memory-bandwidth usage (and bandwidth bottleneck) for the large number of atoms. On the other hand, the wave-function reconstruction approach shows nearly linear scaling, i.e., the time of computations is increased by a factor of ≈30, corresponding to similar growth of the number of atoms in the domain. The timings presented in Fig. 13 were obtained for grid spacing equal to 1 Å. We note that for higher resolution grids the time and memory demand is increased, e.g., by a factor of ≈40% for the 0.9-Å grid spacing.

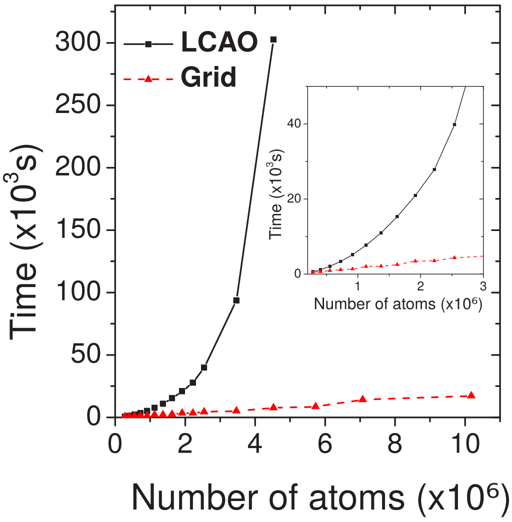

FIG. 13. Time of the Coulomb matrix elements calculation for two different approaches (see the text).

In principle, a further speedup could be achieved: both results were obtained on the same 240-core computer cluster system, however the FFT part of the calculation was run on 48 cores only, due to the limited scaling of the FFT algorithm (as implemented in the FFTW library, which we have used in this work). Artifacts of the FFT parallelization are well visible in Fig. 14 where the total time of the calculation in our approach was divided into the contribution from the reconstruction part and the FFT step of the computations. The reconstruction itself scales practically linear as a function of the number of atoms and in fact the overall time is dominated by this stage. Noticeable steps in time dependence of the FFT part are related to (a) additional padding of the reconstruction domain, to ensure good performance of the FFT, and (b) the load balancing issues on parallel computer system. In fact, the FFT time spent in the 10-million case is smaller than in

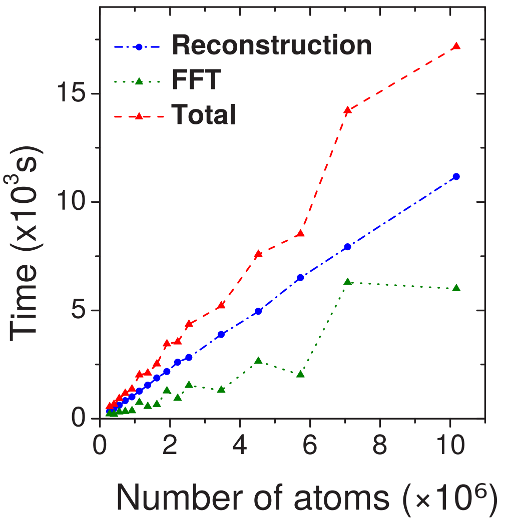

FIG. 14. Times of the grid reconstruction and the FFT stages (see the text) vs the total time of the computation.

the 7-million case, as the grid in the larger case can be more efficiently divided by the number of processors.

We have checked that the quasilinear scaling properties of our approach are not limited to this particular system, but instead, it is a general feature, with similar scaling properties for various nanosystems such as nanocrystals, nanowires, self-assembled and nanowire quantum dots.

VII. SUMMARY

We have presented a method for an efficient, order-N calculation of excitonic spectra of semiconductor nanosystems. Our method is directly applicable to ten-million-atom nanostructures, such as crystal phase quantum dots. The first steps of strain (only for lattice mismatched systems) and the empirical tight-binding calculation are followed by the efficient calculation of Coulomb matrix elements and the configuration interaction approach. Our approach redefines the tight-binding LCAO single-particles wave functions on a three-dimensional, regular grid. The process of wave-function reconstruction can be performed efficiently by relying on a finite extent of basis orbitals, with the cutoff radius smaller than 2.5 nm. The grid step of about 0.8–1.0 Å was found sufficient for the convergence of the key Coulomb and exchange integrals affecting excitonic spectra. The grid reproduced form of the wave function allows for the application of Fourier space methods for the calculation of Coulomb matrix elements that benefit tremendously from the fast Fourier transform algorithm. Our method goes beyond traditional two-center and monopole-monopole approximations; further, our approach can account for different models of dielectric screening function and utilize different basis sets such as those generated by the TB-DFT approach or new schemes of tight-binding parametrization. We studied the role of a basis by comparing results obtained with Slater-type and Herman-Skillman orbitals and the simplified $O(N^2)$ approach. We analyzed the effects of basis locality and orthogonality and found that long-range tails of basis orbitals affect significantly values of important electron-hole exchange integrals. The effect of local basis orthogonality is particularly important for calculations involving bright- and dark-exciton splitting. The straightforward application of highly nonorthogonal Slater-type orbitals would lead to severe overestimation of bright-exciton splitting as compared with other approaches. On the other hand Herman-Skillman orbitals, or recently optimized Slater-type orbitals, should correspond more closely to the actual atomic states, however more accurate calculations of the excitonic fine structure should involve basis sets directly designed for tight-binding calculations.

Finally, we illustrate our approach with the calculation for the crystal phase quantum dots with diameter exceeding 70 nm and number of atoms exceeding 10 million. We demonstrated pronounced diameter dependence of both the effective gap and the excitonic binding energy, with transition to nearly bulklike binding energy for the largest diameter system.

Linear scaling with number of atoms opens an route for semiempirical atomistic calculations of large semiconductor systems with number of atoms reaching 10^8 in the near future. Due to its efficiency, the method should allow for new possible applications such as self-consistent calculations for multimillion-atom nanosystems or accurate nanostructure excitonic calculations involving large many-body basis sets. The range of potential applications varies from quantum dots of different species, through nanowires, up to modeling of electronic properties of single dopants embedded in multimillion atoms transistors.

ACKNOWLEDGMENTS

Support from the Foundation for Polish Science, Homing Plus Programme, cofinanced by the European Union within the European Regional Development Fund, is kindly acknowledged. M.Z. acknowledges support from the Polish Ministry of Science and Higher Education through Research Project No. IP 2012064572 (Iuventus Plus).

---

[1] P. Hawrylak and M. Korkusinski, in *Single Quantum Dots Fundamentals, Applications, and New Concepts*, Topics in Applied Physics, edited by P. Michler (Springer-Verlag, Berlin, Heidelberg, New York, 2003), Vol. 90, pp. 25–92.

[2] G. Bester, S. Nair, and A. Zunger, Phys. Rev. B **67**, 161306 (2003).

[3] G. Klimeck, F. Oyafuso, T. Boykin, R. Bowen, and P. von Allmen, Comput. Model. Eng. Sci. **3**, 601 (2002).

[4] W. Jaskólski, M. Zieliński, G. W. Bryant, and J. Aizpurua, Phys. Rev. B **74**, 195339 (2006).

[5] G. W. Bryant, M. Zieliński, N. Malkova, J. Sims, W. Jaskólski, and J. Aizpurua, Phys. Rev. Lett. **105**, 067404 (2010).

[6] R. Singh and G. Bester, Phys. Rev. Lett. **103**, 063601 (2009).

[7] M. Ediger, G. Bester, A. Badolato, P. M. Petroff, K. Karrai, A. Zunger, and R. J. Warburton, Nat. Phys. **3**, 774 (2007).

[8] E. Engel and R. M. Dreizler, *Density Functional Theory* (Springer, Berlin, 2011).

[9] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).

[10] L. Jacak, P. Hawrylak, and A. Wojs, *Quantum Dots* (Springer, Berlin, 1998).

[11] M. T. Bjork, C. Thelander, A. Hansen, L. Jensen, M. Larsson, L. R. Wallenberg, and L. Samuelson, Nano Lett. **4**, 1621 (2004).

[12] O. Stier, M. Grundmann, and D. Bimberg, Phys. Rev. B **59**, 5688 (1999).

[13] A. Schliwa, M. Winkelnkemper, and D. Bimberg, Phys. Rev. B **76**, 205324 (2007).

[14] D. Bimberg, M. Grundmann, and N. Ledentsov, *Quantum Dot Heterostructures* (Wiley, New York, 1999).

[15] C. Delerue and M. Lannoo, *Nanostrucutres: Theory and Modelling* (Springer, New York, 2004).

[16] M. Bayer, G. Ortner, O. Stern, A. Kuther, A. A. Gorbunov, A. Forchel, P. Hawrylak, S. Fafard, K. Hinzer, T. L. Reinecke, S. N. Walck, J. P. Reithmaier, F. Klopf, and F. Schäfer, Phys. Rev. B **65**, 195315 (2002).

[17] T. Takagahara, Phys. Rev. B **62**, 16840 (2000).

[18] E. Kadantsev and P. Hawrylak, Phys. Rev. B **81**, 045311 (2010).

[19] S. Tomić and N. Vukmirović, J. Appl. Phys. **110**, 053710 (2011).

[20] O. Marquardt, S. Schulz, C. Freysoldt, S. Boeck, T. Hickel, E. P. O'Reilly, and J. Neugebauer, Opt. Quantum Electron. **44**, 183 (2011).

[21] K. F. Karlsson, M. A. Dupertuis, D. Y. Oberli, E. Pelucchi, A. Rudra, P. O. Holtz, and E. Kapon, Phys. Rev. B **81**, 161307 (2010).

[22] M. Zieliński, Phys. Rev. B **88**, 115424 (2013).

[23] J.-W. Luo and A. Zunger, Phys. Rev. B **84**, 235317 (2011).

[24] Y.-M. Niquet and C. Delerue, Phys. Rev. B **84**, 075478 (2011).

[25] F. A. Zwanenburg, A. S. Dzurak, A. Morello, M. Y. Simmons, L. C. L. Hollenberg, G. Klimeck, S. Rogge, S. N. Coppersmith, and M. A. Eriksson, Rev. Mod. Phys. **85**, 961 (2013).

[26] N. Akopian, G. Patriarche, L. Liu, J.-C. Harmand, and V. Zwiller, Nano Lett. **10**, 1198 (2010).

[27] M. B. Bavinck, K. D. Jöns, M. Zieliński, G. Patriarche, J.-C. Harmand, N. Akopian, and V. Zwiller, Nano Lett. **16**, 1081 (2016).

[28] K. Leung and K. B. Whaley, Phys. Rev. B **56**, 7455 (1997).

[29] S. Lee, L. Jönsson, J. W. Wilkins, G. W. Bryant, and G. Klimeck, Phys. Rev. B **63**, 195318 (2001).

[30] R. Santoprete, B. Koiller, R. B. Capaz, P. Kratzer, Q. K. K. Liu, and M. Scheffler, Phys. Rev. B **68**, 235311 (2003).

[31] M. Usman, Phys. Rev. B **86**, 155444 (2012).

[32] S. Schulz, S. Schumacher, and G. Czycholl, Phys. Rev. B **73**, 245327 (2006).

[33] M. Zieliński, M. Korkusiński, and P. Hawrylak, Phys. Rev. B **81**, 085301 (2010).

[34] M. Zieliński, Phys. Rev. B **86**, 115424 (2012).

[35] M. Zieliński, J. Phys.: Condens. Matter **25**, 465301 (2013).

[36] A. J. Williamson, L. W. Wang, and A. Zunger, Phys. Rev. B **62**, 12963 (2000).

[37] G. Bester and A. Zunger, Phys. Rev. B **71**, 045318 (2005).

[38] M. Gong, K. Duan, C.-F. Li, R. Magri, G. A. Narvaez, and L. He, Phys. Rev. B **77**, 045326 (2008).

[39] L. He and A. Zunger, Phys. Rev. B **73**, 115324 (2006).

[40] W. Sheng, S.-J. Cheng, and P. Hawrylak, Phys. Rev. B **71**, 035316 (2005).

[41] M. Korkusiński, M. Zieliński, and P. Hawrylak, J. Appl. Phys. **105**, 122406 (2009).

[42] G. W. Bryant, M. Zieliński, N. Malkova, J. Sims, W. Jaskólski, and J. Aizpurua, Phys. Rev. B **84**, 235412 (2011).

[43] A. Szabo and N. Ostlund, *Modern Quantum Chemistry* (Macmillan, New York, 1983).

[44] A. Zunger, in *Quantum Theory of Real Materials*, edited by J. R. Chelikowsky and S. G. Louie (Kluwer, Boston, 1996).

[45] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).

[46] W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, New York, 1980).

[47] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, Phys. Rev. B **57**, 6493 (1998).

[48] G. Klimeck, R. C. Bowen, T. B. Boykin, and T. A. Cwik, Superlattices Microstruct. **27**, 519 (2000).

[49] P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids **44**, 365 (1983).

[50] T. B. Boykin, G. Klimeck, R. C. Bowen, and F. Oyafuso, Phys. Rev. B **66**, 125207 (2002).

[51] S. Sapra, R. Viswanatha, and D. D. Sarma, J. Phys. D: Appl. Phys. **36**, 1595 (2003).

[52] R. Viswanatha, S. Sapra, T. Saha-Dasgupta, and D. D. Sarma, Phys. Rev. B **72**, 045333 (2005).

[53] J. Cullum and R. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations* (SIAM, Philadelphia, 2002).

[54] R. Lehoucq, D. Sorensen, and C. Yang, *ARPACK User's Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Re- started Arnoldi Methods* (SIAM, Philadelphia, 1998).

[55] A. Franceschetti, L. W. Wang, H. Fu, and A. Zunger, Phys. Rev. B **58**, R13367 (1998).

[56] S. V. Goupalov and E. L. Ivchenko, Phys. Solid State **43**, 1867 (2001).

[57] M. Zieliński, Acta Phys. Pol. A **122**, 312 (2012).

[58] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).

[59] O. F. Sankey and D. J. Niklewski, Phys. Rev. B **40**, 3979 (1989).

[60] E. L. de Oliveira, E. L. Albuquerque, J. S. de Sousa, G. A. Farias, and F. M. Peeters, J. Phys. Chem. C **116**, 4399 (2012).

[61] J. W. Cooley and J. Tukey, Math. Comput. **19**, 297 (1965).

[62] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1992).

[63] G. Makov and M. C. Payne, Phys. Rev. B **51**, 4014 (1995).

[64] A. Franceschetti, H. Fu, L. W. Wang, and A. Zunger, Phys. Rev. B **60**, 1819 (1999).

[65] L. Genovese, T. Deutsch, A. Neelov, S. Goedecker, and G. Beylkin, J. Chem. Phys. **125**, 074105 (2006).

[66] M. E. Mura and N. C. Handy, Theor. Chim. Acta **90**, 145 (1995).

[67] R. Resta, Phys. Rev. B **16**, 2717 (1977).

[68] P. N. Keating, Phys. Rev. **145**, 637 (1966).

[69] R. M. Martin, Phys. Rev. B **1**, 4005 (1970).

[70] C. Pryor, J. Kim, L. W. Wang, A. J. Williamson, and A. Zunger, J. Appl. Phys. **83**, 2548 (1998).

[71] T. Saito and Y. Arakawa, Physica E **15**, 169 (2002).

[72] D. J. Chadi, Phys. Rev. B **16**, 790 (1977).

[73] A. De and C. E. Pryor, Phys. Rev. B **81**, 155210 (2010).

[74] L. Zhang, J.-W. Luo, A. Zunger, N. Akopian, V. Zwiller, and J.-C. Harmand, Nano Lett. **10**, 4055 (2010).

[75] J. C. Slater, Phys. Rev. **36**, 57 (1930).

[76] F. Herman and S. Skillman, *Atomic Structure Calculations* (Prentice-Hall, New York, 1963).

[77] N. Malkova and G. W. Bryant, Phys. Rev. B **82**, 155314 (2010).

[78] R. Benchamekh, F. Raouafi, J. Even, F. Ben Cheikh Larbi, P. Voisin, and J.-M. Jancu, Phys. Rev. B **91**, 045118 (2015).

[79] Y. P. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, Phys. Rev. B **92**, 085301 (2015).

[80] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).

[81] L. Genovese and T. Deutsch, Phys. Chem. Chem. Phys. **17**, 31582 (2015).

[82] V. Mlinar, M. Bozkurt, J. M. Ulloa, M. Ediger, G. Bester, A. Badolato, P. M. Koenraad, R. J. Warburton, and A. Zunger, Phys. Rev. B **80**, 165425 (2009).

[83] M. Zieliński, K. Gołasa, M. R. Molas, M. Goryca, T. Kazimierczuk, T. Smoleński, A. Golnik, P. Kossacki, A. A. L. Nicolet, M. Potemski, Z. R. Wasilewski, and A. Babiński, Phys. Rev. B **91**, 085303 (2015).

[84] J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations: Vol. I: Theory* (Springer, New York, 1985).

[85] M. Zieliński, Phys. Rev. B **88**, 155319 (2013).