# Solving Schrödinger's equation around a desired energy: Application to silicon quantum dots

Lin-Wang Wang and Alex Zunger

Citation: The Journal of Chemical Physics **100**, 2394 (1994); doi: 10.1063/1.466486
View online: http://dx.doi.org/10.1063/1.466486
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/100/3?ver=pdfcov
Published by the AIP Publishing

# Solving Schrödinger's equation around a desired energy: Application to silicon quantum dots

Lin-Wang Wang and Alex Zunger
*National Renewable Energy Laboratory, Golden, Colorado 80401*

(Received 28 June 1993; accepted 26 November 1993)

We present a simple, linear-in-size method that enables calculation of the eigensolutions of a Schrödinger equation in a desired energy window. We illustrate this method by studying the near-gap electronic structure of Si quantum dots with size up to $Si_{1315}H_{460}$ ($\approx$37 Å in diameter) using a plane wave pseudopotential representation.

Modern *ab-initio* electronic structure calculations on large molecules and solids are generally cast in terms of solutions to some effective single-particle Schrödinger equation

$$\hat{H}\psi_i=\epsilon_i\psi_i, \tag{1}$$

e.g., using the local density formalism[1] for $\hat{H}$. These applications can generally be divided into two classes. In the first class one investigates problems in which both the self-consistent potential $v(\mathbf{r})$ and the atomic positions are not known in advance and thus have to be obtained from solutions of *all* occupied $\psi_i$ based on Eq. (1). Examples include surfaces with unsuspected reconstruction geometries[2,3] or crystals and molecules with intricate patterns of charge transfer and hybridization. Here we address the second class of problems, i.e., cases where $v(\mathbf{r})$ and the atomic geometry are either (i) known, or, (ii) can be obtained from small-scale calculations, and one is interested to inspect eigensolutions only in a given energy range, e.g., around a band gap in insulators. An example of (i) includes the study of band gap variation with size in mesoscopic quantum structures,[4] where both the potential and the atomic geometry can be approximated as nearly bulk-like quantities. An example of (ii) is the study of band-gap impurity levels or superlattices, where $v(\mathbf{r})$ and the atomic relaxations are often *localized* near the impurity or at the interface (and thus can be obtained from self-consistent calculations on small systems) but the wave functions extend over many atomic cells.[5]

Most electronic structure methods treat both classes of problems equally. They require solving Eq. (1) for all occupied wave functions $\{\psi_i\}$, even though in "class-two problems" one is interested only in the near band gap solutions. This strategy is inefficient: For a given Hamiltonian $\hat{H}$, the conventional variational method is to minimize the energy $\langle\psi|\hat{H}|\psi\rangle$ by varying the expansion coefficients of $\psi$. Then the first $\psi$ obtained is the lowest energy state of $\hat{H}$. To find a higher state, one needs to orthogonalize $\psi$ to all energy states below it. The effort needed to accomplish this orthogonalization scales as $N^3$ where $N$ is the number of atoms in the system. Consequently, only small systems ($N\leqslant 100$) can be conveniently addressed. Although advances in solving Eq. (1) as a multiparameter minimization problem[6] and progress in parallel computing[3,7] has increased the size of systems amenable to treatment via Eq. (1), fully quantum mechanical mesoscopic problems ($>$1000 atoms) are still outside the scope of such first-principles methods.

We present here an approach which enables calculation of eigensolutions around an interesting energy without having to calculate any of the wave functions below it. The effort involved scales linearly with the system's size thus enabling calculations of band gap properties in mesoscopic systems. The method is exact in that the solutions are identical to those of Eq. (1).

The central point of the present approach is that the eigensolutions $(\epsilon_i, \psi_i)$ of the Eq. (1) also satisfy

$$(\hat{H}-\epsilon_{\text{ref}})^2\psi_i=(\epsilon_i-\epsilon_{\text{ref}})^2\psi_i. \tag{2}$$

Here the spectrum $\{\epsilon_i\}$ of $\hat{H}$ has been folded at the reference point $\epsilon_{\text{ref}}$ into the spectrum $\{(\epsilon_i-\epsilon_{\text{ref}})^2\}$ of $(\hat{H}-\epsilon_{\text{ref}})^2$. The lowest solution of the folded spectrum (2) is the eigenstate with $\epsilon_i$ closest to $\epsilon_{\text{ref}}$. Hence, by placing $\epsilon_{\text{ref}}$ in the physically interesting range, one transforms an arbitrarily high eigensolution into the lowest one, thus obviating the need for orthogonalization. For example, if $\epsilon_{\text{ref}}$ is placed inside an energy gap, minimization of $\langle\psi|(\hat{H}-\epsilon_{\text{ref}})^2|\psi\rangle$ results either in the highest occupied molecular orbital (HOMO) or the lowest unoccupied molecular orbital (LUMO), depending on which is closer to $\epsilon_{\text{ref}}$. Changing $\epsilon_{\text{ref}}$ within the gap region then assures that both the HOMO and the LUMO are found. Because only a few wave functions are calculated, the effort scales linearly with the system's size $N$.[8,9]

Our basic strategy is to solve Eq. (2) by seeking the minimum of

$$F=\int \psi(r)[-\tfrac{1}{2}\nabla^2+v(r)-\epsilon_{\text{ref}}]^2\psi(r)dr \tag{3}$$

in the space of the variational parameters of $\psi$. This requires special treatment, because the use of $(\hat{H}-E_{\text{ref}})^2$ slows down considerably the convergence of standard minimization methods when compared to minimization of $\langle\psi|\hat{H}|\psi\rangle$. We solved this problem by using a plane wave expansion of $\psi$ and minimizing $F$ using a carefully preconditioned conjugate gradient approach. Equation (3) is calculated by applying $[-\frac{1}{2}\nabla^2+v(r)-\epsilon_{\text{ref}}]$ to $\psi=\Sigma_{\mathbf{G}}C_{\mathbf{G}}e^{i\mathbf{G}\mathbf{r}}$ twice. Once $F$ is obtained, we minimize it with respect to the variational wave function coefficients $C_{\mathbf{G}}$, using the conjugate gradient method.[6] This is defined as line mini-

 0021-9606/94/100(3)/2394/4/$6.00 

mizations along the search directions, i.e., finding $\theta$ in $\psi_{\text{new}} = \psi_{\text{old}} \cos(\theta) + P_{\text{search}} \sin(\theta)$ which minimizes $F$. Here, $P_{\text{search}}$ is the normalized search direction which is made orthogonal to $\psi_{\text{old}}$. The search direction $P_{\text{search}}$ is given by the derivative $A \cdot \partial F/\partial \psi$ plus a correction from the search direction of the previous step. We use the Polak–Ribiere formula[9] for this correction. The prefactor $A = \alpha^2/[(\frac{1}{2}\mathbf{G}^2 + v_0 - \epsilon_{\text{ref}})^2 + \alpha^2]$ is used in $\mathbf{G}$ space as a preconditioner, where $v_0$ is the average potential of the system and $\alpha$ is a parameter which is in the same order of the wave function kinetic energy. It is effacacious to test the value of $\alpha$ before doing large scale computations.[10] We will test our method relative to the multiparameter minimization of $\langle \psi | \hat{H} | \psi \rangle$ [Eq. (1)] using the conjugate gradient approach.[6] We will refer to the latter as "conventional" approach only because Eq. (1) rather than Eq. (2) is solved. Note, however, that this approach is considerably more efficient than the (truly conventional) method of directly *diagonalizing* Eq. (1) in a basis.

While this method is quite general, we apply it here to the calculation of LUMO–HOMO band gap of Si quantum dots containing up to $\sim$1300 Si atoms. We use the empirical pseudopotential method (EPM)[11] to describe the system's potential $v(r)$. X-ray diffraction studies[12–14] indicate that the core of Si quantum dots is crystalline with lattice constant close to the bulk value. We thus use the bulk lattice constant, fitting the Si empirical pseudopotential both to the bulk band structure and to the surface work function (4.9 eV).[15] We passivate the surface dangling bonds by hydrogen atoms and model the surface relaxations of the chemisorbed layer according to data for hydrogen-covered (001), (111),[16] and (110)[17] Si *surfaces*. The hydrogen empirical pseudopotential[15] is determined by fitting the calculated surface density of state of these surfaces to experiment.[18] A plane wave basis set with a kinetic energy cutoff of 4.5 Ry is used throughout the calculations. For the largest system this corresponds to a basis set of 100 000 orbitals. The reference energy $\epsilon_{\text{ref}}$ of Eq. (2) is obtained by performing first conventional calculations [Eq. (1)] on small quantum dots and using this $\epsilon_{\text{ref}}$ for all dot sizes. The eigensolutions found in conventional calculations for such small $Si_{13}H_{28}$ and $Si_{59}H_{76}$ clusters were found to be identical to the results of the present method. We then applied our method also to larger rectangular $Si_{163}H_{148}$, $Si_{349}H_{244}$, $Si_{641}H_{364}$, and $Si_{1063}H_{508}$ boxes (Fig. 1). This figure compares the efforts involved in calculating these quantum dots using the conventional approach [Eq. (1)] and the present approach [Eqs. (2)–(3)]. Both calculations use the preconditioned conjugate gradient method[6] with the same convergence tolerance. The dashed line is obtained by extrapolating the actual CPU times for the conventional calculations on the two smallest dots using the expected $N^3$ scaling with size. This extrapolation (Fig. 1) suggests that using the conventional method, calculating the largest quantum dot in the figure would require about two weeks Cray CPU time. The solid line in Fig. 1 shows that the effort involved in *actual* calculations with the current method scales linearly with the system's size and requires less than one CPU hour for the largest

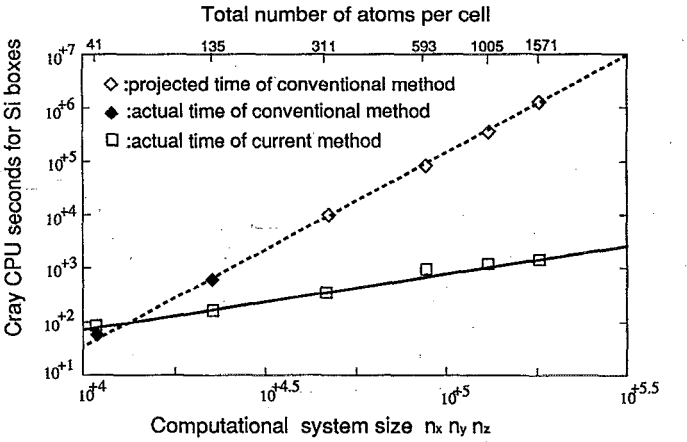


FIG. 1. Computational time $t$ (in units of Cray-YMP CPU seconds) vs the size $n$ of the system measured by the number $n_x n_y n_z$ of fast Fourier transform grid points. The solid line is a fit: $t = 9.0 \times 10^{-3}\, n$. The dashed line depicts a projection of the computer time needed with the conventional conjugate gradient method based on Eq. (1). The clusters here are rectangular shaped.

system. Thus, our method enables efficient direct pseudopotential band gap calculations on $O(10^3)$ atom systems.

Figure 2[19] compares our calculated band gaps vs size variations for *spherical* quantum dots with model calculations [part (a)], other direct calculations [part (b)] and with experimental data [part (c)]. The largest quantum dots calculated here consists of 1315 Si and 460 H atoms (2860 occupied states).

Figure 2(a) compares our results with *model* calculations, i.e., those designed as approximations to direct calculations. These include the effective mass approximation (EMA)[20] and the model used by Rama-Krishna and Friesner (RKF) recalculated with the present pseudopotential.[21] The EMA uses parabolic bands while the method of RKF approximates the states of a quantum dot using a single bulk band. Figure 2 shows that our directly calculated band gaps can be fitted as $E_{\text{gap}}^{\text{bulk}} + 88.34(d/\text{Å})^{-1.37}$ (eV). The EMA, which predicts a $d^{-2}$ dependence, *overestimates* considerably the band gap opening, while the method of RKF *underestimates* it. For example, for a 15 Å particle the method of RKF underestimates the directly calculated band gap by as much as 1 eV out of $\sim$3 eV [Fig. 2(a)], even though the *same* pseudopotential[15] has been used in both calculations (the bulk band structure produced by the present EPM agrees with that of Ref. 21 to within 0.1 eV). Thus, at present, none of the *model* calculations approach the results of the direct calculation with satisfactory precision (more on that later).

Figure 2(b) compares next the present results with those of other direct calculation methods, including the nearest-neighbor tight binding (NN-TB),[22] third neighbor nonorthogonal basis tight binding (TNN-TB),[23] small basis linear-combination-of-atomic-orbital-local-density approximation (LCAO-LDA)[24] and plane-wave-local-density approximation (PW-LDA).[25] Our result agrees closely to TNN-TB result, suggesting that further than nearest-neighbor interactions and nonorthogonal basis

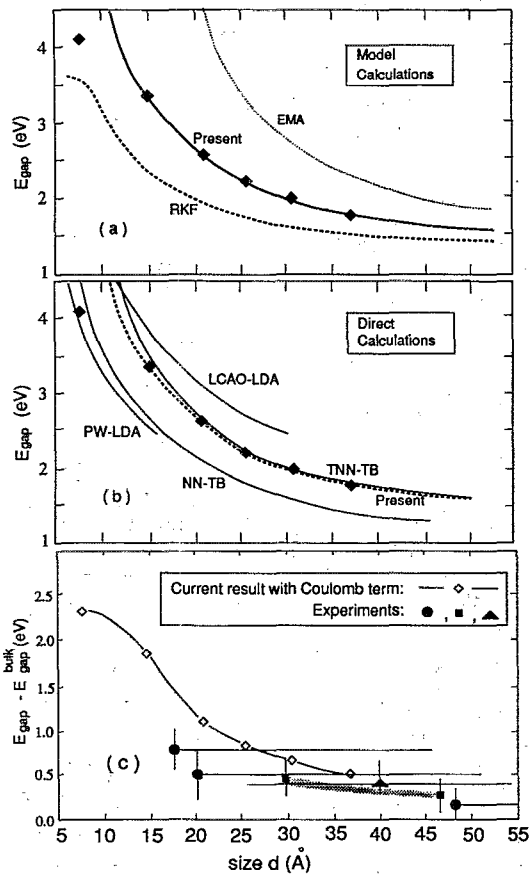


FIG. 2. Band gaps (LUMO–HOMO) of spherical Si quantum dots as a function of the diameter $d=(3/4\pi\times a_0^3/8\times N_{Si})^{1/3}$. Here $a_0$ is the bulk lattice constant and $N_{Si}$ is the number of Si atoms in the quantum dot. (a) Comparison of the present results with *model* calculations, i.e., the EMA (Ref. 20) (with Coulomb interaction being taken out) and the model of RKF [calculated using the present Si EPM (Ref. 15) and the spherical formula of Ref. 21]. (b) Comparison of the present results with other *direct* calculations (see text for references). (c) Comparison of the Coulomb-corrected present result with observed luminescence data. The Coulomb correction is $-3.572/\epsilon d-0.248E_{Ry}$ (a.u.) (Ref. 19), with $\epsilon=11.91$ and $E_{Ry}=0.0082$ (eV). The symbols, ○, △, and □ denote PI data from Refs. 12, 14, and 13, respectively. The vertical lines are the half-height widths of the PL spectra. The horizontal lines denote the size distributions estimated from high-pressure-liquid chromatograph (HPLC), transmission electron microscopic and x-ray experiments. The symbol ○ are placed at the positions of the x-ray data whenever possible. The experiment of Ref. 13 does not report the size distribution. The shaded area represents a range of the experimental points of Ref. 13.

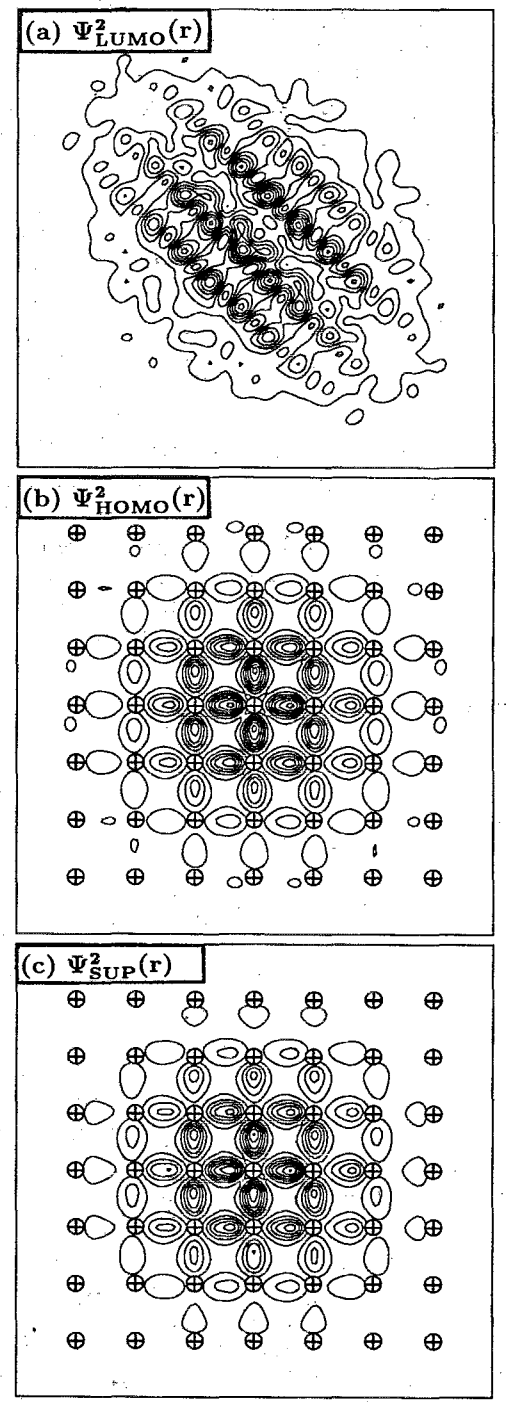


FIG. 3. Wave function contour plots of the rectangular quantum dot described in the text viewed from [001] direction. (a) The resulting charge density after summing along $z$ direction of the LUMO wave function square. (b) The wave function square of HOMO on the $z=d_z/2$ cross section. (c) Square of the composed HOMO wave function based on Eq. (4) on the same cross section. It has the same contour steps as in (b). The crossed circles in (b) and (c) denote the positions of the silicon atoms.

must be important.

Figure 2(c) compares our results with photoluminescence (PL) experiments[12–14] on nearly spherical Si particles. Unfortunately, the existence of a broad size distribution in current Si samples prevents a quantitative comparison with theory. We can only conclude that the current results agree with experiment within the experimental uncertainty.

Figures 3(a) and 3(b) depict the wave function square of the LUMO and HOMO of a rectangular quantum box with sides $d_x=d_y=d_z/\sqrt{2}=23.04$ Å and faces (110), (1$\bar{1}$0), and (001). Note that most of the amplitude is at the *interior* of the dot, so surface perturbations are expected to have minimal effects on these band edge states. It is interesting to analyze the directly calculated wave functions of quantum structure in terms of expansion in bulk Bloch wave functions,[26] thus shedding light on various models. We can expand the HOMO state as

$$\psi^{\text{dot}}_{\text{HOMO}}(\mathbf{r}) = \sum_n \sum_k a_{n,k}\, \psi^{\text{bulk}}_{n,k}(\mathbf{r}). \quad (4)$$

Here, $n$ is the band index and $k$ is a wave vector. This is a generalization of the Luttinger–Kohn model[27] to include interband mixing. Consider, for concreteness, the rectangular quantum box defined above. The wave vectors $\mathbf{k}$ are then quantized as $\pi[\pm j_x d_x^{-1}, \pm j_y d_y^{-1}, \pm j_z d_z^{-1}]$, where $j_x, j_y, j_z$ are positive integers larger than zero. In a particle-in-a-box model, the lowest energy occurs at $j_x = j_y = j_z = 1$ called here $k^*$. We have calculated the projections $P_n = \Sigma_{k^*} |a_{n,k^*}|^2$ of Eq. (4) at the above $k^*$ and found that as much as 93% of the amplitude of $|\psi^{\text{dot}}_{\text{HOMO}}|^2$ comes from the three upper valence bands $n_1, n_2, n_3$ which become triply degenerate at the $\Gamma_{25_v}$ point (the remaining 7% comes primarily from other $k$ points for the same bands). Figure 3(c) shows the approximate $\psi^2_{\text{HOMO}}(r)$ of Eq. (4) constructed from superposition of just these three valence bands at $k^*$. This analysis shows that the HOMO state is "bulklike" in that it can be constructed from just a few bulk states at the "special" $k^*$ points. Equation (4) shows that $\epsilon^{\text{dot}}_{\text{HOMO}} \cong \Sigma_n P_n \epsilon^{\text{bulk}}_n(k^*)/\Sigma_n P_n$. This analysis explains why the method of RKF *underestimates* the band gap: their method limits *ad hoc* the wave function expansion to a *single* band (the highest). The neglect of the other (lighter-mass) bands results therefore in a HOMO that is too high, thus in a band gap that is too small.[28] The significant differences between the results of the direct diagonalization and the RKF method [Fig. 2(a)] thus reflect the neglect of multiband coupling in the latter method. Agreement with experiment then must be fortuitous.

In summary, we have demonstrated a simple, linear-in-size method for solving Schrödinger's equation in a given energy window without having to obtain (and orthogonalize to) the lower eigensolutions. The method enables direct pseudopotential band gap calculations on semiconductor quantum particles with $> 1000$ atoms.

This work was supported by the office of Energy Research, Materials Science Division, U.S. Department of Energy, under Grant No. DE-AC02-83CH100093.

[1] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964); W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
[2] S. Froyen and A. Zunger, Phys. Rev. Lett. **66**, 2132 (1991).
[3] K. D. Brommer, M. Needels, B. E. Larson, and J. D. Joannopoulos, Phys. Rev. Lett. **68**, 1355 (1992).
[4] M. G. Bawendi, M. L. Stiegerwald, and L. E. Brus, Annu. Rev. Phys. Chem. **41**, 477 (1990).
[5] G. A. Baraff, M. Schluter, and G. Allan, Phys. Rev. B **27**, 1010 (1983).
[6] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).
[7] I. Stich, M. C. Payne, R. D. King-Smith, J. S. Lin, and L. J. Clarke, Phys. Rev. Lett. **68**, 1351 (1992).
[8] One can attempt the "inverse iteration" method (Ref. 9) or inverse iteration combined with Lanczos method [J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations* (Birkhauser, Boston, 1985)] to calculate energy eigenstates within an energy window. However, the key is to find a fast and efficient way to solve the large dimension linear equations in the inverse iteration method. To our knowledge this has not been accomplished.
[9] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering, *Numerical Recipes* (Cambridge University, New York, 1989).
[10] The minimization of $F$ of Eq. (3) is analogous to the minimization of $\langle\psi|\hat{H}|\psi\rangle$ in the conventional conjugate gradient method (Ref. 6). There are $N_i$ outside-loop iterations. Inside each iteration, there are $N_c$ steps of conjugate gradient line minimizations, followed by a subspace diagonalization of $(\hat{H}-\epsilon_{\text{ref}})^2$ based on the few wave functions $\psi_i$ computed simultaneously (four wave functions are calculated here). The $N_c$ we used ($\sim 100$) is the square of the typical $N_c$ value used in conventional conjugate gradient methods based on $\hat{H}$ (Ref. 6). The $N_i$ we used is 5, the same as the typical values used in the conventional methods.
[11] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors* (Springer-Verlag, Berlin, 1988).
[12] K. A. Littau *et al.*, J. Phys. Chem. **97**, 1224 (1993).
[13] H. Takagi *et al.*, Appl. Phys. Lett. **56**, 2379 (1990).
[14] W. A. Saunders *et al.*, Mater. Res. Soc. Symp. Proc. **283**, 77 (93).
[15] Standard empirical potentials (Ref. 11) give $V(G)$ only on discrete bulk reciprocal lattice vector $\mathbf{G}$. Calculations on finite objects requires a *continuous* $V(\mathbf{g})$ for all reciprocal vector $g$. Our fit gives: $V_{\text{Si}}(g) = 0.2685\ (g^2-2.19)/(2.06e^{0.487g^2}-1)$; $V_{\text{H}}(g) = -0.1416+9.802*10^{-3}\, g+6.231*10^{-2}g^2-1.895*10^{-2}g^3$ when $g \leqslant 2$ and $V_{\text{H}}(g) = 2.898*10^{-2}/g - 0.3877/g^2+0.9692/g^3-1.022/g^4$ when $g > 2$. The $g$ and the potentials are in atomic units.
[16] E. Kaxiras and J. D. Joannopouplos, Phys. Rev. B **37**, 8842 (1988).
[17] J. E. Northrup, Phys. Rev. B **44**, 1419 (1991).
[18] T. Sakurai and H. D. Hagstrum, Phys. Rev. B **12**, 5349 (1975); J. Vac. Sci. Technol. **13**, 807 (1976); S. Maruno *et al.*, Phys. Rev. B **27**, 4110 (1983).
[19] L. E. Brus, J. Phys. Chem. **90**, 2555 (1986).
[20] T. Takagahara and K. Takeda, Phys. Rev. B **46**, 15578 (1992).
[21] M. V. Rama Krishna and R. A. Friesner, Phys. Rev. Lett. **67**, 629 (1991); J. Chem. Phys. **96**, 873 (1992).
[22] S. Y. Ren and J. D. Dow, Phys. Rev. B **45**, 6492 (1992).
[23] J. P. Proot, C. Delerue, and G. Allan, Appl. Phys. Lett. **61**, 1948 (1992).
[24] B. Delley and E. F. Steigmeier, Phys. Rev. B **47**, 1397 (1993).
[25] M. Hirao, T. Uda, and Y. Murayama, Mater. Res. Soc. Symp. Proc. **283**, 425 (1993).
[26] S. B. Zhang and A. Zunger, Phys. Rev. B **48**, 11204 (1993).
[27] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[28] In our example of the quantum box shown in Fig. 3, we have $\epsilon^{\text{bulk}}_n(k^*) = (-0.205, -0.555, -1.089)$ eV (measured from the top of bulk valence band) and $P_n = (0.774, 0.005, 0.149)$ for $n_1, n_2, n_3$, respectively. Using all three bands gives $\epsilon^{\text{dot}}_{\text{HOMO}} = -0.348$ eV compared with the exact result $-0.338$ eV, while using only $n_1$ gives $\epsilon^{\text{dot}}_{\text{HOMO}} = -0.205$ eV.