P.V. Kamat and D. Meisel (Eds.)
Semiconductor Nanoclusters
Studies in Surface Science and Catalysis, Vol. 103

Pseudopotential Theory of Nanometer Silicon Quantum Dots

Lin-Wang Wang and Alex Zunger

National Renewable Energy Laboratory, Golden, CO 80401

We present a systematic approach to the study of the electronic structure of thousand atom (nanometer scale) quantum structures. This approach uses the empirical pseudopotential method to approximate the Hamiltonian and a plane wave basis to expand the wavefunctions. Two complementary, newly developed methods are used to calculate the electronic structure of the system. The first method solves for the *discrete* near-edge states (the valence band maximum and the conduction band minimum). Its computational time scales *linearly* with the size of the system. The second method calculates *statistically* the electronic density of states and optical absorption spectra. For a given resolution and statistical accuracy, its computational time is independent of the size of the system for systems smaller than $\sim$ 10,000 atoms. The combination of these two methods is used to study the electronic and optical properties of up to thousand Si atom quantum dots passivated by hydrogen. The properties studied include: (1) band gap *vs* size; (2) band gap *vs* shape; (3) analysis of band edge states in terms of bulk Bloch functions; (4) total electronic density of state and optical absorption spectra; (5) dielectric constant *vs* size; (6) photoluminescence radiative lifetime *vs* luminescence photon energy. The results are compared with tight binding and other model calculations. Comparison with experimental data is made whenever possible. Good agreements with experiment are obtained for photoluminescence lifetime and for the ratio between conduction band shift and valence band shift.

1. INTRODUCTION

While electronic structure theory has traditionally focused on the detailed band structure of rather symmetric, simple solids, rapid experimental advances in semiconductor physics are constantly shifting interest to lower symmetry quantum systems with an ever increasing number of atoms. Examples in semiconductor physics include nanometer-sized ($> 10^3$ atoms) quantum dots, films and wires[1, 2], superlattices $A_nB_mA_{n'}B_{m'}...$ with randomly selected layer thicknesses [3], interdiffused ("interfacially rough") quantum wells[4], spatially inhomogeneous alloys exhibiting various degree of long and short range order[5], long period superlattices[6], impurity aggregates, line defects and dislocations in solids[7], amorphous and glassy solids[8] and surfaces with complex reconstructions and step structures[9]. The electronic structure of such low-symmetry nanometer-size systems is often cast in terms of solutions to some effective single-particle Schrodinger's equation:

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

where $\hat{H} = -\frac{1}{2}\nabla^2 + V(\mathbf{r})$ and $V(\mathbf{r})$ is the mean-field potential. Such 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 Eq(1). Examples include surfaces with unsuspected reconstruction geometries[10, 11] or crystals and molecules with intricate patterns of charge transfer and hybridization. This class of problems requires calculating iteratively the ground state energy and density, thus *all* occupied solutions of Eq(1) are needed. Development of iterative diagonalization methods[12], conjugate-gradient minimization of $< \psi_i|\hat{H}|\psi_i >$ [13], and Jacobian update scheme[14] for iterating Eq(1) have dramatically improved the efficiency of such "total energy first-principle calculations". However, the overall effort is still limited by the N^3 scaling of the orthogonalization step, where N is the number of atoms in the system. Consequently, current first-principle calculations are limited to $N < 100$ atoms. While parallel computation has raised the limitation to a few hundred atoms[11, 15], thousand atom computations are still costly. We have recently demonstrated that the orthogonality computation can be reduced using the Lanczos method[16], so that for $N < 1000$ atoms, *all* solutions of Eq(1) can be obtained with an overall scaling of N^2. For larger systems, however, the scaling is still N^3. There are currently a number of proposals for total energy electronic structure calculations with a linear-in-size scaling of the effort [17-22]. These promising approaches are, however, still in their formative stages, and the cross-over size of their cost with respect to the N^3 scaling method is yet unknown.

The second class of problems to which Eq(1) has been applied includes cases where $V(\mathbf{r})$ and the atomic geometry are either known(e,g, large, bulk-like quantum dots), or they can be transported from small-scale first principle calculations. An example of the latter case is the study of band-gap impurity levels in bulk solids, where $V(\mathbf{r})$ and the atomic relaxations are often *localized* near the impurity (and thus can be obtained from self-consistent calculations on small systems) but the wavefunctions extend over many atomic cells[23]. Again, conventional N^3 scaling first-principle methods are limited to $N < 100$ atoms. However, if $V(\mathbf{r})$ and the atomic geometry are known or can be modeled, there is no *need* to calculate all solutions of Eq(1). One could then focus instead on some selected properties of the electronic structure, e.g, the band gap, the density of state and optical absorption spectrum. Here, we provide an effective approach for addressing this class of problems.

Current approaches to "class-two problems" for $\sim 10^3$ atom systems includes the use of the effective mass approximation[2] (EMA) and the tight-binding (TB) approach[24]. The EMA removes the atomic-scale variation of $V(\mathbf{r})$, replacing it by an empty system with walls (e.g, particle-in-a-box) and parabolically varying energy bands. This approach is very useful[2] for quantum systems (dots, wires, films, superlattices) which are larger than $60 - 100\mathring{A}$ (i.e, having $> 10^4$ atoms), but it often fails[25] for intermediate sized systems with $100 - 5000$ atoms ($15 - 60\mathring{A}$) for which the real microscopic potential $V(\mathbf{r})$ can not be renormalized. The TB approach, on the other hand, does include an atomistic $V(\mathbf{r})$, but simplifies the problem by expanding $\psi_i(\mathbf{r})$ in an exceedingly small basis set whose

shape and variational flexibility remain unspecified. Furthermore, $V(\mathbf{r})$ is represented by a few Hamiltonian matrix elements within a finite interaction range. In the present author's view, this restricted variational expression of $\psi_i(\mathbf{r})$ and the inability to test the accuracy of $\psi_i(\mathbf{r})$ and $V(\mathbf{r})$ through explicit real-space comparison with first-principles results constitutes a severe limitation.

We focus here on "class-two problems" whose potential can be represented by a superposition of atom-centered quantities:

$$V(\mathbf{r}) = \sum_{atom} v_{atom}(\mathbf{r} - \mathbf{R}_{atom}), \tag{2}$$

and whose wavefunctions can be conveniently expanded in plane waves

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{q}} a_i(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}}. \tag{3}$$

Here, $\{\mathbf{R}\}_{atom}$ are atomic position vectors and $a_i(\mathbf{q})$ are variationally determined expansion coefficients at the reciprocal lattice vector $\mathbf{q}$ of the supercell. In general, the atom-centered potential v_{atom} need not be spherical, (although in our present work, spherically symmetric potentials are used). Our central approximation is that we use a *fixed* potential $V(\mathbf{r})$ that is not updated to reflect charge rearrangement in differently sized clusters. The plane wave basis is convenient because it permits calculation of the matrix-by-vector products $\hat{H}\psi_i$ in dual space representation[14, 26, 27] using fast Fourier transforms (scaling as $N\ ln\ N$). This avoids the N^2-scaling explicit matrix multiplication using $< \mathbf{q}_1|\hat{H}|\mathbf{q}_2 >$.

Modelling of $\sim 10^3$ atom systems using Eqs(1)-(3) requires three steps:

First, one needs to model the geometry, i.e, to specify $\{\mathbf{R}_{atom}\}$. For "class-two problems", we assume that this is known in advance. For example, experimental structural measurements on > 100 atom Si quantum dots[28] show that the atomic positions and the interatomic distances in the interior of the quantum dots are very close to the values of the extended bulk solid. Atomic relaxations exist only near the surface. But, as will be shown later, one usually has a reasonably good idea on how to model such relaxations using either first-principle calculations or experimental data on the relevant *bulk surfaces*. Another example of knowable structural information is mesoscopic strained layer superlattices, where the atomic structure can be reliably modeled[29] using continuum elasticity or first-principle calculations on *short* period systems. A final example is homogeneous alloys, where relaxations can be obtained from empirical force models such as the valence force field[30]. This *first* step will be demonstrated in Sec.2.1 in the context of Si quantum dots.

The *second* step in using Eqs(1)-(3) requires determination of the atom-centered potentials $v_{atom}(r)$. We use for this purpose the empirical pseudopotential method[31] (EPM). Rather than fit $v_{atom}(\mathbf{G_i})$ at a few *discrete* reciprocal lattice vector $\{\mathbf{G_i}\}$ of the primary unit cell, (as done in classic *bulk* EPM calculations[31]), we will fit a *continuous* $v_{atom}(q)$ to a series of experimental data and detailed first-principles calculations on relevant prototype systems. This will include in the properties that need to be fitted the bulk band structures, the surface work function, and the density of states of chemisorbed surfaces. Unlike the case in tight-binding approaches, we will be able to compare the ensuing potential $V(\mathbf{r})$ with screened first-principles local density approximation (LDA) results. Unlike

the case with the LDA, we will be able to obtain experimentally sensible excitation energies. This *second* step will be demonstrated in Sec.2.2 in the context of Si quantum dots.

The *third* step in using Eqs(1)-(3) involves solving Eq(1) for a given geometry and potential. This is nontrivial for $\sim$ 1000 atom systems since the number of plane wave basis functions in Eq(3) scales as O(N). For most semiconducting systems made up of main group elements, one needs about 50 plane waves per atom. So for $N = 1000$, one typically has a $50,000 \times 50,000$ Hamiltonian problem (the number of FFT grid points is much larger than 50,000 and is around 10^6). The conjugate gradient approaches[13] to this problem scales as N^3, so large systems can not be treated, if all occupied eigenstats of Eq(1) are to be calculated. However, we may not need *all* eigenstates in order to study, e.g, the threshold optical properties of semiconductor quantum structures. What one typically needs to know on such systems includes: (a) the eigenvalues and eigenfunctions of the band edge states (the valence band maximum, VBM and the conduction band minimum, CBM); (b) the total and local electronic density of states; (c) the optical absorption spectra. With these three properties calculated, most of the optical characteristics of the system can be determined. We have designed two new methods[32, 33] to calculate these properties. First, the "folded spectrum method"[32] calculates the band edge states. This will be introduced in Sec.3. Second, the "generalized moments method"[33] calculates the density of states and optical absorption spectra. This will be introduced in Sec.4. As will be demonstrated in Sec.5, using these two methods, it takes 2-3 Cray-YMP cpu hours to calculate the above mentioned electronic and optical properties of a thousand-atom system.

In section 6, we apply our approach to study Si quantum dots. The properties studied include: (a) band gap *vs* quantum dot size; (b) band gap *vs* quantum dot shape; (c) analysis of band edge states in terms of bulk Bloch functions; (d) electronic density of states and optical absorption spectra; (e) dielectric constant *vs* quantum dot size; (f) photoluminescence radiative lifetime *vs* luminescence photon energy. We have also compared our results with tight binding like direct calculations, with model calculations and with experiment. This appears to be the first application of pseudopotential plane wave methods to $\sim 10^3$ atom quantum structures.

2. CONSTRUCTING THE EMPIRICAL PSEUDOPOTENTIAL HAMILTONIAN

In this section we discuss the construction of the empirical pseudopotential $v_{atom}(r)$ of the system studied. We first determined the atomic configuration $\{\mathbf{R}_{atom}\}$ and then construct the atom-centered empirical pseudopotentials $v_{atom}(r)$ of Eq(2). Instead of discussing these steps in general, we will discuss them in the context of Si quantum dots which will be studied in this paper.

2.1. Determination of atomic configurations

Si quantum dots of size 10-40 $\mathring{A}$ can be made by vaporization of Si electrodes[34], microwave plasma decomposition of SiH_4[35], gas-phase pyrolysis of disilene[28], elec-

trochemical reaction [36], amorphous Si thin film crystallization[37], Si^+-implantation of silica glasses[38], and direct atomic force microscope fabrication[39]. Nanometer sized Si quantum dots are also expected to exist in porous Si[40, 41]. The surface of Si quantum dots can be passivated by hydrogen, oxygen or by a host amorphous matrix[37, 38]. X-ray experiments[28] and direct transmission electron microscopy (TEM) observations[35] show that the interior of the quantum dots has a bulk-like structure, with a lattice constant that is within 0.25% of the bulk value. According to the deformation potential[42] of bulk Si, this 0.25% lattice constant uncertainty leads to $< 5meV$ change in the band structure energy. Thus, without introducing any significant errors, we will use in the following the diamond structure and the *bulk* Si lattice constant to describe the interior of the Si quantum dots. We will fully passivate the surface by hydrogen. The main issue here is to model the atomic surface relaxation for the chemisorbed H and for the near surface Si atoms.

We model the H positions at the surface of the quantum dot from the known configurations of H-covered (111), (110) and (100) Si surfaces [43-49]. We thus represent the surface of a convex quantum dot by "patchs" of these three primary flat surfaces. This can be usually achieved rather easily; If it can not, a few atoms are added or removed from the surface to make it so (e.g, we remove all tri-hydride "-SiH_3" group on the surface). The correction is small, so the overall shape of the quantum dot does not change. The reconstructed surface geometries used for the three primary surfaces are: (1x1):H for the (111) oriented film surface[43, 44] [i.e, the surface unit cell is primitive 1x1 in terms of the original ideal surface periodicity, and each surface Si atom has one H atom]; (1x1):H for the (110) oriented film surface[45], and (1x1):2H for the (100) oriented film surface[46, 47] (each surface terminal Si atom has two H atoms). The previously determined relaxations for (111) and (110) surfaces are relatively small[48], so for these surfaces we will use the ideal unrelaxed structures with Si-H bond distance of 1.487$\AA$. There are different relaxation models[49] for the (100) (1x1):2H surface. We will use a "canted dihydride" model of Northrup[49], derived from a LDA total energy minimization. The reconstructed structure of this "canted dihydride" model is shown in Fig.1. The H-Si-H bond angle is 106°, not much different from the ideal tetrahedral bond angle of 109.5°. Note, however, that the H-Si-H group has been rotated, so that the two nearby H atoms from two neighboring H-Si-H groups avoid steric hinderance (they are at a distance of 2.21$\AA$). Figure 1 shows that the structures of the canted dihydride model is different from that of the SiH_4 molecule: there is no tetra-hydride at any surface Si site. Thus the SiH_4 molecular structure can not simulate the relative positions of neighboring SiH_n (n=1,2) groups on the real H-covered silicon surface. (We will see below that the H pseudopotential of the H-covered Si *surface* is also very different from the H potential in SiH_4. This molecule should therefor *not* be used as a paradigm for Si nanostructures).

Following the above procedure, we construct the atomic positions of any, arbitrarily shaped Si quantum dots studied in this paper.

2.2. Constructing the Si and H pseudopotentials

Once the atomic configuration $\{\mathbf{R}_{atom}\}$ has been determined, the next step is to fit the spherically symmetric atomic pseudopotential $v_{atom}(r)$ of Eq(2).

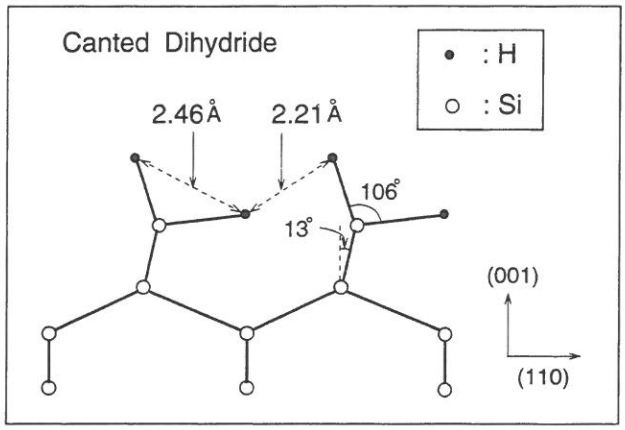

Figure 1: The atomic structure of Canted Dihydride (001) Si surface viewed from (110) as calculated in Ref.[49].

In the classical bulk EPM[31], the pseudopotential $v_{Si}(\mathbf{G})$ is defined only on the *discrete* set of bulk reciprocal lattice vectors $\{\mathbf{G_i}\}$. To describe different *finite* quantum dots inside different computational unit cells, one needs instead a *continuous* momentum space form $v_{Si}(q)$. We represent the Si local pseudopotential in the form[25]:

$$v_{Si}(q) = a_1(q^2 - a_2)/(a_3 e^{a_4 q^2} - 1). \tag{4}$$

The coefficients were fitted to (i) the bulk band structure at high symmetry points [50-54], (ii) the effective masses[55, 56] and (iii) the surface work function[57]. The bulk band structure was calculated in a plane wave basis [Eq(3)] with a energy cut off of 4.5 Ry (the same cut off is used in subsequent calculations) and a lattice constant of 5.43Å. The fit gave $a_1 = 0.2685$, $a_2 = 2.19$, $a_3 = 2.06$ and $a_4 = 0.487$ in atomic units (Hartree for energy, inverse Bohr for q). Table 1 compares the fitted quantities as obtained from the current EPM, the bulk local EPM of Chelikowsky and Cohen[58] and experiment [50-57]. It is clear from Table 1 that the two empirical pseudopotentials have similar quality; the band energies are within 0.1eV of the experimental data (i.e similar to the experimental uncertainty). Although non-local EPM[31] can improve the agreement of the calculated band structure with experiment, we consider the level of agreement reflected in Table 1 as sufficient for the current purposes, especially for studying band *edge* states.

Figure 2 compares the current atomic Si pseudopotential $v_{Si}(q)$ with the Fourier transform of the (self consistently) screened local LDA pseudopotential [59] obtained from bulk calculations. The ability to compare our potential to first principle potential is a significant advantage over tight-binding-like methods. The small, systematic difference between our empirical pseudopotential and the LDA potential reflects the fact that the current potential produces accurate band structures (compared with experiment) while the LDA does not.

Figure 3 compares our calculated bulk density of states (DOS) and optical absorption spectra $\epsilon_2(E)$ with experiment[54, 60]. We have used in this calculation the k.p method[61]

Table 1: Comparison of the bulk Si band structures and effective masses, as obtained in the present EPM[Eq(4)], the bulk local EPM of Chelikowsky and Cohen (Ref.[31]) and in experiments. We use a cut-off energy of 4.5Ryd for the plane wave expansion and a Si bulk lattice constant of 5.43$\AA$. The numbers in the bracket of the experimental data indicate the estimated error in the last digit. $m^{(i)}_{\Gamma-X}(h)$ and $m^{(i)}_{\Gamma-L}(h)$ stand for the non-spin-coupled effective hole mass [defined as $(\hbar k)^2/2\Delta E$] in the $\Gamma-X$ and $\Gamma-L$ directions, where i denotes the band degeneracy. W is the work function. Energies are in eV and effective masses are in the unit of electron mass.

Property	Present EPM	Bulk (Ref.[31]) EPM	Experiment
$\Gamma_{25'v}$	0.	0.	0.
$\Gamma_{1,v}$	-12.57	-12.68	$-12.5(6)^e$
Γ_{15c}	3.24	3.34	$3.35(1)^a$
$\Gamma_{2'c}$	4.12	4.19	$4.15(5)^b$
$L_{2'v}$	-10.19	-10.26	$-9.3(4)^e$
L_{1v}	-7.25	-7.33	$-6.8(2)^e$
$L_{3'v}$	-1.28	-1.27	$-1.2(2)^b$
L_{1c}	2.18	2.13	$2.04(6)^c$
L_{3c}	4.02	3.88	$3.9(1)^b$
X_{4v}	-3.01	-3.03	-2.9^b
X_{1c}	1.32	1.14	$1.13(?)^a$
Σ_{min}	-4.47	-4.55	-4.48^b
E_{gap}	1.167	1.062	1.124^d
W	4.96	–	4.9^h
$m_L(e)$	0.928	0.912	0.916^f
$m_T(e)$	0.199	0.194	0.19^f
$m^{(2)}_{\Gamma-X}(h)$	0.272	0.271	0.34^g
$m^{(1)}_{\Gamma-X}(h)$	0.168	0.170	0.15^g
$m^{(2)}_{\Gamma-L}(h)$	0.669	0.676	0.69^g
$m^{(1)}_{\Gamma-L}(h)$	0.098	0.097	0.11^g

a: from Ref.[50] b: from Ref.[51]
c: from Ref.[52] d: from Ref.[53]
e: from Ref.[54] f: from Ref.[55]
g: from Ref.[56] h: from Ref.[56]

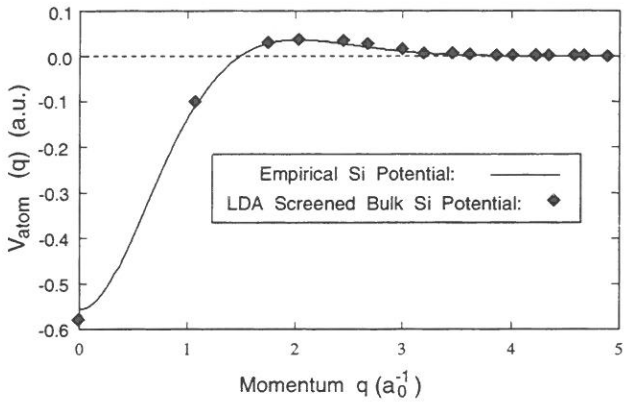

Figure 2: Comparison between the present screened Si pseudopotential $v_{atom}(q)$ and the screened first-principles local pseudopotential obtained from bulk Si self consistent LDA calculations. The screened first principle potential $V(G)$ is decomposed into atomic potentials according to $\sum_{R_i} e^{iR_iG} v_{atom}(G) = V(G)$, where G is the bulk reciprocal lattice vector and R_i is the Si atomic position.

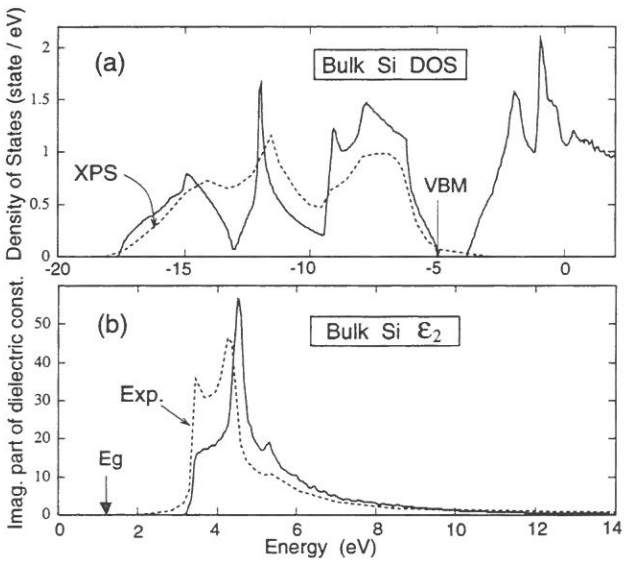

Figure 3: (a) Total density of states (DOS) and (b) optical absorption spectrum (ϵ_2) of bulk Si. The solid lines are calculated values using the present EPM and the dashed lines are experimental data. The XPS photoemission data in (a) is from Ref.[54]. The experimental data in (b) is from Ref.[60]. The lower energy peak in (b) is the exciton transition.

with 10^5 k points. The DOS compares well with experiment. However, as is the case with other single-particle calculations[31] (EPM or LDA), our single-particle Hamiltonian does not reproduce the exciton peak in the absorption spectrum. The static dielectric constant ϵ_s, calculated from our $\epsilon_2(E)$[31], is 10.38, compared with the experimental value[62] of 11.4, the LDA result 12.7[63] and the early EPM value of Walter and Cohen 11.3[64]. In our calculations below, we will correct for the difference between our value of 10.38 and experiment value of 11.4, (caused mostly by the neglect of exciton effects and possible imperfection in the band structure), by scaling our calculated $\epsilon_s - 1$ by a factor $(11.4-1)/(10.38-1) = 1.11$. Recently, we have developed [65] a more general method to construct non-local empirical pseudopotentials that have LDA-quality wavefunctions and experiment-quality band energies. The improvement for CdSe is substantial, but the results for Si are similar to those given here[65].

We next fit the hydrogen pseudopotential. The fitted quantities are the surface local density of states (LDOS) of the three primary [(100),(110),(111)] H covered Si films. We emphasize in particular obtaining the correct energy of the structurally sensitive LDOS peaks corresponding to the Si-H bonds. The experimental data are taken from ultra-violet photoemission spectroscopy[44, 45] (UPS) and angle-resolved electron-energy-loss spectroscopy[46] (AR-EELS). These experiments indicate that the *bonding* Si-H states are located around $E_v - 5\ eV$, where E_v is the valence band maximum. However, there are insufficient experimental data for determining the position of the conduction band Si-H *antibonding* states. We have thus used instead the results of first principle LDA calculations (measured from the conduction band minimum to avoid the LDA error). These calculations and considerations of LDA energy corrections of the unoccupied state[66] suggest that the antibonding surface state at $\bar{\Gamma}$ is at $\geq E_c + 1\ eV$. Similar conclusions apply to other $\bar{k}$ points.

The $v_H(q)$ fitted to the surface LDOS is (in atomic units)

$$v_H(q) = -0.1416 + 9.802 \times 10^{-3} q + 6.231 \times 10^{-2} q^2 - 1.895 \times 10^{-2} q^3; when\ q \leq 2$$

$$= 2.898 \times 10^{-2}/q - 0.3877/q^2 + 0.9692/q^3 - 1.022/q^4; when\ q > 2 \qquad (5)$$

The calculated surface LDOS obtained with this $v_H(q)$ and our $v_{Si}(q)$ of Eq(4) are shown in Fig.4 for the three primary surface orientations. As can be seen, the Si-H bonding states (indicated by the vertical arrows in the figure) are at $E_v - 5\ eV$, while the anti-bonding surface states are at $E_c + 1\ eV$. Figure 5 shows the contour plots of the real-space *total* potential $V(\mathbf{r})$ [Eq(2)] of the H-covered (100) Si film, as produced by our empirical pseudopotentials (right) and by first principle self-consistent local LDA pseudopotential calculation (left). The two potentials are very close, suggesting that we have a realistic description of the total potential $V(\mathbf{r})$. We emphasize here that fitting the electronic structure of H covered Si *surfaces* is crucial to a realistic description of H covered mesoscopic Si *quantum dots*, and that fitting instead the SiH_4 molecule produces Si-H surface states that are incorrectly placed energetically.

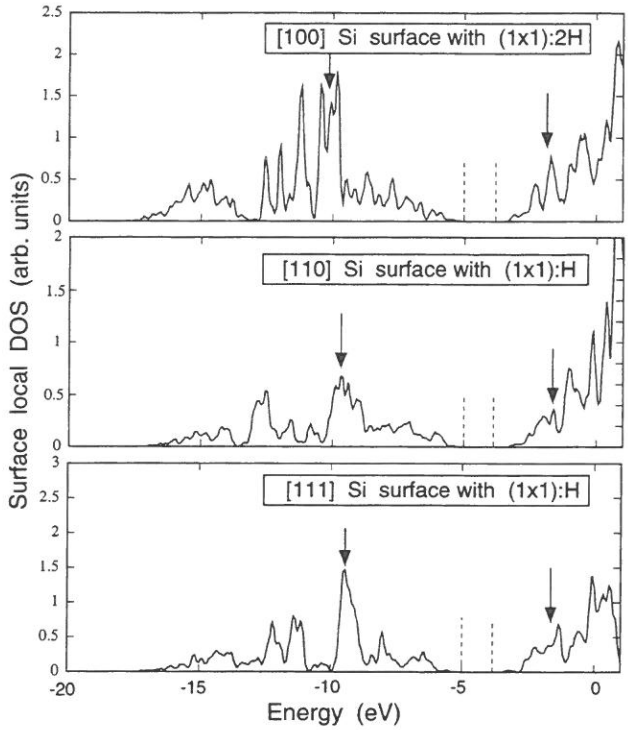

Figure 4: Surface local density of states of three primary hydrogen covered silicon surfaces calculated using the present EPM of Si and H atoms. The vertical lines indicates the energies of the Si-H bonding and antibonding states. The dashed vertical lines show the bulk VBM and CBM.

3. THE FOLDED SPECTRUM METHOD (FSM) FOR CALCULATION OF BAND EDGE STATES

The conventional variational approach to Eq(1) is to minimize the energy $< \psi|\hat{H}|\psi >$ by varying the expansion coefficients $a(\mathbf{q})$ of ψ [Eq(3)]; the first ψ obtained is then the lowest energy eigenstate of $\hat{H}$. To find a higher state, one needs to orthogonalize ψ to all previously converged energy eigenstates below it. The effort needed to accomplish this orthogonalization scales as N^3. Consequently, only small systems ($N \leq 100$) can be conveniently addressed.

We have developed a method that enables calculation of eigensolutions around a given, "interesting" energy, without having to calculate any of the eigensolutions below it. (A brief account was given in Ref.[32]). 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 this approach is that the eigensolutions (ϵ_i, ψ_i) of the Eq(1) also

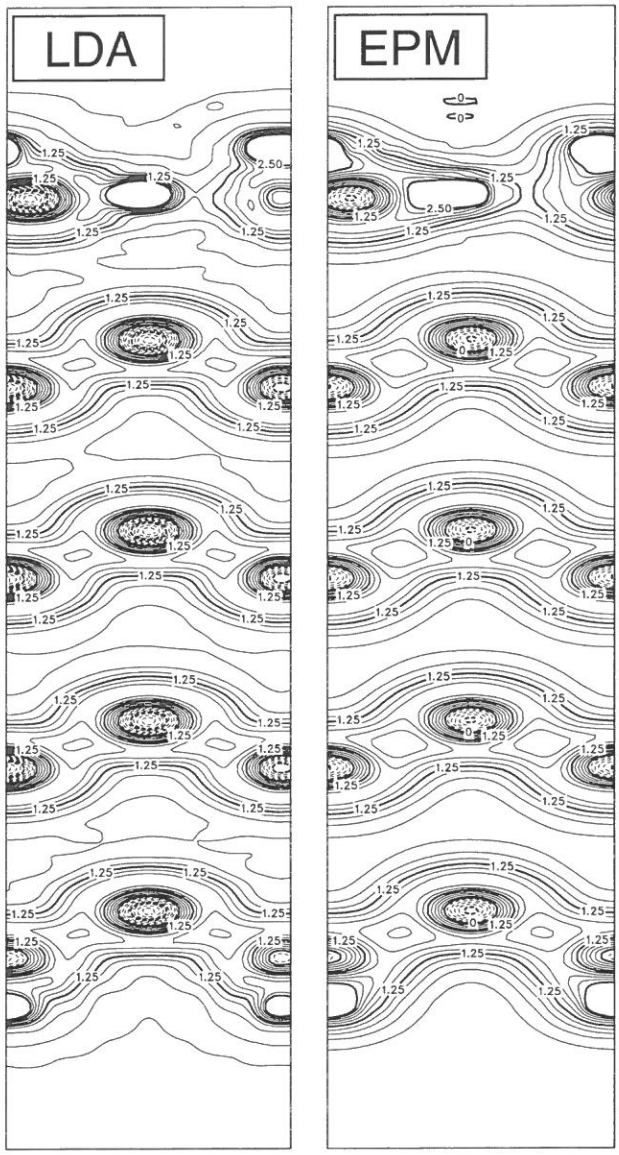

Figure 5: The contour plots of the (110) cross section of the total potential of a H covered (001) Si slab. The EPM potential is produced according to Eq(2) and the LDA screened potential is a self consistent result using *ab initio* atomic local pseudopotentials of Si and H. The contour level interval is 0.25 Hartree.

satisfy

$$(\hat{H} - \epsilon_{ref})^2 \psi_i = (\epsilon_i - \epsilon_{ref})^2 \psi_i. \tag{6}$$

In fact, Bendt and Zunger[14] and Wood and Zunger[12] proposed early on to treat $(\hat{H} - \epsilon)^2$ rather than $\hat{H}$ in the calculation of all occupied eigenstates, using, however different methods to solve for ψ_i. As shown schematically in Fig.6, the spectrum $\{\epsilon_i\}$ of $\hat{H}$ has been folded at the reference point ϵ_{ref} into the spectrum $\{(\epsilon_i - \epsilon_{ref})^2\}$ of $(\hat{H} - \epsilon_{ref})^2$. The lowest solution of the folded spectrum is the eigenstate with ϵ_i closest to ϵ_{ref}. Hence, by placing ϵ_{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 ϵ_{ref} is placed inside an energy gap, minimization of $< \psi|(\hat{H} - \epsilon_{ref})^2|\psi >$ results either in the valence band maximum state or the conduction band minimum state, depending on which is closer to ϵ_{ref}. Changing ϵ_{ref} within the gap region then assures that both the VBM and the CBM are found.

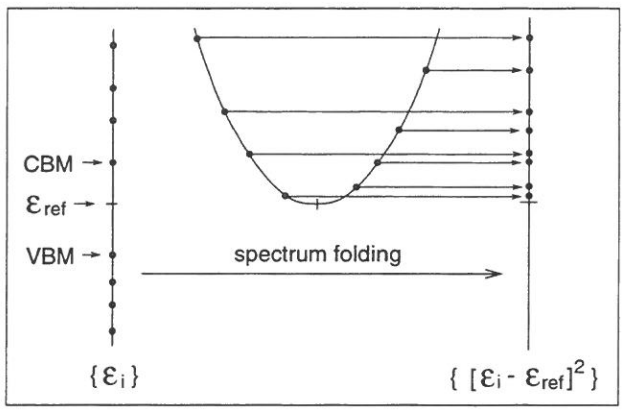

Figure 6: A schematic of the folding process of the spectrum $\{\epsilon_i\}$ to spectrum $\{[\epsilon_i - \epsilon_{ref}]^2\}$.

Our basic strategy is to solve Eq(6) by seeking the minimum of

$$F = \int \psi^*(r)[-\frac{1}{2}\nabla^2 + V(\mathbf{r}) - \epsilon_{ref}]^2 \psi(\mathbf{r}) d^3\mathbf{r} \tag{7}$$

in the space of the variational parameters $a(\mathbf{q})$ of ψ [Eq(3)]. However, comparing to the minimization of $< \psi|\hat{H}|\psi >$, the use of $(\hat{H} - \epsilon_{ref})^2$ slows down considerably the convergence of standard minimization methods. This problem is solved here by using preconditioned conjugate gradient method with large number of conjugate gradient steps. To calculate F, we apply twice $[-\frac{1}{2}\nabla^2 + V(\mathbf{r}) - \epsilon_{ref}]$ to $\psi(\mathbf{r}) = \sum_{\mathbf{q}} a(\mathbf{q}) e^{i\mathbf{qr}}$. The term $-\frac{1}{2}\nabla^2\psi$ is computed in reciprocal space, while $V(\mathbf{r})\psi(\mathbf{r})$ is obtained by using the FFT to transform $a(\mathbf{q})$ to real space $\psi(\mathbf{r})$, then applying V(r) to $\psi(\mathbf{r})$ and transforming the product back to $\mathbf{q}$ space. The result can be casted in the same form as $\sum_{\mathbf{q}} c(\mathbf{q}) e^{i\mathbf{qr}}$ (with the same energy cutoff for $\{\mathbf{q}\}$). Then, $[-\frac{1}{2}\nabla^2 + V(\mathbf{r}) - \epsilon_{ref}]$ is applied again to this

function to get the final result F. Once F is obtained, we minimize it with respect to the variational wavefunction coefficients $a(\mathbf{q})$, using the preconditioned conjugate gradient method[26]. The conjugate gradient method is defined as a series (indexed by $\{j\}$ below) of sequential line minimizations of the task function F. A line minimization implies adding a search wavefunction $P_j(\mathbf{r})$ to the current wavefunction $\psi_j(\mathbf{r})$ and constructing a new wavefunction $\psi_{j+1}(\mathbf{r})$

$$\psi_{j+1}(r) = \psi_j(r)cos(\theta) + P_j(r)sin(\theta) \tag{8}$$

which minimizes F at a value of θ. In this procedure, the search function $P_j(\mathbf{r})$ is made orthogonal to $\psi_j(\mathbf{r})$. The next search direction P_{j+1} is given by

$$P_{j+1}(\mathbf{q}) = A(\mathbf{q})\chi_{j+1}(\mathbf{q}) + \beta_j P_j(\mathbf{q}), \tag{9}$$

where

$$\chi_{j+1}(\mathbf{r}) \equiv \frac{\partial F}{\partial \psi_{j+1}(\mathbf{r})} = [-\frac{1}{2}\nabla^2 + V(\mathbf{r}) - \epsilon_{ref}]^2\psi_{j+1}(\mathbf{r}). \tag{10}$$

The preconditioner $A(\mathbf{q})$ is a $\mathbf{q}$-space function:

$$A(\mathbf{q}) = \frac{E_k^2}{(\frac{1}{2}q^2 + V_0 - \epsilon_{ref})^2 + E_k^2} \quad , \tag{11}$$

where V_0 is the average potential and E_k is the average kinetic energy of the wavefunction ψ. The β_j in Eq(9) is determined using the Polak-Ribiere formula[67]:

$$\beta_j = \frac{\sum_{\mathbf{q}} A(\mathbf{q})[\chi_{j+1}(\mathbf{q}) - \chi_j(\mathbf{q})]\chi_{j+1}(\mathbf{q})}{\sum_{\mathbf{q}} A(\mathbf{q})\chi_j(\mathbf{q})\chi_j(\mathbf{q})}. \tag{12}$$

Usually, a few (4 in our case) wavefunctions ψ are minimized simultaneously while being kept mutually orthogonal. N_l line minimizations steps are carried out for each wavefunction before a subspace diagonalization of these wavefunctions based on $(\hat{H} - \epsilon_{ref})^2$ is carried out. After this, we start another sequence of line minimization iterations. This forms an outside loop. [This algorithmic structure is the same as in the minimization of $< \psi|\hat{H}|\psi >$ in the conventional conjugate gradient method[26]]. Before the final result is obtained, a subspace diagonalization based on $\hat{H}$ is carried out. This gives the eigenenergy E_i. We used $N_l \sim 100$, (i.e, the square of typical N_l values used in conventional conjugate gradient methods based on $\hat{H}$). The number of outside loop N_i is 5, the same as the typical values used in the conventional methods. Following the above procedure, the computational effort to solve for each wavefunction $\psi(r)$ scales linearly with the system's size N. Because only a few wavefunctions need to be calculated, the whole effort of this method also scales linearly with the system's size N. This method will be tested in Sec.5.1.

4. GENERALIZED MOMENTS METHOD (GMM) FOR CALCULATION OF THE DENSITY OF STATES AND OPTICAL ABSORPTION SPECTRUM

While the Folded Spectrum Method allows us to calculate *discrete* band edge states, one often needs some *averaged*, (e.g, statistical) information about other states, noticeably, the density of states and the optical absorption spectrum. In this section, we will describe a generalized moments method to do just that. A preliminary account was given in Ref.[33].

The generalized moments approach for the calculation of the density of states $\rho(E)$ [22,68-72] consists of two steps. *First*, one defines the generalized *moments* I_n of the yet unknown $\rho(E)$ as:

$$I_n \equiv \int_{-1}^{1} T_n(E)\rho(E)dE, \tag{13}$$

where, $T_n(E)$ is a polynomial of power n defined in the interval [-1:1], and the energy of the Hamiltonian $\hat{H}$ has been scaled and shifted, so that all its eigenvalues are inside the [-1:1] interval. In the present study, we will utilize the Chebychev polynomial for $T_n(E)$ which forms an orthogonal polynomial set and affords a linear transformations between I_n and $\rho(E)$. The Chebychev polynomial is defined in the [-1:1] domain as:

$$T_0(E) = 1, \qquad T_1(E) = E$$

$$T_n(E) = 2ET_{n-1}(E) - T_{n-2}(E) \tag{14}$$

and obeys the orthogonality relation:

$$\int_{-1}^{1} T_m(E)T_n(E)(1-E^2)^{-\frac{1}{2}}dE = \frac{\pi}{2}\delta_{nm}(1+\delta_{m0}). \tag{15}$$

In the *second* step, one uses Eq(15) and Eq(13) to reconstruct $\rho(E)$ from its moments I_n as:

$$\rho(E) = \frac{2}{\pi}(1-E^2)^{-\frac{1}{2}}\sum_{n}^{N_{ch}} T_n(E)I_n(1+\delta_{n0})^{-1}, \tag{16}$$

where N_{ch} is the total number of Chebychev moments I_n used. As will be shown later, the property of the Chebychev polynomial will allow us to use the fast Fourier transform to compute Eq(16).

In the first step, we calculate the moments $\{I_n\}$ following Skilling[68] using statistical means. We assume that the Hamiltonian $\hat{H}$ has N_e (= the number of basis functions) eigensolutions $\{\psi_i, E_i\}$ when represented in a finite plane wave basis. A random wavefunction ϕ_0 can be constructed by using random coefficients $\{b(\mathbf{q})\}$ in the plane wave expansion of Eq(3), i.e,

$$\phi_0 = \sum_{\mathbf{q}} b(\mathbf{q})e^{i\mathbf{qr}}. \tag{17}$$

Because the transformation from $\{e^{i\mathbf{qr}}\}$ to the eigenstates $\{\psi_i\}$ is unitary, we can rewrite ϕ_0 as:

$$\phi_0 = \sum_{\mathbf{q}} b(\mathbf{q}) e^{i\mathbf{qr}} = \sum_{i}^{N_e} c_i \psi_i, \tag{18}$$

and $\{c_i\}$ are random numbers just like $\{b(\mathbf{q})\}$ with equal statistical means and zero correlations among them. If we normalize the wavefunction ϕ_0 to N_e, we have:

$$< c_i^* c_j >= \delta_{i,j}, \tag{19}$$

where $<>$ denotes average over different random wavefunctions $\{\phi_0\}$. Following the recursion formula of Eq(14), one can apply $\hat{H}$ to the wavefunctions to generate ϕ_n:

$$\phi_1 = \hat{H}\phi_0$$

$$\phi_n = 2\hat{H}\phi_{n-1} - \phi_{n-2} = T_n(\hat{H})\phi_0. \tag{20}$$

Now, taking products, we have:

$$I'_n = 2 < \phi_0 | \phi_n >= 2 \sum_{i}^{N_e} T_n(E_i) |c_i|^2. \tag{21}$$

If we use many random wavefunctions $\{\phi_0\}$, repeating the above process and taking the average of I'_n over $\{\phi_0\}$, we have:

$$< I'_n >= 2 \sum_{i}^{N_e} T_n(E_i) = \int_{-1}^{1} T_n(E)\rho(E)dE, \tag{22}$$

where we have used Eq(19) and $\rho(E) \equiv 2\sum_i \delta(E - E_i)$. Thus, if we average I'_n over a sufficiently large number N_{ave} of the random wavefunction ϕ_0, we can get a good approximation to I_n. After $\{I_n\}$ are obtained, Eq(16) is used to calculate $\rho(E)$ in the second step.

Equation (21), (22) apply to the total DOS, but with minor changes it can be used to obtain partial DOS. For example, to calculate the projected DOS on a given function χ, i.e, $\rho_p(E) \equiv 2\sum_i \delta(E - E_i)| < \psi_i | \chi > |^2$, simply replace the random wavefunction ϕ_0 by χ and repeat the above operations without averaging. To calculate a local DOS defined by a weighting function $f(r)$, i.e, $\rho_l(E) \equiv 2\sum_i \delta(E - E_i) \int |\psi_i(r)|^2 f(r) d^3r$, replace the product in Eq(21) by $2 < \phi_0 | f | \phi_n >$ and keep everything else the same.

The optical absorption spectra (or the imaginary part of dielectric constant) is defined as:

$$\epsilon_2(E) = A \sum_{i}^{occ} \sum_{f}^{unocc} \frac{1}{(E_f - E_i)^2} | < \psi_f | \hat{\mathbf{p}} | \psi_i > |^2 \delta(E - E_f + E_i), \tag{23}$$

where $A = 8\pi^2 e^2 \hbar^2 / 3m^2 \Omega$, and Ω is the volume of the system while E_i, E_f are the eigenenergies of states i and j, respectively. Note that $| < \psi_f | \hat{\mathbf{p}} | \psi_i > |^2$ is the transition

probability between occupied (occ) state ψ_i and unoccupied (unocc) state ψ_f, and $\hat{\mathbf{p}}$ is the momentum operator $i\vec{\nabla}$. The conventional way to calculate $\epsilon_2(E)$ is to solve for all eigenstates $\{\psi_j\}$, then calculate the transition matrix $< \psi_f|\hat{\mathbf{p}}|\psi_i >$ between ψ_i and ψ_f and sum Eq(23). As we pointed out above, this is impractical for systems with more than thousand atoms. Here, to calculate $\epsilon_2(E)$ we first calculate a two dimensional function:

$$\tau(E_1, E_2) \equiv \sum_{i,j \in all} | < \psi_i|\hat{\mathbf{p}}|\psi_j > |^2 \delta(E_1 - E_i)\delta(E_2 - E_j). \tag{24}$$

After $\tau(E_1, E_2)$ is obtained, $\epsilon_2(E)$ can be calculated (using the original unscaled energy) from:

$$\epsilon_2(E) = \frac{A}{E^2} \int_{-\infty}^{E_F} dE_2 \int_{E_F}^{\infty} dE_1 \tau(E_1, E_2)\delta[E - (E_2 - E_1)], \tag{25}$$

where E_F is the Fermi energy.

To calculate $\tau(E_1, E_2)$, we will first calculate its two-dimensional generalized moments. We generate a random wavefunction ϕ_0 as before [Eq(17)], then calculate:

$$\begin{aligned} \Lambda'_{n,m} &\equiv < \phi_0|\hat{\mathbf{p}}T_n(\hat{H}) \cdot \hat{\mathbf{p}}T_m(\hat{H})|\phi_0 > \\ &= \sum_{i,j,l \in all} c_l^* c_j T_n(E_i) T_m(E_j) < \psi_l|\hat{\mathbf{p}}|\psi_i > \cdot < \psi_i|\hat{\mathbf{p}}|\psi_j > \end{aligned} \tag{26}$$

We now repeat the above products with different random wavefunctions $\{\phi_0\}$ and average them. This gives:

$$\begin{aligned} \Lambda_{n,m} &\equiv < \Lambda'_{n,m} > = \sum_{i,j \in all} T_n(E_i) T_m(E_j)| < \psi_i|\hat{\mathbf{p}}|\psi_j > |^2 \\ &= \int_{-1}^{1} dE_1 \int_{-1}^{1} dE_2 T_n(E_1) T_m(E_2) \tau(E_1, E_2), \end{aligned} \tag{27}$$

where we have used Eq(19) and Eq(24). After obtaining $\Lambda_{n,m}$, the reconstruction of $\tau(E_1, E_2)$ using Eq(15) is straightforward:

$$\tau(E_1, E_2) = (\frac{2}{\pi})^2 (1 - E_1^2)^{-\frac{1}{2}} (1 - E_2^2)^{-\frac{1}{2}} \sum_{n,m}^{N_{ch}} T_n(E_1) T_m(E_2) \Lambda_{n,m} (1 + \delta_{n0})^{-1} (1 + \delta_{m0})^{-1}. \tag{28}$$

The calculation of Eq(26) is carried out by first calculating $\phi_m = T_m(\hat{H})\phi_0$ as defined in Eq(20). After all $\{\phi_m\}$ have been obtained, another set of wavefunction $\phi_n^p(k) \equiv T_n(\hat{H})\hat{p_k}\phi_0$ is obtained using the same recursion relation as in Eq(20). Here, k stands for the directions x,y,z (usually, by symmetry, only one or two k need to be calculated). Then $\Lambda'_{n,m}$ can be calculated as $\Lambda'_{n,m} = \sum_k < \phi_n^p(k)|\hat{p_k}|\phi_m >$. The computational effort of obtaining $\Lambda_{n,m}$ is of the same order as the effort in calculating I_n. The I_n can be calculated when one calculates $\Lambda_{n,m}$ essentially without any extra work. Because there is a large number of moments $\Lambda_{n,m}$ in Eq(28), its direct implementation can be time consuming. However, $T_n(E) = cos(n\theta)$ and $\theta = cos^{-1}(E)$. Substituting this into Eq(28)

we get a Fourier transformation for that equation. Thus, FFTs can be used to evaluate that equation. The same is true for Eq(16).

The idea of using moments to calculate the density of states is not new [22,68-72]. Compared with previous methods, the current approach has the following characteristics: (i) a plane wave basis rather than a tight-binding basis is used; (ii) a high resolution spectrum is obtained; (iii) because of (i) and (ii), a large number (500 or more) of Chebychev moments are calculated; (iv) unlike methods based on moments E^n, the current method is numerically stable even for a large number of Chebychev moments; (v) unlike methods based on the maximum entropy technique[22], the current method uses linear transformation from moments to the spectrum. This allows a large number of moments to be used without having to solve the nonlinear equations needed in the maximum entropy method[22]. The linear transformation we used is in fact a Fourier transformation. Thus FFT technique can be used to carry it out; (vi) most of all, we use a moments method to calculate not just the DOS, but also the optical absorption spectrum.

One interesting aspect of this statistical method is the scaling of its computing time as a function of the system's size N. The resolution ΔE of a constructed spectrum is roughly $2/N_{ch}$ for a given total number N_{ch} of Chebychev moments. The error of the statistical average (i.e, the fluctuation of the magnitude of the spectrum at a given energy) is proportional to $(\Delta N \times N_{ave})^{-1/2}$, where ΔN is the number of eigenstates inside one interval ΔE, and N_{ave} is the number of random wavefunctions used in the statistical average. Since for a fixed ΔE interval, ΔN is proportional to the size N of the system, N_{ave} should be inversely proportional to the size in order to give the same statistical accuracy for different system sizes. This yields a roughly *constant scaling of computing time with the size of the system.* This constant scaling changes to a linear scaling when the system exceeds a size at which $N_{ave} = 1$. In practise, this limit is $> 10,000$ atoms. The conventional methods which solve all the eigenstates then calculate the optical absorption spectrum scales as N^3 of the system's size N. The approximate crossover system size after which the current method becomes faster than the conventional ones, is about 100 atoms, for the same spectral resolution and statistical accuracy used in this paper.

5. TESTING THE FOLDED SPECTRUM AND THE GENERALIZED MOMENTS METHODS

5.1. Testing the folded spectrum method

Since the solutions of Eq(6) are formally identical to those of Eq(1)[73], no tests for the *accuracy* of the FSM is necessary. The test here is for its speed. We will use for this purpose six rectangular Si quantum boxes: $Si_{13}H_{28}$, $Si_{59}H_{76}$, $Si_{163}H_{148}$, $Si_{349}H_{244}$, $Si_{641}H_{364}$ and $Si_{1063}H_{508}$. Their Hamiltonians and atomic structure are generated using the methods described in Sec.2. For comparison with the present folded spectrum method, we independently use the more conventional precoditioned conjugate gradient method[26] to solve Eq(1) for the two smallest systems $Si_{13}H_{28}$ and $Si_{59}H_{76}$ (larger systems are too costly to calculate using this method). We estimate the computing times for larger clusters using the actual time needed in conjugate gradient calculations for these small clusters plus the scalings of various parts of the program. We used in the current folded spectrum

Table 2: Computer time (in Cray YMP cpu seconds) needed to compute four valence and four conduction band eigenstates using the folded spectrum method. N_{occ} is the total number of occupied states. The error tolerance for the eigenstates [defined as $< \psi|(\hat{H} - < \psi|\hat{H}|\psi >)^2|\psi >^{1/2}]$ is 2×10^{-7} Hartree.

system	N_{occ}	FFT grid	cpu time (s)
$Si_{13}H_{28}$	40	$20 \times 20 \times 27$	100
$Si_{59}H_{76}$	156	$25 \times 25 \times 36$	320
$Si_{163}H_{148}$	400	$32 \times 32 \times 45$	720
$Si_{349}H_{244}$	820	$40 \times 40 \times 54$	1760
$Si_{641}H_{364}$	1464	$45 \times 45 \times 64$	2690
$Si_{1063}H_{508}$	2380	$50 \times 50 \times 72$	2980

method the same convergence tolerance as for the conventional conjugate gradient method. Four states near the VBM and four states near the CBM are calculated for each system. The resulting computing times are summarized in Table 2, and illustrated in Fig.7. We see that the computing time of the current method scales linearly as a function of the system's size. For the largest system $Si_{1063}H_{508}$ studied here, the conjugate gradient method would require two weeks Cray-YMP cpu time, while the current method takes less than one hour of Cray-YMP cpu time.

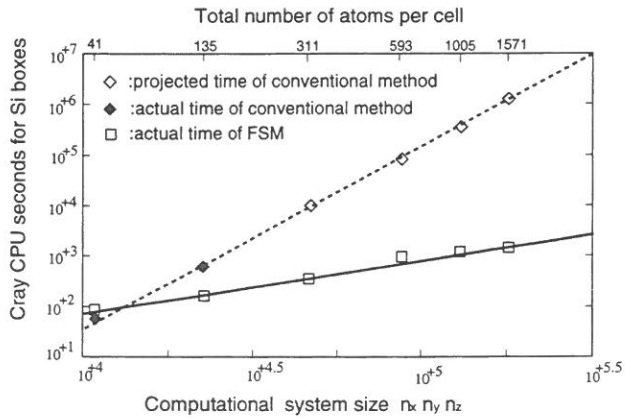

Figure 7: Computational time t (in units of Cray YMP cpu seconds) *vs* the size n of the system measured by the number $n_x n_y n_z$ of fast Fourier transform grid points. The solid line is a fit: $t = 9.0 \times 10^{-3} n$. The dashed line depicts a projection of the computer time needed with the conventional conjugate gradient method based on Eq(1). The system calculated here are rectangular shaped quantum dots.

It is interesting to test the convergence of the folded spectrum algorithm regarding the number N_l of line minimizations, the number N_i of outside loop iterations and the position of the reference energy ϵ_{ref}, (all defined in Sec.3). We find that interchanging N_l (=100) and N_i(=5) slows down the convergence considerably. This means that the

conjugate gradient method (with 100 line minimization steps) does speed up the convergence compare to the simple steepest decent method[67]. The position of ϵ_{ref} also plays a role in the convergence. Placement of ϵ_{ref} at the exact center of the band gap leads to an accidental degeneracy in the folded spectrum between a uper band edge state and a lower band edge state (see Fig.6). This could cause a slow down in the convergence. Also, if ϵ_{ref} is placed very close to the VBM or the CBM, or inside one band, this will also considerably slow down the convergence. All of these problems can be avoided by simply shifting ϵ_{ref}.

5.2. Testing the generalized moments method

Since the generalized moment method is not exact, in order to test its accuracy we need to compare its results to those of direct calculation. However, it is impractical to calculate the DOS of large quantum dots using conventional approaches (which solve for all eigenstates). We will thus use a large *bulk* supercell, containing 1024 primitive unit cells (2048 Si atoms) of bulk Si with periodic boundary conditions. Since the eigenstates of this supercell are identical to those of primitive unit cell of bulk Si (2 atoms/cell) at the corresponding *folded* k points, the exact eigenstates can be calculated easily using direct diagonalization. The results will be compared with GMM calculations of the 2048-atom supercell treated as a new system in its own right. The FFT grid is $64 \times 64 \times 96$ and we use $N_{ch} = 500$ Chebychev polynomial iterations and $N_{ave} = 20$ random wavefunctions averaging. The computation time for the density of states and optical absorption spectrum using a single $\hat{p}_k$ is about 3 Cray-YMP cpu hours.

Figure 8 compares the directly calculated ("exact" for the present purpose) density of states (part a) and optical absorption spectrum (part b) with those found using the GMM. (Note that the directly calculated results in Fig.8 are different from the truly bulk results in Fig.3 because there are only 1024 k points in our current test system). Because we use only a finite number N_{ch} of Chebychev polynomials, this corresponds to a certain convolution (broadening) in the energy space of the spectra obtained from Eq(16) and Eq(28). Thus, using Eq(16) and Eq(28) without modification will lead to undesirable sharp oscillatory features. To avoid this we apply two choices of smooth truncation functions $e^{-(n/0.5N_{ch})^2}$ and $e^{-(n/0.8N_{ch})^8}$ to the moments I_n and $\Lambda_{n,m}$. These two choices correspond to Gaussian and a smooth oscillation broadening in energy space, respectively. In Fig.8, we use the Gaussian broadening. The same Gaussian broadening of width 0.3 eV is used in the direct calculations. We see that the GMM mimics the "exact" results closely. The integrated sum of the DOS produced by the GMM upto the Fermi energy is 99.7% of the exact occupation number.

The static dielectric constant ϵ_s can be calculated from ϵ_2 by:

$$\epsilon_s = 1 + \frac{2}{\pi} \int_0^\infty \frac{\epsilon_2(E)}{E} dE. \tag{29}$$

Using this formula, we get ϵ_s of 10.305 and 10.572 for the direct and the current GMM, respectively. The difference of 2.5% is similar to the estimated 1.5% statistical error caused by the finite N_{ave}. The error in ϵ_s caused by broadening (finite N_{ch}) is about 3%. This is estimated from the difference between the results of the two different broadenings.

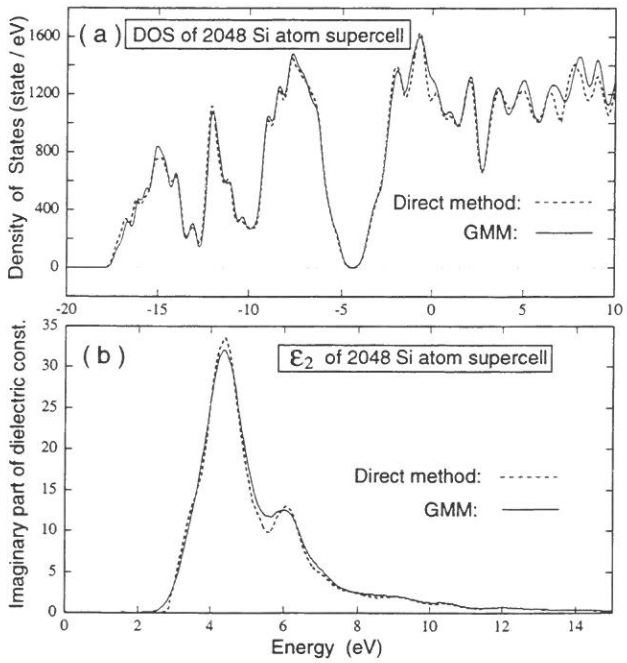

Figure 8: (a) The DOS and (b) the optical absorption spectra of a 2048 Si atom bulk supercell. The Dashed lines indicate "exact" calculations using folded states obtained in a primitive unit cell conjugate gradient calculation, while solid lines give the results of the GMM. The same Gaussian broadening (0.3 eV) is used in both GMM and exact calculations. Note that the exactly calculated curves are different from the truly bulk results in Fig.3, because there are only 1024 folded k points in the current supercell system.

In conclusion, we can calculate a thousand atom system within a few Cray-YMP cpu hours.

6. CALCULATIONS ON Si QUANTUM DOTS

In this section, we will use our approach to study the Si quantum dots. Because of the quantum confinement effect, Si quantum dots with different sizes and shapes can exhibit different colors throughout the visible range[74] . In this work, all quantum dot's atomic configurations and their Hamiltonians will be constructed by the method introduced in Sec.2. The quantum dots are placed in a periodic unit cell with quantum dot's surrounding filled by vacuum. Then the systems are calculated by FSM and GMM methods. Preliminary accounts of these results were given in Ref.[32, 33, 75, 76]. This book chapter provides, however, new data, comparisons, and analyses not given earlier.

6.1. Dependence of the band gap on size for Si spheres, cubes and rectangular boxes

The most commonly addressed question in quantum dot physics[1] is the size dependence of the energy gap. However, experimentally prepared quantum dots not only have different sizes, but for each size there could be a distribution of *shapes* and *surface orientations*. To understand the effects of the shapes on the energy levels of quantum dots, we study here three "prototype shapes": (i) *spherical balls*, (ii) *rectangular boxs* [the surfaces are in the (110),(1$\bar{1}$0) and (001) directions and the lengths of the edges satisfy $d_x = d_y = d_z/\sqrt{2}$] and (iii) *cubic boxs* [the surfaces are in the (001),(010),(100) directions]. To compare the electronic properties of these different prototype shapes, we need a consistent definition of the quantum dot's size. A natural choice is to associate the effective size with the diameter of a sphere which has the mass density ρ of bulk Si and contains the same number N of silicon atoms as the quantum dot in question. Then the diameter is $d(N_{Si}) = (\frac{3}{4\pi\rho})^{\frac{1}{3}} N_{Si}^{\frac{1}{3}} = 3.3685 N_{Si}^{\frac{1}{3}}(\mathring{A})$ and the radius $R = d/2$. Using this definition, the calculated size dependence of the CBM-VBM band gaps of the three prototype quantum dots is reported in Table 3 and depicted as symbols and the solid line in Fig.9. Quite surprisely, the three sets of data corresponding to the three prototype quantum dots collapse into a single, unified curve. Thus, if we measure the effective size by $d \propto N_{Si}^{1/3}$ and vary N_{Si}, the gaps of all prototype shapes (which are not too prolate) fall on a "universal" curve. Express d in $\mathring{A}$, this curve can be fitted as:

$$E_g(d) = 1.167 + \frac{88.34}{d^{1.37}} \quad (eV). \tag{30}$$

The "EMA" and "RKF" curves of Fig.9 will be discussed in sec.6.3.

6.2. Quantum dot wavefunctions and the role of surface atoms

It is commonly thought that since the surface-to-volume ratio increases rapidly as the quantum dot decreases, this must imply a greater role of surface effects on the electronic structure. This is true only if the wavefunction has an amplitude on the surface atoms. We test this next.

Shown in Fig.10(a),(b) are the wavefunction square of the CBM and VBM of the rectangular quantum box with $d = 34.1\mathring{A}(N_{Si} = 1035)$. While other quantum dots represented in Fig.9 may have different wavefunction patterns, in all cases, the VBM and CBM states are found to be *localized in the interior of the quantum dot*, with zero amplitude on the surface. Because of this and the fact that hydrogen potential is of very short range, we find that, as long as all dangling bonds are passivated, the details of the surface passivating atoms play little direct role in the determination of the wavefunctions, hence the band gaps and the oscillator strengths.

Next, we will try to find the relation between the quantum dot wavefunctions and the bulk Bloch states. Both the effective mass[2] and the truncated crystal methods[25] model the states of quantum structure in terms of an expansion in bulk Bloch states. It is thus helpful also to expand our *directly calculated* "exact" wavefunctions in terms of

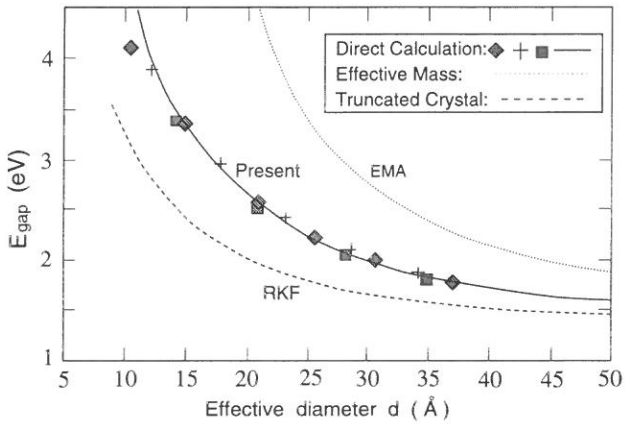

Figure 9: CBM-VBM band gap (without Coulomb corrections) versus the effective diameter $d = 3.3685 N_{Si}^{1/3}(\AA)$ for three prototype quantum dot shapes. The symbols $\diamond$, $+$, and $\square$ stand for the spheres, $(110) \times (1\bar{1}0) \times (001)$ rectanular boxes, and $(100) \times (010) \times (001)$ cubic boxes, respectively. The fitted solid line is given by Eq(30). Also shown are the multiband effective mass result (EMA) (Ref.[79]) and the result of the method of Rama Krishna and Friesner (RKF) (Ref.[78]) [applied here to cubic quantum dots with the present Si pseudopotential of Eq(4)]. In all cases, the excitonic Coulomb energy is excluded.

bulk Bloch wavefunctions:

$$\psi_i^{dot} = \sum_{n,k} a_{n,k}^{(i)} \psi_{n,k}^{bulk} \tag{31}$$

where k and n are the wavevector and the band index of the bulk Bloch wavefunction $\psi_{n,k}^{bulk}$. Consider, as an example, the rectangular quantum box whose directly calculated wavefunctions are shown in Fig.10(a),(b). In a rectangular box, k of Eq(31) is 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 is obtained for $j_x = j_y = j_z = 1$, which corresponds to eight symmetric wavevectors which will be denoted here as k^*. We have calculated the projection $P_n = \sum_{k^*} |a_{n,k^*}|^2$ for the quantum dot VBM state onto bulk wavefunctions and found that as much as 93% of the amplitude of $|\psi_{VBM}^{dot}|^2$ comes from the three upper valence bands n_1, n_2, n_3 at k^* (These three bands become triply degenerated at the $\Gamma_{25,v}$ point). The remaining 7% come primarily from other k points for the same bands. Similar percentage is found in other quantum dots. The projection weights P_{n_1,n_2,n_3} for these three bands are 0.774,0.005,0.149(the small P_{n_2} value in this case is accidental; for most other cases we tested, these three numbers are comparable). Figure 10(c) shows the approximate $\psi_{VBM}^2(r)$, using just these three bulk bands at k^* in Eq(31). We see that this approximation is very close to the directly calculated wavefunction shown in Fig.10(b).

From our analysis of the wavefunctions, we conclude that (i) the band-edge quantum dot wavefunctions are "bulk-like" rather than "surface-like" in that they can be constructed from just a few bulk Bloch wavefunctions; (ii) It is essential to have in Eq(31)

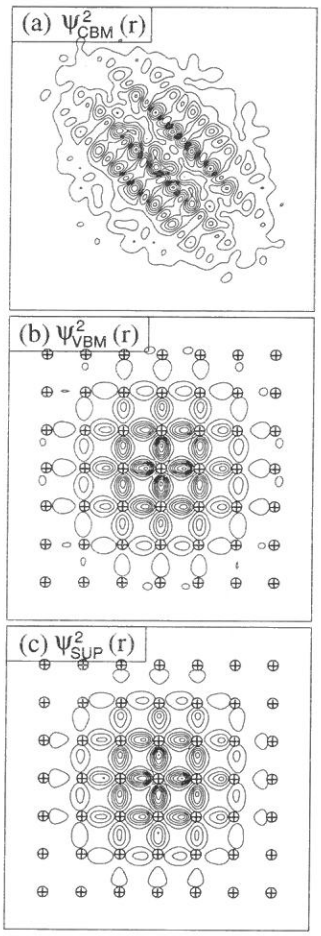

Figure 10: Wavefunction square contour plots of the (110) × (1$\bar{1}$0) × (001) rectangular quantum box with $d = 34.1\mathring{A}$ and $N_{Si} = 1035$ viewed from the [001] direction. (a) The CBM wavefunction square summed along the z direction. (b) The VBM wavefunction square plotted on the $z = d_z/2$ cross section. The crossed circles in (b) and (c) denote the positions of the silicon atoms on that plane. (c) The square of the VBM wave function reconstructed from Eq(31) using just three bulk bands at k^* (see text). This is plotted on the same cross section and has the same contour levels as in (b).

Table 3: Band edge energy shifts $\Delta E_v(d) = E_{VBM}(bulk) - E_{VBM}(d)$ and $\Delta E_c(d) = E_{CBM}(d) - E_{CBM}(bulk)$, for different sized and shapes of the quantum dots. The vacuum level are the same for bulk and quantum dots. Thus, these shifts are also the shifts relative to Si deep core levels (Ref.[100]). $E_{gap} = \Delta E_v + \Delta E_c + E_{gap}(bulk)$. Here $E_{Coulomb} = -3.572/d\bar{\epsilon}_s(d)$, is the electron-hole Coulomb binding energy in a spherical quantum dot. The measurements of Van Buuren etal [100] (see Fig.21 here) refer to $\Delta E_v(d)$ *vs* $\Delta E_c(d) + E_{Coulomb}(d)$. "Sphere" stands for spherical quantum dots; "rectangular" stands for rectangular box quantum dots with their three surfaces in $(110) \times (1\bar{1}0) \times (001)$ and edge length ratios as $d_x : d_y : d_z = 1 : 1 : \sqrt{2}$; "cubic" stands for cubic quantum dots with their three surfaces in $(100) \times (010) \times (001)$.

system	shape	d ($\AA$)	ΔE_v (eV)	ΔE_c (eV)	E_{gap} (eV)	$E_{Coulomb}$ (eV)
$Si_{29}H_{36}$	sphere	10.35	1.311	1.623	4.101	-0.957
$Si_{87}H_{76}$	sphere	14.93	0.989	1.192	3.348	-0.532
$Si_{235}H_{140}$	sphere	20.79	0.648	0.726	2.541	-0.325
$Si_{429}H_{228}$	sphere	25.41	0.513	0.544	2.225	-0.245
$Si_{741}H_{300}$	sphere	30.48	0.416	0.415	1.998	-0.192
$Si_{1315}H_{460}$	sphere	36.90	0.303	0.311	1.781	-0.151
$Si_{47}H_{52}$	rectangular	12.16	1.176	1.547	3.890	–
$Si_{147}H_{116}$	rectangular	17.78	0.834	0.946	2.948	–
$Si_{329}H_{204}$	rectangular	23.25	0.587	0.661	2.415	–
$Si_{617}H_{316}$	rectangular	28.68	0.443	0.487	2.097	–
$Si_{1035}H_{452}$	rectangular	34.07	0.338	0.363	1.868	–
$Si_{75}H_{76}$	cubic	14.21	1.027	1.179	3.373	–
$Si_{239}H_{172}$	cubic	20.90	0.657	0.684	2.509	–
$Si_{577}H_{340}$	cubic	28.04	0.410	0.488	2.065	–
$Si_{1101}H_{532}$	cubic	34.78	0.294	0.356	1.817	–

a band mixing since no single bulk band represents accurately the wavefunction of the quantum dot[77]. We will see in Sec.6.3 that using a single bulk band to represent a dot (Rama-Krishna and Friesner[78]) is insufficient; (iii) The k-point selection rules of particle-in-a-box are a reasonable approximation to the exact results.

6.3. Comparison of the band gaps with previous calculations

Figure 9 compares the results of two model calculations with our *direct* ("exact") calculations. These models includes the multiband effective mass approximation (EMA) of Takagahara and Takeda[79] and the model of Rama Krishna and Friesner(RKF)[78], recalculated here(for consistency of comparison) for cubic boxes using the present pseudopotential of Eq(4).

Comparison with the effective mass method: As we saw in Fig.10 above, the VBM and CBM states found in our direct calculations are not surface states, hence a comparison with the results of the (surfaceless) EMA is warranted. Our "exact" calculation result of Eq(30) gives a $1/d^{1.37}$ size scaling, while the effective mass model predicts a $1/d^2$

scaling. Hence, the parabolic dispersion assumed in the EMA is inadequate in the range of quantum dot sizes studied here($d < 40\mathring{A}$), despite the fact that multiband coupling is correctly included in this EMA calculation. Indeed, the effective mass approximation (which includes only kinetic energy effects neglecting *explicit* potential energy within the dot) exaggerates considerably the increase of band gap$[\Delta E_g(d) = E_g(d) - E_g^{bulk}]$ with reduced size. Replacing in the EMA calculation the infinite wall by a finite barrier reduces $\Delta E_g(d)$ and softens the $1/d^2$ scaling. Solving the EMA equation for a finite barrier of height 4 eV and using an effective mass $m^* = 0.2m$ gives a 10% lower $\Delta E_g(d)$ for $d = 40\mathring{A}$ and a 15% lower $\Delta E_g(d)$ for $d = 25\mathring{A}$. This reduces the EMA error relative to our direct calculations by 20% and 30% for $d = 40$ and $25\mathring{A}$, respectively. The remaining, bigger part of the error must come from (i) the EMA Hamiltonian itself, i.e, from the assumption of parabolic dispersion, and (ii) possible non abruptness of the potential well.

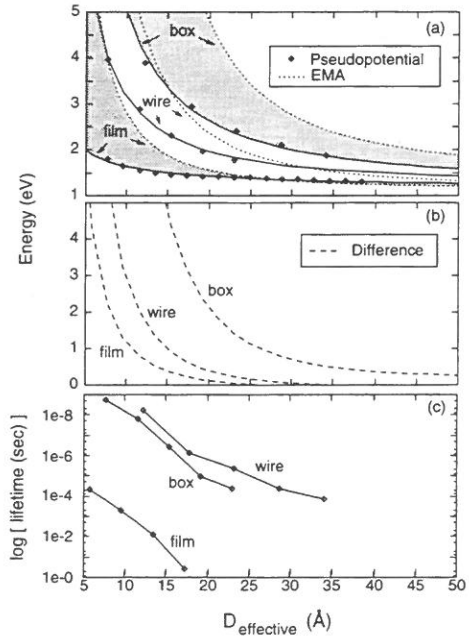

Figure 11: Energy gaps (part a), absolute EMA errors in the band gap (part b) and radiative lifetimes (part c) for H-covered Si films, wires and boxes. The surface orientations are (110) for filmes, (110) $\times$ ($\bar{1}$10) for wires, and (110) $\times$ ($\bar{1}$10) $\times$ (001) for boxes. We use $D_{eff} = D_{110}$ for films, $D_{eff} = D_{110} = D_{1\bar{1}0}$ for wires and $D_{eff} = 3.369 N_{Si}^{1/3}$ for boxes, where N_{Si} is the number of Si atoms and D_{110} is the H surface layer to H surface layer distance in (11) direction. This figure is taken from Ref.[80].

Zunger et.al [80] have extended the comparison of (single band) effective mass *vs* pseudopotential method to Si wires and films. The results are shown in Fig.11. We see that the EMA overestimates significantly the quantum confinement shift $\Delta E_i(d) = E_i(d) - E_i(bulk)$ and that the order of the EMA *errors* is $\delta\Delta E(box) > \delta\Delta E(wire) > \delta\Delta E(film)$

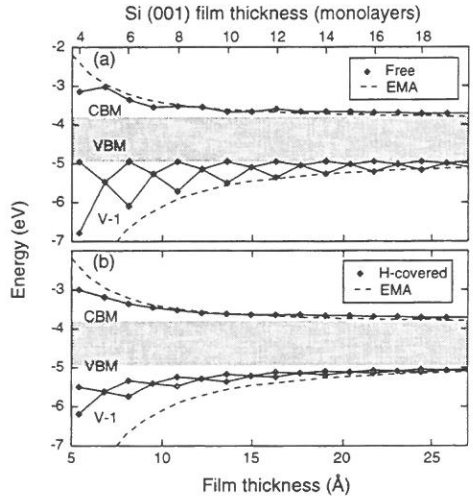

Figure 12: Calculated near-gap energy levels of (001) oriented hydrogen free (part a) and hydrogen covered (part b) Si films. All states shown are bulk-like (i.e, surface states in part a are omitted). Note the oscillations in the highest occupied (VBM) and the next-highest occupied (V-1) valence bands, absent in the EMA description (dashed line). The zero-confinement state having a size independent energy is apparent in part a. This figure is taken from Ref.[80].

[Fig.11(b)]. Note that the error is smallest for 2D films (in vacuum), and is probably still smaller in 2D quantum wells (embedded in a barrier). This explains partly the success of the EMA in quantum wells, but warns against the hope that similar success will be carried over to wire and dots. Note also from Fig.11(a) that the size dependence of the quantum confinement is the weakest fot films (d^{-n} with $n \sim 0.8$ compared to $n \sim 1.3$ for wires). Thus, if one wants to make a 2 eV gap Si system, one needs a $\sim 5\mathring{A}$ film or a $\sim 20\mathring{A}$ wire or a $\sim 32\mathring{A}$ dot.

It is interesting to note that in 2D quantum *films*, the effective mass method leads not only to quantitative but also to *qualitative* errors. This is illustrated in Fig.12 from Ref.[80]. While in the EMA the energy levels vary monotonically with size, the EPM solutions for (001)-oriented quantum films [Fig.12(a)] show non-monotonic (even-odd) oscillations in the valence band energies. The amplitude of these oscillations is reduced somewhat when the film's surface is covered by hydrogen [Fig.12(b)]. No oscillations exist in (110) oriented films. The EMA fails quantitatively in describing the valence band states (in either cases: clean or H covered surfaces) for film's thickness below $\sim 20\mathring{A}$. The reason is that the EMA fails to recognize the changes in symmetry in going from an even to an odd number of atomic layers: EMA has only the film thickness as its parameter. The EPM solutions of a hydrogen-free (001) quantum film exhibit a VBM state whose energy does not change with film thickness [see even-layers in Fig.12(a)]. This "zero confinement state" (ZCS) is missing in the EMA result. The ZCS corresponds to a cosine-type envelope function which is forbidden in EMA. But it is allowed in the EPM calculation because

the boundary condition is satisfied by the Bloch function, not by the envelope function. The ZCS energy becomes unpinned under H chemisorption [Fig.12(b)] and in (110) films. It is absent in wires and boxes.

A recent self-consistent calculation by Delley [81] has reproduced these even-odd oscillations, but with a reduced amplitude. Similarly, recent tight-binding calculation on films by Gavrilenko and Koch [82] noted such oscillations as well as the ZCS.

Comparison with the method of RKF: In the method of Rama-Krishna and Friesner, the cluster eigenvalues E_i^{dot} are approximated by *bulk* band structure energies $E_n^{bulk}(k^*)$ at the k^*-point of a particle-in-a-box model, i.e $E_i^{dot} \cong E_n^{bulk}(k^*)$. Sec.6.2 shows that in general, more than one band is needed to describe a quantum dot wavefunction. Figure 9 shows that the method of RKF[78] which corrects the parabolic scaling of the effective mass model by explicitly using the dispersion relations of the bulk band structure, *underestimates* the band gap opening $\Delta E(d)$. The reason for this is the neglect of band mixing: Eq(31) suggests that $\epsilon_{VBM}^{dot} \cong \sum_n P_n \epsilon_n^{bulk}(k^*) / \sum_n P_n$. The simple truncated crystal method uses the same particle-in-a-box k^* value as in the above analysis, computes the bulk bands from a similar empirical pseudopotential, but assumes *ad-hoc* that a *single* band(the highest) can be used in the sum of Eq(31). The neglect of the other lighter bands results therefore in a VBM that is too high, thus in a band gap that is *too small.*

It is significant that the single-band approximation of RKF *underestimates* the band gap. In a recent paper[83], the authors noted that the experimental observation of an unusually *low* band gap for Si dots (see Fig.19 below) supports their model. However, improving their model by allowing coupling with other bands shifts their band gap upwards significantly (by $\sim 0.6eV$ for $d = 20\mathring{A}$ see Fig.9), thus removing the claimed agreement with experiment.

Comparison with other direct calculations: Figure 11 compares our results for E_g *vs* d with four previous *direct* calculations: the empirically fitted nearest neighbor tight-binding(NN-TB) model of Ren and Dow[84], the empirically fitted third neighbor (nonorthogonal) tight-binding(TNN-TB) model of Proot, Delerue and Allan[85], and two LDA calculations: one which uses a small LCAO basis(LCAO-LDA) by Delley and Steigmeier[86] and one which uses a plane wave basis(PW-LDA) by Hirao, Udo and Murayama[87] (limited to small $N_{Si} \leq 123$ quantum dots). In all calculations, an ideal atomic structures was assumed.

The comparison of Fig.13 shows the following:(i) the PW-LDA calculation underestimates the band gap since the intrinsic LDA band gap error was not corrected. (ii) The small basis LCAO-LDA results appear to be inaccurate. Delley and Steigmeier have subsequently improved their basis set by adding d polarization functions basis [88] and applied a 0.6 eV rigid upward shift to the calculated results to compensate the LDA band gap error. Their new results agree perfectly with our results. (iii) the two TB models differ essentially by a constant shift. Our results agree closely with the TNN-TB, indicating that longer than nearest neighbor interactions and basis set overlap effects are important.

Comparing our valence band shift $\Delta E_v(d)[= E_{VBM}(bulk) - E_{VBM}(d)]$ shown in Table 3 with the valence band shift obtained in the NN-TB model reported in Ref.[84], we find that they differ less than 10%. However the conduction band shift $\Delta E_c(d)[= E_{CBM}(d) - E_{CBM}(bulk)]$ of the NN-TB model reported in Ref.[84] is only one third of our conduction band shift shown in Table 3. This indicates that the band gap error of the NN-TB

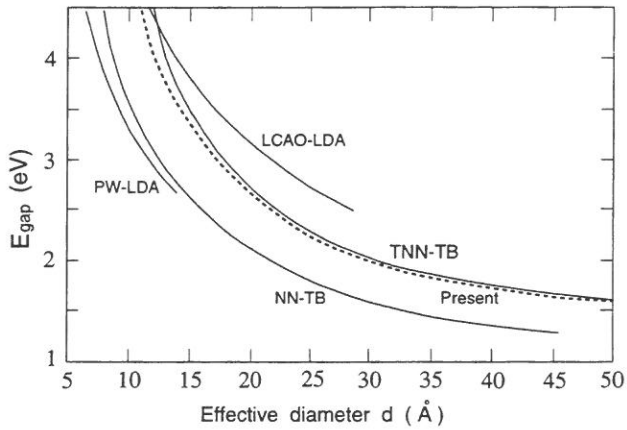

Figure 13: Comparison of calculated CBM-VBM band gaps *vs* size as obtained with different direct calculation methods. The curve representing the current results is the fitted curve of Eq(30). The other results are NN-TB (nearest neighbor tight-bindin) from Ref.[84], TNN-TB (third nearest neighbor nonorthogonal basis tight-binding) from Ref.[85], LCAO-LDA from Ref.[86] and PW-LDA (plane wave LDA) from Ref.[87]. See text for detail discussions.

model comes from the conduction band shift. This is consistent with the fact that simple (without d state) tight-binding models like the NN-TB can not describe the conduction band accurately. It is significant that tight-binding methods using but a few neighbor matrix elements (e.g NN-TB in Fig.13) *underestimate* the calculated band gap relative to better converged TB calculations (e.g, TNN-TB). In this respect we note that the better agreement of the TB calculation of Hill and Whaley [89] with the very low experimental band gap (Fig.19 below) could represent a fortuitous effect of an underconverged TB representation.

6.4. Density of states and optical absorption spectra

The total and surface local density of states of the spherical quantum dots calculated by the generalized moment method are shown in Fig.14, while Fig.15 shows the corresponding optical absorption spectrum. The DOS and optical absorption spectrum of the largest spherical dot $Si_{1315}H_{460}$ already resemble some of the features of their bulk counter parts (Fig.3). On the other hand, the smaller systems show molecular features (more peaks). The calculated *surface* local DOS in Fig.14 show peaks caused by Si-H bonding. As is the case in H covered flat Si surfaces, these peaks are around $E_v - 5\ eV$. The vertical arrows in Fig.14 indicate the VBM and CBM positions calculated by FSM, which are consistent with the band edge positions in the DOS calculated by GMM. The vertical arrows in Fig.15 indicate the band gap energy of the system calculated by FSM. Because the pseudo-direct nature of the band gap, the absorption spectra are very small around the band gap energy. The position of the main peak in the optical absorption spectra shows

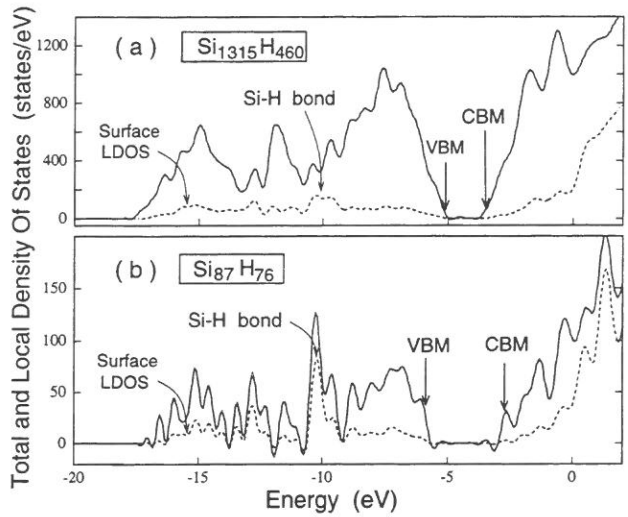

Figure 14: Calculated total and local density of states of spherical Si quantum dots. The vertical arrows denote band edge state positions calculated by FSM. The zero of the energy is the vacuum level. (a) $Si_{1315}H_{460}$, (b) $Si_{429}H_{228}$, (c) $Si_{87}H_{76}$.

a blue shift as the size is reduced. In Fig.15(a), comparing ϵ_2 with the joint density of states (JDOS) shows that the dipole matrix element $< \psi_f|\hat{\mathbf{p}}|\psi_i > /(E_f - E_i)$ controls the shape of the absorption spectra. A constant matrix element approximation (i.e, equating ϵ_2 with the JDOS) is obviously useless here.

An analogous calculation of the density of states has been performed for quantum wires and films (Fig.16). The density of states of films [Fig.16(b)] is rather similar to that of the bulk [Fig.16(a)] while wires [Fig.16(c)] show sharp features which evolve into molecular-like states in q quantum dot [Fig.16(d)]. The electron affinity (distance frin vacuum level 0 to the CBM) decreases in the series film→wire→dot.

6.5. Dielectric constant and exciton screening

The integral of the optical absorption spectra in Fig.15 gives the static dielectric constant ϵ_s [Eq(29)]. Figure 17 shows ϵ_s *vs* the quantum dot diameters. This dielectric constant measures the total polarization respond $\vec{\mathbf{P}}$ of a quantum dot to a constant total electric field $\vec{\mathbf{F}}$: $\epsilon_s \equiv 1 + \vec{\mathbf{P}}/\vec{\mathbf{F}}\Omega$, where Ω is the volume of a quantum dot. While this is useful for the calculation of the macroscopic dielectric constant of materials consisted of these quantum dots (e.g, porous Si), we are more interested in the dielectric constant $\tilde{\epsilon}_s$ which govers the screening of the exciton in a quantum dot. To model that screening, we use a *uncorrelated* electron-hole pair. This uncorrelated pair is an accurate description of the exciton for systems of strong confinement[1], i.e, when the quantum dot dimension is smaller than the free Coulomb exciton Bohr radius. Using this uncorrelated pair, the

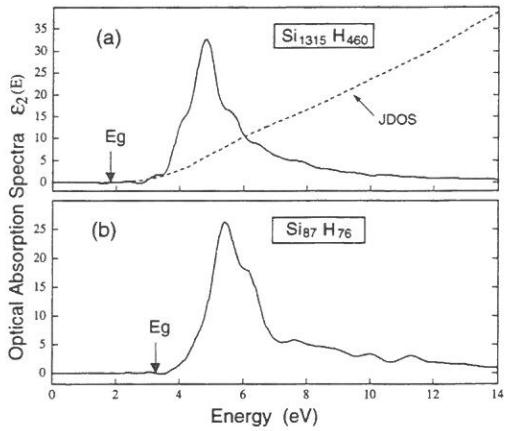

Figure 15: Calculated optical absorption spectra $\epsilon_2(E)$ of spherical Si quantum dots. (a) $Si_{1315}H_{460}$, (b) $Si_{429}H_{228}$, (b) $Si_{87}H_{76}$. The joint density of states (JDOS) in (a) is given in arbitrary units. The vertical arrows denote the band gap values calculated by FSM.

electron (or hole) charge density is:

$$\rho(r) = [(\frac{\pi}{2R^3})^{\frac{1}{2}} \frac{sin(\frac{\pi}{R}r)}{\frac{\pi}{R}r}]^2 \qquad for\ r \leq R, \tag{32}$$

and zero elsewhere, where R is the radius of the spherical quantum dot. Using this charge density, one can get its Coulomb potential (unscreened, external) $v_{ext}(r)$ by solving the Possion's equation $\nabla^2 v_{ext}(r) = 4\pi\rho(r)$ with an boundary condition $v_{ext}(R) = 0$. This external potential will induce a responding potential $v_{ind}(r)$ from the quantum dot, hence results in a final screened potential $v_{scr}(r) = v_{ext}(r) + v_{ind}(r)$. The screening dielectric constant $\tilde{\epsilon}_s$ can be thus defined as the ratio of the electric static energies between the bare external potential and the screened potential:

$$\tilde{\epsilon}_s = \frac{\int v_{ext}(r)\rho(r)d^3r}{\int v_{scr}(r)\rho(r)d^3r}. \tag{33}$$

To simplify the following calculations, we will like to change the shape of $v_{ext}(r)$ to a new shape $v'_{ext}(r)$ so that its resulting screened potential $v'_{scr}(r)$ has the shape of $v_{ext}(r)$. In other words, we can still use Eq(33), but instead of getting $v_{ext}(r)$ by the Possion's equation from $\rho(r)$, we will get $v_{scr}(r)$ by the Possion's equation from $\rho(r)$ [i.e, $\nabla^2 v_{scr}(r) = 4\pi\rho(r)$]. Substituting $v_{ext}(r) = v_{src}(r) - v_{ind}(r)$ in Eq(33), we have:

$$\tilde{\epsilon}_s = 1 - \frac{\int v_{ind}(r)\rho(r)d^3r}{\int v_{scr}(r)\rho(r)d^3r}. \tag{34}$$

Using perturbation theory, the total potential $v_{src}(r)$ will cause a change in quantum dot's charge density:

$$\Delta\rho_{dot}(r) = -\sum_{f\in unocc,\ i\in occ} \frac{4 < f|v_{src}(r)|i >}{E_f - E_i}\ \psi_f(r)\psi_i(r) \tag{35}$$

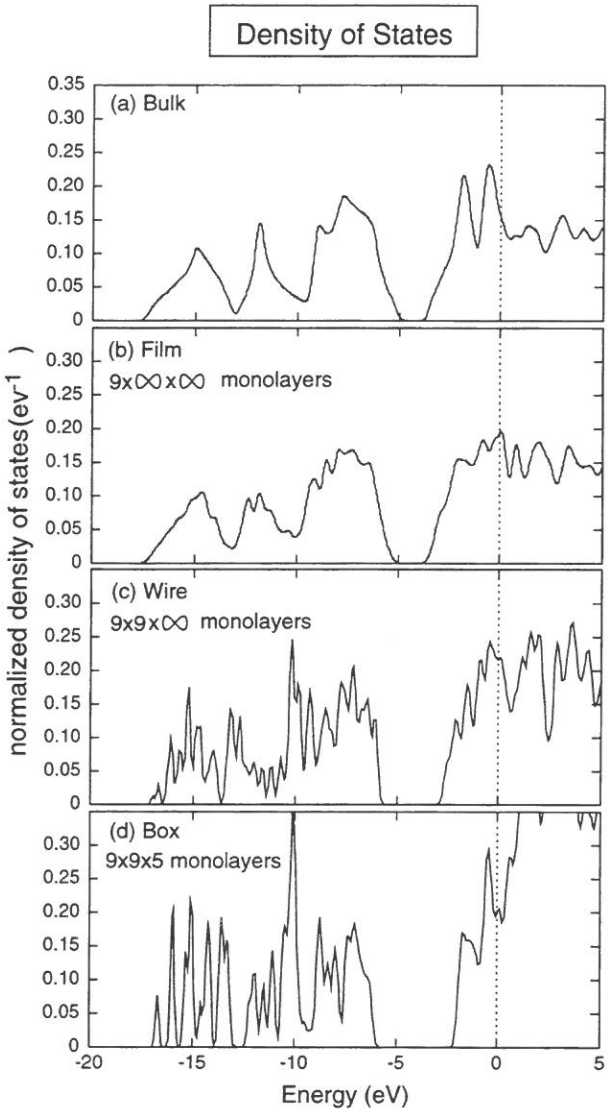

Figure 16: Density of states of H-saturated films, wires and boxes with surface orientations denoted in the caption of Fig.11. They are normalized so that the integral of the valence electrons equals 1. Gaussian broadening is 0.2 eV. This figure is taken from Ref.[80].

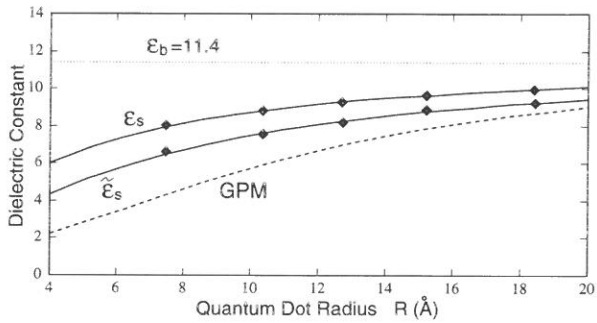

Figure 17: Dielectric constants as a function of spherical quantum dot radius R. Here ϵ_s is for total polarization [Eq(29)] and $\tilde{\epsilon}_s$ is for exciton screening [Eq(37)]. The diamond symbols are the calculated results while the solid lines are the curves fitted to Eq(39). The dashed curve corresponds to the generalized Penn model (GPM) given by Ref.[91]

Then $v_{ind}(r)$ is the potential produced by $\Delta\rho_{dot}(r)$ [i.e, $\nabla^2 v_{ind}(r) = 4\pi\Delta\rho_{dot}(r)$]. Because $v_{scr}(r)$ is the potential produced by $\rho(r)$, we have the identity:

$$\int v_{ind}(r)\rho(r)d^3r = \int \Delta\rho_{dot}(r)v_{scr}(r)d^3r. \tag{36}$$

Substituting Eq(35) and Eq(36) into Eq(34), we have:

$$\tilde{\epsilon}_s = 1 + \frac{2}{\pi}\int_0^\infty \frac{\tilde{\epsilon}_2(E)}{E}dE, \tag{37}$$

where

$$\tilde{\epsilon}_2(E) = \frac{2\pi}{\beta}\sum_f^{unocc}\sum_i^{occ} < f|v_{scr}(r)|i >^2 \delta(E - E_f + E_i) \tag{38}$$

where $\beta = \int v_{scr}(r)\rho(r)d^3r$. We have written the expresion for $\tilde{\epsilon}_s$ and $\tilde{\epsilon}_2(E)$ in the same way as for ϵ_s in Eq(29) and $\epsilon_2(E)$ in Eq(23), so that we can use the generalized moment method to calculate $\tilde{\epsilon}_2(E)$, just by replacing the operator $\hat{\mathbf{p}}$ by $v_{src}(r)$ in Eq(26).

Using this method, we show in Fig.17 the resulting $\tilde{\epsilon}_s(R)$ along with $\epsilon_s(R)$. $\tilde{\epsilon}_s(R)$ is smaller than $\epsilon_s(R)$ as expected from the q dependence of the bulk $\epsilon(q)$[90]. Also plotted in Fig.17 is the dielectric constant predicted from a generalized Penn model[91] (GPM). This model predicts a value much smaller than $\epsilon_s(R)$ and $\tilde{\epsilon}_s(R)$. The result of the GPM can be expressed as

$$\epsilon_s(R) = 1 + \frac{\epsilon_b - 1}{1 + (\alpha/R)^l} \tag{39}$$

with $l = 2$ and $\alpha = 10.93\mathring{A}$, where $\epsilon_b = 11.4$ is the Si bulk dielectric constant. Fitting our directly calculated results to the same form, we find $l = 1.25$, $\alpha = 4.25\mathring{A}$ for our total polarization dielectric constant ϵ_s and $l = 1.37$, $\alpha = 6.9\mathring{A}$ for our screening dielectric constant $\tilde{\epsilon}_s$. [Note, however, that Eq(39) with $l < 2$ can be used only for finite clusters

since $l = 2$ is the rigorously correct scaling for $R \to \infty$.] The difference between our result and the GPM result is highlighted in Fig.18 giving the ratio of free exciton radius a_{eh} to the quantum dot dimension. Using the generalized Penn model, this ratio is very close to one, thus the system is predicted to experience "moderate confinement". On the other hand, the ratio obtained in our microscopic calculation is much larger than one, indicating strong confinements. This strong confinement justifies our use of the uncorrelated electron-hole pair [Eq(32)] for the model of exciton. It also substantiates the use of Brus's[92] simple formula $-1.786/\tilde{\epsilon}_s(R)R$ for the exciton Coulomb binding energy. Recently Lannoo and Allen [93] calculated ϵ *vs* size using a tight-binding based *self-consistent* linear screening method, finding even slightly larger supression of the dielectric constant with size than found here.

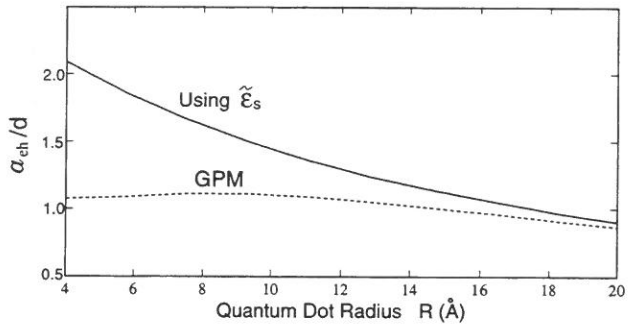

Figure 18: The ratio between the free hydrogenic exciton radius a_{eh} and the quantum dot diameter (d=2R) as a function of quantum dot radius R. The $\tilde{\epsilon}_s$ and GPM values are given in Fig.17.

6.6. Comparison of calculated exciton energy with experiment

Having calculated the intrinsic band gap $E_{gap}(R)$ and the screening dielectric constant $\tilde{\epsilon}_s(R)$ for the Coulomb energy, we are now in a position to compare our exciton energy to experimental data.

We will compare to experiments which measure the band gap of nearly spherical Si quantum dots by photoluminescence (PL) spectra[28, 34, 35, 40]. The size distribution is measured by high-pressure-liquid chromatography (HPLC), transmission electron microscopic (TEM) and X-ray peak width. The results are summarized in Fig.19. There, we use a solid symbol to represent each sample and a crossing horizontal line to indicates the width of the size distribution for that sample. For the data from Ref.[28], we put the symbol at the X-ray position, and mostly ignore the HPLC values because it tends to overestimate the size due to the aggregation of monomers. We also plotted the colloid data from Schuppler, etal[40] using the same measurement methods as in Ref.[28] (the data points from porous Si are not plotted). As can be seen, there is a very wide size distribution in these particles. Although our calculated band gap for the three prototype shapes follows the same curve, to simplify matters, we show in Fig.19 only the results

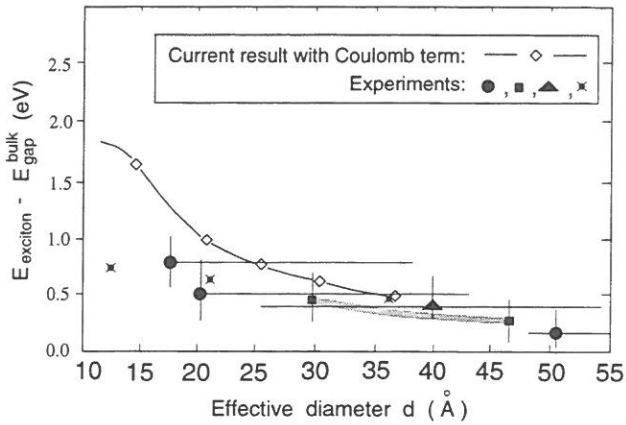

Figure 19: Calculated excitonic energy [Eq(40)] compared with experimental PL data for spherical quantum dots. The symbols ◯, △, □ and × denote PL data from Ref.[28],[34],[35] and [40] respectively. The vertical lines represent the widths at half-maximum of the PL spectra. The horizontal lines denote the size distributions. The size distribution for ◯ is estimated mostly from TEM and X-ray data in Ref.[28]. The experiment of Ref.[34] does not report the size distribution. The shaded area represents a range of experimental points of Ref.[34]. The three × points are the "very small", "small" and "intermediate" colloid data reported in Ref.[40] (the data points of porous Si in Ref.[40] are not used here). The solid line connecting ◇ represents our calculation. Note the quantum dot $Si_{29}H_{36}$ listed in Table 3 is too small to be reliable, thus is not used here.

for the spherical quantum dots. In order to compare with the PL data, the Coulomb interaction energy between the excited electron and the hole is added to the calculated intrinsic band gap E_{gap}. This gives the exciton energy for a sphere as[92] (in atomic unit, Hartree for energy, Bohr radius for R):

$$E_x(R) = E_{gap}(R) - \frac{1.786}{\tilde{\epsilon}_s(R)R} - 0.248E_{Ry}. \tag{40}$$

The second term in Eq(40) is the Coulomb energy (shown in Table 3), while the third term is a correlation energy correction with $E_{Ry} = 8.18\ meV$. The resulting calculated curve is shown in Fig.19 with the experimental value. Unfortunately, the comparison is inconclusive due to the large experimental size uncertainty. However, *for small size quantum dots, it appears that our calculated exciton energy is consistently larger than the experimental photoluminescence energy.* Calculations using but single bulk band coupling (RKF in Ref.[78]) or a variationally restricted TB method [Ref.[89]] produce lower gaps that are in better agreement with experiment. We have seen however, in Sec.6.3 (Figs.9 and 13) that improvement in the theoretical formulation is likely to push the band gaps to higher energies, thus removing the agreement with experiment. What then is the reason for the systematic *overestimation* of the silicon PL energy in Fig.19 ? Before attempting an answer we decided as a "sanity test" to compute the excitonic gap *vs* size[94] for a

system where the *experimental data* is much more accurate – CdSe dots. In the case of CdSe[95], the size can be control within 5% and the exciton energy is measured from optical absorption peak, thus avoids the possible complication of surface states. We use the same methodology, the same type of carefully fitted empirical pseudopotential [65]. The calculated gap *vs* size is shown in Fig.20 where it is compared with the best experimental data [95]. We see that our methodology produces excellent agreement with experiment for CdSe. We conjecture therefor that in Si dots the photoluminescence could come from some persistent (size-independent) *defects or impurities* states (e.g, surface states) [96-99], rather than from intrinsic dot states.

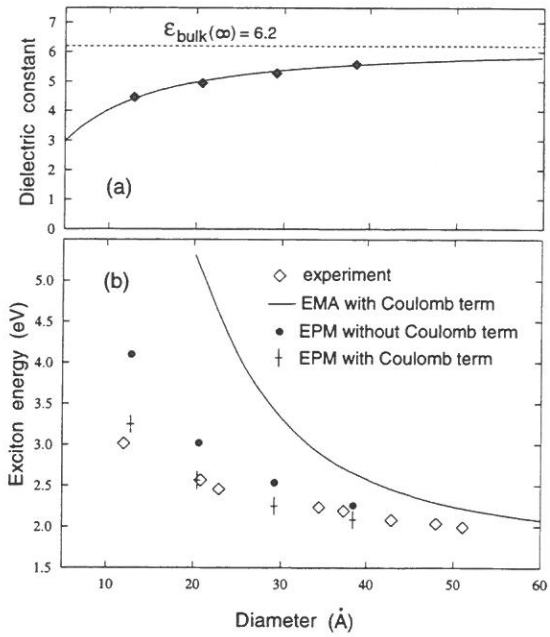

Figure 20: Quantum dot dielectric constant (a) and exciton energies (b) for CdSe quantum dot. The solid line in (a) is the fitted result using Eq(39) with $\alpha = 7.5\AA$ and $l = 1.2$. The experimental data and the effective mass (EMA) curve in (b) are both from Ref.[95]. The screening for the Coulomb potential used in part b also includes ionic contributions in this polar material [94]. Notice the good agreement between the final calculated result (cross) and the experimental result (diamond).

A different type of experiment[100] is, however, uncomplicated by the uncertainty in size distribution and possible surface defect states. In this experiment, the size-induced shift ΔE_v in the top of valence band and the shift ΔE_c in the bottom of conduction band are measured from absorption spectra for porous silicon of different quantum dot sizes:

$$\Delta E_v(d) = E_{VBM}(bulk) - E_{VBM}(d)$$

$$\Delta E_c(d) = -E_{CBM}(bulk) + E_{CBM}(d). \tag{41}$$

Instead of focusing on the highly uncertain size dependence, the ratio ΔE_v *vs* ΔE_c plots are given, thus obviating the need of measuring the size. The calculated band gap edge shifts (ΔE_v, ΔE_c) of the three prototype quantum dots are listed in Table 3. Table 3 shows that as the dot size decreases, $E_v(d)$ moves to deeper binding energies at a similar rate as $E_c(d)$ moves to shallower binding energies. In the experiment, one measures $\Delta E_v(d)$ *vs* $\Delta \bar{E}_c(d) = \Delta E_c(d) - 3.572/d\tilde{\epsilon}_s(d)$. The reason is given in Ref.[101]. $\Delta \bar{E}_c(R)$ and $\Delta E_v(R)$ are plotted in Fig.21. As shown in Fig.21, if the bulk dielectric constant $\epsilon_b = 11.4$ is used instead of $\tilde{\epsilon}_s(R)$, the result deviates considerably from the experimental data. If the $\Delta E_c(d)$ of the NN-TB model (Fig.13) reported in Ref.[84] is used, one finds that the amplitude of the Coulomb energy term $-1.786/\tilde{\epsilon}_s(R)R$ is always larger than the intrinsic $\Delta E_c(d)$. As a result, the corresponding points in Fig.21 will be below zero of the y axis.

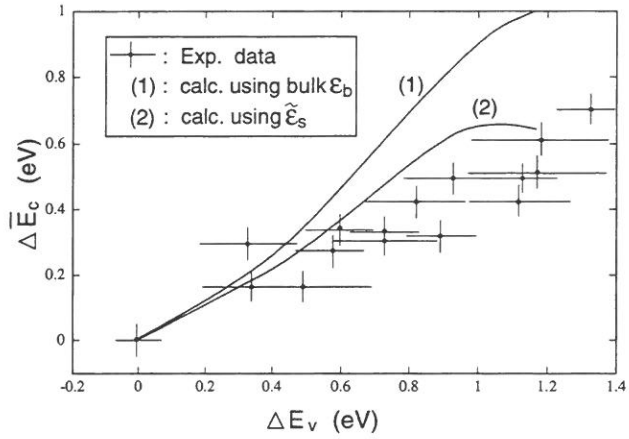

Figure 21: Conduction band edge shifts versus valence band edge shifts. The experimental data is from Ref.[100]. The experimental conduction band shift $\Delta \bar{E}_c = \Delta E_c(d) - 3.572/d\tilde{\epsilon}_s(d)$, where $\Delta E_c(d)$ and the Coulomb energy $-3.572/d\tilde{\epsilon}_s(d)$ are given by Eq(41) and Table 3. See text for detail discussions.

6.7. Radiative recombination rate *vs* exciton energy

Figure 22 depicts the calculated radiative recombination rate versus the exciton energy of Eq(40) for our three prototype-shaped quantum dots. The results are compared with the experimentally measured lifetime of Xie, etal and Vial, etal taken on samples of porous silicon at room temperature[102, 103]. The recombination rate is defined as $1/\tau$, here τ is the radiative lifetime, calculated from:

$$\frac{1}{\tau} = \frac{4}{3}\frac{\alpha\omega n}{m_e^2 c^2} | < \psi_i | \hat{\mathbf{p}} | \psi_f > |^2, \tag{42}$$

where, n(=2.6) is the effective refractive index of Si quantum dot[104], ω is the photon angular frequency, $\alpha = e^2/\hbar c$. The value of τ calculated from Eq(42) strictly between

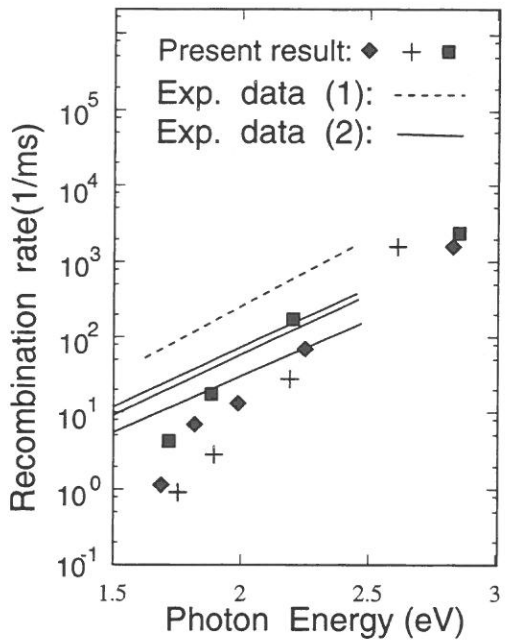

Figure 22: Radiative recombination rate $1/\tau$ [Eq(42)] as a function of the luminescence photon energy [exciton energy, Eq(40)]. Experimental curves (1) and (2) are from Ref.[102] and [103], respectively. They are both for porous Si measured at room temperature. The three curves of (2) are taken from three different samples with different level of oxidizations. The symbols $\diamond$, + and $\square$ represent calculated results for spherical, $(110) \times (1\bar{1}0) \times (001)$ rectangular, and $(100) \times (010) \times (001)$ cubic quantum dots and are for zero-phonon process [Eq(42)] only. Phonon assisted process (Ref.[105]) could shift the positions of the symbols (especially those in the lower energy region) upward by a factor of 3-10. The quantum dot $Si_{29}H_{36}$ listed in Table 3 is too small to be included here.

the VBM (for ψ_f) and CBM (for ψ_i) states fluctuates widely even with small changes in quantum dot's size. To reduce this fluctuation, we have taken an average of $| < \psi_i|\hat{\mathbf{p}}|\psi_f > |^2$ over the four highest occupied states $\{\psi_f\}$ and four lowest unoccupied states $\{\psi_i\}$. The energy spread for these four eigenstates is about 20 meV for the largest quantum dots studied here, thus of the order of kT at room temperature. Figure 23 shows that, unlike the E_g *vs* d curves (Fig.9), which collapse into a single "unified" curve for all prototype quantum dots, the $1/\tau$ *vs* d curve shows more scattering. The difference between the experimental data and the result of Eq(42) is of the same order of magnitude as the difference between the two sets of the experimental data. Yet, our calculated lifetime are systematically *too long* in the small photon energy region. In using Eq(42), we only considered zero-phonon process. As recently pointed out by Hybertsen[105] , phonon-assisted process can be 3-10 times stronger than zero-phonon process [especially for large (40 $\AA$) quantum dots, i.e, in the small photon energy region]. Taking that into account,

Table 4: The band gaps and radiative lifetimes of two cubic quantum dots with different surface orientations. The band gap difference after correcting for the small difference in size is 0.014 eV.

system	orientation	N_{Si}	N_H	size d($\AA$)	$E_{gap}(eV)$	lifetime $\tau(\mu s)$
1	$(100)\times(010)\times(001)$	1101	532	34.783	1.8173	231.2
2	$(110)\times(1\bar{1}0)\times(001)$	1157	492	35.363	1.8237	8647.

our calculated recombination rate in Fig.22 should be moved upward by a factor of 3-10. This will put them right around the experimental data (1) and (2) in Fig.22. This is consistent with a simple picture that the PL of the porous Si comes from the bulk like interior of isolated quantum dots which exist in the porous silicon. However, one must be cautious to explain the experimentally measured PL lifetime of porous Si. As conjectured in Sec.6.6, defect states (e.g, surface states) might play an important role in the PL process. Besides the radiative recombination channel, the nonradiative channel and possible carrier transport process also affect the measured PL lifetime[103]. One must also keep in mind the temperature dependence of the measured PL lifetime in porous Si[106, 103]. So far, a good model of the PL process in porous Si does not exist.

6.8. Surface orientation dependence of the band gap

We have next studied the effect of the surface orientation of the quantum dot on its band gap and recombination rate. To eliminate other effects, we have chosen two quantum dots with the same shape(cubic) and almost the same sizes (differing by 1.7%). One quantum dot has (100),(010) and (001) surfaces and the other has (110),(1$\bar{1}$0) and (001) surfaces. Thus, the latter structure represents a 45^o rotation of the first structure around one of its principle axis. The calculated band gaps and radiative lifetimes are given in Table 4. After correcting the small size difference using the unified curve of Eq(30), we find that the band gap difference for these two orientations is only 0.014 eV. This is only 2% of the band gap blue shift(ΔE_g) and is thus negligible. However, the recombination rate of the [(110),(1$\bar{1}$0),(001)] oriented quantum dot is 40 times smaller than that of the [(100),(010),(001)] oriented quantum dot. We can conclude that the band gap energy has negligibly small dependence on orientation, but the radiative lifetime(recombination rate) is more sensitive to it.

6.9. Shape dependence at constant size

Quantum confinement effects can exist in one dimension(film), two dimension(wire) and three dimension(particle) systems. In Fig.9, we showed that if the effective size is measured as $d \propto N_{Si}^{1/3}$, the band gap *vs* size(or N_{Si}) curves are similar for three prototype shapes, for which the structures are not too prolate. It would be interesting to study the band gap change when a quantum dot goes through an extreme shape change, e.g, from a film-like object to a wire-like object. We examined this by changing the aspect ratio $d_z/d_x = d_z/d_y$ of a rectangular box(see inserts to Fig.23): When the ratio $d_z/d_x \ll 1$,

the quantum dot is film-like, when $d_z/d_x = 1$, the quantum dot is cubic, while when $d_z/d_x \gg 1$, the quantum dot is wire-like. To eliminate the effects of orientation, we studied boxes with fixed surface orientations [(100),(010),(001)] for all d_z/d_x ratios. To eliminate the effect of size N_{Si}, we studied quantum dots having almost the same number $N_{Si} = 1108 \pm 13$ of Si atoms. Figure 23 depicts the recombination rate(part a) and band gap energy(part b) versus the ratio d_z/d_x. Note that different shapes at $N_{Si} \cong const$ can have gaps that differ by as much as 0.8 eV ! The structure with the smallest band gap(i.e, weakest quantum confinement effect) occurs when $d_z/d_x = 1$(cubic). From $d_z/d_x = 4$ to $d_z/d_x = 14$, there is a switch between the near VBM states, as a result, the trend of the recombination rate has been changed. While the most elongated 1100-atom bar gives the largest blue shift, interestingly the cube has a *faster* decay (i.e, the transition is "more allowed") than this most elongated bar. Thus a larger blue shift comes, unfortunately, with a weaker transition.

It is interesting to compare the band gap and radiative lifetime of rectangular boxs(finite d_z) to those of infinitely long($d_z = \infty$) quantum wires. In both cases, we use structures with cross section [(110),(1$\bar{1}$0)]. The results of our rectangular box were given in Fig.9 and Fig.22. The results of the quantum wire are from Ref.[107] (the VBM and CBM for these quantum wires have $k_z = 0$). The comparison in Fig.24 shows that: (i) Given the same [(110),(1$\bar{1}$0)] cross sections for the wire and rectangular box, the box has a band gap 0.2-0.3 eV higher than that of the infinite wire (in the cross section size range of $10-20\mathring{A}$). (This is different from Fig.23, where the total number of Si, not the cross section size are the same for the box and wire). (ii) The radiative lifetime of the box is about twice the radiative lifetime of wire. The fact that the gap of a dots is larger than that of the infinite wire is consistent with the larger quantum confinement for finite d_z(i.e. box). The slightly shorter lifetime of the wire is probably related to the fact that both the CBM and VBM of the wire has $k_z = 0$. As a result, the overlap between the wire CBM and VBM in z direction is large in the matrix element ($< i|p|j >^2$) calculation. As discussed by Yeh, etal[107], it is reasonable to assume that porous Si consists of both quantum wires and quantum dots. Thus, the quantitative results presented here for different shapes of the quantum dots should be useful in detail analysis of the experimental data.

7. CONCLUSIONS

We have introduced a new approach for electronic structure calculations of nanoscale quantum systems. A plane wave basis is used to describe the wavefunctions and accurate empirical pseudopotentials are used to approximate the Hamiltonian. The empirical pseudopotentials (of continuous forms in reciprocal space) are fitted to the bulk band structure and surface density of states. The shapes of first principle LDA screened potentials are also taken into account in the empirical pseudopotential fitting. As a result, our fitted potential $V(\mathbf{r})$ is very close to LDA screened potential (Fig.5), except that our potential gives the correct bulk band gap and band structure, while the LDA potential does not. Our central approximation is the use of a *fixed*, non self consistent potential for all dots. Two newly developed methods, the folded spectrum method and the generalized moments method, are used to calculate the band edge states, density of states and optical absorption spectra of thousand atom systems. These three quantities are sufficient to determine

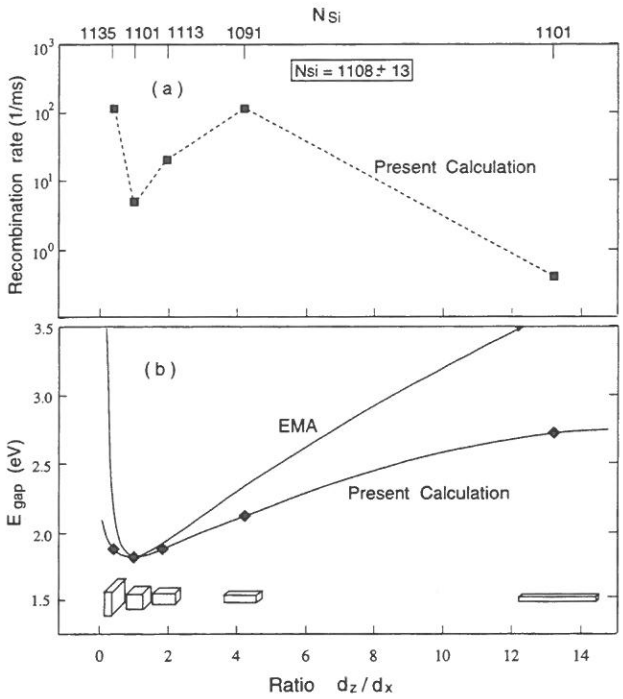

Figure 23: Dependence of the band gap and recombination rate $(1/\tau)$ on the shape $(d_z/d_x = d_z/d_y)$ of Si quantum boxes. The box has the $(100) \times (010) \times (001)$ orientation. The ratio d_z/d_x changes from 0.38 to 13.2 and the box changes from filmlike to wirelike. (a) The radiative recombination rate $(1/\tau)$ *vs* ratio d_z/d_x. (b) The band gap *vs* the ratio d_z/d_x. Fitting the $d_z/d_x = 1$ point to an EMA formula shows that when d_z/d_x differs from 1, the effective mass formula overestimates the quantum confinement effects. This is consistents with the results of Fig.9 which indicates that the smaller the length d_x or d_z, the larger the error of the EMA.

most optical characteristics of the system. The whole procedure is designed as a more reliable and realistic alternative to the tight-binding method. The current approach is illustrated for Si quantum dots with surface passivation of H atoms. The interiors of such systems still maintain a bulk-like structure. The surface atomic configuration is modeled using the experimental data and first principle calculations on three primary H covered Si surfaces.

We have calculated the following properties of Si quantum dots: (1) band gap *vs* size and shape of the quantum dots; (2) photoluminescence lifetime *vs* size and shape of the quantum dots; (3) total electronic density of state and optical absorption spectra of the quantum dots; (4) dielectric constant *vs* size of the quantum dots. The band-edge wavefunctions are analyzed in terms ot bulk Bloch wave functions and found to be bulk like. The results are compared to tight-binding like direct calculations and model

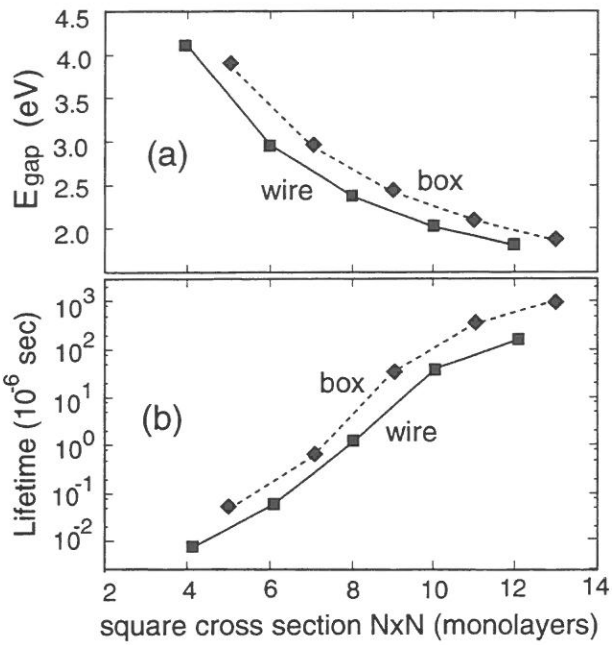

Figure 24: Comparison of the band gap and radiative lifetime (τ) of finite rectangular quantum boxes of length $d_z = \sqrt{2}d_x = \sqrt{2}d_y$ and infinitely long ($d_z = \infty$) quantum wires. The cross sections perpendicular to the z direction are the same for rectangular box and the wire, which is [(100),(1$\bar{1}$0)] with N Si monolayer on each sides. The distance between two Si monolayers is 1.92 $\mathring{A}$. The quantum wire results are taken from Ref.[107] with adjustments so that same way of calculating τ is used as described in text. (The lifetimes of quantum wire shown in Fig.4 of Ref.[107] and in Fig.8 of Ref.[75] are in error: They should be divided by a factor of 2).

calculations. We find that next nearest neighbor interaction and nonorthogonal basis are necessary for tight-binding method to get accurate results, and the model calculations are not correct quantitatively. Good agreement between our results and experiments are found for the photoluminescence lifetime and the ratio between conduction band shift and valence band shift.

All these demonstrate the usefulness of the current approach in describing quantum nanoscale electronic structure. Recently, this approach has also been used in the study of $\sim$ 300 nm disorder superlattices[108] and CdSe quantum dots [94].

ACKNOWLEDGEMENTS

We would like to thank S.B. Zhang and C.Y. Yeh for many helpful discussions. This work was supported by the office of Energy Research, Materials Science Division, U.S. Department of Energy, under Grant DE-AC02-83CH10093.

References

[1] A.D. Yoffe, Adv. Phys. **42**, 173 (1993).

[2] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures*, (Les editions de physique, Less Ulis) 1988.

[3] A. Sasaki, J. Crystal Growth. **115**, 490 (1991).

[4] M.H. Bode and A. Ourmazd, J. Vac. Sci. Technol. B **10**, 1787 (1992).

[5] A. Zunger and S. Mahajan, *Atomic Ordering and Phase Separation in Epitaxial III-V Alloys*, Handbook on Semiconductors, Vol. 3, 2nd ed., (Elsevier, Amsterdam, 1994).

[6] R.Merlin, K. Bajemu and R. Clarke, Phys. Rev. Lett. **55**, 1768 (1985).

[7] G. Benedek, A. Cavallini and W. Schroter, *Point and Extended Defects in Semiconductors*, (Plenum Press, New York, 1989).

[8] M.H. Brodsky, *Amorphous Semiconductors*, (Springer-Verlag, Berlin, 1979).

[9] M.D. Pashley, K.W. Haberern and J.M. Gaines, Appl. Phys. Lett. **58**, 406 (1991); S.B. Zhang and A. Zunger (unpublished).

[10] S. Froyen and A. Zunger, Phys. Rev. Lett. **66**, 2132 (1991).

[11] K.D. Brommer, M. Needels, B.E. Larson and J.D. Joannopoulos, Phys. Rev. Lett. **68**, 1355 (1992).

[12] D.M. Wood and A. Zunger, J. Phys. A **18**, 1343 (1985).

[13] M.P. Teter, M.C. Payne and D.C. Allan, Phys. Rev. B **40**, 12255 (1989).

[14] P. Bendt and A. Zunger, Solar Energy Research Institute Tech. Rep., TP-212-1698, 1982; Phys. Rev. B **26**, 3114 (1982) describe the Jacobian update method.

[15] I. Stich, M.C. Payne, R.D. King-Smith, J.S. Lin and L.J. Clarke, Phys. Rev. Lett. **68**, 1351 (1992).

[16] L.W. Wang and A. Zunger, Comp. Mat. Sci. **2**, 326 (1994).

[17] L.W. Wang, Ph.D thesis, (Cornell Univ) 1991.

[18] W. Yang, Phys. Rev. Lett. **66**, 1438 (1991).

[19] S. Baroni and P. Giannozzi, Europhys. Lett. **17**, 547 (1992).

[20] G. Galli and M. Parrinello, Phys. Rev. Lett. **69**, 3547 (1992); F. Mauri, G. Galli and R. Car, Phys. Rev. B **47**, 9973 (1993).

[21] X.P. Li, R.W. Nunes and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).

[22] D.A. Drabold, O.F. Sankey, Phys. Rev. Lett. **70**, 3631 (1993).

[23] G.A. Baraff, M. Schluter and G. Allan, Phys. Rev. B **27**, 1010 (1983).

[24] W.A. Harrison, *Electronic Structure and the Properties of Solids*, (Freeman, San Francisco, 1980).

[25] S.B. Zhang and A. Zunger, Phys. Rev. B **48**, 11204(1993); D.M. Wood and A. Zunger (unpublished).

[26] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias and J.D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[27] J.L. Martins and M.L. Cohen, Phys. Rev. B **37**, 6134 (1988).

[28] K.A. Littau, P.J. Szajowski, A.J. Muller, A.R. Kortan and L.E. Brus, J. Phys. Chem. **97**, 1224 (1993).

[29] R.G. Dandrea, J.E. Bernard, S.H. Wei and A. Zunger, Phys. Rev. Lett. **64**, 36(1990); R.G. Dandrea, C.B. Dukex and A. Zunger, J. Vac. Sci. Technol. B **10**, 1744 (1992).

[30] P.N. Keating, Phys. Rev. **145**, 637 (1966); R.M. Martin, Phys. Rev. B. **1**, 4005 (1970).

[31] M.L. Cohen and J.R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors*, (Springer-Verlag, Berlin, 1988).

[32] L.W. Wang and A. Zunger, J. Chem. Phys. **100**, 2394 (1994).

[33] L.W. Wang, Phys. Rev. B, **49**, 10154 (1994).

[34] W.A. Saunder, H.A. Atwater, K.J. Vahala, R.C. Flagan and P.C. Sercel, Mat. Res. Soc. Symp. Proc. **268**, 118 (1993).

[35] H. Takagi, H. Ogawa, Y. Yamazaki, A. Ishizaki and T. Nakagiri, Appl. Phys. Lett. **56**, 2379 (1990).

[36] A. Fojtik, M. Gersing and A. Henglein, Ber. Bunsenges. Phys. Chem. **97**, 1493 (1993).

[37] X.W. Zhao, O. Schoenfeld, Y. Aoyagi and T. Sugano, J. Phys. D: Appl. Phys. **27**, 1575 (1994).

[38] T. Shimizu-Iwayama, K. Fujita, S. Nakao, K. Saitoh, T. Fujita and N Itoh, J. Appl. Phys. **75**, 7779 (1994).

[39] E.S. Snow and P.M. Campbell, Appl. Phys. Lett. **64**, 1932 (1994).

[40] S. Schuppler, S.L. Friedman, M.A. Marcus, D.L. Adler, Y.H. Xie, T.D. Harris, W.L. Brown, Y.J. Chebal, L.E. Brus and P.H. Citrin, Phys. Rev. Lett. **72**, 2648, 1994.

[41] Y.M. Weng, X.J. Yang, Y.F. Chen and X.F. Zong, Chin. Phys. Lett. **11**, 390 (1994).

[42] C.O. Rodriguez, E.L.Peltzer Y Blanca and O.M. Cappannini, Solid Stat. Com. **56**, 575 (1985).

[43] K.C. Pandey, Phys. Rev. B **14**, 1557 (1976).

[44] T. Sakurai and H.D. Hagstrum, Phys. Rev. B **12**, 5349 (1975).

[45] T. Sakurai and H.D. Hagstrum, J. Vac. Sci. Technol. **13**, 807 (1976).

[46] S. Maruno, H. Iwasaki, K. Horioka, S.T. Li and S. Nakamura, Phys. Rev. B **27**, 4110 (1983)

[47] J.J. Boland, Phys. Rev. Lett. **65**, 3325 (1990).

[48] E. Kaxiras and J.D. Joannopouplos, Phys. Rev. B **37**, 8842 (1988).

[49] J.E. Northrup, Phys. Rev. B **44**, 1419 (1991).

[50] M. Welkowsky and R. Braunstein, Phys. Rev. B **5**, 497 (1972).

[51] W.E. Spicer and R.C. Enden, In *Proceedings of the Ninth International Conference of the Physics of Semiconductors*; Edited by S.M. Ryvkin; Nauka, Moscow 1968; Vol. 1, p65.

[52] R. Hulthen and N.G. Nilsson, Solid State Commun. **18**, 1341 (1976).

[53] W. Bludau, A. Onton and W. Heinke, J. Appl. Phys. **45**, 1846 (1974).

[54] L. Ley, S.P. Kowalcyzk, R.A. Pollak and D.A. Shirley, Phys. Rev. Lett. **29**, 1088 (1972).

[55] J.C. Hensel, H. Hasegawa and M. Nakayama, Phys. Rev. **138**, A225 (1965).

[56] G. Dzesselhaus, A.F. Kip and C. Kittel, Phys. Rev. **98**, 368 (1955).

[57] F.G. Allen, J. Phys. Chem. Solids **8**, 119 (1959).

[58] J.R. Chelikowsky and M.L. Cohen, Phys. Rev. B **10**, 5095 (1974).

[59] D.R. Hamann, M. Schluter and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979); D.R. Hamann, Phys. Rev. B **40**, 2980 (1989).

[60] H.R. Philipp and H. Ehrenreich, Phys. Rev. **120**, 1550 (1963).

[61] D. Buss and N.J. Parada, Phys. Rev. B **1**, 2692 (1970).

[62] R.A. Faulkner, Phys. Rev. **184**, 713 (1969); H.W. Icenogle, B.C. Platt and W.L. Wolfe, Appl. Opt. **15**, 2348 (1976).

[63] S. Baroni and R. Resta, Phys. Rev. B **33**, 7017 (1986).

[64] J.P. Walter and M.L. Cohen, Phys. Rev. B **2**, 1821 (1970).

[65] L.W. Wang and A. Zunger, Phys. Rev. B **51**, 17398 (1995).

[66] M.S. Hybertson, S.G. Louie, Phys. Rev. B, **34**, 5390 (1986); R.W. Goodby, M. Schulter, L.J. Sham, Phys. Rev. B bf 37, 10159 (1988).

[67] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vettering, *Numerical Recipes* (Cambridge University, New York, 1989).

[68] J. Skilling, in *Maximum Entropy and Bayesian Methods*, edited by J. Skilling (Kluwer, Dordrecht, Holland, 1983).

[69] N.I. Akhiezer, *The Classical Moment Problem*, (Hafner, Publ. Co. N.Y. 1965).

[70] F. Ducastelle, F. Cyrot-Lackmann, J. Phys. Chem. Solids. **31**, 1295 (1970).

[71] I. Turek, J. Phys. C. **21**, 3251 (1988).

[72] J.C. Wheeler, C. Blumstein, Phys. Rev. B **6**, 4380 (1972); J.C. Wheeler, M.G. Prais and C Blumstein, Phys. Rev. B **10**, 2429 (1974).

[73] One might have thought that application of ∇^2 *twice* [as needed in Eq(6)] would have required specification of additional boundary conditions relative to Eq(1) or would have doubled the number of eigensolutions relative to Eq(1). If so, how can we be sure that the solutions we get from Eq(6) are the *same* as those of Eq(1) ? The fact is that we used a *fixed* basis of N functions both for solving Eq(1) and Eq(6). Thus we have N solutions for both Eq(1) and Eq(6). Because each one of these N solutions of Eq(1) is also a solution of Eq(6) [this is easy to prove by applying Eq(1) twice], thus the N solutions of Eq(1) must equal the N solutions of Eq(6). As a result, each one of the N solutions of Eq(6) must also be a solution of Eq(1). We thank Prof. Alexi Efros for bringing up this question.

[74] M.A. Reed, Sci. Am. **268**, 118 (1993).

[75] L.W. Wang and A. Zunger, J. Phys. Chem. **98**, 2158 (1994). The lifetime values of wire shown in Fig.8 should be divided by 2.

[76] L.W. Wang and A. Zunger, Phys. Rev. Lett. **73**, 1039 (1994).

[77] M. Willatzen, T. Tanaka, Y. Arakawa and J. Singh, IEEE J. Quantum Electonics, **30**, 640 (1994).

[78] M.V. Rama Krishna and R.A. Friesner, Phys. Rev. Lett. **67**, 629 (1991); J. Chem. Phys. **96**, 873 (1992).

[79] T. Takagahara and K. Takeda, Phys. Rev. B **46**, 15578 (1992).

[80] A. Zunger, C.Y. Yeh, L.W. Wang and S.B. Zhang, p.1763, *Proceedings of the 22nd International Conference on the Physics of Semiconductors*, Vancouver, Canada, edited by D.J. Lockwood, (World Scientific, Singapore, 1995).

[81] B. Delley, (private communication).

[82] V.I. Gavrilenko and F. Koch, J. Appl. Phys. **77**, 3288 (1995).

[83] A. Tomasulo and M.V. Rama Krishna, (unpublished).

[84] S.Y. Ren and J.D. Dow, Phys. Rev. B **45**, 6492 (1992).

[85] J.P. Proot, C. Delerue and G. Allan, Appl. Phys. Lett. **61**, 1948 (1992).

[86] B. Delley and E.F. Steigmeier, Phys. Rev. B **47**, 1397 (1993).

[87] M. Hirao, T. Uda and Y. Murayama, Mat. Res. Soc. Symp. Proc. **283**, 425 (1993); M. Hirao and T. Uda, Surf. Sci. *306*, 87 (1994); T. Uda and M. Hirao, J. Phys. Soc. Jpn. **63**, suppl. B 97 (1994).

[88] B. Delley and E.F. Steigmeier, (unpublished).

[89] N.A. Hill and K.B. Whaley, Phys. Rev. Lett. **75**, 1130 (1995).

[90] The $\epsilon_s(R)$ is analogy to $\epsilon_{bulk}(q = 0)$ in the bulk and $\tilde{\epsilon}_s(R)$ is analogy to $\epsilon_{bulk}(q \approx \pi/R)$ in the bulk.

[91] R. Tsu, L. Ioriatti, J.F. Harvey, H. Shen and R.A. Lux, Mat. Res. Soc. Symp. Proc. **283**, 437 (1993).

[92] L.E. Brus, J. Phys. Chem. **90**, 2555 (1986).

[93] M. Lannoo and G. Allan, Phys. Rev. Lett. **74**, 3415 (1995).

[94] L.W. Wang and A. Zunger, (unpublished).

[95] C.B. Murray, D.J. Norris and M.G. Bawendi, J. Am. Chem. Soc. **115**, 8706 (1993).

[96] F. Koch, V. Petrova-Koch, T. Muschik, A. Nikolov and V. Gavrilenko, Mat. Res. Soc. Symp. Proc. **283**, 197 (1993).

[97] Y. Kanemitsu, Phys. Rev. B **49**, 16845 (1994).

[98] H. Tamura, M. Ruckschloss, T. Wirschem and S. Veprek, Appl. Phys. Lett. **65**, 1537, (1994).

[99] T. Shimizu-Iwayama, S. Nakao, K. Saitoh and N. Itoh, J. Phys.: Condens. Matter. **6**, L601, 1994.

[100] T. van Buuren, T. Tiedje, J.R. Dahn and B.M. Way, Appl. Phys. Lett. **63**, 2911 (1993).

[101] The reason for measuring $\Delta\tilde{E}_c(d)$ is that a X-ray photon excites one electron from 2s2p core state to conduction band and the edge of this absorption spectrum is used to determine the conduction band position. Because the excited electron and hole of the core state are both inside the quantum dot, a Coulomb attraction term $-3.572/d\tilde{\epsilon}_s(d)$ between them (as shown in Table 3) should be added to the calculated intrinsic conduction band shift $\Delta E_c(d)$ of Table 3. On the other hand, the valence band shift is measured by X-ray photoemission of the valence band electron. Because a Si 2p core state photoemission spectrum of the same quantum dot is used as a reference point, any Coulomb effect will be canceled out between the reference energy and the valence band photoemission energy. Thus the measured shift is only the shift of the valence band energy caused by the quantum dot size, which is $\Delta E_v(d)$.

[102] Y.H. Xie, W.L. Wilson, H.M. Ross, J.A. Mucha, M.A. Fitzgerald, J.M. Macaulay and T.D. Harris, J. Appl. Phys. **71**, 2403 (1992).

[103] J.C. Vial, A. Bsiesy, F. Gaspard, R. Herino, M. Ligeon, F. Muller and R. Romestain, Phys. Rev. B **45**, 14171 (1992).

[104] W.B. Jackson, N.M. Johnson, Mat. Res. Soc. Symp. Proc. **46**, 545 (1985).

[105] M.S. Hybertsen, Phys. Rev. Lett. **72**, 1514 (1994).

[106] P.D.J. Calcott, K.J. Nash, L.T. Canham, M.J. Kane and D. Brumhead, J. Phys. Condens. Matter, **5**, L91 (1993).

[107] C.Y. Yeh, S.B. Zhang and A. Zunger, Appl. Phys. Lett. **63**, 3455 (1993). The lifetime values reported in Fig.4 here should be divided by 2.

[108] K.A. Mader, L.W. Wang and A. Zunger, Phys. Rev. Lett. **74**, 2555 (1995).