# Electronic-Structure Theory of Semiconductor Quantum Dots

Alex Zunger

## The "Standard Model" of the Electronic Structure of Dots

Progress made in the growth of "free-standing" (e.g., colloidal) quantum dots[1,2] (see also articles in this issue by Nozik and Mićić, and by Alivisatos) and in the growth of semiconductor-embedded ("self-assembled") dots[3,4] (see also the article by Bimberg, Grundmann, and Ledentsov in this issue) has opened the door to new and exciting spectroscopic studies of quantum structures. These have revealed rich and sometimes unexpected features such as quantum-dot *shape*-dependent transitions, size-dependent (red) shifts between absorption and emission, emission from high excited levels, surface-mediated transitions, exchange splitting, strain-induced splitting, and Coulomb-blockade transitions. These new observations have created the need for developing appropriate theoretical tools capable of analyzing the electronic structure of $10^3$–$10^6$-atom objects. The main challenge is to understand (a) the way the *one-electron levels* of the dot reflect quantum size, quantum shape, interfacial strain, and surface effects and (b) the nature of *"many-particle" interactions* such as electron-hole exchange (underlying the "red shift"), electron-hole Coulomb effects (underlying excitonic transitions), and electron-electron Coulomb (underlying Coulomb-blockade effects).

Interestingly, while the electronic-structure theory of periodic solids has been characterized since its inception by a diversity of approaches (all-electron versus pseudopotentials; Hartree Fock versus density-functional; computational schemes creating a rich "alphabetic soup," such as APW, LAPW, LMTO, KKR, OPW, LCAO, LCGO, plane waves, ASW, etc.), the theory of *quantum nanostructures* has been dominated mainly by a single approach so widely used that I refer to it as the "Standard Model": the effective-mass approximation (EMA) and its extension to the "$\mathbf{k}\cdot\mathbf{p}$"[5–7] (where $\mathbf{k}$ is the wave vector and $\mathbf{p}$ is the momemturn). In fact, speakers at nanostructure conferences often refer to it as "theory" without having to specify what is being done. The audience knows.

The essential idea[5–7] is sweeping in its simplicity: The single-particle wave functions $\psi_i(\mathbf{r})$ of three-dimensional (3D)-periodic bulk, two-dimensional (2D)-periodic film/well, one-dimensional (1D)-periodic wires, or zero-dimensional (OD)-periodic dot are expanded by a handful of 3D-periodic Bloch orbitals taken from the Brillouin zone center (I' point) of the underlying bulk solid. The physical accuracy of this representation is naturally highest for systems closest to the reference from which the basis functions are drawn ($\Gamma$ point of the 3D bulk). It decreases as one wanders away from the Brillouin zone center and as dimensionality (D) is reduced in the sequence 3D→2D→1D→ OD. For example," reproducing the energy of the $X_{1c}$ 3D bulk *state* in GaAs within 1 meV requires $N_b$ = 150 (T-like) Bloch bands. However if only $N_b$ = 10 (I-like) bands are used, the error increases to 300 meV. If $N_b$ is further reduced to eight bands, the error in $X_{1c}$ goes up to 20 eV and the curvatures (hence, effective masses) of the bulk valence bands develop an unphysical negative sign.[8] Application of direct diagonalization and "first-principles $\mathbf{k}\cdot\mathbf{p}$" to the 2D GaAs/AlAs superlattices[9] showed that the $\mathbf{k}\cdot\mathbf{p}$ errors versus $N_b$ parallel those in bulk GaAs. The severity of such errors is mitigated by the central feature of the Standard Model: Instead of computing the parameters of the Hamiltonian from the basis set used, they are fit to the observed band properties of the bulk[5–7] (bandgaps, masses) or the nanostructures themselves.[10,11]

The Standard Model has been eminently successful in describing the spectroscopy and transport of high-dimensional nanostructures such as 2D quantum-well structures[7] and laser devices.[12] Its success in describing lower dimensional nanostructures such a OD quantum dots, while very impressive (e.g., see the good agreement with the excited-state energies of the CdSe dot[10,11]), is sometimes clouded by the fact that the parameters were fitted to the data on the nanostructure itself,[10,11] and that improvements via an increase in the number of basis functions come with an unpleasant increase in the number of adjustable parameters. Consequently in designing an alternative theoretical description of the electronic structure of nanostructures, we set the following requirements:

- No adjustable parameters, except for the 3D bulk, in which "local-density-approximation (LDA) errors"[13] must be corrected.
- The accuracy of the physical description should be the same for nanostructures of all dimensions (3D, 2D, 1D, and OD). Likewise, zone-center ($\Gamma$) and off-r states should be described equivalently.
- . The atomistic symmetry of the object at hand should be preserved. This is pertinent because previous Standard-Model calculations showed that the difference between odd and even symmetries in an $(AlAs)_n(GaAs)_n$ (001) superlattice[9] or in $(Si)_n$ film[14] (as the number $n$ of monolayers changes) is absent (see Figure 1). Furthermore the distinction between the unequal symmetries of the (110) and ($\bar{1}$10) faces of InAs pyramidal dots[15,16] is also lost (Figure 2). These misrepresentations of the true, *atomistic* symmetries by the continuum approach underlying the Standard Model introduced errors in the energy levels.[9,14–16]
- The real atomistic *surface*[17] of the nanostructure should be included in the description rather than an (infinite) "potential barrier" lacking "chemical personality."
- Flexibility: The basic constructs determining the electronic structures should be incorporable in a flexible/modular manner and on equal footing. This includes the ability to incorporate different chemical species (e.g., dots made of either ionic or covalent materials), arbitrary

shapes of the nanostructure, crystal-field and spin-orbit splittings, and the response to pressure and strain. Our approach, satisfying these requirements, consists of two steps described in the following two sections.

## The New Model, Step 1: The Single-Particle Problem

Instead of inventing a specialized approach for nanostructures, we formulate the problem such that the same conceptual methods and sophistication with which bulk solids have been successfully treated in the past can be applied to nanostructures. We first surround the nanostructure, which is periodic in $n$ dimensions, by a few "monolayers of vacuum" in the remaining $3-n$ dimensions. For example a dot is surrounded by vacuum in three dimensions, while a film (periodic in n = 2 dimensions) is surrounded by vacuum in the one remaining dimension. We then place this (nanostructure + vacuum) object in a periodically repeated "supercell." Because this creates an artificial 3D periodic lattice, ordinary band-theoretic methodology can now be applied. As the thickness $d$ of the vacuum layers increases, the unphysical interaction between the periodically repeated nanostructures diminishes, and we approach the desired limit of an isolated nanostructure.

The potential felt by the electrons is represented by a superposition of atom-centered screened potentials?

$$V(r) = \sum_{atom} v_{atom}(\mathbf{r} - \mathbf{R}_{atom}), \qquad (1)$$

and the wave functions are conveniently expanded in plane waves:

$$\psi_i(\mathbf{r}) = \sum_{\mathbf{q}} a_i(\mathbf{q}) e^{i\mathbf{q}\cdot\mathbf{r}}. \qquad (2)$$

Here $\{R\}_{atom}$ are the atomic-position vectors, and a,(q) are the variationally determined expansion coefficients at the reciprocal lattice vector q of the supercell. The single-particle problem is then addressed by explicitly solving

$$\{-\tfrac{1}{2}\nabla^2 + V(\mathbf{r})\}\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}), \qquad (3)$$

The terms in the curly brackets constitute the Hamiltonian $\hat{H}$. The matrix elements of $\nabla^2$ and V(r) in the basis (2) are computed analytically and via a numerical Fourier transform, respectively.

Modeling of $\geq 10^3$ atom systems using Equations 1-3 requires three stages. We will use Figures 1 and 3 to illustrate them.

*First,* one needs to model the atomic geometry-that is, to specify $\{\mathbf{R}_{atom}\}$. There are two cases here.

(1) For *"freestanding" nanostructures* representing: colloidal structures,[1,2] the atomic positions of the core are essentially known in advance. For example, experimental structural measurements on Si quantum dots with more than

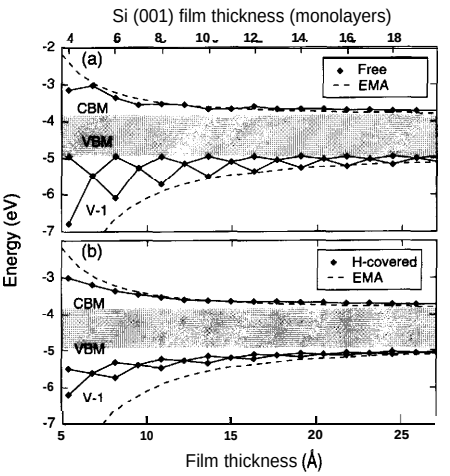


*Figure 1. Calculated near-gap energy levels of (001)-oriented hydrogen-free (a) and hydrogen-covered (b) Si films. Note the oscillations in the highest occupied (valence-band maximum, VBM) and next-to-highest occupied (V-l) valence bands, absent in the effective-mass- approximation (EMA) description (dashed line). The shaded area denotes the bulk bandgap region. The zero-confinement VBM state having a size-independent energy is apparent in (a). From Reference 74.*

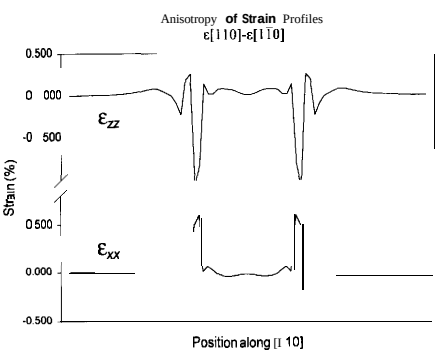


*Figure 2. Difference in atomistic strain energy $\varepsilon_{\alpha\beta}[110] - \varepsilon_{\alpha\beta}[\bar{1}10]$ on the two opposite faces of an InAs pyramid. In simple continuum models, this difference vanishes. From Reference 15.*

100 atoms[19] 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. However one usually has a reasonably good idea how to model such relaxations using either first-principle calculations or experimental data on the relevant *bulk surfaces.*[18] Arbitrary shapes (spheres, cubes, plates, pancake, cigars) of colloidal dots can be treated in this method, simply by positioning the atoms in the desired places. Free surfaces of colloidal dots are passivated by adatoms that remove surface states from the bandgap.[17]

(2) For *embedded nanostructures* representing-for example—"self-assembled"[3,4] structures, we surround the "well material" (e.g., InAs) by the "barrier material" (e.g., GaAs) and permit relaxation of all atoms to minimize the strain energy represented by an atomistic force field. We use for this purpose Keating's valence force field.[20] For example the center of Figure 3 shows a 45-Å-tall InAs pyramid (containing 3273 InAs molecules) with (110) and $(\bar{1}10)$ faces, lying on top of a l-monolayer InAs "wetting layer" and capped by GaAs. Relaxing all (InAs and GaAs) atomic positions produces the strain asymmetry depicted in Figure 2. The symmetry of the pyramid made of zinc-blende material is $C_2$, not the ideal $C_4$ symmetry of an ideal pyramid. This lower symmetry is missed by continuum elasticity models.[15] As will be described later, this misrepresentation of the symmetry can affect the energy levels.

The second stage in using Equations 1–3 requires determination of the atom-centered potentials $v_{atom}(\mathbf{r})$. We use for this purpose the empirical pseudopotential method.[21] Rather than fit $v_{atom}(\mathbf{G}_i)$ at a few *discrete* reciprocal lattice vectors $\{\mathbf{G}_i\}$ of the primary unit cell (as is done in classic calculations[21] for *bulk* solids), we fit[18] a *continuous* $v_{atom}(\mathbf{q})$ to a series of experimental data values and to detailed first-principles calculations on relevant prototype systems. This includes the bulk band structures, effective masses, deformation potentials, the surface work function, interfacial band offsets, and the density of states of chemisorbed surfaces. Unlike the case in tight-binding approaches,[22] we are able to compare[24] the ensuing potential V(r) with screened first-principles LDA results. Unlike the case with LDA,[13] experimentally sensible excitation energies are obtained. Strain effects are explicitly "felt" by the potential through its dependence on the atomic positions. Thus "crystal-field splitting"

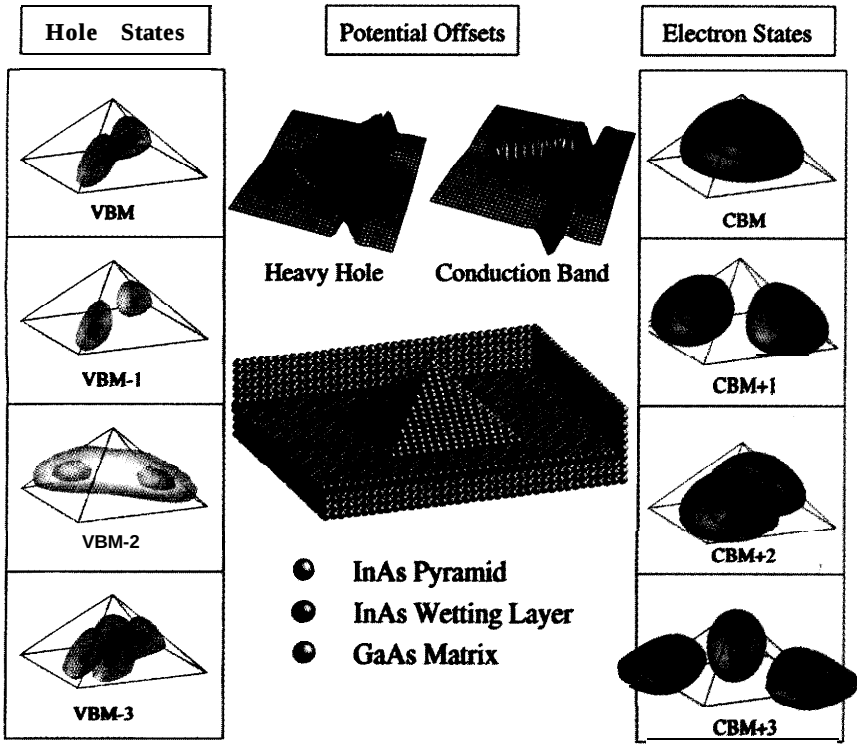


*Figure 3. The electronic structure of a 45-A -high, 90-Å base, strained InAs (1 10) pyramidal quantum dot embedded within GaAs. The strain-modified band offsets (for holes and electrons) are shown above the atomic structure. They exhibit a well for both heavy holes and electrons. These are localized within the pyramid and wetting layer as shown by the blue raised (lowered) triangle and ridge (trough), respectively. Isosurface plots of the four highest hole states and four lowest electron states, as obtained from pseudopotential calculations, appear on the left and right. The lowest electron state-the conduction-band minimum (CBM)-is s-like, while the next two states (CBM+ 1 and CBM+2) are nondegenerate p-like. From J. Kim, L. W. Wang, A. J. Williamson, and A. Zunger (unpublished). The calculation was performed using the Cray T3E at the National Energy Research Scientific Computing Center, University of California -Berkeley.*

(as well as spin-orbit splitting) is represented directly and nonperturbatively. This is illustrated in Figure 3 showing the strain-modified confining "conduction-band (CB) potential offset" and "heavy-hole potential offset." We see that the wetting layer and the pyramid itself pose an electron-attracting potential for the conduction electron and a hole-attracting potential for the valence holes.

The *third* stage of using Equations 1-3 involves solving Equation 3 for a given geometry and given pseudopotential. This is nontrivial for N $\sim 10^3$–$10^6$-atom systems because the number of plane-wave basis functions in Equation 2 scales ıs the order of N [denoted O(N)]. For nost semiconducting systems made up ɔf main-group elements, one needs about 50 plane waves per atom. So for N = 1,000, one typically has a 50,000 *X* 50,000 Hamiltonian problem. (With inclusion of the "vacuum region" in the cell, the basis is much larger.) The conventional variational approach to Equation 3 is to minimize the energy $\langle\psi|\hat{H}|\psi\rangle$ by varying the expansion coefficients $a(\mathbf{q})$ of $\psi$ (Equation 2); the first $\psi$ obtained is then the lowest energy eigenstate of $\hat{H}$. To find a higher state, one needs to orthogonalize $\psi$ to all previously converged energy eigenstates below it. The effort needed to accomplish this orthogonalization scales $N^3$. Consequently only small systems ($N \leq 100$ atoms) can be conveniently addressed by such conventional methods. However in order to study-for example-the near-edge optical properties of semiconductor quantum structures, one does not really need to know the $\sim 10^3$–$10^6$ lower states. What one typically needs to know about such systems includes the eigenvalues and eigenfunctions within -1 eV of the band-edge states (the valence-band maximum [VBM] and the conduction-band minimum [CBM]). To this end, we have developed a novel method[23,24] that enables calculation of eigensolutions around a given, "interesting" reference energy without the need to calculate any of the eigensolutions below it. The effort involved scales linearly with the system's size, thus enabling calculations of bandgap properties in mesoscopic systems. The method is exact in that the solutions are identical to those of Equation 3. The central point of this approach[23,24] is that the eigensolutions $(\psi_i \varepsilon_i)$ of Equation 3 also satisfy

$$(\hat{H} - \varepsilon_{\text{ref}})^2 \psi_i = (\varepsilon_i - \varepsilon_{\text{ref}})^2 \psi_i \ . \qquad (4)$$

Here the original spectrum $\{\varepsilon_i\}$ of $\hat{H}$ (Equation 3) has been folded at the reference point into the spectrum $\{(\varepsilon_i - \varepsilon_{\text{ref}})^2\}$ of $(\hat{H} - \varepsilon_{\text{ref}})^2$. The lowest solution of this "folded" spectrum is the eigenstate with $\varepsilon_i$ closest to $\varepsilon_{\text{ref}}$. Hence by placing $\varepsilon_{\text{ref}}$ in the physically interesting range, one transforms an arbitrarily high eigensolution into the lowest one, thus obviating the need for orthogonalization. For example, if $\varepsilon_{\text{ref}}$ is placed inside the energy window comprising the bandgap, solving Equation 4 results either in the VBM state or the CBM state, depending on which is closer to $\varepsilon_{\text{ref}}$. Changing $\varepsilon_{\text{ref}}$ within the bandgap region then assures that both the VBM and the CBM are found. The effort scales as O(N), so large systems can be calculated. Figure 3 illustrates the wave-function-square isosurfaces obtained for the GaAs-embedded InAs pyramid of a 90-A base size.

We see that the approach described in the last three steps satisfies the design criteria listed in the first section:

(1) Adjustable parameters underlying $v_{\text{atom}}(\mathbf{r})$ are fixed once and for all for the bulk solid. The bulk band structure $\varepsilon_{n\mathbf{k}}$ is hence "exact" throughout the Brillouin zone.

(2) The "universal" basis set of Equation 2 assures equivalent treatment of nanostructures of all dimensions.

(3) The full atomic symmetry is retained via specification of $\{\mathbf{R}_{\text{atom}}\}$.

(4) $v_{\text{atom}}(\mathbf{r} - \mathbf{R}_{\text{atom}})$ encodes explicitly spin-orbit effects, and through $\{\mathbf{R}_{\text{atom}}\}$ the effects of strain, pressure, crystal structure, and crystal field are all treated nonperturbatively.

Furthermore the approach is simple to use: Once you have the pseudopotentials and you specify the atomic positions, solving a ~1,000-atom problem requires less than an hour on a table-top Risc workstation. The results are energies,

wave functions, transition probabilities, etc. On a massively parallel CRAY T3E, we routinely handle up to a million atoms.

So what is the weakness/approximation involved? Our central approximation is that the screened atomic pseudopotential $v_{atom}(\mathbf{r})$ is *fixed* at the outset, and is not allowed to respond self-consistently to charge rearrangements underlying either electronic excitations or changes in size/shape. This is the price we pay for not calculating *all* occupied orbitals. At this point, this appears to be a reasonable tradeoff.

We next provide three illustrations for applications of "Step 1" of the model to a film (Figure 1), a freestanding quantum box (Figure 4), and an embedded quantum pyramid (Figure 3).

- The solid lines in Figure 1 show the pseudopotential calculated energy levels of a Si(OO1) quantum film as a function of its thickness.[14] All states shown are film-interior, not surface states. The dashed lines are EMA results. The shaded areas denote the bulk bandgap region. We see that (1) the exact VBM of the film (a) has no size dependence. We call this a "zero-confinement state." This behavior [14] is missed by the EMA. (2) The state next to the VBM ("V-l") exhibits odd-even oscillations. This too is missed by the continuumlike EMA that lacks atomistic information on odd versus even numbers of monolayers.
- Figure 4 shows the calculated[25] single-particle wave functions of the VBM and the CBM of a *freestanding* 6,000-atom GaAs quantum box (left), compared with the EMA results (right). We see that the pseudopotential wave functions are more extended than the EMA wave functions and do not exhibit the simple, sinelike envelope function predicted by the EMA. These differences will affect profoundly the electron-hole interactions discussed in the next section.
- Figure 3 shows the calculated single-particle wave functions of the valence and conduction states of a "self-assembled" InAs pyramidal dot embedded within GaAs. The lowest conduction band (CBM) is s-like, located 177 meV above the bulk GaAs CBM. The next two states (CBM + 1 and CBM + 2) are p-like, located 113 meV above the s-state and split (by 28 meV) via the $C_2$ symmetry. The VBM is located 265-meV below the bulk GaAs VBM. The next three valence states are also shown. By calculating larger pyramidal dots, we find that (1) the s-p splitting in the CB is reduced as the pyramid-base size (90 Å in Figure 3) increases, (2) the splitting between the two $p$ states is almost independent of the pyramid size but is reduced as the facet orientation (highest-to-base ratio) is lowered from (110) to (113), and finally (3) the confinement energies of the VBM and the CBM s-states scale within the *volume* of the dot.

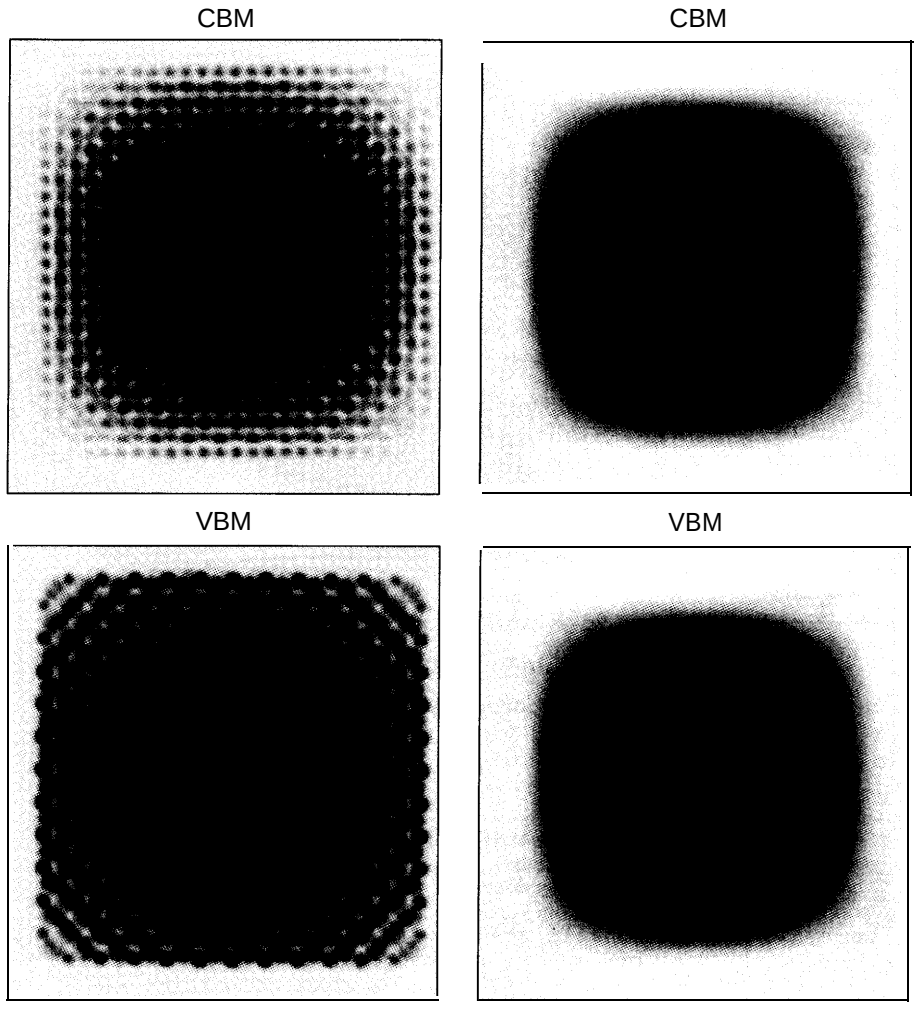


*Figure 4. The VBM and CBM wave functions of a 6,000-atom (110) × (1$\bar{1}$0) x (001) GaAs quantum dot, as obtained from a plane-wave pseudopotential calculation (left-hand side), are compared with the EMA (right-hand side). The wave-function amplitude, averaged along the [001] direction, is plotted in the (001) plane. From Reference 25.*

## The New Model, Step 2: Electron-Hole (Exciton) and Electron -Electron Interact ions

The ground state $\Phi_0$ of the nanostructure is obtained when all single-particle valence (v) orbitals $\{\psi_{i,v}\}$ fully occupied and all conduction (c) orbitals $\{\psi_{j,c}\}$ are empty. These single-particle states $\{\psi_i\}$ and energy levels $\{\varepsilon_i\}$ obtained from Equation 3 represent *average* interelectronic interactions. Upon excitation the hole in $\psi_{i,v}$ and the electron in $\psi_{j,c}$ interact both via Coulomb, exchange, and correlation interactions missing from Equation 3. The low-lying excited states of a quantum dot are calculated by a configuration-interaction expansion of the many-particle wave function in terms of single-substitution Slater determinants ((a,,). These are obtained[26] from the ground-state $\Phi_0$ by exciting an electron from the valence state $\psi_{i,v}(\mathbf{x})$ of energy $\varepsilon_{i,v}$ to the conduction $\psi_{j,c}(\mathbf{x})$ of energy $\varepsilon_{j,c}$. [Here x = $(\mathbf{r},\sigma)$ where $\sigma = \uparrow\downarrow$ is the spin variable.] By taking the ground-state energy as energy zero, the matrix elements of the many-particle Hamiltonian $\mathcal{H}$ in the representation $\{\Phi_{ij}\}$ are

$$\langle\Phi_{ij}|\mathcal{H}|\Phi_{kl}\rangle = (\varepsilon_{j,c} - \varepsilon_{i,v})\delta_{i,k}\delta_{j,l} - J_{ij,kl} - K_{ij,kl} \quad (5)$$

where $J$ and $K$ are the Coulomb and exchange integrals:

$$J_{ij,kl} = \int\int \frac{\psi^*_{i,v}(\mathbf{x}_1)\psi^*_{l,c}(\mathbf{x}_2)\psi_{k,v}(\mathbf{x}_1)\psi_{j,c}(\mathbf{x}_2)}{\bar{\varepsilon}(\mathbf{r}_1,\mathbf{r}_2)\,|\mathbf{r}_1-\mathbf{r}_2|}\,d\mathbf{x}_1 d\mathbf{x}_2 \quad (6)$$

and

$$K_{ij,kl} = \int\int \frac{\psi^*_{i,v}(\mathbf{x}_1)\psi^*_{l,c}(\mathbf{x}_2)\psi_{j,c}(\mathbf{x}_1)\psi_{k,v}(\mathbf{x}_2)}{\bar{\varepsilon}(\mathbf{r}_1,\mathbf{r}_2)\,|\mathbf{r}_1-\mathbf{r}_2|}\,d\mathbf{x}_1 d\mathbf{x}_2 . \quad (7)$$

The screening of the e-h interaction caused by the polarization of the medium is described by the microscopic dielectric constant $\varepsilon(\mathbf{r}_1,\mathbf{r}_2)$, which depends on the electron/hole positions. $\bar{\varepsilon}(\mathbf{r}_1,\mathbf{r}_2)$ consists of an electronic piece and an ionic piece.[26] The former depends on the macroscopic dielectric constant $\varepsilon_s(R)$, which is a function of the dot size[27] $R$ (not the electron position r), reflecting the total polarization response of a quantum dot to a constant field. Figure 5b shows[26] the microscopic dielectric constant $\bar{\varepsilon}(S)$ as a function of the electron-hole separation $|\mathbf{r}_e-\mathbf{r}_h|$ = S for a fixed dot size. When the electron and the hole are in the same Wigner-Seitz cell, they "see" a much reduced screening relative to the bulk value.

The integrals appearing in Equations 6 and 7 are computed directly in real space via multigrid or multipolar expansions[26] using the solutions of Equation 3. A diagonalization of the Hamiltonian matrix of Equation 5 then yields the low-energy excitonic states of the nanostructures. Four effects are included here: (1) The first term of Equation 5 represents pure single-particle effects (discussed in the previous section), and (2) electron-hole Coulomb interactions $J$ shift the single-particle energy difference (by a level-dependent amount), while (3) electron-hole exchange interactions $K$ split the levels. Finally (4) the inclusion of numerous electron-hole configurations in the Hamiltonian matrix introduces electron-electron and hole-hole correlation. (The effect is small for the "strong-confinement" limit where the size of the nanostructure is smaller than the bulk excitonic radius.)

## Electron-Hole Coulomb and Exchange Effects on the Spectra of Dots

Despite the great importance[28,29] of electron-hole Coulomb and exchange effects, only the highly simplified one-band EMA has been almost universally used to estimate these quantities even when the calculation of the single-particle energy gap requires more sophisticated and reliable methods. In fact the EMA provides simple, analytical expressions for the Coulomb energy and the exchange energy: Assuming an infinite potential barrier at the boundaries of the quantum dot and using the envelope functions of a noninteracting electron-hole pair, one obtains the well-known equations[30]

$$J^{\mathrm{EMA}}_{\mathrm{Coul}} = C_{\mathrm{Coul}}\frac{e^2}{\varepsilon R} \quad (8)$$

and

$$K^{\mathrm{EMA}}_{\mathrm{exch}} = C_{\mathrm{exch}}\left(\frac{a_x}{R}\right)^3 E_x \quad (9)$$

where $R$ is the dimension of the quantum dot; $E_x$ and $a_x$ are the bulk exciton exchange splitting and exciton radius, respectively; and $C_{\mathrm{Coul}}$, $C_{\mathrm{exch}}$ are dimensionless constants that depend only on the shape of the quantum dot. For example in the case of a spherical dot of radius $R$, the EMA yields for Equations 8 and 9 $C_{\mathrm{Coul}} = 1.786$ and $C_{\mathrm{exch}} = 2.111$. Using however microscopic pseudopotential wave functions from Equation 3, the *unscreened* Coulomb interaction (taking $\varepsilon = 1$ in Equation 6) is different from the commonly used EMA values of Equation 8: The ratio of EMA to pseudopotential unscreened Coulomb energies for say the lowest exciton in rectangular GaAs quantum dots ranges from 1.39 to 1.20 for box sizes 9.8-46 Å, respectively.[25] Furthermore the size dependence of the Coulomb energy is $R^{-\alpha}$ with $\alpha < 1$, while EMA (Equation 8) gives $\alpha = 1$. There are two main reasons for the 20–40% overestimation of the Coulomb energy by the EMA. First the EMA envelope functions are required unrealistically to vanish exactly at the boundary of the quantum dot, while the pseudopotential wave functions are allowed to decay variationally and spill out into the vacuum region (Figure 4). Second the contribution to the Coulomb energy resulting from the microscopic oscillations of the wave function (Figure 4) are neglected in the EMA. These effects lead to an overestimation of up to 40% of the electron-hole Coulomb interactions of the standard expression (Equation 8).

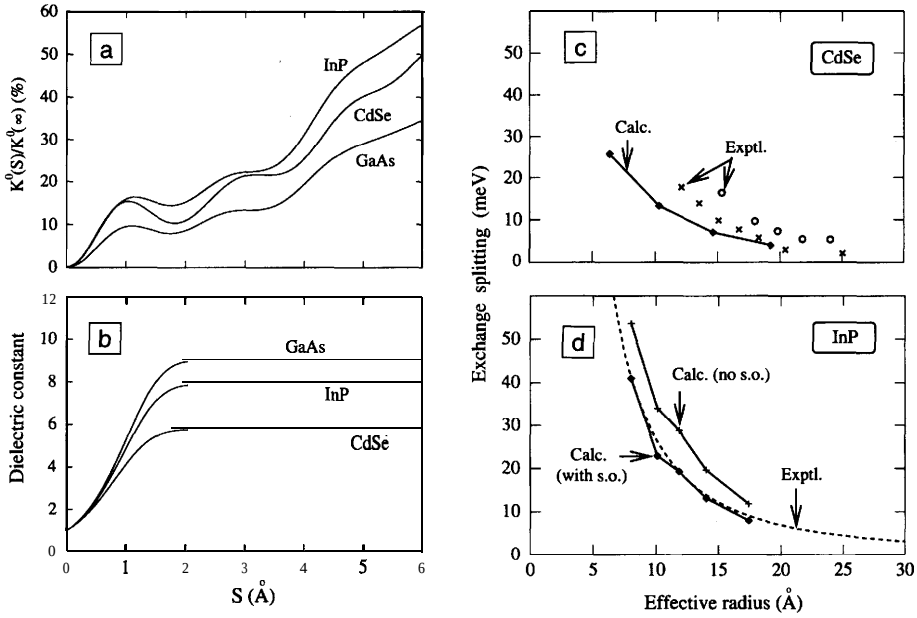


*Figure 5. (a) The unscreened exchange integral $K^0_{vc,vc}(S)$ (normalized by its converged value at S→∞) appears in part (a) as a function of the e-h distance S for GaAs (R = 22.5 A), InP (R = 17.4 A), and CdSe (R = 19.2 A) spherical quantum dots. The asymptotic values of $K^0_{vc,vc}(\infty)$ are 18.1, 26.9, and 18.2 meV, respective/y. (b) The distance-dependent dielectric constant $\bar{\varepsilon}(S)$. From Reference 26. (c) and (d) Electron-hole splitting of InP and CdSe spherical quantum dots as a function of the dot radius R. The experimental data for CdSe were taken from Reference 29 (crosses) and Reference 28 (circles). The dotted line in the InP plot is a fit to the experimental results of Reference 32. Calculated results (Reference 26) are shown both with and without spin-orbit coupling.*

How about electron-hole exchange interactions? In general the exchange interaction contains a short-range (SR) component that decays exponentially with the e-h separation $S = |\mathbf{r}_e - \mathbf{r}_h|$ and a long-range (LR) component that decays as a power law. Conventional wisdom[30,31] suggests that the LR exchange interaction in quantum dots originates, as in bulk semiconductors, from dipole-dipole coupling of the transition density between unit cells. Under this assumption, the LR contribution to the exchange splitting of s-like excitons in spherical quantum dots vanishes.[31] In the EMA, the e-h exchange is thus described[29] by an SR term of Equation 9 with an $R^{-3}$ size scaling, while the LR contribution is set to zero. While this approach fits well the observed red shift in CdSe nanocrystals,[28,29] in the case of spherical zinc-blende quantum dots, Equation 9 predicts a $1/R^3$ scaling of the red shift with size—which is not observed in either [32]InP or InAs nanocrystals. In both cases, the observed scaling is $R^{-2}$.

Direct calculations[26] of the excitonic spectra using our approach in Equations l-7 on CdSe, InP, and GaAs dots have shown the following:

(1) The e-h exchange interaction has a sizable LR contribution, comparable in magnitude with the SR contribution. This is evident in Figure 5a, which shows the (unscreened) exchange integral [26]$K$ (Equation 7) as a function of the electron-hole separation S. We see that the SR exchange contained in a Wigner-Seitz cell ($S \leq 2.5$ Å) is only -15% of the total exchange ($S \to \infty$).

(2) The LR component does not originate however from dipole-dipole interactions between unit cells, as in the case of bulk excitons, but from monopole-monopole interactions that are peculiar to quantum-confined systems and scale as $R^{-1}$, not $R^{-3}$.

(3) The calculated[26] screened exchange splitting (Figures 5c and 5d) is in reasonably good agreement with available experimental results,[29,32] even though no empirical adjustments are used.

(4) The reason why the phenomenological model of Equation 9, which neglects LR interactions, agrees with experiment for CdSe quantum dots[29] (but not for [32]InP or InAs dots) is that the SR exchange parameter used in Reference 29 is roughly twice as large as our directly calculated SR parameter.

(5) For zinc-blende InP dots, both the directly calculated exchange splitting and the experimentally measured red shifts[32] scale as $\sim 1/R^2$ with the dot radius $R$. This is in contrast with the $1/R^3$ scaling law predicted by Equation 9.

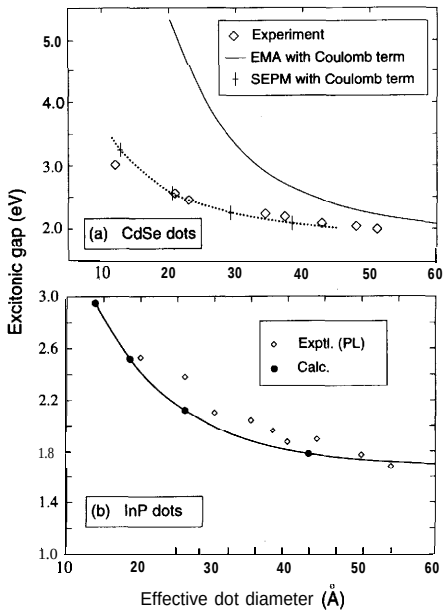


*Figure 6. Calculated [17,34] and measured [33,35] excitonic gaps of (a) CdSe [34,35] and (b) InP [17,33] dots.*

## Dependence of Bandgaps of Dots on Quantum Size and Shape

Figure 6b depicts our calculated[17] variation of the excitonic bandgap of freestanding, spherical InP dots with size. The calculated bandgap is seen to agree well with experiments.[33] Similar results were presented by us for CdSe dots[34] (Figure 6a)[35] and for Si dots.[24] This confirms the validity of our approach to the single-particle problem and to electron-hole interactions.

How well does the Standard Model describe band-edge states? Figure 7a compares the results of the present direct-diagonalization pseudopotential approach[36] (solid lines) and 6 X 6 k • p (dotted lines) for the energies of the VBM and CBM of InP dots. The k • p equations for dots are solved via the spherical-wave representation using a 6 X 6 k • p model fully parallel with the work of Norris and Bawendi.[10] The striking feature of Figure 7a is that the k • p approach produces (1) an incorrect order of the valence states: The state of (envelope function) $p$ symmetry is above that of s symmetry. This order occurred in other 6 X 6 k • p calculations of dots with small spin-orbit energies-for example in[37] CdS and in[38] InP. Because the lowest dot conduction state always has s symmetry, the 6 X 6 k • p method predicts that the lowest transition ($p \to s$) is forbidden in one-photon-absorption experiments while a direct diagonalization predicts that the lowest transition ($s \to s$) is one-photon-allowed. Thus k • p predicts a *large* red shift between absorption and emission that is not seen experimentally.[32] (2) The k . p energy levels are considerably deeper (larger confinement) than the exact result. For a dot with a 20-Å diameter, the k • p error for valence states is ~600 meV and (3) the curvature of conduction ener-

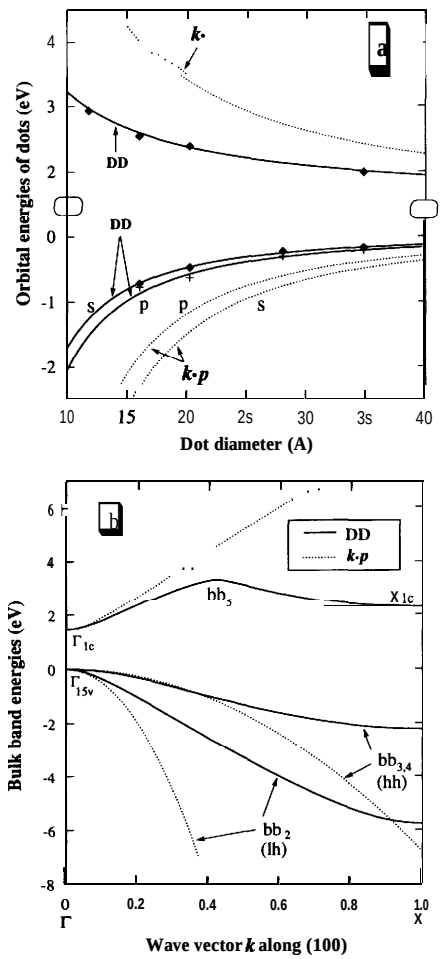


*Figure 7. (a) The orbital energies of the lowest conduction state and of the two highest valence states in InP dots. Solid lines: direct diagonalization (DD), dotted lines: k • p mode/. (b) Band dispersions of bulk InP along the ΓX direction as calculated by pseudopotential direct diagonalization (DD, solid lines) and by the 6 x 6 (for valence) and 2 x 2 (for conduction) k • p model (dotted lines). The notation bb, (n = 2 to 5) indicates the second to fifth bulk bands in the order of increasing energy All curves are calculated with the spin-orbit parameter $\Delta_0 = 0$. From Reference 36.*

gies versus size is consistently too large in k • p.

A recent analysis[36] explained these discrepancies in the **k** . **p** model. It turns out that these errors can be traced to **k** • **p** errors in the bulk *band structure (see* Figure 7b). The **k** . **p** produces a too-deep bulk light-hole band ($bb_2$ in Figure 7b), and exaggerates the off-Γ dispersion of the bulk heavy hole ($bb_{3,4}$ in Figure 7b) and conduction bands. Calculating the projections of the dot wave functions $\psi_i(\mathbf{r})$ onto the bulk Bloch bands then reveals the following:

(1) The s-like dot valence state has a large contribution from the bulk light-hole band, while the p-like dot valence state has no contribution from the bulk light-hole band. Given that the **k** • **p** approximation places the bulk light-hole band at spuriously deep energies (see Figure 7b),[36] we expect that the **k** • **p** will also place the dot's s-like state at too deep an energy (overconfinement). This is indeed shown to be the case by Figure 7a.

(2) The **k** points that contribute most significantly to the dot p-like state are generally more distant from **k** = **0** than those **k** points contributing significantly to the dot s-like state. Given that the k • **p** approach does not describe well the bulk dispersion *away* from Γ, we expect that the **k** . p model will not describe well the p-like dot valence state either. This is also shown to be true by Figure 7a.

(3) The s-like dot valence state has a larger contribution from the bulk conduction band than the p-like dot valence state, indicating that the s-like dot valence state is more affected by the neglect of coupling with the CBs in the standard k • **p** model.

Effects (1)–(3) explain why the 6 *X* 6 k . **p** produces an incorrect order of s and *p* valence states for small dots while overestimating the global confinement.

(4) The lowest s-like dot conduction state has a large contribution from the lowest bulk CB away from the k = 0 zone center. Since the **k** . **p** overestimates significantly the up-dispersion of the bulk CB, we expect it to **also** overestimate the dot's conduction-state energy. This is indeed shown to be true by our direct calculation (Figure 7a).

The analysis just described suggests that in order to improve the description of the nanostructure energy levels in the Standard Model, two main points should be taken into consideration: (1) a correct description of the bulk dispersion $\varepsilon_{n\mathbf{k}}$ and (2) introduction of interband coupling within a significant set of bulk bands.

Surprisingly bandgaps depend strongly not only on size but also on shape. If one considers a Si quantum dot with fixed number of atoms (say,[24] 1,100) and fixed interatomic distances but arranges them in different shapes, pseudopotential calculations (Figure 23 in Reference 24) show that the largest bandgap can be achieved using an elongated box while the fastest radiative lifetime is achieved with a near-cubic shape.

## Prediction of Charge Separation in GaAs/AlAs "Russian Dolls"

Our previous discussion showed that capturing the physics of quantum dots requires a realistic description of *interband coupling.* Models that describe the dot using just a few bulk bands are often insufficient. This is illustrated by the existence of charge separation in "Russian Dolls,"[39] an effect that vanishes when interband coupling is ignored.

It is well-known that in a sequence of *flat,* type-1 $(GaAs)_m/(AlAs)_n/(GaAs)_p/(AlAs)_q\ldots$ multiple quantum wells (MQWs), the wave functions of both the VBM and the CBM are localized on the widest well. Thus electron-hole charge separation is not possible. On the other hand, for short-period superlattices ("type II"), the electron and hole are localized on different materials (electron on AlAs and hole on GaAs) and different band-structure valleys (hole at the Brillouin zone center at Γ and electron at the Brillouin zone corner at X). Using our plane-wave pseudopotential direct-diagonalization approach, we predict[39] that electron-hole charge separation on different layers of the *same* material (GaAs) and *same* valley (Γ) is possible in *curved* (but not in flat) geometries. This is predicted for a set of concentric, nested cylinders ("Russian Doll") of GaAs and AlAs (Figure 8). Because the flat multiple-quantum-well (MQW) structure and the Russian-Doll structure with the same layer thicknesses have the same band offset diagram, the difference in behavior is not due to the potential. Rather it reflects different interband coupling induced by the *curvature* present in the Russian-Doll geometry but absent in the MQW. This identifies a new geometric degree of freedom-curvature-that can be used to tailor electronic properties.

## Conclusions

We have shown that a theory of the electronic structure of quantum dots requires primarily (1) a correct depiction of the dispersion relation in the underlying bulk material and (2) incorporation of coupling between a variationally significant number of bulk states. The Standard Model is lacking in both respects but can be improved along these lines.

Our model, based on considering the dot as a new structure in its own right (rather than viewing it as a perturbation on the bulk material), is made possible computationally by a series of innovations rendering a $10^3$–$10^6$-atom problem tractable within a pseudopotential framework using common workstations. The approach includes no adjustable parameters outside the *bulk* band structure; treats nanostructures of all dimensionality (including the bulk) on equal footing; captures the correct atomistic structure, strain, and symmetry (including surface effects); and incorporates Coulomb and exchange effects without any further approximations. It can be applied to "free-standing" (e.g., colloidal) dots as well as to embedded ("self-assembled") dots. We have applied it to Si,[14,24] InP,[17,18,36] InAs/GaAs,[15,16] CdSe,[34,36] GaAs AlAs,[25,26,39] and[40] InP/GaP nanostructures. These studies revealed the dependence of bandgaps on size and shape,[14–17,24,34] the origin of the red-shifted emission,[17,32] the microscopic origin of electron-hole exchange,[26] and size-dependent screening,[27] the scaling of the Coulomb interaction,[25] the existence of "zero-confinement states"[14] and odd-even oscillations[14] in films, the nature of states in strained pyramids,[16] and charge-separation in Russian Dolls.[39]

## Acknowledgments

I thank my collaborators, Alberto Franceschetti, Huaxiang Fu, Jeongnim Kim, Lin-Wang Wang, Andrew Williamson, Su-Huai Wei, David Wood, and Shengbai Zhang. This work was supported by the Office of Energy Research, Basic Energy Science, Division of Materials Science (OER-BES-DMS).

## References

1. H. Weller and A. Eychmiiller, in Semiconductor *Nanoclusters,* edited by P.V. Kamat and D. Meisel, vol. 103 (Elsevier, New York, 1996) p. 5.
2. J.T.G. Overbeek, *Adv. Colloid I. Sci.* 15 (1982) p. 251.
3. W. Seifert, N. Carlsson, M. Miller, M.E. Pistol, L. Samuelson, and L. Wallenberg, *Prog. Cryst.* Growth *Charact.* 33 (1966) p. 423.
4. M. Tabuchi, S. Noda, and A. Sasaki, in *Science and Technology of Mesoscopic Sfrucfures,* edited by S. Namba, C. Hamaguchi, and T. Ando (Springer, Tokyo, 1992) p. 379.
5. J.M. Luttinger and W. Kohn, *Phys. Rev.* 97 (1955); p. 869; E.O. Kane,]. *Phys. Chem. Solids 1* (1957) p. 249; M. Cardona and F.H. Pollak, *Phys. Rev.* 142 (1966) p. 530.
6. E.O. Kane, in *Physics of III-V Compounds,* edited by R.K. Willardson and A.C. Beer, *Semiconductors and Semimefals,* vol. 1 (Academic Press, New York, 1966) p. 75.

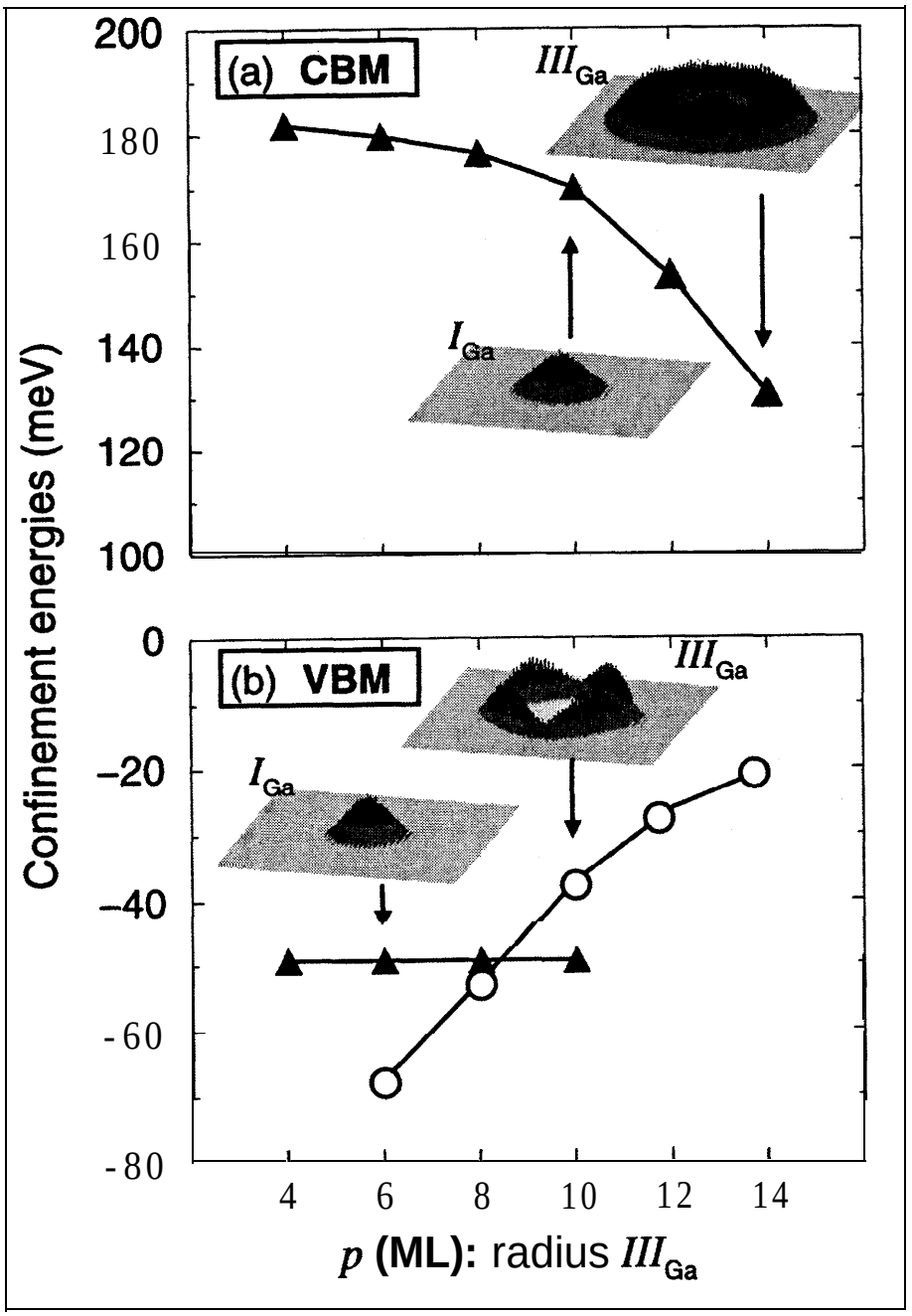


*Figure 8. Confinement energies of the (a) CBM and (b) two highest valence bands for cylindrical Russian Dolls versus the thickness p(GaAs).* The structure *is* made *of a IO-ML GaAs core, followed by 4 ML of AlAs,* p *ML of GaAs, and 8 ML of AlAs. Wave-function amplitudes, averaged along the wire direction, appear as insets for a few structures. From Reference 39.*


7. G. Bastard, J.A. Brum, and R. Ferreira, in *Solid* Sfafe *Physics,* edited by D. Turnbull and H. Ehrenreich, vol. 44 (Academic Press, New York, 1991) p. 229.
8. L.W. Wang and A. Zunger, Phys. *Rev. B* 54 (1996) p. 11417.
9. D.M. Wood, D. Gershoni, and A. Zunger, *Europhys. Lett.* 33 (1996) p. 383; D.M. Wood and A. Zunger, *Phys. Rev. B* 53 (1996) p. 7949.
10. D.J. Norris and M.G. Bawendi, *Phys.* Rev. B 83 p. 16338.
11. 0. Wind, F. Gindell, and U. Waggon, *J. Lu*min. 72-74 (1997) p. 300.
12. D. Gershoni, C.H. Henry, and G.A. Baraff, IEEE *J.* Quantum Electron. 29 (1993) p. 2433.
13. J.P. Perdew and A. Zunger, Phys. *Rev. B* 23 (1981) p. 5048.
14. A. Zunger, C.Y. Yeh, L.W. Wang, and S.B. Zhang, in *Znt. Conf. Phys. Semicond.* (World Scientific, Singapore, 1994) p. 1763; S.B. Zhang and A. Zunger, *Appl.* Phys. Left. 63 (1993) p. 1399; S.B. Zhang, C.Y. Yeh, and A. Zunger, *Phys.* Rev. *B* 48 (1993) p. 11204.
15. C. Pryor, J. Kim, L.W. Wang, A. Williamson, and A. Zunger, *J. Appl. Phys.* 83 (5) (1998).
16. J. Kim, L.W. Wang, and A. Zunger, Phys. Rev. B (in press).
17. H. Fu and A. Zunger, Phys. Rev. B 56 (1997) p. 1496.
18. L.W. Wang and A. Zunger, *Phys. Rev. B* 51 (1995) p. 17398; H. Fu and A. Zunger, *Phys. Rev. B* 55 (1997) p. 1642.
19. K.A. Littau, P.J. Szajowski, A.J. Muller, A.R. Kortan, and L.E. Brus. *J. Phys. Chem.* 97 (1993) p. 1224.
20. P.N. Keating, *Phys. Rev.* 145 (1966) p. 637.
21. D. Brust, J.C. Phillips, and F. Bassani, Phys. *Rev.* Left. 9 (1962) p. 94; M.L. Cohen and T.R. Bergstresser, *Phys. Rev.* 141 (1966) p. 789.
22. S.Y. Ren and J.D. Dow, Phys. *Rev. B* 45 (1992) p. 6492; J.P. Proot, C. Delerue, and G. Allen, *Appl.* Phys. Left. 61 (1992) p. 1948.
23. L.W. Wang and A. Zunger, *J. Chem.* Phys. 100 (1994) p. 2394; L.W. Wang and A. Zunger, *J. Chem.* Phys. 94 (1994) p. 2158.
24. L.W. Wang and A. Zunger, in *Semiconduc*for Nanoclusters: Studies in *Surface Science and* Catalysis, edited by P.V. Kamat and D. Meisel, vol. 103 (Elsevier, New York, 1996) p. 161.
25. A. Franceschetti and A. Zunger, Phys. *Rev.* Left. 78 (1997) p. 915.
26. A. Franceschetti, H. Fu, L.W. Wang, and A. Zunger (unpublished manuscript).
27. L.W. Wang and A. Zunger, *Phys. Rev. Lett.* 73 (1994) p. 1039.
28. M. Chamarro, C. Gourdon, I? Lavallard, 0. Lublinskaya, and A.I. Ekimov, Phys. Rev. *B* 53 (1996) p. 1336.
29. M. Nirmal, D.J. Norris, M. Kuno, M.G. Bawendi, A. Efros, and M. Rosen, *Phys. Rev. Lett.* 75 (1995) p. 3728; ibid., Phys. *Rev. B* 54 (1996) p. 4843; ibid. 53 (1996) p. 16347.
30. R.S. Knox, Solid *State* Phys. 5 (1963).
31. T. Takagahara, *Phys. Rev. B* 47 (1993) p. 4569.
32. 0.1. Mićić, H.M. Cheong, H. Fu, A. Zunger, J.R. Sprague, A. Mascarenhas, and A.J. Nozik, *J.* Phys. *Chem.* BlOl(l997) p. 4904.
33. 0.1. Mićić, J. Sprague, Z. Lu, and A.J. Nozik, *Appl.* Phys. Left. 68 (1996) p. 3150; 0.1. Mićić, C.J. Curtis, K.M. Jones, J.R. Sprague, and A.J. Nozik, *J.* Phys. Chem. 98 (1994) p. 4966.
34. L.W. Wang and A. Zunger, Phys. Rev. B 53 (1996) p. 9579.
35. C.B. Murray, D.J. Norris, and M.G. Bawendi, *J.* Am. *Chem. Soc.* 115 (1993) p. 8706.
36. H. Fu, L.W. Wang, and A. Zunger, *Appl. Phys. Lett.* 71 (1997) p. 3433.
37. G.B. Grigoryan, E.M. Kazaryan, A.L. Efros, and T.V. Yazeva, Sov. Phys. Solid State 32 (1990) p. 1031.
38. T. Richard, I'. Lefebre, H. Mathieu, and J. Allegre, Phys. *Rev. B* 53 (1996) p. 7287.
39. J. Kim, L.W. Wang, and A. Zunger, Phys. *Rev. B Rapid Commun. 56 (1997)* p. R15541.
40. A. Williamson, H. Fu, and A. Zunger, *ibid. 57 (7) (1998).*