1.07 Quantum Dots: Theory

N Vukmirovic and L-W Wang, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
© 2011 Elsevier B.V. All rights reserved.

1.07.1 Introduction

Since the early 1980s, remarkable progress in technology has been made, enabling the production of nanometer-sized semiconductor structures. This is the length scale where the laws of quantum mechanics rule and a range of new physical effects are manifested. Fundamental laws of physics can be tested on the one hand, while on the other hand many possible applications are rapidly emerging for nanometer-sized semiconductors.

The ultimate nanostructure where carriers are confined in all the three spatial dimensions is called a quantum dot. In the last 15 years, quantum dots have been produced in several different ways in a broad range of semiconductor material systems. The properties of quantum dots and their possible applications are largely dependent on the method they have been obtained with, which can, therefore, be used as a criterion for classification of different types of quantum dots:

**Electrostatic quantum dots.** One can fabricate quantum dots by restricting the two-dimensional (2D) electron gas in a semiconductor heterostructure laterally by electrostatic gates or vertically by etching techniques [1,2]. The properties of this type of quantum dots, sometimes termed as electrostatic quantum dots, can be controlled by changing the applied potential at gates, the choice of the geometry of gates, or external magnetic field. The typical size of these dots is of the order of 100 nm.

Self-assembled quantum dots. Self-assembled quantum dots are obtained in heteroepitaxial systems with different lattice constants. During the growth of a layer of one material on top of another, the formation of nanoscale islands takes place [3], if the width of the layer (the so-called wetting layer) is larger than a certain critical thickness. This growth mode is called Stranski–Krastanov mode. The most common experimental techniques of the epitaxial nanostructure growth are molecular beam epitaxy (MBE) and metalorganic chemical vapor deposition (MOCVD) [4,5]. Since the quantum dot material is embedded in another material, we refer to these dots also as embedded quantum dots. Self-assembled quantum dots typically have lateral dimensions of the order of 15–30 nm and height of the order 3–7 nm.
Colloidal quantum dots. A very different approach to obtain quantum dots is to synthesize single crystals of the size of a few nanometers through chemical methods. The dots obtained in this way are called nanocrystals or colloidal quantum dots [6]. Their size and shape can be controlled by the duration, temperature, and ligand molecules used in the synthesis [7]. Colloidal quantum dots are typically of spherical shape. They are often smaller than embedded quantum dots with the diameter sometimes as low as 2–4 nm.

Quantum dots have enabled the study of many fundamental physical effects. Electrostatic quantum dots can be controllably charged with a desired number of electrons and therefore the whole periodic system [8] of artificial atoms created, providing a wealth of data from which an additional insight into the many-body physics of fermion systems could be obtained [1]. Single-electron transport and Coulomb blockade effects on the one hand, and the regime of Kondo physics on the other hand, have been investigated [9,10].

One of the most exciting aspects of quantum dot research is certainly the prospect of using the state of the dot (spin state, exciton, or charged exciton) as a qubit in quantum information processing. Coherent control of an exciton state in a single dot selected from an ensemble of self-assembled quantum dots as well as the manipulation of the spin state in electrostatic quantum dots [12,13] have been achieved [11]. The theoretical and experimental progress in the field of spin-related phenomena in quantum dots has been reviewed in Refs. [1 and 14]. These results appear promising, although the control of a larger number of quantum dot qubits is not feasible yet, mainly due to difficulty in controlling qubit–qubit interactions.

The practical applications of quantum dots certainly do not lag behind these exciting areas of fundamental science with quantum dots. For example, colloidal quantum dots have found several cutting-edge applications such as fluorescent biological labels [15], highly efficient photovoltaic solar cells [16], and nanocrystal-based light-emitting diodes [1]. Self-assembled quantum dots find the main application as optoelectronic devices – lasers [17], optical amplifiers [18], single-photon sources [19,20], and photodetectors [21,22,23].

This chapter focuses on theoretical methods used for calculation of physical properties of self-assembled and colloidal quantum dots.

### 1.07.2 Single-Particle Methods

While quantum dots seem to be small and simple objects, a look at their structure from the atomic side reveals their high complexity. Bearing in mind that the lattice constants of the underlying semiconductor materials are typically of the order of 0.5 nm, one can estimate that a single self-assembled quantum dot contains \( \sim 10^6 \) nuclei and even a larger number of electrons interacting among each other with long-range Coulomb forces. Even the smallest colloidal quantum dots contain thousands of atoms.

This clearly indicates that direct solution of the many-body quantum dot Hamiltonian is not a feasible approach and that smart and efficient methods need to be developed. Here, methods that reduce the problem to an effective single-particle equation are reviewed.

More than two decades ago, Brus introduced [24–25] a simple effective mass method to calculate ionization energies, electron affinities, and optical transition energies in semiconductor nanocrystals. Within Brus’s model, the single-particle (electron or hole) energies \( E \) and wave functions \( \psi(\mathbf{r}) \) satisfy the Schrödinger’s equation given as

\[
\left[-\frac{1}{2m^*} \nabla^2 + P(\mathbf{r}) \right] \psi(\mathbf{r}) = E \psi(\mathbf{r})
\]

where \( m^* \) is the electron or hole effective mass. The system of atomic units where the reduced Planck’s constant \( \hbar \), the electron mass \( m_e \), and the electron charge \( e \) are all equal to 1 was used in equation 1 and will be used in what follows. For simplicity, equation 1 assumes that the particle must be confined within the dot, that is, that the potential outside the dot is infinite. This simplifying assumption can be easily relaxed by adding a more realistic confining potential \( V_{\text{conf}}(\mathbf{r}) \).

\( P(\mathbf{r}) \) in equation 1 is the additional potential caused by the presence of the surface of the quantum dot. It has a certain analogy with electrostatic image potentials in the case when a charge is near the surface of the metal or the interface between two dielectrics. It can be obtained by calculating the interaction energy between a bare electron and its induced screening potential. The extra interaction energy of an electron at \( \mathbf{r} \) inside the quantum dot compared to the corresponding value in bulk is then \( P(\mathbf{r}) \).

To model the two-particle excitations (such as electron + hole = exciton), Brus introduced an
electrostatic interaction energy term among these particles as

$$V(r_1, r_2) = \pm \frac{e^2}{\varepsilon |r_1 - r_2|} \pm P_M(r_1, r_2) + P(r_1) + P(r_2)$$  \hspace{1cm} (2)$$

where $\varepsilon$ is the dielectric constant, $P_M$ corresponds to the interaction of the charge of one particle with surface-induced polarization potential of the other particle, while the $P$-terms describe the interaction of the charge of one particle with its own surface-induced polarization potential, as previously described. The plus (minus) sign is for the two particles of the same (opposite) charge. The effective exciton Hamiltonian is then given as

$$H_{\text{exciton}} = -\frac{1}{2m_e} \nabla^2 + \frac{1}{2m_h} \nabla^2 - \frac{e^2}{|r_1 - r_2|} - P_M(r_1, r_2) + P(r_1) + P(r_2)$$  \hspace{1cm} (3)$$

The solution of the eigenvalue problem of this Hamiltonian can be written down analytically as

$$E^* = E_0 + \frac{\pi^2}{2R^2} \left[ \frac{1}{m_e} + \frac{1}{m_h} \right] - \frac{a_c e^2}{\varepsilon R} + \text{small term}$$  \hspace{1cm} (4)$$

where $a_c = 2 - \frac{\text{Si}(2\pi)}{\pi} + \frac{\text{Si}(4\pi)}{2\pi} \approx 1.8$ and $\text{Si}(x)$ is the sine integral function, $\text{Si}(x) = \int_0^x \frac{\sin t}{t} dt$. The last term in equation 4 originates from the last three terms in equation 3. One should note that $P(r) = P_M(r, r)/2$; therefore, $P_M(r_1, r_2)$ and $P(r_1) + P(r_2)$ cancel out exactly when $r_1 = r_2$ and lead to a small term when $r_1$ and $r_2$ are not equal. This small term can often be ignored in practice for spherical quantum dots.

The cancellation of the polarization terms gives us a guide for a general approach for calculating the excitons in nanocrystals. As a first step, one calculates the single-particle energies from equation 1 without the polarization term. As a second step, the screened electron–hole interaction is added perturbatively. One should, however, have in mind that such an approach is an approximation based on classical electrostatic consideration. It ignores the effects such as dynamic screening and the local-field effects of the dielectric function. The single-particle states obtained in this way are not the quasiparticles from the usual GW formalism. (The eigenenergies of equation 1 with the $P$-term are the quasiparticle energies that correspond to the electron affinity and ionization potential.) However, such single-particle states are the natural extension of single-particle states considered in other nanostructures, such as quantum wells and superlattices. These are also fully in line with eigenstates defined in the density functional theory (DFT) discussed below. This section is, therefore, completely devoted to the theoretical frameworks and methodologies for calculating these states.

### 1.07.2.1 Density Functional Theory

Within the DFT [26], the many-body Hamiltonian problem reduces to a set of single-particle Kohn–Sham equations [27] that read as

$$\left(-\frac{1}{2} \nabla^2 + V_{\text{ion}} + V_{\text{H}} + V_{\text{XC}}\right)\psi_i(r) = \epsilon_i \psi_i(r)$$  \hspace{1cm} (5)$$

In equation 5, $\psi_i(r)$ and $\epsilon_i$ are the wave functions and energies of Kohn–Sham orbitals, $V_{\text{ion}}(r)$ is the potential of all nuclei in the system, and $V_{\text{H}}(r)$ is the Hartree potential of electrons given as

$$V_{\text{H}}(r) = \int \rho(r') \frac{\rho(r)}{|r - r'|}$$  \hspace{1cm} (6)$$

where

$$\rho(r) = \sum_i |\psi_i(r)|^2$$  \hspace{1cm} (7)$$

is the electronic charge density of the system. The summation in equation 7 goes over all occupied Kohn–Sham orbitals. The exchange-correlation potential $V_{\text{XC}}$ in equation 5 is supposed to take into account all the other effects of electron–electron interactions beyond the simple Coulomb repulsion (described in $V_{\text{H}}$). The exact form of this potential is not known, and it needs to be approximated. The most widely used approximation is the local density approximation (LDA) where it is assumed that $V_{\text{XC}}$ depends only on the local electronic charge density and takes the same value as in the free electron gas of that density [27]. Equations 5 and 7 need to be solved self-consistently until the convergence is achieved.

DFT calculations are still computationally demanding, partly due to the necessity of self-consistent calculations. One also needs to calculate all the orbitals $\psi_i$ in each iteration, while in semiconducting systems one is often interested in only a few states in the region around the gap that determine the optical and transport properties of the system.

An alternative approach that avoids the full self-consistent calculation without loss in accuracy is the charge patching method (CPM) [28,29,30–32,33]. The basic assumption of the CPM is that the charge density around a given atom depends only on the local atomic environment around the atom. This is true if there is no long-range external electric field
that causes long-range charge transfer. This is often satisfied if there is a band gap in the material. Based on this assumption, the idea is to calculate (e.g., using DFT in LDA) the charge density of some small prototype system, \( \rho_{\text{LDA}}(r) \), decompose it into contributions from individual atoms (charge density motifs), and then use these motifs to patch the charge density of a large system.

In particular, charge density motifs are calculated from the charge density of the prototype system as

\[
m_{\alpha}(r - R_\alpha) = \rho_{\text{LDA}}(r) \frac{w_\alpha(|r - R_\alpha|)}{\sum_{\alpha'} w_{\alpha'}(|r - R_{\alpha'}|)}
\]

where \( R_\alpha \) is the position of atom type \( \alpha \) and \( m_{\alpha}(r - R_\alpha) \) is the charge density motif of this atom type, \( w_\alpha(r) \) is an exponentially decaying function that defines the partition function \( \frac{w_\alpha(|r - R_\alpha|)}{\sum_{\alpha'} w_{\alpha'}(|r - R_{\alpha'}|)} \) that divides the space into (mutually overlapping) regions assigned to each atom. \( m_{\alpha}(r - R_\alpha) \) is therefore, a localized function that can be stored in a fixed-size numerical array. \( l_\alpha \) denotes the atomic bonding environment of the atom \( \alpha \). After the charge density motifs are obtained from small prototype systems, the total charge density of the large nanosystem is obtained simply as the sum of motifs assigned to each of the atoms:

\[
\rho_{\text{patch}}(r) = \sum_{\alpha} m_{\alpha}(r - R_\alpha)
\]

Once the charge density is obtained using the charge patching procedure, the single-particle Hamiltonian can be generated by solving the Poisson equation for the Hartree potential and using the LDA formula for the exchange-correlation potential. The energies and wave functions of a few states around the gap can then be found using the methods developed to find a few eigenvalues of the Hamiltonian only, such as the folded spectrum method (FSM) [34] (that is described in Section 1.07.2.2).

The CPM was used to generate the charge densities of carbon fullerenes [33], semiconductor alloys [28], semiconductor impurities [29], organic molecules and polymers [35], and semiconductor quantum dots [32]. The resulting patched charge density is typically within 1\% of the self-consistently calculated LDA charge density, and the corresponding energies are within 30 meV. Typical numerical uncertainty (due to basis function truncations and different nonlocal pseudopotential treatments) of an LDA calculation is about the same order of magnitude. Therefore, the CPM can be considered to be as accurate as the direct "ab initio" calculations.

There are, however, cases where CPM cannot be used. One example is the total dipole moment of an asymmetric quantum dot [36]. Such a dipole moment can also induce an internal electric field and cause the long-range charge transfer in the system. It is, therefore, necessary to solve the charge density self-consistently, which can be done using the DFT/LDA method. However, a much more efficient linear scaling method to do such calculations has been developed recently: the linear scaling three-dimensional fragment (LS3DF) method [37]. Within the LS3DF method, the system is divided into many small fragments. The wave functions and charge densities of each fragment are calculated separately, each within the standard DFT/LDA method, using a group of a small number of computer processors. After the fragment charge densities are obtained, they are patched together to get the charge density of the whole system using a novel scheme that ensures that the artificial surface effects due to the system subdivision will be cancelled out among the fragments. The patched charge density is then used to solve a global Poisson equation for the global potential. An outside loop is iterated, which yields the self-consistency between the global charge density and the input potential. Due to the use of this novel patching scheme, the LS3DF is very accurate, with its results (including the dipole moments) essentially the same as the original direct DFT calculation results [37], but with potentially 1000 times speed-ups, for systems with more than 10 000 atoms. As the system grows larger, there are more fragments (while the fragment size is fixed); thus, more processor groups can be used to solve them. This provides a perfect parallelization to the number of processors. Meanwhile, the total computational cost is proportional to the number of fragments, and consequently the total number of atoms.

A well-known problem of the LDA-based calculations is that the band gap is severely underestimated [38,39]. DFT is rigorously valid only for ground-state properties, and there is no physical meaning for the Kohn–Sham eigen energies [27]. This conceptual difficulty can be circumvented by using time-dependent DFT, which is discussed later. In practice, however, one often restricts to the simple empirical ways to correct the band-gap error. One such way is to slightly modify the LDA Hamiltonian to fit the crystal bulk bandstructure, which can be done, for example, by changing the s, p, and d functions.
nonlocal pseudopotentials [29] to move the position of the conduction band while keeping the position of the valence band unchanged. This approach is based on the assumption that the valence band alignment predicted by the LDA is reliable.

For the treatment of colloidal quantum dots, one also has to take care of the quantum dot surface. The surface of an unpassivated nanocrystal consists of dangling bonds that introduce band-gap states. One way to remove these states is to pair the dangling bonds that introduce band-gap states. One passivating agent should provide $2 - m/4$ additional electrons. To keep the system locally neutral, there must be a positive $2 - m/4$ nuclear charge nearby. The simplest passivation agent can, therefore, be a hydrogen-like atom with $2 - m/4$ electrons and a nuclear charge $Z = 2 - m/4$. For IV–IV group materials like Si, this is a $Z = 1$ hydrogen atom. For III–V and II–VI systems, the resulting atoms have a noninteger $Z$; consequently, these are pseudohydrogen atoms. These artificial pseudohydrogen atoms do describe the essential features of good passivation agents and serve as simplified models for the real passivation situations, where organic molecules with complicated and often unknown structure are involved.

**1.07.2.2 Empirical Pseudopotential Method**

The empirical pseudopotential method (EPM) was introduced in the 1960s by Cohen et al. [40,41] to fit the bandstructure of bulk semiconductors. Within the EPM, the Schrödinger equation is given as

$$
\left[ -\frac{1}{2} \nabla^2 + V(r) \right] \psi(r) = E \psi(r)
$$

with

$$
V(r) = \sum_{\text{atom}} \psi_{\text{atom}}(|r - \mathbf{R}_{\text{atom}}|)
$$

where $\mathbf{R}_{\text{atom}}$ are the positions of the atoms and $\psi_{\text{atom}}(r)$ are spherical atomic potentials that in an effective manner take into account the effects of nuclei, core, and valence electrons. The great success of the EPM was that it was actually possible to fit the bandstructure of the semiconductors using this single-particle approach.

In the EPM calculations, the plane-wave representation is typically used, that is, the wave function is expanded as a linear combination of plane waves, where the summation is restricted only to reciprocal lattice $\mathbf{q}$ vectors with kinetic energy smaller than a certain predefined value $E_{\text{cut}}$. To evaluate the resulting Hamiltonian matrix in plane-wave representation, Fourier transforms of atomic potentials $\psi_{\text{atom}}(|\mathbf{q}|)$ are needed. Only a few of these are nonzero. These are used as adjustable parameters to fit the semiconductor bandstructure.

To apply the EPM to nanostructures, one needs to have a continuous $\psi_{\text{atom}}(\mathbf{q})$ curve, since the supercell is very large, and consequently $\mathbf{q}$ points are very dense. The continuous $\psi_{\text{atom}}(\mathbf{q})$ can be represented by a function of four parameters $a_1 - a_4$

$$
\psi(\mathbf{q}) = \frac{a_1 (\mathbf{q}^2 - a_2)}{a_3 e^{a_4 \mathbf{q}^2} - 1}
$$
or a sum of Gaussians

$$
\psi(\mathbf{q}) = \sum a_i e^{-i(\mathbf{q} - \mathbf{k})^2}
$$

For a full description of the colloidal quantum dots, the pseudopotentials of surface passivating hydrogen or pseudohydrogen atoms need to be fitted as well. The pseudopotentials are fitted to experimental data and first-principles calculations of bulk bandstructures, clean surface work function, and the density of states of chemisorbed surfaces.

Another approach to fit the pseudopotentials is to fit them directly to the LDA-calculated potential [42] and then modify them slightly to correct the band-gap error. The $\psi_{\text{atom}}(\mathbf{q})$ obtained in such a manner are able to fit the band structure within 0.1 eV and have in the same time a 99% overlap with the original LDA wave function. This approach, called the semiempirical pseudopotential method (SEPM), has been applied to CdSe [42], InP [43], and Si [42] nanostructures, representing II–VI, III–V, and IV–IV semiconductor systems, respectively.

With the empirical or semiempirical pseudopotentials at hand, one is able to construct the single-particle Hamiltonian. The diagonalization of this Hamiltonian is a routine task in the case of bulk semiconductors due to a small number of atoms in a supercell. However, this is no longer the case in quantum dots that contain a large number of atoms. Even the conventional conjugate gradient method [44] that is often used in ab initio calculations cannot be used since it scales as $O(N^3)$ due to an orthogonalization step, which is a necessary part of the algorithm. Fortunately, for the analysis of most electronic, transport, and optical properties of
semiconductor nanosystems, only the states in the spectral region close to the band gap are relevant, and there is no need to find all the eigenstates of the Hamiltonian. The FSM, specialized to find the eigenstates in a certain spectral region only, has therefore been developed by Wang and Zunger [34]. The method is based on the fact that the Hamiltonians \( H \) and \((H - E_{\text{ref}})^2\) have the same eigenvectors and that the few lowest eigenvectors of \((H - E_{\text{ref}})^2\) are the eigenvectors of \( H \) closest to the energy \( E_{\text{ref}} \). The lowest eigenstates of \((H - E_{\text{ref}})^2\) are then solved using the conjugate gradient method. It turns out that the use of \((H - E_{\text{ref}})^2\) compared to \( H \) slows down the convergence but with the use of preconditioners and a large number of iterations, convergence can still be achieved. The FSM within the plane-wave representation has been implemented in the parallel code Parallel Energy SCAN (PESCAN) [45]. It can be routinely used to calculate systems with a few thousand atoms, or even near-million atom systems [46]. Since only a few wave functions are calculated, the computational effort scales linearly to the size of the system. Linear scaling method for the calculation of the total and local electronic density of states and the optical absorption spectrum has been developed by Wang [47]. The reader is referred to Ref. [47] for the description of this method, called the generalized moments method (GMM).

Another method for solving of the EPM Hamiltonian is the linear combination of bulk Bloch bands (LCBBs) method [48]. The disadvantage of the plane-wave expansion is that it does not lend itself to systematic approximations. A basis set in which such approximations can be naturally made is the basis of full-zone bulk Bloch states. In this basis, the wave function expansion reads

\[
\psi(r) = \sum_{n} \sum_{k} G_{kn} \phi_{kn}(r)
\]

(12)

where

\[
\phi_{kn}(r) = \frac{1}{\sqrt{N}} e^{i k \cdot r} \phi_{n}(r)
\]

(13)

is the bulk Bloch function of the constituent bulk solid, where \( n \) is the band index, \( k \) is the supercell reciprocal-lattice vector, \( N \) is the number of primary cells in the supercell. The LCBB expansion allows one to select the physically important bands \( n \) and \( k \)-points. As a result, the number of basis functions can be significantly reduced compared to the plane-wave basis. It turns out that it is possible to use a fixed number of basis functions to achieve the same degree of accuracy for different system sizes, in contrast to the plane-wave basis where the number of basis functions scales linearly with system size. The origin of this effect is the fact that when the size of the system increases, the envelope function of the electronic state becomes smoother, and therefore the maximum value of the \( k \)-vector needed to represent it becomes smaller. This makes the LCBB method ideal for studying very large systems such as embedded quantum dots.

1.07.2.3 Tight-Binding Methods

The tight-binding (TB) method [49] is the simplest method that still includes the atomic structure of a quantum dot in the calculation [50,51,52,53]. In the TB method, one selects the most relevant atomic-like orbitals \(|i\alpha\rangle\) localized on atom \( i \), which are assumed to be orthonormal. The single-particle wave function is expanded on the basis of these localized orbitals as

\[
|\psi\rangle = \sum_{i\alpha} \epsilon_{i\alpha} |i\alpha\rangle
\]

(14)

and therefore the TB single-particle Hamiltonian is of the form

\[
H = \sum_{i \alpha} \epsilon_{i \alpha} |i \alpha\rangle \langle i \alpha| + \frac{1}{2} \sum_{i \alpha j \beta} t_{i \alpha j \beta} |i \alpha\rangle \langle j \beta|
\]

(15)

where \( \epsilon_{i \alpha} \) are the energies of the orbitals (the on-site energies), while \( t_{i \alpha j \beta} \) are the hopping integrals between different orbitals, which can be restricted to include only nearest neighbors or next-nearest neighbors. For the sake of notational simplicity, the form that does not include the spin–orbit interaction and therefore does not mix the states of different spin was presented. The extension to include spin orbit interaction is straightforward. The most popular flavor of TB is the empirical TB where the parameters of the Hamiltonian are treated as phenomenological and fitted to reproduce the bulk band structure obtained from experiment or higher level calculations. In such an approach, the atomic orbitals are not treated explicitly, since the whole spectrum of the single-particle Hamiltonian is determined by the onsite energies and hopping integrals. The wave function is represented by the coefficients \( \epsilon_{i\alpha} \) that slowly vary from site to site.
In the TB method, one restricts the atomic orbitals to include only a few for each atom. Since one is typically interested in states around the energy gap, one has to select the orbitals that define these states. In III–V, IV–IV, and II–VI semiconductors, these are typically the \( s, p_x, p_y, p_z \) orbitals and sometimes \( d \) orbitals. Quite often, an additional \( s \)-like orbital called \( s^* \) is added to provide an additional degree of freedom in fitting the TB parameters, which leads to models such as \( sp^3s^* \) \cite{54} and \( sp^3d^5s^* \) \cite{55}. In these models, the size of the resulting Hamiltonian matrix is \( nN \times nN \), where \( N \) is the number of atoms and \( n \) is the number of orbitals per atom \((n = 10 \text{ for } sp^3s^* \text{ with spin and } n = 20 \text{ for } sp^3d^5s^* \text{ with spin})\). Due to nearest-neighbor approximation, the matrix is sparse, and efficient methods for the diagonalization of sparse matrices can therefore be exploited.

One problem of the TB method is the lack of explicit basis functions. Although these can be added after the TB eigenstates have been calculated, these basis functions are not an intrinsic part of the TB Hamiltonian and its fitting process; thus, their compatibility is an issue. This causes problems to calculate physical properties such as dipole transitions and Coulomb and exchange interactions. Another issue in treating quantum dot heterostructures is how to choose the parameters at the interface of two materials, since only the parameters for bulk materials are available. An approximation needs to be introduced, usually by assuming the parameters at the interface as a certain average of the TB parameters of the two materials. In colloidal quantum dots, the surface has to be passivated. Here, we give an example of how this is done in the case of Si nanocrystals. The surface is passivated by H atoms, where the TB nearest-neighbor matrix elements \( V_{11-Si} \) between H and Si are scaled from the Si–Si matrix elements \( V_{Si-Si} \) according to the Harrison’s rule \cite{56}: \( V_{11-Si} = V_{Si-Si}(d_{Si-Si}/d_{11-Si})^3 \), where \( d_{Si-Si} \) and \( d_{11-Si} \) are the bond distances \cite{57}. Another way to treat the surface passivation is simply to remove the dangling-bond states from the calculated results or even from the Hamiltonian before the matrix is diagonalized. This is done by removing the hybrid \( sp^3 \) dangling-bond orbital from the TB Hamiltonian basis set (e.g., by removing the Hamiltonian matrix columns and rows expanded by these \( sp^3 \) bases) \cite{58}. This is a unique way of artificial passivation only applicable to TB calculations. The ability to describe the surface atomistically is a big advantage of the TB model compared to the \( k \cdot p \) model, which is described next.

### 1.07.2.4 \( k \cdot p \) Method

The previously described methods treat explicitly the atomistic details of the nanostructure, which therefore leads to their high accuracy and reliability but also to a significant computational cost. In the \( k \cdot p \) method, central quantities are the slowly varying envelope functions that modulate the rapidly varying atomistic wave function. Historically, the \( k \cdot p \) method was introduced to describe the bulk band structure around a certain special point in the Brillouin zone, and later on it was extended to describe heterostructures.

Let the Hamiltonian of an electron in a semiconductor be

\[
\hat{H} = \frac{\hat{p}^2}{2} + V_0(\mathbf{r}) + \hat{H}_0 \tag{16}
\]

where \( \hat{p} \) is the momentum operator, \( V_0(\mathbf{r}) \) the crystal potential (including nuclei, core electrons, and self-consistent potential of valence electrons), and \( \hat{H}_0 \) the spin–orbit interaction Hamiltonian arising from relativistic corrections to Schrödinger equation given by

\[
\hat{H}_0 = \frac{\alpha^2}{4} [\nabla V_0(\mathbf{r}) \times \hat{p}] \sigma \tag{17}
\]

where \( \alpha \) is the fine structure constant, and \( \sigma \) is a vector of Pauli matrices:

\[
\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{18}
\]

The envelope representation of the wave function of an electron is given by

\[
\Psi(\mathbf{r}) = \sum_u \psi_u(\mathbf{r}) u_i(\mathbf{r}) \tag{19}
\]

where \( u_i(\mathbf{r}) \) form the complete orthonormal set of functions with periodicity of the Bravais lattice, and \( \psi_u(\mathbf{r}) \) are slowly varying envelope functions. The most common choice of the functions \( u_i \) are bulk Bloch functions at the \( \Gamma \) point. After the replacement of equation 19 in the Schrödinger equation and making an approximation that eliminates the nonlocal terms that appear in the derivation, one arrives at [59,60]

\[
-\frac{1}{2} \nabla^2 \psi_u(\mathbf{r}) + \sum_u (-i) p_{uu'} \nabla \psi_u(\mathbf{r}) + \sum_u H_{uu'}(\mathbf{r}) \psi_u(\mathbf{r}) = E \psi_u(\mathbf{r}) \tag{20}
\]
Since the second term in equation 20 is crucial in determining the Hamiltonian matrix (and \((-i\nabla)\) becomes the \(k\)-vector if, e.g., the envelope function is expanded in plane waves), the method being described is called the \(k \cdot p\) method. The terms in the previous equation are given by

\[
P_{nm} = \frac{1}{\Omega} \int u_n(r)^\dagger \hat{H} u_m(r) d^3 r \tag{21}
\]

where the integration goes over the volume of the crystal unit cell \(\Omega\), and \(H_{nm}(r)\) is the term that, away from the interfaces, reduces to the bulk matrix elements of the Hamiltonian

\[
H_{nm} = \frac{1}{\Omega} \int u_n(r)^\dagger \hat{H} u_m(r) d^3 r = E_m \delta_{mn} \tag{22}
\]

where \(E_m\) is the band edge of band \(m\). In practice, one has to restrict to a finite number of bands. Historically, the \(k \cdot p\) method was first applied to valence band (six-band Hamiltonian) \([61–62]\), and later on the conduction band was added (eight-band Hamiltonian) \([63]\).

The explicit form of the eight-band Hamiltonian for the crystals with zinc-blende structure (such as, InAs, GaAs, AlSb, CdTe, GaP, GaSb, InP, InSb, ZnS, ZnSe, and ZnTe) is given below. This Hamiltonian also perturbatively includes the effect of remote bands. Since the point \(T_d\) symmetry group of the zinc-blende crystal is a subgroup of the diamond group of Ge and Si, the same \(k \cdot p\) Hamiltonian can be applied to these semiconductors, as well. On the basis \([\mathcal{T}_d]\) that diagonalizes the bulk Hamiltonian at \(k = 0\)

\[
|1\rangle = \begin{bmatrix} 1/2 \end{bmatrix} \begin{bmatrix} 1 \ -1/2 \end{bmatrix} = |S \downarrow\rangle
\]

\[
|2\rangle = \begin{bmatrix} 1/2 \end{bmatrix} \begin{bmatrix} 1 \ -1/2 \end{bmatrix} = |S \uparrow\rangle
\]

\[
|3\rangle = \begin{bmatrix} 1/2 \end{bmatrix} \begin{bmatrix} 3/2 \ -1/2 \end{bmatrix} = -\frac{i}{\sqrt{6}}[(X + iY) \downarrow] + i\sqrt{2/3}|Z \downarrow\rangle
\]

\[
|4\rangle = \begin{bmatrix} 3/2 \end{bmatrix} \begin{bmatrix} 3/2 \ 1/2 \end{bmatrix} = \frac{i}{\sqrt{2}}[(X + iY) \uparrow]\]

\[
|5\rangle = \begin{bmatrix} 3/2 \end{bmatrix} \begin{bmatrix} 1/2 \ -3/2 \end{bmatrix} = -\frac{i}{\sqrt{2}}[(X-iY) \downarrow]
\]

\[
|6\rangle = \begin{bmatrix} 3/2 \end{bmatrix} \begin{bmatrix} 1/2 \ -1/2 \end{bmatrix} = \frac{i}{\sqrt{6}}[(X-iY) \uparrow] + i\sqrt{2/3}|Z \downarrow\rangle
\]

\[
|7\rangle = \begin{bmatrix} 1/2 \end{bmatrix} \begin{bmatrix} 1/2 \ -1/2 \end{bmatrix} = -\frac{i}{\sqrt{3}}[(X-iY) \uparrow] + i\sqrt{3}|Z \downarrow\rangle
\]

\[
|8\rangle = \begin{bmatrix} 1/2 \end{bmatrix} \begin{bmatrix} 1/2 \ -1/2 \end{bmatrix} = -\frac{i}{\sqrt{3}}[(X+iY) \downarrow] - i\sqrt{2/3}|Z \downarrow\rangle \tag{23}
\]

the eight-band \(k \cdot p\) Hamiltonian reads (where the definition \(k = -i\nabla\) was introduced)

\[
H_k = \begin{bmatrix}
A & 0 & V^+ & 0 & \sqrt{3}V & -\sqrt{2}U & -U & \sqrt{2}V^+
0 & A & -\sqrt{2}U & -\sqrt{3}V^+ & 0 & -V & \sqrt{2}V & U \\
V & -\sqrt{2}U & -P + Q & -S^+ & R & 0 & \sqrt{3}S & -\sqrt{2}Q \\
0 & -\sqrt{3}V & -S & -P - Q & 0 & R & -\sqrt{2}R & 1/\sqrt{2}S \\
\sqrt{3}V^+ & 0 & R^+ & 0 & -P - Q & S^+ & 1/\sqrt{2}S^+ & \sqrt{2}R^+ \\
-\sqrt{2}U & -V^+ & 0 & R^+ & S & -P + Q & \sqrt{2}Q & \sqrt{3}2S^+ \\
-U & \sqrt{2}V^+ & \sqrt{3}2S^+ & -\sqrt{2}R^+ & 1/\sqrt{2}S & \sqrt{2}Q & -P - \Delta & 0 \\
\sqrt{2}V & U & -\sqrt{2}Q & 1/\sqrt{2}S^+ & \sqrt{2}R & \sqrt{3}2S & 0 & -P - \Delta
\end{bmatrix} \tag{24}
\]
where

\[ A = E_C + A' k_z^2 + \frac{k^2}{2} \]

\[ U = \frac{1}{\sqrt{3}} P_0 k_z \]

\[ V = \frac{1}{\sqrt{6}} P_0 (k_x - ik_y) \]

\[ P = -E_F + \gamma_1 k_z^2 \frac{1}{2} \]

\[ Q = \gamma_1 \left( \frac{k_z^2 + k_y^2}{2} - 2k_z^2 \right) \]

\[ R = -\sqrt{\frac{3}{2}} \left[ \gamma_1 (k_z^2 - k_y^2) - 2i\gamma_3 k_x k_y \right] \]

\[ S = \sqrt{3}\gamma_1 k_z (k_x - ik_y) \]

In previous equations \(|S\rangle, |X\rangle, |Y\rangle, \) and \(|Z\rangle\) are the bulk Bloch functions that transform as \(s, x, y\), and \(z\) under the action of the symmetry group

\[
P_0 = -i\langle S|\hat{\epsilon}_s|X\rangle = -i\langle S|\hat{\epsilon}_x|Y\rangle = -i\langle S|\hat{\epsilon}_z|Z\rangle
\]

(25)

is the interband matrix element of the velocity operator [64] usually reported in energy units as \(E_F = 2P_0^2\), the parameter \(A'\) is related to the conduction band effective mass as

\[
A' = \frac{1}{2m^*} - \frac{P_0^2}{E_x + \frac{1}{4} \Delta} - \frac{1}{2}
\]

(26)

\(\Delta\) is the spin–orbit splitting, \(E_x\) is the energy gap (the actual energy gap after the effect of \(\Delta\) was taken into account) equal to \(E_x = E_C - E_{\text{f}}\), while \(\gamma_1, \gamma_2, \) and \(\gamma_3\) are the Luttinger parameters [61] of the eight-band model that can be expressed in terms of the parameters of the six-band model \(\gamma_1', \gamma_2',\) and \(\gamma_3'\)

\[
\gamma_1 = \gamma_1' - \frac{E_p}{3E_x + \Delta}
\]

\[
\gamma_2 = \gamma_2' - \frac{1}{2} \frac{E_p}{3E_x + \Delta}
\]

\[
\gamma_3 = \gamma_3' - \frac{1}{2} \frac{E_p}{3E_x + \Delta}
\]

Since material parameters in the Hamiltonian of a quantum dot are position dependent and the \(k\) operators do not commute with coordinate operators, an ambiguity arises about the proper choice of operator ordering. It is necessary to choose the ordering in such a way that the Hamiltonian remains hermitian; however, this condition still does not give a unique choice. The most widely used [65,66,67,68,69] operator ordering in \(k \cdot p\)-based quantum dot electronic structure calculations is heuristic, symmetrical arrangement of operators

\[
f(r)k_k \rightarrow \frac{1}{2} (k_{\|} f(r) + k_{\perp} f(r))
\]

(27)

It has been pointed out that such ordering of operators can lead to unphysical solutions in some circumstances [70]. One can derive the appropriate form of the envelope function Hamiltonian with proper operator ordering starting from the empirical pseudopotential [59] or the LDA Hamiltonian [71]; however, such Hamiltonians are still not widely used.

A variety of numerical methods can be used to solve the \(k \cdot p\) Hamiltonian; these include the finite-difference methods [66,65,72,73] and the wave function expansion methods, where the basis functions can be plane waves [74,75,76,77,69,78], the eigenfunctions of the particle in a cylinder with infinite walls [79,80,81], or eigenfunctions of a harmonic oscillator [82].

While the \(k \cdot p\) model can be quite reliable for large embedded quantum dots, the colloidal quantum dots are often only a few nanometers in size. In reciprocal space, this could correspond to the \(k\) point at 1/3 toward the Brillouin zone boundary, where the \(k \cdot p\) might no longer be adequate. Indeed, it was found that the \(k \cdot p\) result compared to the result of a more accurate calculation might differ by 50% in the confinement energy [83], and sometimes it could change the ordering of the states [84]. Without care, spurious states in the energy gap might appear in \(k \cdot p\) calculations [85]. These states appear as the consequence of the fact that \(k \cdot p\) Hamiltonian does not correctly represent the bandstructure for \(k\)-vectors far away from \(\Gamma\) point and can give states in the gap for these \(k\)-vectors. The finite-difference method is, in particular, susceptible to the appearance of these states. The wave function expansion methods are less susceptible to this [86] since by the expansion in a finite basis set, the high \(k\) components of the envelope function are effectively filtered out. Another issue is that the \(k \cdot p\) Hamiltonian with a limited number of bands has a larger symmetry group than the true symmetry group of the system. This weakness from the fundamental point of view can be turned into a strength from the computational point of view, as it allows for block diagonalization of the Hamiltonian and therefore a more efficient solution of the problem [87,77,81].
1.07.2.5 The Effect of Strain

In previous sections, it was assumed that positions of atoms in a quantum dot are known \textit{a priori} and that local arrangement is the same as in the bulk crystal. However, in real structures this is certainly not the case. Self-assembled quantum dots are grown by depositing layers of material with a different lattice constant than the substrate, and therefore the quantum dot is strained. In colloidal quantum dots, there is also some relaxation of atoms close to the surface. It is well understood that strain has a strong effect on the electronic structure of semiconductors. Therefore, in this section, we describe how the effect of strain can be included in each of the methods described previously.

Within the framework of DFT, the effect of strain appears naturally in the formalism itself. One starts with a reasonable initial guess for the positions of atoms in the structure, then self-consistently solves the Kohn–Sham equations and obtains the forces on all atoms. One then moves the atoms in the direction of forces and obtains the new atomic configuration. The valence band maximum state introduces the so-called derivative motifs, defined as the change in the motif due to a particular bond length or angle change. These motifs can also be extracted from small-system calculations on prototype structures with slightly changed bond lengths or angles. Once the motifs and derivative motifs are obtained, the total charge density is constructed and the calculation of the electronic structure can be performed as previously described: charge patching, empirical pseudopotentials, and TB.

It has been pointed out in Section 1.07.2.1 that the charge density motifs used in the CPM depend on the local environment of the atom. In strained structures, bond lengths and angles change compared to the ideal ones, which therefore represents the change in the environment that affects the motifs. To include this effect, one introduces the so-called derivative motifs, defined as the change in the motif due to a particular bond length or angle change. These motifs can also be extracted from small-system calculations on prototype structures with slightly changed bond lengths or angles. Once the motifs and derivative motifs are obtained, the total charge density is constructed and the calculation of the electronic structure can be performed as previously described.

It might seem at first sight that it is not necessary to introduce any modifications to the empirical pseudopotential Hamiltonian to include the effects of strain. However, it turns out that within such an approach it would be difficult to correctly describe the dependence of the valence band maximum state on the hydrostatic strain. Therefore, a strain-dependent term is introduced for the local part of the pseudopotential of the atom of type \( \alpha \) in the form

\[
\epsilon_{i\alpha}^{\text{loc}}(\mathbf{r}, \epsilon) = \epsilon_{i\alpha}^{\text{loc}}(\mathbf{r}, 0)[1 + \gamma_{\alpha} \Tr(\epsilon)]
\]

(30)

where \( \gamma_{\alpha} \) is a fitting parameter and \( \Tr(\epsilon) = \epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz} \) is the trace of the strain tensor. The SEPM strain-dependent Hamiltonian obtained this way can be solved by representing it on the basis of plane waves or bulk Bloch bands. The extension of the unstrained cases to the strained cases for the basis of plane waves is straightforward. On the other hand, this is not true if
two-center integrals from which the hopping bond lengths is modeled by scaling the Slater–Koster bond lengths and bond angles. The dependence on is through the dependence of hopping integrals on 

\[ V = V_0 \left( \frac{d_0}{d} \right)^\eta \]  

which is a generalization [52,93] of Harrison’s \( d^{-2} \) rule [56,50]. In the above equation, \( V_0 \) is the integral for equilibrium bond length \( d_0 \) and \( V \) the integral when the bond length is \( d \). The change in bond angles in the system leads to different relative orientation of orbitals of neighboring atoms and consequently to a different hopping integral. This effect is naturally included through the Slater–Koster [92] tables of matrix elements in terms of the two-center integrals and direction cosines. Furthermore, there is a question whether the influence of strain on onsite energies should also be included. This is indeed done in many recent works [94,93,95], although different methods are used. Currently, there does not seem to exist a unique and simple model for the inclusion of this dependence as for the hopping integrals.

In \( \mathbf{k} \cdot \mathbf{p} \) models, the effect of strain is included through the bulk deformation potential parameters that can be either measured or determined from \( \textit{ab initio} \) calculations. In the case of eight-band Hamiltonian for zinc-blende crystals, the strain contribution to the Hamiltonian reads

\[
\hat{H}_S = \begin{pmatrix}
  a_e & -v^+ & 0 & -\sqrt{3}v & \sqrt{2}u & u & \sqrt{2}v^+ \\
  0 & a_e & \sqrt{2}u & \sqrt{3}v^+ & v & -\sqrt{2}v & -u \\
  -v & \sqrt{2}u & -p+q & -s^+ & r & 0 & \sqrt{3}t - \sqrt{2}q \\
  0 & \sqrt{3}v & -s & -p-q & 0 & -\sqrt{2}r & \frac{1}{\sqrt{2}}s \\
  -\sqrt{3}v^+ & 0 & r^+ & 0 & -p-q & s^+ & \frac{1}{\sqrt{2}}s^+ \sqrt{2}r^+ \\
  \sqrt{2}u & v^+ & 0 & r^+ & s & -p+q & \sqrt{2}q \sqrt{3}t^+ \\
  u & -\sqrt{2}v^+ & \sqrt{3}t^+ & -\sqrt{2}r^+ & \frac{1}{\sqrt{2}}s & \sqrt{2}q & -p & 0 \\
  -\sqrt{2}v & -u & -\sqrt{2}q & \frac{1}{\sqrt{2}}s^+ & \sqrt{2}r & \sqrt{3}t & 0 & -p \\
\end{pmatrix} 
\]

where

\[
\begin{align*}
  e &= e_{11} + e_{22} + e_{33} \\
  p &= a_e e \\
  q &= b \left[ e_{33} - \frac{1}{2} (e_{11} + e_{22}) \right] \\
  r &= \frac{\sqrt{3}}{2} b (e_{11} - e_{22}) - i d e_{12}
\end{align*}
\]

\[
\begin{align*}
  s &= -d (e_{13} - ie_{23}) \\
  u &= \frac{1}{\sqrt{3}} P_0 \sum_{j=1}^{3} e_{ij} k_j \\
  v &= \frac{1}{\sqrt{6}} P_0 \sum_{j=1}^{3} (e_{ij} - ie_{ij}) k_j
\end{align*}
\]

where \( a_e \) and \( a_v \) are the conduction and valence band hydrostatic deformation potentials, respectively, and
and \( d \) are the shear deformation potentials. The strain tensor that enters the Hamiltonian (32) can be obtained either from the VFF model (previously described) or from the continuum mechanical (CM) model.

In the CM model, the quantum dot structure is modeled by an elastic classical continuum medium whose elastic energy is given by

\[
W = \frac{1}{2} \sum_{ijkl} \int d^3r \lambda_{ijkl} \varepsilon_{ij} \varepsilon_{kl}
\]

where \( \lambda_{ijkl} \) is the elastic tensor relating the stress and strain tensor by Hooke’s law

\[
\sigma_{ij} = \sum_{kl} \lambda_{ijkl} \varepsilon_{kl}
\]

In the crystals with zinc-blende lattice, the elastic tensor is of the form

\[
\lambda_{ijkl} = C_{12} \delta_{ij} \delta_{kl} + C_{44} (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) + C_{an} \sum_{p=1}^{3} \delta_{ip} \delta_{jp} \delta_{ip}
\]

where \( C_{12}, C_{44}, \) and \( C_{an} = C_{11} - C_{12} - 2C_{44} \) are the elastic constants. The finite element discretization and minimization of the functional (33) leads to a system of linear equations that can be efficiently solved.

There have been several comparisons in the literature between the VFF and CM models [66,90,96]. While certain differences have been obtained, the results of the two models give overall agreement, as can be seen from a comparison between strain distribution in a pyramidal \( \text{InAs/GaAs} \) quantum dot from Ref. [90] that is given in Figure 1. From the fundamental point of view, the advantage of the VFF model is that it captures the atomistic symmetry of the system, while the CM models have a higher

**Figure 1** Strain profiles of \( \text{InAs/GaAs} \) square-based pyramidal quantum dots and the differences along the \( z \)-direction through the pyramidal tip (left-hand side). The \( \varepsilon_{xx}, \varepsilon_{zz} \) and \( \text{Tr}(\varepsilon) = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} \) components of strain are shown respectively in parts (a), (b) and (c). The difference between these components calculated using continuum elasticity (CE) and atomistic elasticity (AE) is given in parts (d), (e) and (f) respectively. The discrepancy is the largest around the interfaces, while the strains in the barrier (GaAs substrate and capping layer) agree well within 0.5%. A significant difference is also found inside the quantum dot where the InAs experience large compressive strains at about 7% due to the lattice mismatch. Reproduced with permission from Pryor C, Kim J, Wang LW, Williamson AJ, and Zunger A (1998) Comparison of two methods for describing the strain profiles in quantum dots. *Journal of Applied Physics* 83: 2548–2554.
symmetry group. From the computational point of view, the VFF model is more demanding as the displacement of each atom is considered, in contrast to the CM models where a grid of the size of lattice constant or even larger may be used, leading to a smaller number of variables to be handled. In several important cases, there are analytical or nearly analytical solutions of the CM model [97,98]. However, these advantages of the CM models are becoming less important as modern computers can handle the VFF calculations quite easily.

The nonself-consistent methods described above do not allow for long-ranged charge redistributions and therefore neglect the effects such as piezoelectricity where charge is moved due to strain. The piezoelectric potential then has to be calculated independently and added as an additional potential. The components of piezoelectric polarization in a crystal of arbitrary symmetry are given as

$$P_i = \sum_{k,l=1}^{3} \varepsilon_{ikl} \epsilon_{kl}$$

where $\varepsilon_{ikl}$ are the piezoelectric constants of the material. In a crystal with zinc-blende symmetry, the only nonzero components of $\varepsilon_{ikl}$ are

$$\varepsilon_{121} = \varepsilon_{132} = \varepsilon_{213} = \varepsilon_{231} = \varepsilon_{312} = \varepsilon_{321}$$

The charges induced by piezoelectric polarization can then be calculated, and the additional piezoelectric potential is obtained from the solution of Poisson's equation. It has been recently realized that in addition to the first-order piezoelectric effect given by equation 36, second-order piezoelectric effects might be important as well [99].

### 1.07.3 Many-Body Approaches

The methods presented in Section 1.07.2 give a strategy for calculating the single-particle states. These can be very useful for calculating the optical properties, as demonstrated, for example, in Section 1.07.4.3. Nevertheless, there are cases when the many-body nature of electron–hole excitations should be directly considered. The approaches along this line are described in Section 1.07.3.

#### 1.07.3.1 Time-Dependent DFT

Within the time-dependent DFT (TDDFT) [100,101], one solves the time-dependent Kohn–Sham equations

$$i \frac{\partial}{\partial t} \psi_i(r, t) = \left[ -\frac{1}{2} \nabla^2 + V(r, t) \right] \psi_i(r, t)$$

where

$$\rho(r, t) = \sum_{i=1}^{M} |\psi_i(r, t)|^2$$

The potential $V(r, t)$ should depend, in principle, on charge density in all times before $t$. A widely used approximation is the adiabatic LDA in which it is assumed that $V(r, t)$ depends only on $\rho(r, t)$, and that the functional form of this dependence is the same as in LDA in time independent DFT. We refer to this approximation as time-dependent LDA (TDLDA).

The TDLDA can be used to calculate the optical absorption spectrum by adding the external electromagnetic field perturbation potential to equation 38 and solving the equations by explicit integration in time [102]. Another approach is to assume that the perturbation is small and use the linear response theory. The exciton energy can then be found from the eigenvalue problem

$$\sum_{k} [(\varepsilon_i - \varepsilon_k) \delta_{ik} + (f_i - f_k) K_{ik} \delta(\omega)] C_{ik} = \omega C_{ik}$$

where $\varepsilon_i$ and $\varepsilon_k$ are the LDA ground-state Kohn–Sham eigen energies, and $f_i$ and $f_k$ are the occupation number of Kohn–Sham eigen states $\psi_i$ and $\psi_k$. Within TDLDA, $K_{ik,\delta}$ becomes independent of $\omega$ and is given as

$$K_{ik,\delta} = \int d\tau_1 d\tau_2 \psi_i(\tau_1) \psi_k(\tau_2) \left[ \frac{1}{|\tau_1 - \tau_2|} \right]$$

$$+ \delta(\tau_1 - \tau_2) \frac{\partial f_{\text{LDA}}(\rho(\tau_1))}{\partial \rho(\tau_1)} \psi_k(\tau_2) \psi_i(\tau_1)$$

where $f_{\text{LDA}}(\rho)$ is the LDA exchange-correlation potential. The first term in equation 41 is the exchange interaction, while the second can be called the screened Coulomb interaction. The justification of this assignment would require a comparison with equations from other approaches, such as the configuration interaction and GW+Bethe–Salpeter equation (BSE). It might also seem surprising that the screened Coulomb interaction is not a nonlocal integral between $r_1$ and $r_2$. This is because in the LDA, the exchange-correlation term is a local functional of charge density.
TDLDA appears to work quite well for optical spectra of small clusters and molecules. The results of TDLDA can then agree quite well with experimental measurements, as shown, for example, for the case of SiH$_4$ in Ref. [103]. These results are significantly improved compared to bare LDA results, which is due to exchange interaction in equation 40, which can be quite strong in such small systems. The screened Coulomb interaction in equation 40, however, does not play a significant role then, as also shown in Ref. [103].

On the other hand, the TDLDA is not as accurate for larger systems. For a bulk system, it is known that the TDLDA band gap will be the same as the LDA band gap [104,49]. The TDLDA does not provide a better bulk optical absorption spectrum than the LDA, as shown in Ref. [103]. The origin of these problems is the screened Coulomb interaction in equation 40, however, does not play a significant role then, as also shown in Ref. [103].

There are arguments that the exchange interaction $K_{v^v,\psi'}$ should not be screened, which come from the two-particle Green’s function construction, where screening of the exchange term would cause double counting [107]. Nevertheless, in practice, it is found that the exchange consists of a long-range term that should be unscreened and a short-range term [108] that should be screened by the bulk dielectric function [109,110]. The effective dielectric function $\tilde{\varepsilon}(r_1, r_2)$ used in equation 44 incorporates this because $\tilde{\varepsilon}(r_1, r_2) \rightarrow 1$ for $|r_1 - r_2| \rightarrow 0$. The seeming contradiction to the Green’s function argument can be resolved by realizing that if only a limited configuration space is used in equation 43, the effect of other unused configurations can be included in the exchange screening term [107].

The CI equation 43 has the same form as the corresponding equation in the TDLDA (equation 40), with the difference in the expressions for the exchange

\[
\sum_{v'c'} H_{vc,v'c'} \Phi_{v'c'} = \sum_{v'c'} [(E_v - E_{v'}) \delta_{vc,vc'} + K_{vc,v'c'} - J_{vc,v'c'}] \Phi_{v'c'} = EC_{vc} \quad (43)
\]

where $E_v$ and $E_{v'}$ are the single-particle eigenenergies, $E$ is the exciton energy, and $K_{vc,v'c'}$ and $J_{vc,v'c'}$ are the exchange and Coulomb interactions, respectively.

The effective dielectric screening used in equations 44 and 45 is of the form

\[
\frac{1}{\tilde{\varepsilon}(r_1, r_2)|r_1 - r_2|} = \int \varepsilon_{\text{bulk}}^{-1}(r_1, r) \frac{1}{|r - r_2|} \, dr \quad (46)
\]

where $\varepsilon_{\text{bulk}}^{-1}(r_1, r)$ is the bulk inverse dielectric function that differs from the one of the quantum dot nanostructure, which also contains the surface polarization $P$ discussed in Section 1.07.2. The use of bulk inverse dielectric function is in line with the fact that the single-particle energies $E_v$ and $E_{v'}$ are obtained from the Schrödinger equation that does not contain the surface polarization term. If the surface polarization term is used in the single-particle equation, then the full inverse dielectric function should be used and the surface polarization terms will roughly cancel out.

1.07.3.2 Configuration Interaction Method

When the single-particle states are obtained, one can form many-body excitations by creating Slater determinants out of these single-particle states. One can then diagonalize the many-body Hamiltonian in the Hilbert space formed from a restricted set of such determinants. This approach is called the configuration interaction (CI) method. When one is interested in excitons, the wave function is assumed as

\[
\Psi = \sum_{\nu_1=1}^{N_\nu} \sum_{\nu=1}^{N\nu} C_{\nu_1 \nu} \Phi_{\nu_1 \nu} \quad (42)
\]

where $\Phi_{\nu_\nu}$ is the Slater determinant when the electron from the valence band state $\nu$ is excited to

the conduction band state $\nu$. The eigenvalue problem of the Hamiltonian then reads

\[
\sum_{\nu'c'} H_{\nu\nu',cc'} C_{\nu'c'} = \sum_{\nu'c'} [(E_v - E_{\nu'}) \delta_{\nu\nu',cc'} + K_{\nu\nu',cc'} - J_{\nu\nu',cc'}] C_{\nu'c'} = EC_{\nu\nu} \quad (43)
\]
and the Coulomb integrals. On top of a single-particle calculation, the CI method was used to calculate very large systems, such as pyramidal quantum dots with near by one million atoms [111]. It was also used to calculate many-body excitations, such as multie excitons, and few electron excitations, and to study Auger effects [112]. All these calculations are made possible by selecting a limited window of single-particle states used in these configurations. It is difficult or impossible to study such systems using TDLDA or GW + BSE. One should nevertheless be cautious about the models used for screening in these multiparticle excitations.

**1.07.3.3 GW and BSE Approach**

Within this approach, one first calculates the quasi-particle excitations, which is somewhat analogous to single-particle calculations in Section 1.07.2. These are then used to solve the BSE for excitons, which is in some sense similar to CI equations of the previous section.

A quasiparticle is defined as the pole in frequency space in the single-particle Green’s function

$$G(\mathbf{r}, \mathbf{r}'; \omega) = -\frac{1}{\pi} \int |M| T \psi(\mathbf{r}) \psi^\dagger(\mathbf{r}'') |M\rangle$$

(47)

where $\psi(\mathbf{r})$ is the particle creation operator, $|M\rangle$ the $M$ particle ground state, and $T$ the time-ordering operator. Quasiparticle energies correspond to energies for adding or removing one electron from the system [113]. Within the GW approximation [114], the appropriate single-particle equation reads

$$\left[ \frac{1}{2} \nabla^2 + \sum_{\text{atom}} V_{\text{bare}}(\mathbf{r} - \mathbf{R}_{\text{atom}}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \right] \psi_i(\mathbf{r}) + \int \sum_{\mathbf{r}, \mathbf{r}'} \epsilon_i(\mathbf{r}') \psi_i(\mathbf{r}') d\mathbf{r}' = \epsilon_i \psi_i(\mathbf{r})$$

(48)

where

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = -\sum \psi_i(\mathbf{r}) \psi_i^\dagger(\mathbf{r}')$$

$$\times \left[ f_0 W(\mathbf{r}, \mathbf{r}', \epsilon_i - \omega) + \frac{1}{\pi} \int \frac{\text{Im} W(\mathbf{r}, \mathbf{r}', \omega')}{\omega - \epsilon_i - \omega' + i\delta} d\omega' \right]$$

(49)

is the self-energy potential that replaces the LDA exchange-correlation potential of LDA single-particle equations.

$$W(\mathbf{r}, \mathbf{r}', \omega) = \int \epsilon^{-1}(\mathbf{r}, \mathbf{r}_1, \omega) \frac{1}{|\mathbf{r}_1 - \mathbf{r}'|} d\mathbf{r}_1$$

(50)

is the dynamically screened interaction, where $\epsilon^{-1}(\mathbf{r}, \mathbf{r}_1, \omega)$ is the inverse dielectric function.

While equations 48 and 49 should, in principle, be solved self-consistently, one usually replaces the self-energy term with its expectation value with respect to the LDA Kohn–Sham wave functions $\langle \psi_i | \Sigma | \psi_i \rangle$, which constitutes the zeroth-order approximation of the GW procedure. It has been shown that the self-consistent calculations [115,116,117,118,119] make the spectral properties worse. Such calculations are performed with the use of pseudopotentials. It is possible that self-consistency will not make the results worse if all-electron calculation is performed.

The two-particle Green’s function defined as

$$G(\mathbf{r}_1, \mathbf{r}_2; \mathbf{r}_1', \mathbf{r}_2') = -\langle \psi(\mathbf{r}_1) \psi(\mathbf{r}_2) \psi^\dagger(\mathbf{r}_1') \psi^\dagger(\mathbf{r}_2') | M \rangle$$

(51)

contains the information about the exciton energies. These can be retrieved by taking $t_1 = t_1' + 0^-$ and $t_2 = t_2' + 0^-$ and transforming to frequency space to obtain (in the condensed notation) $G_2(\omega)$, whose poles are the exciton energies. The Dyson equation for $G_2(\omega)$ reads [120,113]

$$G_2(\omega) = G_2^{(0)}(\omega) + G_2^{(0)}(\omega) K'(\omega) G_2(\omega)$$

(52)

where $G_2^{(0)}(\omega)$ is the noninteracting two-particle Green’s function, and $K'(\omega)$ is an electron–hole interaction kernel. Equation 52 for the electron–hole pair is called the BSE [120]. It can be solved by expanding the exciton wave function as

$$|M, S\rangle = \sum_{\mathbf{r}_1} \sum_{\mathbf{r}_2} C_{\mathbf{r}_1, \mathbf{r}_2} a_{\mathbf{r}_1}^\dagger b_{\mathbf{r}_2}^\dagger |M\rangle$$

(53)

where $a_{\mathbf{r}_1}^\dagger$ and $b_{\mathbf{r}_2}^\dagger$ are the hole and electron creation operators. The equations for $C_{\mathbf{r}_1, \mathbf{r}_2}$ coefficients are then given as

$$(\epsilon_{\mathbf{r}_1} - \epsilon_{\mathbf{r}_2}) C_{\mathbf{r}_1, \mathbf{r}_2} + \sum_{\mathbf{r}_1', \mathbf{r}_2'} (K_{\mathbf{r}_1, \mathbf{r}_1'; \mathbf{r}_2, \mathbf{r}_2'} - J_{\mathbf{r}_1, \mathbf{r}_1'; \mathbf{r}_2, \mathbf{r}_2'}) C_{\mathbf{r}_1, \mathbf{r}_2'} = \Omega_S C_{\mathbf{r}_1, \mathbf{r}_2}$$

(54)

where $\epsilon_{\mathbf{r}_1}, \epsilon_{\mathbf{r}_2}$ are the quasiparticle eigenenergies obtained from equation 48, and $\Omega_S$ is the exciton energy. The $K_{\mathbf{r}_1, \mathbf{r}_1'; \mathbf{r}_2, \mathbf{r}_2'}$ is the same as in equation 44 without the screening $\overline{\epsilon}(\mathbf{r}_1, \mathbf{r}_2)$. The $J_{\mathbf{r}_1, \mathbf{r}_1'; \mathbf{r}_2, \mathbf{r}_2'}$, on the other hand, reads

$$J_{\mathbf{r}_1, \mathbf{r}_1'; \mathbf{r}_2, \mathbf{r}_2'} = \int d\mathbf{r}_1' \psi_i^\dagger(\mathbf{r}_1) \psi_i(\mathbf{r}_2) \psi_{i'}^\dagger(\mathbf{r}_1') \psi_{i'}(\mathbf{r}_2')$$

$$\times \left[ \frac{i}{2\pi} \int d\omega e^{-i\omega t_1} W(\mathbf{r}, \mathbf{r}', \omega) \right]$$

$$\times \left[ (\Omega_S - \omega - (\epsilon_{\mathbf{r}_1} - \epsilon_{\mathbf{r}_2} + \overline{\epsilon})^{-1} \right]$$

(55)

where $W(\mathbf{r}, \mathbf{r}', \omega)$ is the screened Coulomb interaction given by equation 50.
The GW + BSE approach is thought to be one of the most reliable methods for the calculation of the optical absorption spectra and excited-state electronic structures. It has been used to calculate small molecules and bulk crystals. Unfortunately, its use for the larger systems is hindered by the significant computational cost.

Excellent agreement with experimental results was obtained for the optical absorption spectrum of bulk Si calculated using the GW + BSE approach [121]. Within GW + BSE, the lower energy peak originating from the excitonic binding effect was obtained for the first time. In contrast, previous LDA and TDLDA results were unable to predict this peak due to the inadequacy of the local approximation for the Coulomb interaction. Delerue et al. [123] have shown numerically that the Coulomb correction term cancels the polarization term in the self-energy of the quasiparticle eigenenergy. Consequently, the results of the GW + BSE are expected to be similar to the results of the CI where \( J \) is screened by the bulk dielectric function. In the TDLDA, where the Coulomb interaction is local, it is also screened by the bulk dielectric function, in line with the fact the LDA single-particle states used in equation 40 do not include the surface polarization term. The different cancellation schemes in TDLDA, CI, and GW + BSE (equations 40, 43, and 54) can be illustrated by comparing the calculated absorption spectra in these methods with the one obtained from single-particle energies. It was shown in Ref. [124] that the BSE absorption spectrum of small clusters of \( SiHm \) is red-shifted from the calculated single-particle spectrum, which is mostly due to negative surface polarization energies in the Coulomb interaction \( J \). On the other hand, it was shown in Ref. [103] that the TDLDA spectrum is blue-shifted from the single-particle LDA spectrum.

There is no surface polarization in \( J \) or single-particle energies then; thus, the exchange interaction dominates the spectrum shift. However, if total LDA energy differences for adding or removing an electron are used in equation 40, then the surface polarization must be considered [125] and Coulomb interaction cannot be calculated from equations 45 and 46. The above-discussed cancellations are only good for spherical quantum dots. For quantum rods, wires, and the nanostructures of other shapes, the GW + BSE-like quantum dots. For quantum rods, wires, and the nanostructures of other shapes, the GW + BSE-like quantum dots.

### 1.07.3.4 Quantum Monte Carlo Methods

Within the quantum Monte Carlo (QMC) method [126], the whole system is described by a many-body wave function and the many-body Schrödinger equation is solved using some of the Monte Carlo techniques such as variational Monte Carlo method (VMC) [127,128] or diffusion Monte Carlo method (DMC) [129,130].

In the VMC, the variational form of the many-body wave function \( \Psi(X) \) is assumed as a Slater determinant multiplied by a Jastrow term

\[
\Psi(X) = D^i (R) D^j (R) \exp \left[ \sum_{i=1}^{M} \chi_i (r_i) - \sum_{i<j}^{M} u(|r_i - r_j|) \right] \tag{56}
\]

where \( X = \{ r_i, s_i \} \) for \( i = 1, M \). The Slater determinant \( D \) is usually constructed from single-particle
LDA or Hartree–Fock wave functions, while parameterized forms are used to express $\chi$ and $n$. The total energy of the system is found by minimizing the expectation value of the many-body Hamiltonian $H$, given as

$$E = \langle \Psi | H | \Psi \rangle / \langle \Psi | \Psi \rangle$$

The last expression can be rewritten as

$$E = \frac{\int \Psi(X) H \Psi(X) dX}{\int |\Psi(X)|^2 dX} = \frac{\int \left[ \frac{H \Psi(X)}{\Psi(X)} \right] |\Psi(X)|^2 dX}{\int |\Psi(X)|^2 dX}$$

The last integral can be viewed as the average value of the quantity $H \Psi(X) / \Psi(X)$ distributed in multi-dimensional space with a probability $|\Psi(X)|^2$. It can be found using the Metropolis algorithm. In this algorithm, one simulates the path of the walker in the multidimensional space. The jump of the walker from $X$ to $X'$ is accepted if $\mu = |\Psi(X')|^2 / |\Psi(X)|^2 > 1$ and accepted with probability $\mu$ if $\mu < 1$. The average value of $H \Psi(X) / \Psi(X)$ along the path of the walker gives $E$ in equation 57.

In the DMC, one treats the many-body imaginary time Schrödinger equation as the classical diffusion equation [129,130]. In this method, the many-body wave function (not its square) corresponds to the equilibrium distribution of Monte Carlo walkers. However, for fermion system, antisymmetry is required for the many-body wave function. This causes a sign problem, which is usually approximately solved by a fixed nodal approximation where an auxiliary wave function is used to define the fixed nodal hypersurface for the DMC wave function. Usually, the VMC wave function of equation 56 is used as the auxiliary wave function.

With the use of pseudopotentials [131], both VMC and DMC methods have been used for systems up to a dozen atoms. Williamson et al. [132] showed that QMC methods can be used for exciton energies. This is done by replacing one single-particle valence band wave function with a conduction band wave function in the Slater determinant. The DMC with a nodal hypersurface defined by this new Slater determinant is performed then, and it fully takes into account the resulting correlation effects. This approach gives the Si band structure that agrees well with the experiments. QMC is one of the most reliable methods for small-system calculations.

The development of a linear scaling QMC method [133] extended its applicability from a dozen atoms to a few hundred atoms. Within the linear scaling QMC method, Slater determinants are represented on a basis of localized Wannier functions. This makes the Slater determinant sparse and therefore the calculation time is proportional to the size $N$, instead of $N^3$ in the old scheme. Consequently, this allows the QMC calculation of a few hundred atoms and makes possible the use of the QMC method for small quantum dots [134].

### 1.07.4 Application to Different Physical Effects: Some Examples

#### 1.07.4.1 Electron and Hole Wave Functions

The shape of the single-particle wave functions and their energies determine many physical properties of quantum dots. This section is, therefore, devoted to the analysis of electron and hole wave functions. The wave functions of the lowest four states in the conduction band and the top four states in the valence band of a pyramidal [119]-faceted InAs/GaAs quantum dot are presented in Figure 2. The results presented in Figure 2 were obtained using the EPM, including the effect of spin–orbit interaction.

The lowest conduction state is an $s$-like state, while the next two conduction states are $p$-like states oriented in the directions of base diagonals, with nodal planes perpendicular to these directions. These are followed by $d$-like states. Due to lateral dimension larger than the quantum dot height, none of the nodal planes is parallel to the pyramid base. There are, therefore, only two $p$-like states, in contrast to spherical quantum dots where there are three $p$-like states.

The two $p$-states are relatively close in energy and their splitting is caused by several effects. To discuss each of these effects, we first assume that the structure is unstrained.

1. In the simplest single-band effective mass model, these states are degenerate and can be split if the base of the pyramid acquires a shape different than the square. The same is the case for the four-band $k \cdot p$ model (i.e., eight-band model without spin–orbit effects). Addition of spin–orbit effects to four-band $k \cdot p$ splits these levels by a small (less than 1 meV) amount [77,135]. Atomistic methods predict the correct symmetry of the system and split the $p$-states, as well as $k \cdot p$ models with larger number of bands (14, for example).
2. When strain (without piezoelectricity) is included in the \( \mathbf{k} \cdot \mathbf{p} \) model within CM approach, it cannot cause the splitting, while the VFF model, due to its atomistic nature, splits the p-states.

3. Piezoelectricity added to any of the models also causes the splitting of the p-states.

The splitting of the p-states is therefore caused by the shape anisotropy, spin–orbit effect, atomistic (a)symmetry, strain, and piezoelectricity. It is amazing that a single quantity is determined by such a large number of effects. Unfortunately, in a given quantum dot, all these effects are present and cannot be probed separately.

The conduction band states are formed essentially of a single envelope function and therefore these can be classified as being s, p, and d-like. On the other hand, the band mixing of the valence states is much stronger and such a simple classification is not possible. The valence band functions actually have no nodal planes. (This becomes obvious from Figure 2, when the isosurface values are additionally reduced for valence band maximum (VBM-1) and VBM-2.) The approximation of using a single heavy hole band to describe the valence state, which is often used in quantum wells, is therefore not applicable to quantum dots due to stronger heavy and light hole mixing.

As the dot size is reduced, the valence band energies become lower and the conduction band energies higher. The bound states are then less confined and the effective energy gap increases. With the reduction in quantum dot dimensions, some bound states become mixed with wetting layer or continuum states and the number of truly bound states decreases.

### 1.07.4.2 Intraband Optical Processes in Embedded Quantum Dots

Most of the semiconductor optoelectronic devices utilize transitions between the conduction-band states and the valence band states. The operating wavelength of these devices is mainly determined...
Quantum Dots: Theory

by the band gap of the materials employed and is therefore limited to the near-infrared and visible part of the spectrum. However, if one wishes to access longer wavelengths, a different approach is necessary, that is, the transitions within the same band have to be used. These transitions are called intraband transitions. Intraband optical transitions in bulk are not allowed and therefore low-dimensional nanostructures have to be used. Therefore, in the past two decades, semiconductor nanostructures, such as quantum wells, wires, and dots, have been recognized as sources and detectors of electromagnetic radiation in the mid- and far-infrared region of the spectrum.

When the use of nanostructures as detectors is concerned, several limitations of quantum well infrared photodetectors (QWIPs) have motivated the development of quantum dot infrared photodetectors (QDIPs). The main origin of the undesirable dark current in QWIPs is thermal excitation (due to interaction with phonons) of carriers from the ground state to the continuum states. The discrete electronic spectrum of quantum dots as opposed to continuum spectrum of quantum wells significantly reduces the phase space for such processes and therefore reduces the dark current. Higher operating temperatures of QDIPs are therefore expected. Due to optical selection rules, QWIPs based on intersubband transitions in the conduction band interact only with radiation having the polarization vector in the growth direction. This is not the case in quantum dots since these are 3D objects where the corresponding selection rules are different.

For the QDIP applications, it is essential to understand the quantum dot absorption spectrum. The simplest model that is sufficient to qualitatively understand the quantum dot intraband absorption spectrum is the parabolic dot model, where the potential is assumed in a separable form $V(\mathbf{r}) = V_x(x, y) + V_z(z)$, where $V_x(x, y) = \frac{1}{2} m^* \omega^2 (x^2 + y^2)$ is the potential of a 2D harmonic oscillator, and $V_z(z)$ is the potential of a quantum well confining the electrons in the $z$-direction. The solutions are of the form $\Psi(\mathbf{r}) = b_{n_x}(x) b_{n_y}(y) \psi_{n_z}(z)$, where $b_{n_q}(q)$ is the wave function of a 1D harmonic oscillator, and $\psi_{n_z}(z)$ are solutions of the 1D Schrödinger equation with potential $V_z(z)$ (corresponding to energies $\varepsilon_{n_z}$). The eigenenergies are then of the form $E(n_x, n_y, n_z) = \hbar \omega (n_x + n_y + 1) + \varepsilon_{n_z}$. The factor $\hbar \omega$ corresponds to the transition energy from the ground to first excited state, and for modeling realistic quantum dots it should be set to $\hbar \omega \sim 40 - 70$ meV. Typical quantum dots are wide in the $xy$-plane (diameters of the order of 20 nm and more) and have very small height (of the order of 3–7 nm) in the $z$-direction; therefore, the effective potential well representing the $z$-direction confinement is narrow (see Figure 3). In a typical case, therefore, $\varepsilon_{n_z}$ is of the order of at least 100 meV.

The optical absorption matrix elements on the transitions between states are proportional to the matrix elements of coordinate operators; therefore, by calculating the latter, one obtains the following selection rules on the transitions between states:

- $\Delta n_x = \pm 1$, $\Delta n_y = 0$, $\Delta n_z = 0$, for $x$-polarized radiation;
- $\Delta n_y = \pm 1$, $\Delta n_x = 0$, $\Delta n_z = 0$, for $y$-polarized radiation; and
- $\Delta n_z = 0$, $\Delta n_y = 0$, for $z$-polarized radiation.

The transitions from the ground state are of primary importance for QDIPs. From the selection rules obtained, one concludes that only the transition to a pair of degenerate first excited states is allowed for in-plane polarized radiation, while in the case of $z$-polarized radiation, only the transitions to higher excited states are allowed, as demonstrated in Figure 3.

Although the model presented considers the quantum dot band structure in a very simplified manner, it is excellent for understanding the results of more realistic models. The strict selection rules from this model are then relaxed, and strictly forbidden transitions become weakly allowed. Nevertheless, qualitatively, the absorption spectrum retains the same features as in this model.

![Figure 3](attachment:figure3.png) Scheme of energy levels and allowed optical transitions in a parabolic quantum dot model with infinite potential barriers. Only the levels with $n_x + n_y \leq 2$ and $n_z \leq 1$ are shown.
The absorption spectrum obtained by the eight-band \( k \cdot p \) model for an InAs/GaAs quantum dot of conical shape with the diameter \( D = 25 \text{ nm} \) and height \( h = 7 \text{ nm} \) is presented in Figure 4. The dimensions were chosen to approximately match those reported for quantum dots in a QDIP structure in Ref. [136]. The optical absorption spectrum in the case of \( z \)-polarized radiation is shown in Figure 4(a). The experimental intraband photocurrent spectrum exhibits the main peak at 175 meV and a much smaller peak at 115 meV, in excellent agreement with the results obtained for \( z \)-polarized incident radiation where the corresponding peaks occur at 179 and 114 meV, respectively. The corresponding absorption spectrum for in-plane polarized incident radiation is presented in the inset of Figure 4(a). There is a single peak in the spectrum, which is due to the transition from the ground state to a pair of nearly degenerate first excited states (see Figure 4(b)).

The results presented and other similar calculations suggest that the in-plane polarized radiation causes nonnegligible transitions only between the ground and first excited states, these being located in the region 40–80 meV in the far-infrared. On the other hand, \( z \)-polarized radiation causes the transition in the \( \sim \)100–300 meV region in the mid-infrared. The best way to understand the origin of such behavior is through a simplified parabolic model presented. Such behavior can be altered only if the dot dimension in the \( z \)-direction becomes comparable to the in-plane dimensions.

1.07.4.3 Size Dependence of the Band Gap in Colloidal Quantum Dots

The size dependence of the band gap is the most prominent effect of quantum confinement in semiconductor nanostructures. The band gap increases as the nanostructure size decreases. Many of quantum dot applications rely on the size dependence of the optical properties. Therefore, studying the size dependence of the band gap is one of the most important topics in semiconductor nanocrystal research.

According to a simple effective-mass approximation model, the band-gap increase of spherical quantum dots from the bulk value is

\[
\Delta E_g = \frac{2\hbar^2 \pi^2}{m^* d^2} \quad (58)
\]

where \( d \) is the quantum dot diameter and

\[
\frac{1}{m^*} = \frac{1}{m_e^*} + \frac{1}{m_h^*} \quad (59)
\]

with \( m_e^* \) and \( m_h^* \) being the electron and hole effective masses.

The experiments usually measure the optical gap of a quantum dot. Therefore, in addition to the difference in single-particle energies, one has to include the interaction between created electron and hole, in order to calculate the optical gap. One simple approach to do this is to calculate the exciton energy by including the electron–hole interaction on top of the single-particle gap. This procedure ignores the electron–hole exchange interaction and possible correlation effects. However, in the strong confinement
regime, which is present in most colloidal nanocrystals and embedded quantum dots, these effects are very small. Under this approximation, the exciton energy can be expressed as

\[ E_{\text{ex}} = \varepsilon_e - \varepsilon_v - E_{\text{C}} \]

where \( \varepsilon_e \) and \( \varepsilon_v \) are the single-particle CBM and VBM energies, and \( E_{\text{C}} \) is the electron–hole Coulomb energy calculated as

\[ E_{\text{C}} = \int \int d\mathbf{r}_1 d\mathbf{r}_2 |\psi_e(\mathbf{r}_1)|^2 |\psi_h(\mathbf{r}_2)|^2 \varepsilon(\mathbf{r}_1 - \mathbf{r}_2)|\mathbf{r}_1 - \mathbf{r}_2| \]

where \( \psi_e(\mathbf{r}) \) and \( \psi_h(\mathbf{r}) \) are the electron and hole wave functions, and \( \varepsilon(\mathbf{r}_1 - \mathbf{r}_2) \) is a distance-dependent screening dielectric function, which can be modeled as described in Ref. 32.

The dependence of calculated optical gap on CdSe nanocrystal size is presented in Figure 5. A fit of the theoretical results to the dependence yields values quite different from the simple \( d^{-2} \) law predicted from the effective mass approach. In the case of CdSe, \( \alpha = 1.18 \). The \( \alpha \) parameter is material-dependent and its values for III–V and II–VI semiconductors typically fall in the range of 1.1–1.7.

1.07.4.4 Excitons

In the previous section, we have presented the exciton calculations based on a simple, but useful approach. For the calculation of excitons, the methods in Section 1.07.3 must be used in principle. The results of these methods are shown in Figure 6 for H-passivated Si quantum dots [138]. The DMC method and GW–BSE method produce almost the same band gap for the smallest quantum dots. The DMC result is about 1 eV above all the other results for somewhat larger quantum dots with the diameter up to 1.6 nm. It remains to be seen how accurate is this DMC result, for example, when compared with well-controlled experiments (perhaps for other material quantum dots like CdSe). The TDLDA method gives almost the same results as the LDA Kohn–Sham energy difference. This suggests that both the exchange and Coulomb interactions in the TDLDA results have a very small contribution. Besides TDLDA, TDDFT-B3LYP was used in Refs. [138,139]. The TDDFT-B3LYP band gap is below the DMC result, especially for relatively large quantum dots. However, in Ref. [139], it was shown that for small molecules, the TDDFT-B3LYP result agrees with the MR-MP2 quantum chemistry calculations. The TB and EPM results in Figure 6 can be considered as the lowest order results of the CI equation 43, where only the zero-order screened Coulomb interactions between the VBM and CBM states are taken into account. These agree well with each other. However, they are between the TDLDA and TDDFT-B3LYP results.

To summarize these results, the DMC result is above all the other methods for \( d = 1.5 \) nm Si quantum dots. The LDA and TDLDA have the lowest band gap, followed by the TB and EPM-limited CI results and the TDDFT-B3LYP results. For very small quantum dots, the DMC results agree well with the GW-BSE results.

![Figure 5](image-url) Comparison of the exciton energy shift from its bulk value of CdSe quantum dots (QDs) between experiment, charge patching method (with band-gap corrected pseudopotentials) (LDA + C), and SEPM calculations. Coulomb energies are considered in this calculation. Experimental data is from Ref. [137]. Reproduced with permission from Li J and Wang L-W (2005) Band-structure-corrected local density approximation study of semiconductor quantum dots and wires. Physical Review B 72: 125325.
1.07.4.5 Auger Effects

Auger effects play a crucial role in carrier dynamics in nanostructures when both types of carriers (electrons and holes) are present. They become important, in particular, in quantum dots that have discrete electronic levels, which implies that the competing phonon-assisted relaxation processes are strongly reduced. Different types of Auger processes are schematically illustrated in Figure 7.

According to Fermi’s golden rule, the formula for the Auger rate is given as

$$W_i = \frac{2\pi}{\hbar} \sum_{u} |\langle \bar{\nu} | \Delta H | \nu \rangle|^2 \delta(E_{\bar{\nu}} - E_i)$$

where $|\bar{\nu}\rangle$ and $|\nu\rangle$ are the initial and final Auger states, $E_i$ and $E_{\bar{\nu}}$ their energies, and $\Delta H$ is the Coulomb interaction. At a temperature $T \neq 0$, the Boltzmann average over the initial states has to be taken. It would seem at first sight that the discreteness of quantum dot energy levels and the requirement for energy conservation in the process would not allow for efficient Auger processes. However, other excitations, such as phonons, can be involved as well and help satisfy the energy conservation. Their effect can then be phenomenologically modeled by Lorentzian broadening of the delta function in Fermi’s golden rule expression as

$$\delta(E_{\bar{\nu}} - E_i) \rightarrow \frac{\Gamma}{2\pi} \frac{1}{(E_{\bar{\nu}} - E_i)^2 + (\Gamma/2)^2} \quad (63)$$

The most important step in the electron cooling process involves the transition of the electron from the $p$-level ($e_p$) to the ground $s$ electronic state ($e_s$). This process is mediated by a transition of the hole from $b_s$ to $b_p$. The calculated Auger lifetime for this process is shown in Figure 8. Its value is of the order of 0.1–0.5 ps, in agreement with experimental results [140]. This result suggests that Auger processes are sufficient to explain electron cooling in quantum dots, although other mechanisms are not necessarily ruled out.

The same process can take place in the presence of an electron and a hole that act only as spectators. It is very interesting that the electron lifetime increases by an order of magnitude in those cases, as demonstrated in Figure 8.
Figure 7  Illustration of different Auger processes: (a) electron thermalization; (b) thermalization with spectator; (c) bi-exciton recombination; (d) tri-exciton recombination; (e) negative trion recombination; (f) positive trion recombination. In each of the figures, the initial state in the process is shown in the left-hand side and the final one in the right-hand side. Reproduced with permission from Wang L-W, Califano M, Zunger A, and Franceschetti A (2003) Pseudopotential theory of Auger processes in CdSe quantum dots. Physical Review Letters 91: 056404.

Figure 8  Electron cooling. Auger lifetimes at $T = 300$ K calculated with EPM within the single-particle (SP) approximation (solid line), and with CI, both in the absence (dashed line) and in the presence (long-dashed line) of a spectator ground-state exciton. The initial states include all three electron p-states and both hole s-states, and the final states $\epsilon_p$ and 30 hole states $h_n$ with energy centered around $E_n = \epsilon_h - \epsilon_p + \epsilon_a$. Reproduced with permission from Wang L-W, Califano M, Zunger A, and Franceschetti A (2003) Pseudopotential theory of Auger processes in CdSe quantum dots. Physical Review Letters 91: 056404.
1.07.4.6 Electron–Phonon Interaction

The theory and results presented so far covered only the stationary electronic structure of quantum dots when atoms are in their equilibrium positions. However, at finite temperatures the vibrations of atoms around their equilibrium positions (phonons) create additional potential that perturbs otherwise stationary electronic states and causes transitions among them.

Phonons in quantum dots can be treated at various levels of approximations. The approximation that is often used for large quantum dots is that the phonons are the same as in bulk material. The strongest coupling between electrons and phonons in polar crystals is polar coupling to longitudinal optical (LO) phonons, while deformation potential coupling to longitudinal acoustic (LA) phonons might also sometimes be important.

In order to calculate the transition rates among different electronic states due to the interaction with LO phonons, it is tempting to apply Fermi’s Golden rule, which is a good approximation in quantum wells, for example, [141]. However, its direct application to quantum dots leads to the result that transition rates are zero unless two levels are separated by one LO phonon energy exactly [142]. Such an approach treats the electron and phonon systems separately with their interaction being only a perturbation. It is currently known that electrons and phonons in quantum dots form coupled entities – polarons. Polarons in self-assembled quantum dots have so far been evidenced experimentally by optical means in the intraband magneto-optical spectrum [143,144], magneto-photoluminescence spectrum [145], and Raman scattering [146], and it has been suggested theoretically that they could have transport signatures as well [147]. Polarons are usually evidenced by anticrossing of electron energy levels when these are gradually changed, such as, for example, by magnetic field. We illustrate this here by a numerical experiment where the energies of the pair of first excited states are shifted in opposite directions by the same amount $\Delta E$, which is varied. The polaron states were calculated by direct diagonalization of the electron–phonon interaction Hamiltonian. The polaron energy levels that contain a contribution from at least one of the electronic states larger than 10% are shown by circles in Figure 9. Anticrossing features in polaron spectrum are clearly visible.

![Figure 9](https://example.com/figure9.png)

**Figure 9** Dependence of the polaron energy levels obtained by direct diagonalization (circles) on the artificial shift $\Delta E$. The corresponding single-particle levels calculated using the eight-band $k \cdot p$ model are shown by full lines. Lens-shaped quantum dot with the diameter of 20 nm and the height of 5 nm is considered.

There is therefore a widespread thought that carrier relaxation in quantum dots should be treated by considering the carriers as polarons. The polaron lifetime is then determined by anharmonic decay of an LO phonon into two low-energy phonons [148,143,149,150,151]. It is thought that the physical process responsible for that decay process is the decay either to two LA phonons [152,148] or to one acoustic and one optical phonon [151]. Within such assumptions, the polaron lifetime is in the case of a two-level system given by [148]

$$W = \frac{\Gamma}{b} - \frac{\sqrt{2(R - X)}}{b}$$

where $R = \sqrt{X^2 + Y^2}$, $X = g^2 + (\Delta^2 - \Gamma^2)/4$, $Y = \Gamma \Delta/2$, $\Delta = E_i - E_f - \hbar \omega_{LO}$, $g$ is the electron–phonon coupling strength, $\Gamma/b$ the phonon decay rate, and $E_i$ and $E_f$ the energies of the single-particle states. Equation 64 has been used in several occasions to fit the experimental results on intraband carrier dynamics in quantum dots [149,153].

The approximation of bulk phonon modes certainly fails in small quantum dots. In that case, one should use the atomistic description of phonons. To calculate the phonon frequencies and displacements, one needs a force field that describes the vibrations of atoms around their equilibrium positions. VFF, for example, can be used for that purpose. To calculate the electron–phonon coupling, one needs to be able to calculate the change in single-particle Hamiltonian due to atomic displacements.
Any of the single-particle methods described in Section 1.07.2 can, in principle, be used for that purpose. However, if some of the empirical methods are used, one should be sure that the fitted parameters are appropriate for this purpose as well. Due to large number of atoms and consequently phonon modes, such calculations could be quite expensive nevertheless they are sometimes practiced. For example, Delerue et al. [154] calculated the phonon modes in a Si nanocrystal using a VFF model, and the coupling between the phonon modes and the transition electronic states explicitly using the Harrison's rule [56] for changes of TB parameters following the atomic displacements. Most recently, Chelikowsky et al. [155] calculated the phonons of Si quantum dots using direct DFT calculations.

1.07.5 Conclusions

We have given an overview of theoretical methods used for electronic structure calculations in quantum dots. We have emphasized the weaknesses and strengths of each of the methods. An interested reader can therefore choose the method of choice depending on the desired application, the degree of accuracy required, and the available computational resources.

For the treatment of single-particle states, the simplest effective mass method is excellent for pedagogical purposes to illustrate the effect of quantum confinement. It is often even used in research when one wishes to qualitatively take into account the effect of quantum confinement and the details of the electronic structure are not essential. The multiband $k \cdot p$ method gives a more quantitative description, especially for large quantum dots. It is widely used in modeling of optical and transport processes in optoelectronic devices. Atomistic methods give a very detailed description of quantum dot electronic structure and are clearly the best choice in research for understanding the new physical effects.

For the treatment of excitations in quantum dots, Section 1.07.3 gives an overview of the methods that can be applied in principle. For application of these methods to quantum dots, linear scaling of the method is an essential requirement. CI approach satisfies this but it is based on classical model derivations and physical intuitions. QMC also appears to be promising. However, the method is relatively new, when the calculations of excited states and large systems are concerned. A deeper understanding of the accuracy, that is, the quality of the variational form of the wave function or the nodal hypersurfaces, is required. Where the GW–BSE approach is concerned, it is a challenge to make it scalable to larger systems.

Acknowledgment

This work was supported by US Department of Energy BES/SC under contract number DE-AC02-05CH11231.

References


78. Tomic S (2006) Electronic structure of In_{x}Ga_{1-x}As_{y}N_{1-y}GaAs(N) quantum dots by ten-band k·p theory. Physical Review B 73: 125348.


