

# Electronic Structure of Clusters and Nanocrystals

James R. Chelikowsky\* and Yousef Saad†

\*Department of Chemical Engineering and Materials Science

†Department of Computer Science

Institute for the Theory of Advanced Materials in Information Technology,

Digital Technology Center

University of Minnesota

Minneapolis, Minnesota 55455 USA

\*Email: [jrc@msi.umn.edu](mailto:jrc@msi.umn.edu), †Email: [saad@cs.umn.edu](mailto:saad@cs.umn.edu),

February 1, 2004



# Contents

§	<b>Electronic Structure of Clusters and Nanocrystals</b>	<b>5</b>
§.1	Introduction . . . . .	5
§.2	Quantum descriptions of matter . . . . .	8
§.2.1	The Hartree approximation . . . . .	11
§.2.2	The Hartree-Fock approximation . . . . .	14
§.3	Density functional approaches . . . . .	18
§.3.1	Free electron gas . . . . .	18
§.3.2	Hartree-Fock exchange in a free electron gas . . . . .	21
§.3.3	Density functional theory . . . . .	23
§.3.4	Time-dependent density functional theory . . . . .	28
§.4	Pseudopotentials . . . . .	42
§.5	Solving the eigenvalue problem . . . . .	57
§.6	Properties of confined systems: clusters . . . . .	66
§.6.1	Structure . . . . .	67
§.6.2	Photoemission spectra . . . . .	72
§.6.3	Vibrational modes . . . . .	78

§.6.4	Polarizabilities . . . . .	85
§.6.5	Optical spectra . . . . .	88
§.7	Quantum confinement in nanocrystals . . . . .	94
§.7.1	The role of oxygen in silicon quantum dots . . . . .	113
§.7.2	Doping quantum dots . . . . .	124
§.8	Acknowledgments . . . . .	135

# Chapter §

## Electronic Structure of Clusters and Nanocrystals

### §.1 Introduction

Frequently one categorizes material properties at the macroscopic scale by terms such as *intensive* and *extensive*. An intensive property is not dependent on the sample size, or mass. For example, if one specifies the temperature of a sample as that of room temperature, the description is complete. One does not need to specify the size of the sample. Other intensive properties include pressure and density. In contrast, an extensive property does depend on the size of the sample. For example, the volume of a sample depends on the size. Two moles of a gas occupies twice the volume as one mole. Therefore, one often specifies volume in terms of the volume per mole, or the specific

volume. Other extensive properties include thermodynamics properties like the heat capacity, enthalpy, entropy and free energy of the system.

*Matter at the nanoscale is different.* Properties that are intensive at the macroscopic scale may not be intensive at the “nano-” or “subnano-” scale. In fact, such properties may be hard to define at very small length scales. Consider a small cluster of atoms, *e.g.*, a dozen silicon atoms. Such an ensemble contains so few atoms that it is difficult to define a property such as temperature or volume. While extensive properties such as the heat capacity or free energy of the system remain so in the sense of changing with the size of the system, such properties may no longer scale linearly with size. For example, the free energy of two dozen atoms of silicon may not be twice the free energy of one dozen atoms of silicon.

In this latter example, one can make some simple arguments to explain this behavior. Suppose we consider the scaling of the free energy for a spherical sample of matter whose radius is  $R$ . The volume energy term would scale as  $R^3$ ; the surface term as  $R^2$ . As  $R$  tends towards the nanoscale, the surface terms can become the dominant term in the free energy. This surface term does not scale linearly with the volume or mass of the sample. This different scaling between volume and surface terms is well known in nucleation theory. A manifestation of this effect is that particles must exceed a certain size before they are stable.

One definition of the nanoscale is the size at which deviations from intrinsic intensive and extensive properties at the macroscopic scale occur.

Consider an intensive electronic property such as the band gap of silicon. For specificity, let us consider a spherical sample of silicon whose radius is  $R$ . If  $R$  changes from 10 cm to 1 cm, the optical gap will remain unchanged, as expected for an intensive property. However, if one considers values of  $R$  changing from 10 nm to 1 nm, the optical gap will be strongly altered. In this size regime, the optical gap is no longer an intensive property; the properties of such systems are said to characterize nanoscale. One of the first manifestations of this effect was observed in porous silicon, which exhibits remarkable room temperature luminescence [1]. This is in strong contrast to crystalline silicon, which is optically inactive. It is widely accepted that the localization of optical excitations at the nanoscale results in the luminescence of porous silicon.

In this chapter, the cornerstones of theoretical methods for understanding the electronic and structural properties of matter at the nanoscale will be reviewed.

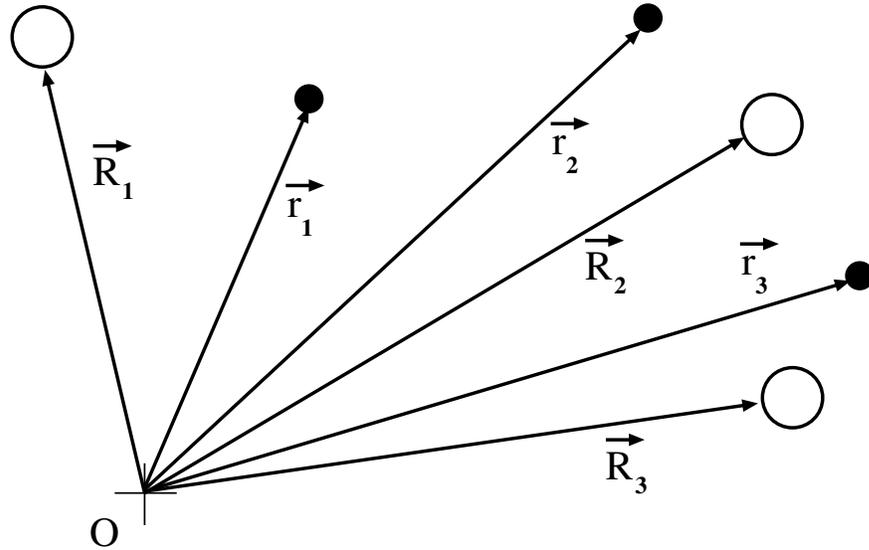


Figure §.1: Atomic and electronic coordinates. The electrons are illustrated by filled circles; the nuclei by open circles.

## §.2 Quantum descriptions of matter

Quantum mechanical laws that describe the behavior of matter at the nanoscale were discovered in the early part of the twentieth century. These laws mark one of the greatest scientific achievements of humankind. Using these laws, it is possible to predict the electronic properties of matter from the nanoscale to the macroscale, at least in principle.

While it is relatively easy to write down the Hamiltonian for interacting fermions, obtaining a solution to the problem that is sufficiently accurate to make predictions is another matter.

Consider  $N$  nucleons of charge  $Z_n$  at positions  $\{\vec{R}_n\}$  for  $n=1, \dots, N$  and  $I$  electrons at positions  $\{\vec{r}_i\}$  for  $i=1, \dots, M$ . This is shown schematically in

Figure §.1. The Hamiltonian for this system in its simplest form can be written as

$$\begin{aligned} \mathcal{H}(\vec{R}_1, \vec{R}_2, \vec{R}_3, \dots; \vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) = & \sum_{n=1}^N \frac{-\hbar^2 \nabla_n^2}{2\mathcal{M}_n} + \frac{1}{2} \sum_{n,m=1, n \neq m}^N \frac{Z_n Z_m e^2}{|\vec{R}_n - \vec{R}_m|} \\ & + \sum_{i=1}^M \frac{-\hbar^2 \nabla_i^2}{2m} - \sum_{n=1}^N \sum_{i=1}^M \frac{Z_n e^2}{|\vec{R}_n - \vec{r}_i|} + \frac{1}{2} \sum_{i,j=1, i \neq j}^M \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \quad (\S.1) \end{aligned}$$

$\mathcal{M}_n$  is the mass of the nucleon,  $\hbar$  is Planck's constant divide by  $2\pi$ ,  $m$  is the mass of the electron. This expression omits some terms such as those involving relativistic interactions, but captures the essential features for nanoscale matter.

Using the Hamiltonian in Eq. §.1, the quantum mechanical equation known as the Schrödinger equation for the electronic structure of the system can be written as:

$$\begin{aligned} \mathcal{H}(\vec{R}_1, \vec{R}_2, \vec{R}_3, \dots; \vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) \Psi(\vec{R}_1, \vec{R}_2, \vec{R}_3, \dots; \vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) \\ = E \Psi(\vec{R}_1, \vec{R}_2, \vec{R}_3, \dots; \vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) \quad (\S.2) \end{aligned}$$

where  $E$  is the total electronic energy of the system and  $\Psi$  is the many body wavefunction.

Soon after the discovery of the Schrödinger equation, it was recognized that this equation provided the means of solving for the electronic and nuclear degrees of freedom. Using the variational principle which states that an

approximate wave function will always have a less favorable energy than the true ground state energy, one had an equation and a method to test the solution. One can estimate the energy from

$$E = \frac{\int \Psi^* \mathcal{H} \Psi \, d^3 R_1 \, d^3 R_2 \, d^3 R_3 \dots \, d^3 r_1 \, d^3 r_2 \, d^3 r_3 \dots}{\int \Psi^* \Psi \, d^3 R_1 \, d^3 R_2 \, d^3 R_3 \dots \, d^3 r_1 \, d^3 r_2 \, d^3 r_3 \dots} \quad (\S.3)$$

However, a solution of Eq. §.2 for anything more complex than a few particles becomes problematic even with the most powerful computers. Obtaining an approximate solution for systems with many atoms is difficult, but considerable progress has been made since the advent of reliable digit computers.

A number of highly successful approximations have been made to solve for the both the ground state and excited state energies. For the most part, these approximations used are to remove as many “irrelevant” degrees of freedom from the system as possible. One common approximation is to separate the nuclear and electronic degrees of freedom. Since the nuclei are considerably more massive than the electrons, it can be assumed that the electrons will respond “instantaneously” to the nuclear coordinates. This approximation is called the *Born-Oppenheimer* or adiabatic approximation. It allows one to treat the nuclear coordinates as classical parameters. For most condensed matter systems, this assumption is highly accurate [2, 3].

### §.2.1 The Hartree approximation

Another common approximation is to construct a specific form for the many body wavefunction. If one can obtain an accurate wave function, then via the variational principle an accurate estimate for the energy will emerge. The most difficult part of this exercise is to use physical intuition to define a trial wave function close to the true wave function.

One can utilize some simple limiting cases to illustrate the construction of many body wave functions. Suppose one considers a solution for *non-interacting electrons*, *i.e.*, in Eq. §.1 the last term in the Hamiltonian is ignored. In this limit, it is possible to write the many body wave function as a sum of independent Hamiltonians. Using the adiabatic approximation, the *electronic* part of the Hamiltonian becomes:

$$\mathcal{H}_{el}(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) = \sum_{i=1}^M \frac{-\hbar^2 \nabla_i^2}{2m} - \sum_{n=1}^N \sum_{i=1}^M \frac{Z_n e^2}{|\vec{R}_n - \vec{r}_i|} \quad (\S.4)$$

Let us define a nuclear potential,  $V_N$ , which the  $i$ th electron sees as

$$V_N(\vec{r}_i) = - \sum_{n=1}^N \frac{Z_n e^2}{|\vec{R}_n - \vec{r}_i|} \quad (\S.5)$$

One can now rewrite our simplified Schrödinger equation as

$$\mathcal{H}_{el}(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) \psi(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) = \sum_{i=1}^M H^i \psi(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) \quad (\S.6)$$

where the Hamiltonian for the  $i$ th electron is

$$H^i = \frac{-\hbar^2 \nabla_i^2}{2m} + V_N(\vec{r}_i) \quad (\S.7)$$

For this simple Hamiltonian, let us write the many body wave function as

$$\psi(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) = \phi_1(\vec{r}_1) \phi_2(\vec{r}_2) \phi_3(\vec{r}_3) \dots \quad (\S.8)$$

The  $\phi_i(\vec{r})$  orbitals can be determined from a “one-electron” Hamiltonian

$$\begin{aligned} H^i \phi_i(\vec{r}) &= \left( \frac{-\hbar^2 \nabla^2}{2m} + V_N(\vec{r}) \right) \phi(\vec{r}) \\ &= E_i \phi_i(\vec{r}) \end{aligned} \quad (\S.9)$$

The index  $i$  for the orbital  $\phi_i(\vec{r})$  can be taken to include the spin of the electron plus any other relevant quantum numbers. This type of Schrödinger equation can be easily solved for fairly complex condensed matter systems. The many body wave function in Eq. §.8 is known as the *Hartree* wave function. If one uses this form of the wave function as an approximation to solve the Hamiltonian *including* the electron-electron interactions, this is known as the *Hartree Approximation*. By ignoring the electron-electron terms, the Hartree approximation treats the electrons moving independently in the nuclear potential. The total energy of the system is given by the sum of the eigenvalues,  $E_i$ .

To obtain a realistic Hamiltonian, the electron-electron interactions must

be reinstated in Eq. §.6:

$$\mathcal{H}_{el}(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) \psi(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) = \sum_{i=1}^M \left( H^i + \frac{1}{2} \sum_{j=1, j \neq i}^M \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \right) \psi(\vec{r}_1, \vec{r}_2, \vec{r}_3 \dots) \quad (\S.10)$$

In this case, the individual orbitals,  $\phi_i(\vec{r})$ , can be determined by using the minimizing the total energy as per (Eq. §.3) with the constraint that the wave function be normalized. This minimization procedure results in the following *Hartree* equation:

$$H^i \phi_i(\vec{r}) = \left( \frac{-\hbar^2 \nabla^2}{2m} + V_N(\vec{r}) + \sum_{j=1, j \neq i}^M \int \frac{e^2 |\phi_j(\vec{r}')|^2}{|\vec{r} - \vec{r}'|} d^3 r' \right) \phi_i(\vec{r}) = E_i \phi_i(\vec{r}) \quad (\S.11)$$

Using the orbitals,  $\phi(\vec{r})$ , from a solution of Eq. §.11, the Hartree many body wave function can be constructed and the total energy determined from (Eq. §.3).

The Hartree approximation is useful as an illustrative tool, but it is not a very accurate approximation. A significant failing of the Hartree wave function is that it does not reflect the anti-symmetric nature of the electrons as required by the Pauli principle [4]. Moreover, the Hartree equation is difficult to solve. The Hamiltonian is orbitally dependent because the summation in Eq. §.11 does not include the  $i$ th orbital. This means that if there are  $M$  electrons, then  $M$  Hamiltonians must be considered and Eq. §.11 solved for each orbital.

### §.2.2 The Hartree-Fock approximation

It is possible to write down a many body wavefunction that reflects the antisymmetric nature of the wave function. In this approach, the spin coordinate of each electron is explicitly treated. The coordinates of an electron may be specified by  $\vec{r}_i s_i$  where  $s_i$  represents the spin coordinate. Starting with one-electron orbitals,  $\phi_i(\vec{r}s)$ , the following form can be invoked:

$$\Psi(\vec{r}_1 s_1, \vec{r}_2 s_2, \vec{r}_3 s_3, \dots) = \begin{vmatrix} \phi_1(\vec{r}_1 s_1) & \phi_1(\vec{r}_2 s_2) & \dots & \dots & \phi_1(\vec{r}_M s_M) \\ \phi_2(\vec{r}_1 s_1) & \phi_2(\vec{r}_2 s_2) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \phi_M(\vec{r}_1 s_1) & \dots & \dots & \dots & \phi_M(\vec{r}_M s_M) \end{vmatrix} \quad (\S.12)$$

This form of the wave function is called a *Slater determinant*. It reflects the proper symmetry of the wave function and the Pauli principle. If two electrons occupy the same orbit, two rows of the determinant will be identical and the many body wave function will have zero amplitude. Likewise, the determinant will vanish if two electrons occupy the same point in generalized space (*i.e.*,  $\vec{r}_i s_i = \vec{r}_j s_j$ ) as two columns of the determinant will be identical. If two particles are exchanged, this corresponds to a sign change in the determinant. The Slater determinant is a convenient representation, but it is an *ansatz*. It is probably the simplest many body wave function that incorporates the required symmetry properties for fermions, or particles with non-integer spins.

If one uses a Slater determinant to evaluate the total electronic energy and maintains wave function normalization, the orbitals can be obtained from the following *Hartree-Fock* equations:

$$H^i \phi_i(\vec{r}) = \left( \frac{-\hbar^2 \nabla^2}{2m} + V_N(\vec{r}) + \sum_{j=1}^M \int \frac{e^2 |\phi_j(\vec{r}')|^2}{|\vec{r} - \vec{r}'|} d^3 r' \right) \phi_i(\vec{r}) - \sum_{j=1}^M \int \frac{e^2}{|\vec{r} - \vec{r}'|} \phi_j^*(\vec{r}') \phi_i(\vec{r}') d^3 r' \delta_{s_i, s_j} \phi_j(\vec{r}) = E_i \phi_i(\vec{r}) \quad (\S.13)$$

It is customary to simplify this expression by defining an electronic charge density,  $\rho$ :

$$\rho(\vec{r}) = \sum_{j=1}^M |\phi_j(\vec{r})|^2 \quad (\S.14)$$

and an orbitally dependent *exchange-charge density*,  $\rho_i^{HF}$  for the  $i$ th orbital:

$$\rho_i^{HF}(\vec{r}, \vec{r}') = \sum_{j=1}^M \frac{\phi_j^*(\vec{r}') \phi_i(\vec{r}') \phi_i^*(\vec{r}) \phi_j(\vec{r})}{\phi_i^*(\vec{r}) \phi_i(\vec{r})} \delta_{s_i, s_j} \quad (\S.15)$$

This “density” involves a spin dependent factor which couples only states  $(i, j)$  with the same spin coordinates  $(s_i, s_j)$ .

With these charge densities defined, it is possible to define corresponding potentials. The *Coulomb* or *Hartree* potential,  $V_H$ , is defined by

$$V_H(\vec{r}) = \int \rho(\vec{r}') \frac{e^2}{|\vec{r} - \vec{r}'|} d^3 r' \quad (\S.16)$$

and an *exchange* potential can be defined by

$$V_x^i(\vec{r}) = - \int \rho_i^{HF}(\vec{r}, \vec{r}') \frac{e^2}{|\vec{r} - \vec{r}'|} d^3r' \quad (\S.17)$$

This combination results in the following Hartree-Fock equation:

$$\left( \frac{-\hbar^2 \nabla^2}{2m} + V_N(\vec{r}) + V_H(\vec{r}) + V_x^i(\vec{r}) \right) \phi_i(\vec{r}) = E_i \phi_i(\vec{r}) \quad (\S.18)$$

Once the Hartree-Fock orbitals have been obtained, the total Hartree-Fock electronic energy of the system,  $E_{HF}$ , can be obtained from

$$E_{HF} = \sum_i^M E_i - \frac{1}{2} \int \rho(\vec{r}) V_H(\vec{r}) d^3r - \frac{1}{2} \sum_i^M \int \phi_i^*(\vec{r}) \phi_i(\vec{r}) V_x^i(\vec{r}) d^3r \quad (\S.19)$$

$E_{HF}$  is not a sum of the Hartree-Fock orbital energies,  $E_i$ . The factor of  $\frac{1}{2}$  in the electron-electron terms arises because the electron-electron interactions have been double counted in the Coulomb and exchange potentials. The Hartree-Fock Schrödinger equation is only slightly more complex than the Hartree equation. Again, the equations are difficult to solve because the exchange potential is orbitally dependent.

There is one notable difference in the Hartree-Fock summations compared to the Hartree summation. The Hartree-Fock sums include the  $i = j$  terms in Eq. §.13. This difference arises because the exchange term corresponding to  $i = j$  cancels an equivalent term in the Coulomb summation. The  $i = j$  term in both the Coulomb and exchange term is interpreted as a “self-screening” of

the electron. Without a cancellation between Coulomb and exchange terms a “self-energy” contribution to the total energy would occur. Approximate forms of the exchange potential often do not have this property. The total energy then contains a self-energy contribution which one needs to remove to obtain a correct Hartree-Fock energy.

The Hartree-Fock equation is an approximate solution to the true ground state many body wave functions. Terms not included in the Hartree-Fock energy are referred to as *correlation* contributions. One definition for the correlation energy,  $E_{corr}$  is to write it as the difference between the exact total energy of the system,  $E_{exact}$  and the Hartree-Fock energies:  $E_{corr} = E_{exact} - E_{HF}$ . Correlation energies may be included by considering Slater determinants composed of orbitals which represent excited state contributions. This method of including unoccupied orbitals in the many body wave function is referred to as *configuration interactions* or “CI” [5].

Applying Hartree-Fock wave functions to systems with many atoms is not routine. The resulting Hartree-Fock equations are often too complex to be solved for extended systems, except in special cases. The number of electronic degrees of freedom grows rapidly with the number atoms often prohibiting an accurate solution, or even one’s ability to store the resulting wave function. As such, it has been argued that a “wave function” approach to systems with many atoms does not offer a satisfactory approach to the electronic structure problem. An alternate approaches is based on density functional theory.

### §.3 Density functional approaches

Descriptions of quantum states based on a knowledge of the electronic charge density (Eq. §.14) have existed since the 1920's. For example, the Thomas-Fermi description, based on a knowledge of  $\rho(\vec{r})$ , was on the first attempts at quantitative theory for the electronic structure of atoms [6–8]. However, most treatments of density functional theory begin by considering a free electron gas of uniform charge density. The justification for this starting point comes from the observation that simple metals like aluminum and sodium have properties that appear to resemble those of a free electron gas. A “free electron” model cannot be applied to systems with localized electrons such as highly covalent materials like carbon or highly ionic materials like sodium chloride.

#### §.3.1 Free electron gas

Perhaps the simplest descriptions of a condensed matter system is to imagine non-interacting electrons contained within a box of volume,  $\Omega$ . The Schrödinger equation for this system is similar to Eq. §.9 with the potential set to zero.

$$\frac{-\hbar^2 \nabla^2}{2m} \phi(\vec{r}) = E \phi(\vec{r}) \quad (\S.20)$$

Ignoring spin for the moment, the solution of Eq. §.20 is

$$\phi(\vec{r}) = \frac{1}{\sqrt{\Omega}} \exp(i\vec{k} \cdot \vec{r}) \quad (\S.21)$$

$\vec{k}$  is called a *wave vector*. The energy is given by  $E(k) = \hbar^2 k^2 / 2m$  and the charge density by  $\rho = 1/\Omega$ .

A key issue in describing systems with a very large number of atoms is to account properly for the number of states. In the limit of systems corresponding to crystalline states, the eigenvalues are closely spaced and essentially “infinite” in number. For example, if one has a mole of atoms, then one can expect to have  $\sim 10^{23}$  occupied states. In such systems, the number of states per unit energy is a natural measure to describe the energy distribution of states.

It is easy to do this with *periodic boundary conditions*. Suppose one considers a one dimensional specimen of length  $L$ . In this case the wave functions obey the rule:  $\phi(x + L) = \phi(x)$  as  $x + L$  corresponds in all physical properties to  $x$ . For a free electron wave function, this requirement can be expressed as  $\exp(ik(x + L)) = \exp(ikx)$  or as  $\exp(ikL) = 1$  or  $k = 2\pi n/L$  where  $n$  is an integer.

Periodic boundary conditions force  $k$  to be a discrete variable with allowed values occurring at intervals of  $2\pi/L$ . For very large systems, one can describe the system as continuous in the limit of  $L \rightarrow \infty$ . Electron states can be defined by a *density of states* defined as follows:

$$\begin{aligned} D(E) &= \lim_{\Delta E \rightarrow 0} \frac{N(E + \Delta E) - N(E)}{\Delta E} \\ &= \frac{dN}{dE} \end{aligned} \tag{§.22}$$

where  $N(E)$  is the number of states whose energy resides below  $E$ . For the one-dimensional case,  $N(k) = 2k/(2\pi/L)$  (the factor of two coming from spin) and  $dN/dE = (dN/dk) \cdot (dk/dE)$ . Using  $E(k) = \hbar^2 k^2/2m$ , we have  $k = \sqrt{2mE}/\hbar$  and  $dk/dE = \frac{1}{2}\sqrt{2m/E}/\hbar$ . This results for this one-dimensional density of states as

$$D(E) = \frac{L}{2\pi\hbar} \sqrt{2m/E} \quad (\S.23)$$

The density of states for a one-dimensional system diverges as  $E \rightarrow 0$ . This divergence of  $D(E)$  is not a serious issue as the integral of the density of states remains finite. In three dimensions, it is straightforward to show

$$D(E) = \frac{\Omega}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \sqrt{E} \quad (\S.24)$$

The singularity is removed, although a discontinuity in the derivative exists as  $E \rightarrow 0$ .

One can determine the total number of electrons in the system by integrating the density of states up to the highest occupied energy level. The energy of the highest occupied states is called the *Fermi level* or *Fermi Energy*,  $E_f$ :

$$N = \frac{\Omega}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2} \int_0^{E_f} \sqrt{E} dE \quad (\S.25)$$

and

$$E_f = \frac{\hbar^2}{2m} \left(\frac{3\pi^2 N}{\Omega}\right)^{2/3} \quad (\S.26)$$

By defining a *Fermi* wave vector as  $k_f = (3\pi^2 n_{el})^{1/3}$  where  $n_{el}$  is the electron

density,  $n_{el} = N/\Omega$ , of the system, one can write:

$$E_f = \frac{\hbar^2 k_f^2}{2m} \quad (\S.27)$$

It should be noted that typical values for  $E_f$  for simple metals like sodium or potassium are on the order of several  $eV$ 's. If one defines a temperature,  $T_f$ , where  $T_f = E_f/k_b$  and  $k_b$  is the Boltzmann constant, typical values for  $T_f$  might be  $10^4 K$ . At ambient temperatures one can often neglect the role of temperature in determining the Fermi Energy.

### §.3.2 Hartree-Fock exchange in a free electron gas

For a free electron gas, it is possible to evaluate the Hartree-Fock exchange energy directly [9,10]. The Slater determinant is constructed using free electrons orbitals. Each orbital is labeled by a  $\vec{k}$  and a spin index. The Coulomb potential for an infinite free electron gas diverges, but this divergence can be removed by imposing a compensating uniform positive charge. The resulting Hartree-Fock eigenvalues can be written as

$$E_k = \frac{\hbar^2 k^2}{2m} - \frac{1}{\Omega} \sum_{\vec{k}' < k_f} \frac{4\pi e^2}{|\vec{k} - \vec{k}'|^2} \quad (\S.28)$$

where the summation is over occupied  $\vec{k}$ -states. It is possible to evaluate the summation by transposing the summation to an integration. This transposition is often done for solid state systems as the state density is so high that

the system can be treated as a continuum:

$$\frac{1}{\Omega} \sum_{k, k' < k_f} \frac{4\pi e^2}{|\vec{k} - \vec{k}'|^2} = \frac{1}{(2\pi)^3} \int_{k' < k_f} \frac{4\pi e^2}{|\vec{k} - \vec{k}'|^2} d^3k \quad (\S.29)$$

This integral can be solved analytically [10]. The resulting eigenvalues are given by

$$E_k = \frac{\hbar^2 k^2}{2m} - \frac{e^2 k_f}{\pi} \left( 1 + \frac{1 - (k/k_f)^2}{2(k/k_f)} \ln \left| \frac{k + k_f}{k - k_f} \right| \right) \quad (\S.30)$$

Using the above expression and Eq. §.19, the total electron energy,  $E_{HF}^{FEG}$ , for a free electron gas within the Hartree-Fock approximation is given by:

$$E_{HF}^{FEG} = 2 \sum_{k < k_f} \frac{\hbar^2 k^2}{2m} - \frac{e^2 k_f}{\pi} \sum_{k < k_f} \left( 1 + \frac{1 - (k/k_f)^2}{2(k/k_f)} \ln \left| \frac{k + k_f}{k - k_f} \right| \right) \quad (\S.31)$$

The factor of 2 in the first term comes from spin. In the exchange term, there is no extra factor of 2 because one can subtract off a “double counting term” (see Eq. §.19). The summations can be executed as per Eq. §.29 to yield:

$$E_{HF}^{FEG}/N = \frac{3}{5} E_f - \frac{3e^2}{4\pi} k_f \quad (\S.32)$$

The first term corresponds to the average energy per electron in a free electron gas. The second term corresponds to the exchange energy per electron. The exchange energy is attractive and scales with the cube root of the average density. This form provides a clue as to what form the exchange energy might take in an interacting electron gas or non-uniform electron gas.

Slater was one of the first to propose that one replace  $V_x^i$  in Eq. §.18 by a term that depends only on the cube root of the charge density [11–13]. By analogy to Eq. §.32, he suggested that  $V_x^i$  be replaced by

$$V_x^{Slater}[\rho(\vec{r})] = -\frac{3e^2}{2\pi}(3\pi\rho(\vec{r}))^{1/3} \quad (\S.33)$$

This expression is not orbitally dependent. As such, a solution of the Hartree-Fock equation (Eq. §.18) using  $V_x^{Slater}$  is much easier to implement. Although Slater exchange was not rigorously justified for non-uniform electron gases, it has been quite successful in replicating the essential features of atomic and molecular systems as determined by Hartree-Fock calculations [11–13].

### §.3.3 Density functional theory

In a number of classic papers Hohenberg, Kohn and Sham established a theoretical basis for justifying the replacement of the many body wavefunction by one-electron orbitals [8, 14, 15]. In particular, they proposed that the charge density play a central role in describing the electronic structure of matter. A key aspect of their work is the *local density approximation* or LDA. Within this approximation, one can express the exchange energy as

$$E_x[\rho(\vec{r})] = \int \rho(\vec{r})\mathcal{E}_x[\rho(\vec{r})] d^3r \quad (\S.34)$$

where  $\mathcal{E}_x[\rho]$  is the exchange energy per particle of uniform gas at a density of  $\rho$ . Within this framework, the exchange potential in Eq. §.18 is replaced by a potential determined from the functional derivative of  $E_x[\rho]$ :

$$V_x[\rho] = \frac{\delta E_x[\rho]}{\delta \rho} \quad (\S.35)$$

One serious issue is the determination of the exchange energy per particle,  $\mathcal{E}_x$ , or the corresponding exchange potential,  $V_x$ . The exact expression for either of these quantities is unknown, save for special cases. If one assumes the exchange energy is given by Eq. §.32, *i.e.*, the Hartree-Fock expression for the exchange energy of the free electron gas. Then one can write:

$$E_x[\rho] = -\frac{3e^2}{4\pi}(3\pi^2)^{1/3} \int [\rho(\vec{r})]^{4/3} d^3r \quad (\S.36)$$

and taking the functional derivative, one obtains:

$$V_x[\rho] = -\frac{e^2}{\pi}(3\pi^2\rho(\vec{r}))^{1/3} \quad (\S.37)$$

Comparing this to the form chosen by Slater, this form, known as *Kohn-Sham* exchange,  $V_x^{KS}$  differs by a factor of 2/3, *i.e.*  $V_x^{KS} = 2V_x^{Slater}/3$ . For a number of years, some controversy existed as to whether the Kohn-Sham or Slater exchange was more accurate for realistic systems [8]. Slater suggested that a parameter be introduced which would allow one to vary the exchange between the Slater and Kohn-Sham values [13] The parameter,  $\alpha$ , was often

placed in front of the Slater exchange:  $V_{x\alpha} = \alpha V_x^{Slater}$ .  $\alpha$  was often chosen to replicate some known feature of an exact Hartree-Fock calculation, such as the total energy of an atom or ion. Acceptable values of  $\alpha$  were viewed to range from  $\alpha = 2/3$  to  $\alpha = 1$ . Slater's so called " $X_\alpha$ " method was very successful in describing molecular systems [13]. Notable drawbacks of the  $X_\alpha$  method center on its *ad hoc* nature through the  $\alpha$  parameter and the omission of an explicit treatment of correlation energies.

In contemporary theories,  $\alpha$ , is taken to be  $2/3$ , and correlation energies are explicitly included in the energy functionals [8]. Numerical studies have been performed on uniform electron gases resulting in local density expressions of the form:  $V_{xc}[\rho(\vec{r})] = V_x[\rho(\vec{r})] + V_c[\rho(\vec{r})]$  where  $V_c$  represents contributions to the total energy beyond the Hartree-Fock limit [16]. It is also possible to describe the role of spin explicitly by considering the charge density for up and down spins:  $\rho = \rho_\uparrow + \rho_\downarrow$ . This approximation is called the *local spin density approximation* [8].

The *Kohn-Sham equation* [15] for the electronic structure of matter is given by

$$\left( \frac{-\hbar^2 \nabla^2}{2m} + V_N(\vec{r}) + V_H(\vec{r}) + V_{xc}[\rho(\vec{r})] \right) \phi_i(\vec{r}) = E_i \phi_i(\vec{r}) \quad (\S.38)$$

This equation is usually solved "self-consistently." An approximate charge is assumed to estimate the exchange-correlation potential and this charge is used to determine the Hartree potential from Eq. §.16. These approximate

potentials are inserted in the Kohn-Sham equation and the total charge density determined as in Eq. §.14. The “output” charge density is used to construct new exchange-correlation and Hartree potentials. The process is repeated until the input and output charge densities or potentials are identical to within some prescribed tolerance.

Once a solution of the Kohn-Sham equation is obtained, the total energy can be computed from

$$E_{KS} = \sum_i^M E_i - 1/2 \int \rho(\vec{r}) V_H(\vec{r}) d^3r + \int \rho(\vec{r}) (E_{xc}[\rho(\vec{r})] - V_{xc}[\rho(\vec{r})]) d^3r \quad (\S.39)$$

where  $E_{xc}$  is a generalization of Eq. §.34, *i.e.*, the correlation energy density is included. The electronic energy as determined from  $E_{KS}$  must be added to the ion-ion interactions to obtain the structural energies. This is a straightforward calculation for confined systems. For extended systems such as crystals, the calculations can be done using Madelung summation techniques [17].

Owing to its ease of implementation and overall accuracy, the local density approximation is a popular choice for describing the electronic structure of matter. It is relatively easy to implement and surprisingly accurate. Recent developments have included so-called gradient corrections to the local density approximation. In this approach, the exchange-correlation energy depends on the local density the gradient of the density. This approach is called the generalized gradient approximation or GGA [18].

When first proposed, density functional theory was not widely accepted within the chemistry community. The theory is not “rigorous” in the sense that it is not clear how to improve the estimates for the ground state energies. For wave function based methods, one can include more Slater determinants as in a configuration interaction approach. As the wave functions improve via the variational theorem, the energy is lowered. The Kohn-Sham equation is variational, but owing to its approximate form it need not approach the true ground state energy. This is not a problem provided that one is interested in *relative* energies and any inherent density functional errors cancel in the difference. For example, if the Kohn-Sham energy of an atom is 10% too high and the corresponding energy of the atom in a crystal is also 10% too high, the cohesive energies which involve the difference of the two energies can be better than the nominal 10% error of the absolute energies. An outstanding fundamental issue of using density functional theory is obtaining an *ab initio* estimate of the cancellation errors.

In some sense, density functional theory is an *a posteriori* theory. Given the transference of the exchange-correlation energies from an electron gas, it is not surprising that errors would arise in its implementation to highly non-uniform electron gas systems as found in realistic systems. However, the degree of error cancellations is rarely known *a priori*. The reliability of density functional theory has been established by numerous calculations for a wide variety of condensed matter systems. For example, the cohesive energies, compressibility, structural parameters and vibrational spectra of el-

emental solids have been calculated within the density functional theory [19]. The accuracy of the method is best for systems in which the cancellation of errors is expected to be complete. Since cohesive energies involve the difference in energies between atoms in solids and in free space, error cancellations are expected to be significant. This is reflected in the fact that historically cohesive energies have presented greater challenges for density functional theory: the errors between theory and experiment are typically  $\sim 5-10\%$ , depending on the nature of the density functional. In contrast, vibrational frequencies which involve small structural changes within a given crystalline environment are easily reproduced to within 1-2 %.

### §.3.4 Time-dependent density functional theory

One of the most significant limitations of “conventional” density functional formalism is its inability to deal with electronic excitations. Within time-independent, or static, density functional theory, a quantum mechanical system is described through the ground state electronic charge density. While this approach can be accurate for the ground state of a many-electron system, unoccupied electronic states cannot be identified as those belonging to electronic, or quasi-particle, excitations [20, 21]. The inability of “conventional” density functional theory to describe excitations severely restricts its range of applications as many important physical properties such as optical absorption and emission are associated with excited states.

Explicit calculations for excited states present enormous challenges for

theoretical methods. Accurate calculations for excitation energies and absorption spectra typically require computationally intensive techniques, such as the configuration interaction method [22, 23], quantum Monte Carlo simulations [24–26] or the Green’s function methods [27–29]. While these methods describe electronic excitations properly, they are usually limited to very small systems because of high computational demands. An alternative approach is to consider methods based on time dependent density functional theory such as those using the time dependent local density approximation (TDLDA) [20, 21, 30–38].

The TDLDA technique can be viewed as a natural extension of the ground state density-functional LDA formalism, designed to include the proper representation of excited states. TDLDA excitation energies of a many-electron system are usually computed from conventional, time independent Kohn-Sham transition energies and wave functions. Compared to other theoretical methods for excited states, the TDLDA technique requires considerably less computational effort. Despite its relative simplicity, the TDLDA method incorporates screening and relevant correlation effects for electronic excitations. [20, 21, 30, 31] In this sense, TDLDA represents a fully *ab initio* formalism for excited states.

As in the case of time-independent density functional theory, the time-dependent formalism reduces the many-electron problem to a set of self-

consistent single particle equations, [39, 40]

$$\left(-\frac{\nabla^2}{2} + v_{\text{eff}}[\rho(\mathbf{r}, t)]\right) \psi_i(\mathbf{r}, t) = i \frac{\partial}{\partial t} \psi_i(\mathbf{r}, t). \quad (\S.40)$$

In this case, the single particle wave functions,  $\psi_i(\mathbf{r}, t)$ , and the effective potential,  $v_{\text{eff}}[\rho(\mathbf{r}, t)]$ , explicitly depend on time. The effective potential is given by

$$v_{\text{eff}}[\rho(\mathbf{r}, t)] = \sum_a v_{\text{ion}}(\mathbf{r} - \mathbf{r}_a) + v_H[\rho(\mathbf{r}, t)] + v_{\text{xc}}[\rho(\mathbf{r}, t)]. \quad (\S.41)$$

The three terms on the right side of Eq. (§.41) describe the external ionic potential, the potential Hartree, and the exchange-correlation potential, respectively. In the adiabatic approximation, which is local in *time*, the exchange-correlation potential and its first derivative can be expressed in terms of the *time-independent* exchange-correlation energy,  $E_{\text{xc}}[\rho]$ ,

$$v_{\text{xc}}[\rho(\mathbf{r}, t)] \cong \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})}, \quad \frac{\delta v_{\text{xc}}[\rho(\mathbf{r}, t)]}{\delta \rho(\mathbf{r}', t')} \cong \delta(t - t') \frac{\delta^2 E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r}) \delta \rho(\mathbf{r}')}. \quad (\S.42)$$

The LDA makes a separate local approximation, *i.e.*, within the LDA, the exchange-correlation energy density is local in *space*.

While the local density approximation in time dependent density function theory has proven itself for molecules, clusters and small quantum dots and small clusters, several questions remain as areas of active research. The application of TDLDA to large, extended systems remains problematic. It

is widely accepted that TDLDA as outlined here will approach the LDA results for extended systems and, consequently suffer the flaws of LDA such as exhibiting band gaps much smaller than experiment [41–43].

Most implementations of time dependent density functional theory are based on the local density approximation or the generalized gradient approximation [44, 45]. However, these approximations are known to have the wrong asymptotic behavior, *e.g.* the potential does not scale as  $1/r$  for large distances. It is widely believed that more accurate TDLDA methods will necessitate other forms of the density functional. Examples of such an approach are the asymptotically corrected local density approximations introduced by Casida and Salahub [46], and by Leeuwen and Baerends [47]. These potentials have recently been investigated using the current formalism [48].

The linear response formalism within TDDFT provides a theoretical basis for the TDLDA method. In this section, we illustrate how TDLDA excitation energies and oscillator strengths are derived from single-electron Kohn-Sham eigenvalues and eigen wave functions. A comprehensive analysis of time-dependent density functional response theory can be found elsewhere [20, 21, 30, 31]. The notation in the work by Casida [30, 31] is implemented within this section.

The response of the Kohn-Sham density matrix within TDDFT is obtained by introducing a time-dependent perturbation  $\delta v_{\text{appl}}(\mathbf{r}, t)$ . Due to the self-consistent nature of the Kohn-Sham Hamiltonian, the *effective* pertur-

bation includes the response of the self-consistent field,  $\delta v_{\text{SCF}}[\rho(\mathbf{r}, t)]$ ,

$$\delta v_{\text{eff}}[\rho(\mathbf{r}, t)] = \delta v_{\text{appl}}(\mathbf{r}, t) + \delta v_{\text{SCF}}[\rho(\mathbf{r}, t)], \quad (\S.43)$$

where the self-consistent field is given by the last two terms in Eq. (§.41):

$$v_{\text{SCF}}[\rho(\mathbf{r}, t)] = \int \frac{\rho_v(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{\text{xc}}[\rho(\mathbf{r}, t)]. \quad (\S.44)$$

With the frequency domain, the response of the Kohn-Sham density matrix,  $\delta \mathbf{P}(\omega)$ , to the perturbation can be derived using a generalized susceptibility,  $\chi(\omega)$ . For quasi-independent Kohn-Sham particles, the sum-over-states representation of the generalized susceptibility is given by

$$\chi_{ij\sigma,kl\tau}(\omega) = \delta_{i,k} \delta_{j,l} \delta_{\sigma,\tau} \frac{\lambda_{lk\tau}}{\omega - \omega_{lk\tau}}, \quad (\S.45)$$

where  $\lambda_{lk\tau} = n_{l\tau} - n_{k\tau}$  is the difference between the occupation numbers, and  $\omega_{lk\tau} = \epsilon_{k\tau} - \epsilon_{l\tau}$  is the difference between the eigenvalues of the  $l$ -th and  $k$ -th single particle states. The susceptibility in Eq. (§.45) is expressed in the basis of the unperturbed Kohn-Sham orbitals  $\{\psi_{i\sigma}\}$  and the indices  $i, j$ , and  $\sigma$  ( $k, l$ , and  $\tau$ ) refer to space and spin wave components, respectively. The linear response of the density matrix is

$$\delta P_{ij\sigma}(\omega) = \sum_{kl\tau} \chi_{ij\sigma,kl\tau}(\omega) \delta v_{kl\tau}^{\text{eff}}(\omega) = \frac{\lambda_{ji\sigma}}{\omega - \omega_{ji\sigma}} \left[ \delta v_{ij\sigma}^{\text{appl}}(\omega) + \delta v_{ij\sigma}^{\text{SCF}}(\omega) \right]. \quad (\S.46)$$

Equation (§.46) is, however, complicated by the fact that  $\delta v_{\text{SCF}}(\omega)$  depends on the response of the density matrix,

$$\delta v_{ij\sigma}^{\text{SCF}}(\omega) = \sum_{kl\tau} K_{ij\sigma,kl\tau} \delta P_{kl\tau}(\omega), \quad (\S.47)$$

where the coupling matrix  $\mathbf{K}$  describes the response of the self-consistent field to changes in the charge density. Within the adiabatic approximation, this matrix is frequency-independent. The analytical expression for the adiabatic coupling matrix,  $K_{ij\sigma,kl\tau} = \partial v_{ij\sigma}^{\text{SCF}} / \partial P_{kl\tau}$ , can be derived from Eq. (§.44) by making use of the functional chain rule,

$$K_{ij\sigma,kl\tau} = \iint \psi_{i\sigma}^*(\mathbf{r}) \psi_{j\sigma}(\mathbf{r}) \left( \frac{1}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta^2 E_{\text{xc}}[\rho]}{\delta \rho_{\sigma}(\mathbf{r}) \delta \rho_{\tau}(\mathbf{r}')} \right) \psi_{k\tau}(\mathbf{r}') \psi_{l\tau}^*(\mathbf{r}') d\mathbf{r} d\mathbf{r}'. \quad (\S.48)$$

The functional derivative in Eq. (§.48) is evaluated with respect to the *unperturbed* charge densities. By using the coupling matrix, Eq. (§.46) can be rewritten as

$$\sum_{kl\tau}^{\lambda_{kl\tau} \neq 0} \left[ \delta_{i,k} \delta_{j,l} \delta_{\sigma,\tau} \frac{\omega - \omega_{lk\tau}}{\lambda_{lk\tau}} - K_{ij\sigma,kl\tau} \right] \delta P_{kl\tau}(\omega) = \delta v_{ij\sigma}^{\text{appl}}(\omega). \quad (\S.49)$$

Since the summation in Eq. (§.49) is performed over all occupied and unoccupied orbitals, it contains both the particle-hole and hole-particle contributions. These contributions can be written as two separate equations: the

particle-hole part of  $v_{\text{appl}}(\omega)$  is given by

$$\sum_{kl\tau}^{\lambda_{kl\tau}>0} \left[ \delta_{i,k} \delta_{j,l} \delta_{\sigma,\tau} \frac{\omega - \omega_{lk\tau}}{\lambda_{lk\tau}} - K_{ij\sigma,kl\tau} \right] \delta P_{kl\tau}(\omega) - \sum_{kl\tau}^{\lambda_{kl\tau}>0} K_{ij\sigma,kl\tau} \delta P_{kl\tau}(\omega) = \delta v_{ij\sigma}^{\text{appl}}(\omega) \quad (\S.50)$$

and the hole-particle part of  $v_{\text{appl}}(\omega)$  is

$$\sum_{kl\tau}^{\lambda_{kl\tau}>0} \left[ \delta_{i,k} \delta_{j,l} \delta_{\sigma,\tau} \frac{\omega - \omega_{kl\tau}}{\lambda_{kl\tau}} - K_{ji\sigma,kl\tau} \right] \delta P_{kl\tau}(\omega) - \sum_{kl\tau}^{\lambda_{kl\tau}>0} K_{ji\sigma,kl\tau} \delta P_{kl\tau}(\omega) = \delta v_{ji\sigma}^{\text{appl}}(\omega). \quad (\S.51)$$

Combining Eqs. (§.50) and (§.51), one can separate the real and imaginary parts of the density matrix response,  $\delta\mathbf{P}(\omega)$ . If the basis functions  $\{\psi_{i\sigma}\}$  in Eq. (§.48) are real, the coupling matrix  $\mathbf{K}$  is also real and symmetric with respect to the interchange of space indices  $i \leftrightarrow j$  and  $k \leftrightarrow l$ . Since  $\delta\mathbf{P}(\omega)$  is hermitian (*i.e.*  $\delta P_{ji\sigma} = \delta P_{ij\sigma}^*$ ), the real part of  $\delta\mathbf{P}(\omega)$  for a real perturbation  $v_{\text{appl}}(\omega)$  is given by

$$\sum_{kl\tau}^{\lambda_{kl\tau}>0} \left[ \frac{\delta_{i,k} \delta_{j,l} \delta_{\sigma,\tau}}{\lambda_{kl\tau} \omega_{kl\tau}} (\omega^2 - \omega_{kl\tau}^2) - 2K_{ij\sigma,kl\tau} \right] \Re(\delta P_{kl\tau})(\omega) = \delta v_{ij\sigma}^{\text{appl}}(\omega), \quad (\S.52)$$

where  $\Re(\delta P_{ij\sigma})(\omega)$  denotes the Fourier transform of the *real* part of  $\delta P_{ij\sigma}(t)$ .

Equation (§.52) can be used to obtain the density functional expression for the dynamic polarizability. This is accomplished by introducing a perturbation  $\delta \hat{v}_{\text{appl}}(t) = \hat{\gamma} \mathcal{E}_\gamma(t)$ , where  $\mathcal{E}_\gamma$  is an external electric field applied along the  $\gamma$ -axis,  $\gamma = \{x, y, z\}$ . The linear response of the dipole moment,  $\delta\mu(\omega)$ ,

is expressed through of the real part of  $\delta\mathbf{P}(\omega)$  as

$$\delta\mu_\beta(\omega) = -2 \sum_{ij\sigma}^{\lambda_{ij\sigma}>0} \beta_{ji\sigma} \Re(\delta P_{ij\sigma})(\omega) \quad , \quad \beta = \{x, y, z\}. \quad (\S.53)$$

The components of the dynamic polarizability tensor are given by

$$\alpha_{\beta\gamma}(\omega) = \frac{\delta\mu_\beta(\omega)}{\mathcal{E}_\gamma(\omega)} = -\frac{2}{\mathcal{E}_\gamma(\omega)} \sum_{ij\sigma}^{\lambda_{ij\sigma}>0} \beta_{ji\sigma} \Re(\delta P_{ij\sigma})(\omega) \quad , \quad \beta, \gamma = \{x, y, z\}, \quad (\S.54)$$

Solving Eq. (§.52) with respect to  $\Re(\delta P_{ij\sigma})(\omega)$  and substituting the result into Eq. (§.54), one obtains the following matrix equation for the polarizability components:

$$\alpha_{\beta\gamma}(\omega) = 2 \hat{\beta} \mathbf{R}^{1/2} [\mathbf{Q} - \omega^2 \mathbf{1}]^{-1} \mathbf{R}^{1/2} \hat{\gamma} \quad , \quad \beta, \gamma = \{x, y, z\}, \quad (\S.55)$$

where the matrices  $\mathbf{R}$  and  $\mathbf{Q}$  are given by

$$R_{ij\sigma,kl\tau} = \delta_{i,k} \delta_{j,l} \delta_{\sigma,\tau} \lambda_{kl\tau} \omega_{kl\tau}, \quad (\S.56)$$

$$Q_{ij\sigma,kl\tau} = \delta_{i,k} \delta_{j,l} \delta_{\sigma,\tau} \omega_{kl\tau}^2 + 2 \sqrt{\lambda_{ij\sigma} \omega_{ij\sigma}} K_{ij\sigma,kl\tau} \sqrt{\lambda_{kl\tau} \omega_{kl\tau}}. \quad (\S.57)$$

The TDLDA expressions for excitation energies and oscillator strengths can be derived by comparing Eq. (§.55) with the general sum-over-states formula for the average dynamic polarizability,  $\langle\alpha(\omega)\rangle = \text{tr}(\alpha_{\beta\gamma}(\omega))/3 = \sum_I f_I / (\Omega_I^2 - \omega^2)$ . The true excitation energies,  $\Omega_I$ , which correspond to the *poles* of the dynamic polarizability, are obtained from the solution of the

eigenvalue problem,

$$\mathbf{Q} \mathbf{F}_I = \Omega_I^2 \mathbf{F}_I. \quad (\S.58)$$

The oscillator strengths,  $f_I$ , which correspond to the *residues* of the dynamic polarizability, are given by

$$f_I = \frac{2}{3} \sum_{\beta=1}^3 |\mathbf{B}_\beta^T \mathbf{R}^{1/2} \mathbf{F}_I|^2, \quad (\S.59)$$

where  $\mathbf{F}_I$  are the eigenvectors of Eq. (§.58),  $(B_\beta)_{ij} = \int \psi_i^* r_\beta \psi_j d\mathbf{r}$ , and  $\{r_1, r_2, r_3\} = \{x, y, z\}$ .

The adiabatic TDLDA calculations for optical spectra require only the knowledge of the time-independent single-electron Kohn-Sham transition energies and wave functions. The most computationally demanding part in such calculations is the evaluation of the coupling matrix given by Eq. (§.48). This equation can be split into two parts:  $\mathbf{K} = \mathbf{K}^{(I)} + \mathbf{K}^{(II)}$ . The first term represents a double integral over  $1/|\mathbf{r} - \mathbf{r}'|$ . Instead of performing the costly double integration by direct summation, we calculate this term by solving the Poisson equation within the boundary domain. The conjugate-gradient method is employed to solve

$$\nabla^2 \Phi_{ij\sigma}(\mathbf{r}) = -4\pi \psi_{i\sigma}(\mathbf{r}) \psi_{j\sigma}(\mathbf{r}). \quad (\S.60)$$

The first term in Eq. (§.48) is calculated as

$$K_{ij\sigma,kl\tau}^{(I)} = \int \Phi_{ij\sigma}(\mathbf{r})\psi_{k\tau}(\mathbf{r})\psi_{l\tau}(\mathbf{r})d\mathbf{r}. \quad (\S.61)$$

The Poisson equation method provides a considerable speed-up as compared to the direct summation. The second term in Eq. (§.48) represents a double integral over the functional derivative of the exchange-correlation energy,  $\delta^2 E_{xc}[\rho] / \delta\rho_\sigma(\mathbf{r})\delta\rho_\tau(\mathbf{r}')$ . Within the *local* approximation of the exchange-correlation potential this term is reduced to a single integral,

$$K_{ij\sigma,kl\tau}^{(II)} = \int \psi_{i\sigma}(\mathbf{r})\psi_{j\sigma}(\mathbf{r})\frac{\delta^2 E_{xc}[\rho]}{\delta\rho_\sigma(\mathbf{r})\delta\rho_\tau(\mathbf{r})}\psi_{k\tau}(\mathbf{r})\psi_{l\tau}(\mathbf{r})d\mathbf{r}, \quad (\S.62)$$

where the LDA exchange-correlation energy,  $E_{xc}[\rho]$ , in analogy to Eq. (§.34).

Eq. (§.62) requires the evaluation of the second derivatives for the LDA exchange-correlation energy with respect to spin-up and spin-down charge densities. The LDA *exchange* energy per particle is normally approximated by that of the homogeneous electron gas, [49]

$$\epsilon_x[\rho_\sigma(\mathbf{r})] = -\frac{3}{4\pi} (6\pi^2\rho_\sigma(\mathbf{r}))^{1/3}, \quad \sigma = \{\uparrow, \downarrow\}. \quad (\S.63)$$

The first derivative of the total exchange energy determines the LDA exchange potential,

$$\frac{\delta E_x[\rho]}{\delta\rho_\sigma} = v_x[\rho_\sigma] = -\frac{1}{\pi} (6\pi^2\rho_\sigma)^{1/3}, \quad \sigma = \{\uparrow, \downarrow\}. \quad (\S.64)$$

The second derivatives are:

$$\frac{\delta^2 E_x[\rho]}{\delta\rho_\uparrow\delta\rho_\uparrow} = -\left(\frac{2}{9\pi}\right)^{1/3} \rho_\uparrow^{-2/3}, \quad \frac{\delta^2 E_x[\rho]}{\delta\rho_\uparrow\delta\rho_\downarrow} = 0. \quad (\S.65)$$

A parameterized form of Ceperley-Alder functional [16,50,51] can be used for the LDA *correlation* energy. This functional is based on two different analytical expressions for  $r_s < 1$  and  $r_s \geq 1$ , where  $r_s = (3/4\pi\rho)^{1/3}$  is the local Seitz radius and  $\rho = \rho_\uparrow + \rho_\downarrow$ . One can adjust the parametrization for  $r_s < 1$  to guarantee a continuous second derivative of the correlation energy. The adjusted interpolation formula for the correlation energy per particle is given by [52]

$$\epsilon_c^{U,P} = \begin{cases} A \ln r_s + B + C r_s \ln r_s + D r_s + X r_s^2 \ln r_s, & r_s < 1 \\ \gamma / (1 + \beta_1 \sqrt{r_s} + \beta_2 r_s), & r_s \geq 1, \end{cases} \quad (\S.66)$$

where two separate sets of coefficients are used for the *polarized spin* ( $P$ ) and *unpolarized spin* ( $U$ ) cases. The numerical values of all fitting parameters appearing in Eq. (§.66) can be found in Ref. [53]. The adjusted interpolation formula for the correlation energy is continuous up to its second derivative, while the original Perdew-Zunger parametrization is not [51].

Equations (§.63) – (§.66) describe only the cases of the completely polarized and unpolarized spin. For intermediate spin polarizations, the corre-

lation energy can be obtained with a simple interpolation formula,

$$\epsilon_c = \epsilon_c^U + \xi(\rho)[\epsilon_c^P - \epsilon_c^U], \quad (\S.67)$$

where

$$\xi(\rho) = \frac{1}{1 - 2^{-1/3}} \left( x_\uparrow^{4/3} + x_\downarrow^{4/3} - 2^{-1/3} \right) \quad ; \quad x_\uparrow = \frac{\rho_\uparrow}{\rho}, \quad x_\downarrow = \frac{\rho_\downarrow}{\rho}, \quad (\S.68)$$

The expression for the second derivative of the correlation energy in case of an arbitrary spin polarization can be written as

$$\begin{aligned} \frac{\delta^2 E_c[\rho]}{\delta \rho_\sigma \delta \rho_\tau} &= \frac{\delta^2 E_c^U}{\delta \rho^2} + \xi(\rho) \left( \frac{\delta^2 E_c^P}{\delta \rho^2} - \frac{\delta^2 E_c^U}{\delta \rho^2} \right) + \\ &\left( \frac{\partial \xi(\rho)}{\partial \rho_\sigma} + \frac{\partial \xi(\rho)}{\partial \rho_\tau} \right) \left( \frac{\delta E_c^P}{\delta \rho} - \frac{\delta E_c^U}{\delta \rho} \right) + \frac{\partial^2 \xi(\rho)}{\partial \rho_\sigma \partial \rho_\tau} \rho (\epsilon_c^P - \epsilon_c^U) \quad , \quad \sigma, \tau = \{\uparrow, \downarrow\}, \end{aligned} \quad (\S.69)$$

where the spin polarization function,  $\xi(\rho)$ , and its derivatives are given by

$$\frac{\partial \xi(\rho)}{\partial \rho_\uparrow} = \frac{4}{3\rho(1 - 2^{-1/3})} \left( x_\uparrow^{1/3} - x_\uparrow^{4/3} - x_\downarrow^{4/3} \right), \quad (\S.70)$$

$$\frac{\partial^2 \xi(\rho)}{\partial \rho_\uparrow \partial \rho_\uparrow} = \frac{4}{9\rho^2(1 - 2^{-1/3})} \left( x_\uparrow^{-2/3} - 8x_\uparrow^{1/3} + 7(x_\uparrow^{4/3} + x_\downarrow^{4/3}) \right), \quad (\S.71)$$

$$\frac{\partial^2 \xi(\rho)}{\partial \rho_\uparrow \partial \rho_\downarrow} = \frac{4}{9\rho^2(1 - 2^{-1/3})} \left( 7(x_\uparrow^{4/3} + x_\downarrow^{4/3}) - 4(x_\uparrow^{1/3} + x_\downarrow^{1/3}) \right). \quad (\S.72)$$

The TDLDA formalism presented in previous sections can be further simplified for systems with the unpolarized spin. In this case, the spin-up

and spin-down charge densities are equal,  $\rho_\uparrow = \rho_\downarrow$ , and Eqs. (§.68), (§.70) – (§.72) yield

$$\begin{aligned} \xi(\rho) = 0, \quad \frac{\partial^2 \xi(\rho)}{\partial \rho_\uparrow \partial \rho_\uparrow} &= \frac{4}{9\rho^2(2^{1/3} - 1)}, \\ \frac{\partial \xi(\rho)}{\partial \rho_\uparrow} = 0, \quad \frac{\partial^2 \xi(\rho)}{\partial \rho_\uparrow \partial \rho_\downarrow} &= -\frac{4}{9\rho^2(2^{1/3} - 1)}. \end{aligned} \quad (\S.73)$$

Since the coordinate parts of spin-up and spin-down Kohn-Sham wave functions for systems with the unpolarized spin are identical,  $\psi_{i\uparrow} = \psi_{i\downarrow}$ , it follows that  $Q_{ij\uparrow,kl\uparrow} = Q_{ij\downarrow,kl\downarrow}$  and  $Q_{ij\uparrow,kl\downarrow} = Q_{ij\downarrow,kl\uparrow}$ . This allows us to separate “singlet” and “triplet” transitions by representing Eq. (§.58) in the basis set of  $\{\mathbf{F}_+, \mathbf{F}_-\}$ , chosen as

$$F_{ij}^{\{+,-\}} = \frac{1}{\sqrt{2}} (F_{ij\uparrow} \pm F_{ij\downarrow}). \quad (\S.74)$$

In this basis, the matrix  $\mathbf{Q}$  becomes

$$Q_{ij,kl}^{\{+,-\}} = \delta_{i,k} \delta_{j,l} \omega_{kl}^2 + 2\sqrt{\lambda_{ij}\omega_{ij}} K_{ij,kl}^{\{+,-\}} \sqrt{\lambda_{kl}\omega_{kl}}, \quad (\S.75)$$

where  $K_{ij,kl}^{\{+,-\}} = K_{ij\uparrow,kl\uparrow} \pm K_{ij\uparrow,kl\downarrow}$ . The components of  $\mathbf{K}^{\{+,-\}}$  in their explicit

form are given by

$$\begin{aligned}
 K_{ij,kl}^+ &= 2 \iint \frac{\psi_i(\mathbf{r})\psi_j(\mathbf{r})\psi_k(\mathbf{r}')\psi_l(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \\
 &+ 2 \int \psi_i(\mathbf{r})\psi_j(\mathbf{r}) \left( \frac{\delta^2 E_c^U}{\delta \rho^2} - \frac{1}{(9\pi)^{1/3} \rho^{2/3}} \right) \psi_k(\mathbf{r})\psi_l(\mathbf{r}) d\mathbf{r}, \quad (\S.76)
 \end{aligned}$$

$$K_{ij,kl}^- = 2 \int \psi_i(\mathbf{r})\psi_j(\mathbf{r}) \left( \frac{4(\epsilon_c^P - \epsilon_c^U)}{9(2^{1/3} - 1)} - \frac{1}{(9\pi)^{1/3} \rho^{2/3}} \right) \psi_k(\mathbf{r})\psi_l(\mathbf{r}) d\mathbf{r}. \quad (\S.77)$$

For most practical applications, only “singlet” transitions represented by the  $\mathbf{F}_+$  basis vectors are of interest. Triplet transitions described by the  $\mathbf{F}_-$  vectors have zero dipole oscillator strength and do not contribute to optical absorption. By solving Eq. (§.58) for the  $\mathbf{F}_+$  vectors only, one can reduce the dimension of the eigenvalue problem by a factor of two. Eqs. (§.75) – (§.77), however, can only be applied to systems with the unpolarized spin. In case of an arbitrary spin polarization, the general form of the matrix  $\mathbf{Q}$  presented by Eq. (§.57) with the coupling matrix given by Eq. (§.48) and the functional derivatives given by Eqs. (§.64) through (§.72) must be used.

Other than the adiabatic local density approximation, no other approximations have been made. The exact solution of the matrix equation (§.58) incorporates all relevant correlations among single-particle transitions.

The frequency-domain approach presented here lends itself naturally to a massively parallel solution within the real-space grid. A practical solution begins with solving the Kohn-Sham equation (Eq. §.38). Once the Kohn-

Sham equation is solved, the wave functions and eigenvalues can be used to set up the TDLDA equation for the excited states. Forming the coupling matrix (Eq. §.48) is the most computationally intensive step. Fortunately, each  $ij\sigma$  combination defines a row in the matrix as  $\psi_{ij\sigma}$  is fixed within each row of the matrix as such it is easy to parallelize a solution to Eq. §.60. Once this step is accomplished, one may evaluate the integrals of Eqs. §.61 and §.62 as simple sums, for each element within a matrix row. The evaluation of each matrix row is completely independent of the evaluation of another row, leading to embarrassingly simple parallelization, with no need for communication between processors working on different matrix rows. Once, the matrix is diagonalized using any off-the-shelf diagonalization approach, e.g., QR factorization. Further implementation details can be found elsewhere [54].

## §.4 Pseudopotentials

The pseudopotential model of a solid has led the way in providing a workable model for computing the electronic properties of materials [55]. For example, it is now possible to predict accurately the properties of complex systems such as quantum dots or semiconductor liquids with hundreds, if not thousands of atoms.

The pseudopotential model treats matter as a sea of valence electrons moving in a background of ion cores (Fig. §.2). The cores are composed of nuclei and inert inner electrons. Within this model many of the complexities

of an all-electron calculation are avoided. A group IV solid such as C with 6 electrons is treated in a similar fashion to Pb with 82 electrons since both elements have 4 valence electrons.

Pseudopotential calculations center on the accuracy of the valence electron wave function in the spatial region away from the core, *i.e.*, within the “chemically active” bonding region. The smoothly varying pseudo wave function is taken to be identical to the appropriate all-electron wave function in the bonding regions. A similar construction was introduced by Fermi in 1934 [56] to account for the shift in the wave functions of high lying states of alkali atoms subject to perturbations from foreign atoms. In this remarkable paper, Fermi introduced the conceptual basis for both the pseudopotential and the scattering length. In Fermi’s analysis, he noted that it was not necessary to know the details of the scattering potential. Any number of potentials which reproduced the phase shifts of interest would yield similar scattering events.

A significant advance in the construction of pseudopotentials occurred with the development of density functional theory [8,14,15]. Within density functional theory, the many body problem is mapped on to a one-electron Hamiltonian. The effects of exchange and correlation are subsumed into a one electron potential that depends only on the charge density. This procedure allows for a great simplification of the electronic structure problem. Without this approach, most electronic structure methods would not be feasible for systems of more than a few atoms. The chief limitation of density functional

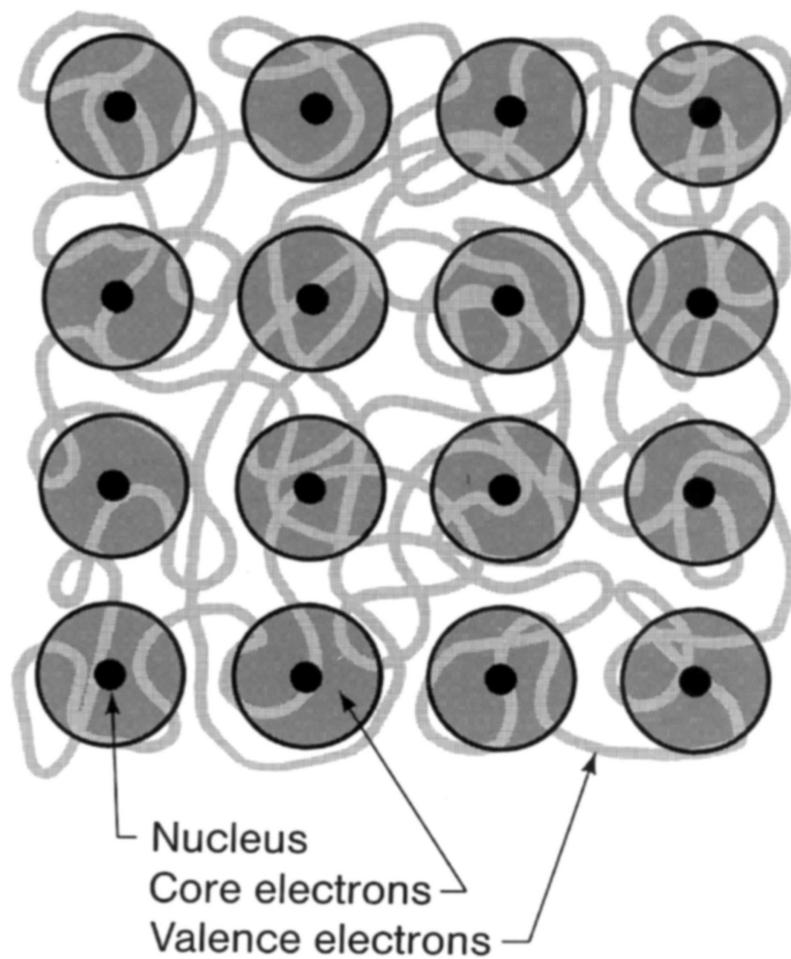


Figure §.2: Standard pseudopotential model of a solid. The ion cores composed of the nuclei and tightly bound core electrons are treated as chemically inert. The pseudopotential model describes only the outer, chemically active, valence electrons.

methods is that they are appropriate only for the ground state structure and cannot be used to describe excited states without other approximations.

In this section, the procedure for constructing an *ab initio* pseudopotential within density functional theory will be illustrated. Using the approach Kohn and Sham [15], one can write down a Hamiltonian corresponding to a one-electron Schrödinger equation. It is not difficult to solve the Kohn-Sham equation (Eq. §.87) for an atom as the atomic charge density is taken to be spherically symmetric. The problem reduces to solving a one-dimensional equation. The Hartree and exchange-correlation potentials can be iterated to form a self-consistent field. Usually the process is so quick that it can be done on desktop or laptop computer in a matter of seconds. In three dimensions, as for a complex atomic cluster, the problem is highly non-trivial. One major difficulty is the range of length scales involved. For example, in the case of a multi-electron atom, the most tightly bound, core electrons can be confined to within  $\sim 0.01$  Å whereas the outer valence electron may extend over  $\sim 1-5$  Å. In addition, the nodal structure of the atomic wave functions are difficult to replicate with a simple basis, especially the wave function cusp at the origin where the Coulomb potential diverges. The pseudopotential approximation eliminates this problem and is quite efficacious when combined with density functional theory. However, it should be noted that the pseudopotential approximation is not dependent on the density functional theory. Pseudopotentials can be created without resort to density functional theory, *e.g.*, pseudopotentials can be created within Hartree-Fock theory.

To illustrate the construction of an *ab initio* pseudopotential, let us consider a sodium atom. Extensions to more complex atoms is straightforward. The starting point of any pseudopotential construction is to solve for the atom within density functional theory, *i.e.*, the electronic structure problem for the Na atom is solved, including the core and valence electrons. In particular, one extracts the eigenvalue,  $\epsilon_{3s}$ , and the corresponding wave function,  $\psi_{3s}(r)$  for the single valence electron. Several conditions for the Na pseudopotential are employed: (1) The potential binds only the valence electron: the 3s-electron. (2) The eigenvalue of the corresponding valence electron be identical to the full potential eigenvalue. (The full potential is also called the *all-electron* potential.) (3) The wave function be nodeless and *identical* to the “all electron” wave function outside the core region. For example, we construct a pseudo-wave function,  $\phi_{3s}(r)$  such that  $\phi_{3s}(r) = \psi_{3s}(r)$  for  $r > r_c$  where  $r_c$  defines the size spanned by the *ion core*, *i.e.*, the nucleus and core electrons. For Na, this means knowing the spatial extent of the “size” of the ion core, which includes the  $1s^2 2s^2 2p^6$  states and the nucleus. Typically, the size of the ion core is taken to be less than the distance corresponding of the maximum of the valence wave function as measured from the nucleus, but greater than the distance of the outermost node from the nucleus. This valence wave functions are depicted in Fig. §.3.

The pseudo-wave function,  $\phi_p(r)$ , must be identical to the all electron wave function,  $\psi_{AE}(r)$ , outside the core:  $\phi_p(r) = \psi_{AE}(r)$  for  $r > r_c$ . This condition will guarantee that the pseudo-wave function possesses identical

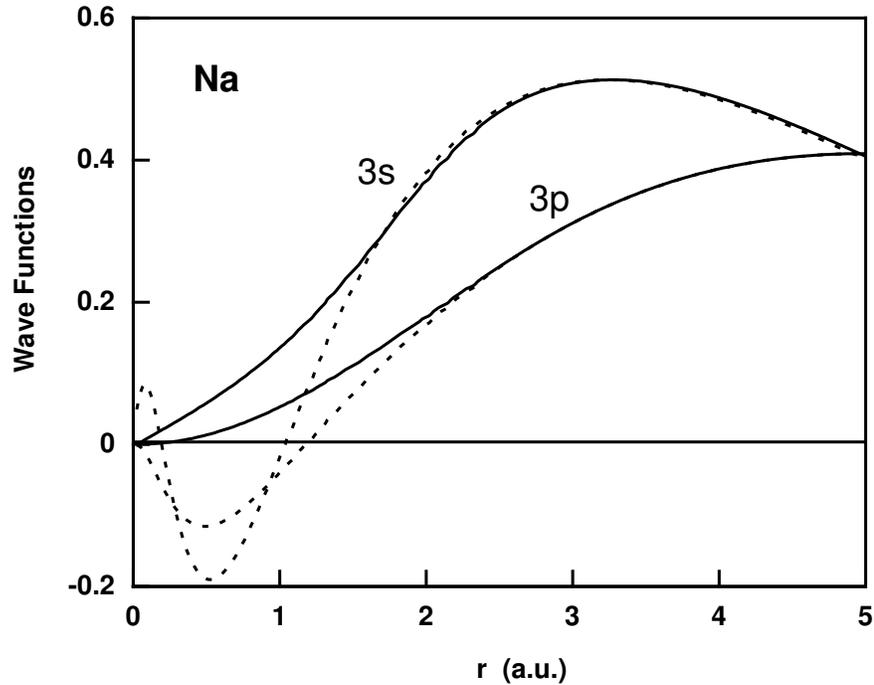


Figure §.3: Pseudopotential wave functions compared to all-electron wave functions for the sodium atom. The all-electron wave functions are indicated by the dashed lines.

properties as the all electron wave function. For  $r < r_c$ , one may alter the all-electron wave function as desired within certain limitations, *e.g.*, the wave function in this region must be smooth and nodeless.

Another very important criterion is mandated. Namely, the integral of the pseudocharge density, *i.e.*, square of the wave function  $|\phi_p(r)|^2$ , within the core should be equal to the integral of the all-electron charge density. Without this condition, the pseudo-wave function can differ by a scaling

factor from the all-electron wave function, that is,  $\phi_p(r) = C \times \psi_{AE}(r)$  for  $r > r_c$  where the constant,  $C$ , may differ from unity. Since we expect the chemical bonding of an atom to be highly dependent on the tails of the valence wave functions, it is imperative that the normalized pseudo wave function be identical to the all-electron wave functions. The criterion by which one insures  $C = 1$  is called *norm conserving* [57]. Some of the earliest *ab initio* potentials did not incorporate this constraint [58]. These potentials are not used for accurate computations. The chemical properties resulting from these calculations using these non-norm conserving pseudopotentials are quite poor when compared to experiment or to the more accurate norm conserving pseudopotentials [55].

In 1980, Kerker [59] proposed a straightforward method for constructing local density pseudopotentials that retained the norm conserving criterion. He suggested that the pseudo-wave function have the following form:

$$\phi_p(r) = r^l \exp(p(r)) \quad \text{for } r < r_c \quad (\S.78)$$

where  $p(r)$  is a simple polynomial:  $p(r) = -a_0r^4 - a_1r^3 - a_2r^2 - a_3$  and

$$\phi_p(r) = \psi_{AE}(r) \quad \text{for } r > r_c \quad (\S.79)$$

This form of the pseudo-wave function for  $\phi_p$  assures us that the function will be nodeless and have the correct behavior at large  $r$ . Kerker proposed criteria for fixing the parameters ( $a_0$ ,  $a_1$ ,  $a_2$  and  $a_3$ ). One criterion is that

the wave function be norm conserving. Other criteria include: (a) The all electron and pseudo-wave functions have the same valence eigenvalue. (b) The pseudo-wave function be nodeless and be identical to the all-electron wave function for  $r > r_c$ . (c) The pseudo-wave function must be continuous as well as the first and second derivatives of the wave function at  $r_c$ .

Other local density pseudopotentials include those proposed by Hamann, Schluter, and Chiang [57], Bachelet, Hamann, and Schluter [60] and Green-side and Schluter [61]. These pseudopotentials were constructed from a different perspective. The all-electron potential was calculated for the free atom. This potential was multiplied by a smooth, short range cut-off function which removes the strongly attractive and singular part of the potential. The cut-off function is adjusted numerically to yield eigenvalues equal to the all-electron valence eigenvalues, and to yield nodeless wave functions converged to the all electron wave functions outside the core region. Again, the pseudo-charge within the core is constrained to be equal to the all-electron value.

As indicated, there is some flexibility in constructing pseudopotentials. While all local density pseudopotentials impose the condition that

$$\phi_p(r) = \psi_{AE}(r)$$

for  $r > r_c$ , the construction for  $r < r_c$  is not unique. The non-uniqueness of the pseudo wave function was recognized early in its inception [55]. This attribute can be exploited to optimize the convergence of the pseudopotentials

for the basis of interest. Much effort has been made to construct “soft” pseudopotentials. By soft, one means a rapidly convergent calculation using a simple basis such as plane waves. Typically, soft potentials are characterized by a large core radius,  $r_c$ . As the core radius becomes larger, the convergence between the all electron and pseudo wave functions is postponed to larger distances. The quality of the pseudo wave functions starts to deteriorate and the transferability of the pseudopotential between the atom and complex environments such as a cluster of atoms becomes limited.

Several schemes have been developed to generate soft pseudopotentials for species which extend effectively the core radius while preserving transferability. The primary motivation for such schemes is to reduce the size of the basis. One of the earliest discussions of such issues is from Vanderbilt [62]. A common measure of pseudopotential softness is to examine the behavior of the potential in reciprocal space. For example, a hard core pseudopotential, *i.e.* one that scales as  $1/r^2$  for small  $r$  will decay only as  $1/q$  in reciprocal space. This rate of decay is worse than using the bare coulomb potential which scales as  $1/q^2$ .

The Kerker pseudopotential [59] does no better than the coulomb potential as the Kerker pseudopotential has a discontinuity in its third derivative at the origin and at the cut-off radius. This gives rise to a slow  $1/q^2$  decay of the potential, although one should examine each case, as the error introduced by truncation of such a potential in reciprocal space may still be acceptable in terms of yielding accurate wave functions and energies. Hamann-Schluter-

Chiang [57] potentials often converge better than the Kerker potentials [59] in that they contain no such discontinuities.

An outstanding issue that remains unresolved is the “best” criterion to use in constructing an “optimal” pseudopotential. An optimal pseudopotential is one that minimizes the number of basis functions required to achieve the desired goal; it yields a converged total energy yet does not sacrifice transferability.

One straightforward approach to optimizing a pseudopotential is to build additional constraints into the polynomial given in Eq. (§.78). For example, suppose we write

$$p(r) = c_o + \sum_{n=1}^N c_n r^n \quad (\S.80)$$

In Kerker’s scheme,  $N=4$ . However, there is no compelling reason for demanding that the series terminate at this particular point. If we extend the expansion, we may impose additional constraints. For example, we might try to constrain the reciprocal space expansion of the pseudo-wave function so that beyond some momentum cut-off the function vanishes. A different approach has been suggested by Troullier and Martins, 1991 [63]. They write Eq. §.80 as

$$p(r) = c_o + \sum_{n=1}^6 c_{2n} r^{2n} \quad (\S.81)$$

As usual, they constrained the coefficients to be norm conserving. In addition, they demanded continuity of the pseudo-wave functions and the first four derivatives at  $r_c$ . The final constraint was to demand zero curvature of

the pseudopotential at the origin. These potentials tend to be quite smooth and converge very rapidly in reciprocal space.

Once the pseudo wave function is defined as in Eqs. (§.78,§.79) one can *invert* the Kohn-Sham equation and solve for the ion core pseudopotential,  $V_{ion,p}$ :

$$V_{ion,p}^n(\vec{r}) = E_n - V_H(\vec{r}) - V_{xc}[\vec{r}, \rho(\vec{r})] + \frac{\hbar^2 \nabla^2 \phi_{p,n}}{2m\phi_{p,n}} \quad (§.82)$$

This potential, when self-consistently screened by the pseudo-charge density:

$$\rho(\vec{r}) = -e \sum_{n,occup} |\phi_{p,n}(\vec{r})|^2 \quad (§.83)$$

will yield an eigenvalue of  $E_n$  and a pseudo wave function  $\phi_{p,n}$ . The pseudo wave function by construction will agree with the all electron wave function away from the core.

There are some important issues to consider about the details of this construction. First, the potential is *state* dependent as written in Eq. (§.82), *i.e.*, the pseudopotential is dependent on the quantum state  $n$ . This issue can be handled by recognizing the nonlocality of the pseudopotential. The potential is different for an  $s$ -,  $p$ -, or  $d$ -electron. The nonlocality appears in the angular dependence of the potential, but not in the radial coordinate.

A related issue is whether the potential is highly dependent on the state energy, *e.g.*, if the potential is fixed to replicate the  $3s$  state in Na, will it also do well for the  $4s$ ,  $5s$ ,  $6s$ , etc.? Of course, one could also question how dependent the pseudopotential is on the atomic state used for its construc-

tion. For example, would a Na potential be very different for a  $3s^1 3p^0$  versus a  $3s^{1/2} 3p^{1/2}$  configuration? Finally, how important are loosely bound core states in defining the potential? For example, can one treat the  $3d$  states in copper as part of the core or part of the valence shell?

Each of these issues has been carefully addressed in the literature. In most cases, the separation between the core states and the valence states is clear. For example, in Si there is no issue that the core is composed of the  $1s^2 2s^2 2p^6$  states. However, the core in Cu could be considered to be the  $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10}$  configuration with the valence shell consisting of the  $4s^1$  state. Alternatively, one could well consider the core to be the  $1s^2 2s^2 2p^6 3s^2 3p^6$  configuration with the valence shell composed of the  $3d^{10} 4s^1$  states. It is quite apparent that solely treating the valence state in Cu as a  $4s$  state cannot be correct, otherwise K and Cu would be chemically similar as K has the same electron core as Cu, save the  $3d$  shell. It is the outer  $3d$  shell that distinguishes Cu from K. Such issues need to be traditionally considered on a case by case situation. It is always possible to construct different pseudopotentials one for each core-valence dichotomy. One can examine the resulting electronic structure for each potential, and verify the role of including a questionable state as a valence or core state.

Another aspect of this problem centers on “core-valence” exchange-correlation. In the all electron exchange-correlation potential, the charge density is composed of the core and valence states; in the pseudopotential treatment only the valence electrons are included. This separation neglects terms that may

arise between the overlap of the valence and core states. There are well defined procedures for including these overlap terms. It is possible to include a fixed charge density from the core and allow the valence overlap to be explicitly included. This procedure is referred to as a *partial core correction* [64]. This correction is especially important for elements such as Zn, Cd and Hg where the outermost filled d-shell can contribute to the chemical bonding. Again, the importance of this correction can be tested by performing calculations with and without the partial core. Of course, one might argue that the most accurate approach would be to include any “loosely bound” core states as valence states. This approach is often not computationally feasible nor desirable. For example, the Zn core without the  $3d$  states results in dealing with an ion core pseudopotential for  $\text{Zn}^{+12}$ . This results in a very strong pseudopotential, which is required to bind 12 valence electrons. The basis must contain highly localized functions to replicate the  $d$ -states plus extended states to replicate the  $s$ -states. Moreover, the number of occupied eigenstates increases by a factor of six. Since most “standard” algorithms to solve for eigenvalues scale superlinearly in time with the number of eigenvalues (such as  $N_{eig}^2$  where  $N_{eig}$  is the number of require eigenvalues), this is a serious issue.

With respect to the state dependence of the pseudopotential, these problems can be overcome with little computational effort. Since the core electrons are tightly bound, the ion core potential is highly localized and is not highly sensitive to the ground state configuration used to compute the pseu-

dopotential. There are well-defined tests for assessing the accuracy of the pseudopotential, especially in terms of the phase shifts [55]. Also, it should be noted that higher excited states sample the tail of the pseudopotential. This pseudopotential should converge to the all electron potential outside of the core. A significant source of error here is the local density approximation. The LDA yields a potential that scales exponentially at large distances and not as one would expect for an image charge, *i.e.*, intuitively, the true potential should incorporate an image potential such that  $V_{xc}(r \rightarrow \infty) \rightarrow -e^2/r$ .

Nonlocality in the pseudopotential is often treated in Fourier space, but it may also be expressed in real space. The interactions between valence electrons and pseudo-ionic cores may be separated into a local potential and a Kleinman and Bylander [65] form of a nonlocal pseudopotential in *real space* [63],

$$V_{ion}^p(\vec{r})\phi_n(\vec{r}) = \sum_a V_{loc}(|\vec{r}_a|)\phi_n(\vec{r}) + \sum_{a, n, lm} G_{n,lm}^a u_{lm}(\vec{r}_a)\Delta V_l(r_a), \quad (\S.84)$$

$$K_{n,lm}^a = \frac{1}{\langle \Delta V_{lm}^a \rangle} \int u_{lm}(\vec{r}_a)\Delta V_l(r_a)\psi_n(\vec{r})d^3r, \quad (\S.85)$$

and  $\langle \Delta V_{lm}^a \rangle$  is the normalization factor,

$$\langle \Delta V_{lm}^a \rangle = \int u_{lm}(\vec{r}_a)\Delta V_l(r_a)u_{lm}(\vec{r}_a)d^3r, \quad (\S.86)$$

where  $\vec{r}_a = \vec{r} - \vec{R}_a$ , and the  $u_{lm}$  are the atomic pseudopotential wave functions of angular momentum quantum numbers  $(l, m)$  from which the  $l$ -dependent

ionic pseudopotential,  $V_l(r)$ , is generated.  $\Delta V_l(r) = V_l(r) - V_{loc}(r)$  is the difference between the  $l$  component of the ionic pseudopotential and the local ionic potential.

In the case of Na, one might choose the local part of the potential to replicate only the  $l = 0$  component as defined by the  $3s$  state. The nonlocal parts of the potential would then contain only the  $l = 1$  and  $l = 2$  components. For simple metals like Na or electronic materials such as Si and GaAs, the angular momentum components for  $l = 3$  (or higher) are not significant in the ground state wave functions. In these systems, one can treat the summation over  $l = 0, 1, 2$  to be complete.

The choice of the angular component for the local part of the potential is somewhat arbitrary. It is often convenient to choose the highest  $l$ -component present. This avoids the complex projections with the highest  $l$ . These issues can be tested by choosing different components for the local potential.

In Fig. §.4, the ion core pseudopotential for Na is presented using the Troullier-Martins formalism for creating pseudopotentials. The nonlocality of the potential is evident by the existence of the three potentials corresponding to the  $s$ -,  $p$ - and  $d$ -states.

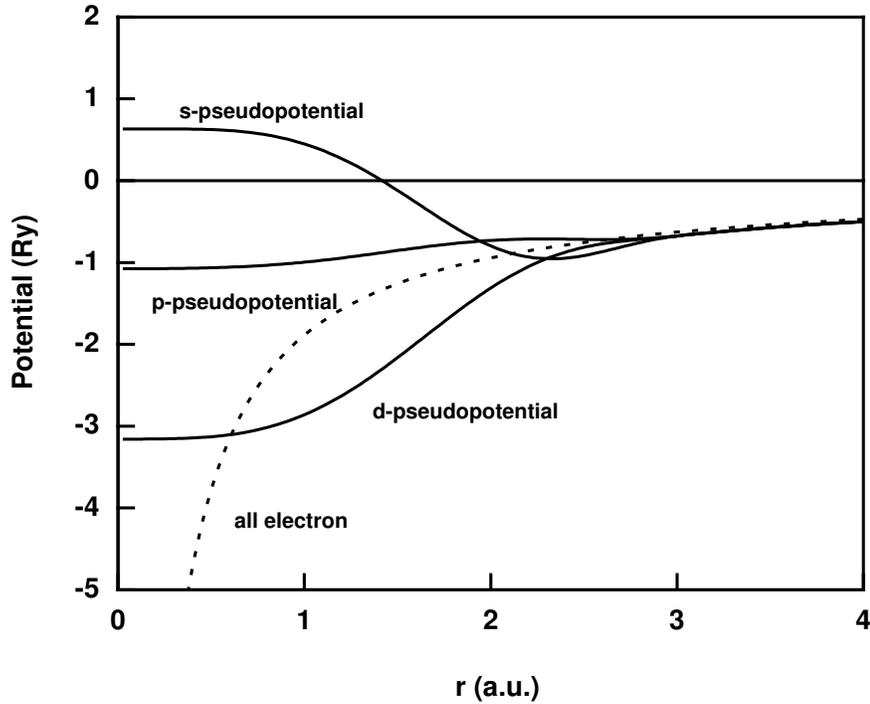


Figure §.4: Pseudopotential compared to the all electron potential for the sodium atom.

## §.5 Solving the eigenvalue problem

Once the pseudopotential has been determined, the resulting eigenvalue problem needs to be solved for the system of interest:

$$l\left( \frac{-\hbar^2 \nabla^2}{2m} + V_{ion}^p(\vec{r}) + V_H(\vec{r}) + V_{xc}[\vec{r}, \rho(\vec{r})] \right) \phi_n(\vec{r}) = E_n \phi_n(\vec{r}) \quad (\S.87)$$

where  $V_{ion}^p$  is the ionic pseudopotential for the system. Since the ion cores can be treated as chemically inert and highly localized, it is a simple matter

to write:

$$V_{ion}^p(\vec{r}) = \sum_{\vec{R}_a} V_{ion,a}^p(\vec{r} - \vec{R}_a) \quad (\S.88)$$

where  $V_{ion,a}^p$  is the ion core pseudopotential associated with the atom,  $a$ , at a position  $\vec{R}_a$ .

A major difficulty in solving the eigenvalue problem in Eq. (§.87) are the length and energy scales involved. The inner (core) electrons are highly localized and tightly bound compared to the outer (valence electrons). A simple basis function approach is frequently ineffectual. For example, a plane wave basis might require  $10^6$  waves to represent converged wave functions for a core electron whereas only  $10^2$  waves are required for a valence electron [55]. The pseudopotential overcomes this problem by removing the core states from the problem and replacing the all electron potential by one that replicates only the chemically active, valence electron states [55]. By construction, the pseudopotential reproduces the valence state properties such as the eigenvalue spectrum and the charge density outside the ion core.

Since the pseudopotential is weak, simple basis sets such as a plane wave basis can be quite effective for crystalline matter. For example, in the case of crystalline silicon only 50-100 plane waves need to be used. The resulting matrix representation of the Schrödinger operator is dense in Fourier (plane wave) space, but it is not formed explicitly. Instead, matrix-vector product operations are performed with the help of fast Fourier transforms. This plane wave approach is akin to spectral techniques used in solving certain types of

partial differential equations [66]. The plane wave method uses a basis of the form:

$$\psi_{\vec{k}}(\vec{r}) = \sum_{\vec{G}} \alpha(\vec{k}, \vec{G}) \exp(i(\vec{k} + \vec{G}) \cdot \vec{r}) \quad (\S.89)$$

where  $\vec{k}$  is the wave vector,  $\vec{G}$  is a reciprocal lattice vector and  $\alpha(\vec{k}, \vec{G})$  represent the coefficients of the basis. In a plane wave basis, the Laplacian term of the Hamiltonian is represented by a diagonal matrix. The potential term  $V_{tot}^p$  gives rise to a dense matrix.

In real space, it is trivial to operate with the potential term which is represented by a diagonal matrix, and in Fourier space it is trivial to operate with the Laplacian term, which is also represented by a diagonal matrix. The use of plane wave bases also leads to natural preconditioning techniques that are obtained by simply employing a matrix obtained from a smaller plane wave basis, neglecting the effect of high frequency terms on the potential. For periodic systems, where  $\vec{k}$  is a good quantum number, the plane wave basis coupled to pseudopotentials is quite effective. However, for non-periodic systems such as clusters, liquids or glasses, the plane wave basis must be combined with a *supercell method* [55]. The supercell repeats the localized configuration to impose periodicity to the system. This preserves the validity of  $\vec{k}$  and Bloch's theorem, which Eq. (§.89) obeys. There is a parallel to be made with spectral methods that are quite effective for simple periodic geometries, but lose their superiority when more generality is required. In addition to these difficulties the two fast Fourier transforms performed at

each iteration can be costly, requiring  $n \log n$  operations, where  $n$  is the number of plane waves, versus  $O(N)$  for real space methods where  $N$  is the number of grid points. Usually, the matrix size  $N \times N$  is larger than  $n \times n$  but only within a constant factor. This is exacerbated in high performance environments where fast Fourier transforms require an excessive amount of communication and are particularly difficult to implement efficiently.

Another popular basis employed with pseudopotentials include Gaussian orbitals [67–70]. Gaussian bases have the advantage of yielding analytical matrix elements provided the potentials are also expanded in Gaussians. However, the implementation of a Gaussian basis is not as straightforward as with plane waves. For example, numerous indices must be employed to label the state, the atomic site, and the Gaussian orbitals employed. On the positive side, a Gaussian basis yields much smaller matrices and requires less memory than plane wave methods. For this reason Gaussians are especially useful for describing transition metal systems.

An alternative approach is to avoid the use of an explicit basis. For example, one can use a real space method that avoids the use of plane waves and fast Fourier transforms altogether. This approach has become popular and different versions of this general approach have been implemented [66,71–82].

A real space approach overcomes many of the complications involved with non-periodic systems, and although the resulting matrices can be larger than with plane waves, they are sparse and the methods are easier to parallelize. Even on sequential machines, real space methods can be an order of magni-

tude faster than the traditional approach.

Real space algorithms avoid the use of fast Fourier transforms by performing all calculations in real physical space instead of Fourier space. Fast Fourier transforms require global communications; as such, they are not efficient for implementation on multi-processor platforms. The only global operation remaining in *real space* approaches is that of the inner products. These inner products are required when forming the orthogonal basis used in the generalized Davidson procedure as discussed below.

The simplest real space method utilizes finite difference discretization on a cubic grid. A key aspect to the success of the finite difference method is the availability of *higher order finite difference expansions* for the kinetic energy operator, *i.e.*, expansions of the Laplacian [83]. Higher order finite difference methods significantly improve convergence of the eigenvalue problem when compared with standard finite difference methods. If one imposes a simple, uniform grid on our system where the points are described in a finite domain by  $(x_i, y_j, z_k)$ , we approximate  $\frac{\partial^2 \psi}{\partial x^2}$  at  $(x_i, y_j, z_k)$  by

$$\frac{\partial^2 \psi}{\partial x^2} = \sum_{n=-M}^M C_n \psi(x_i + nh, y_j, z_k) + O(h^{2M+2}), \quad (\S.90)$$

where  $h$  is the grid spacing and  $M$  is a positive integer. This approximation is accurate to  $O(h^{2M+2})$  upon the assumption that  $\psi$  can be approximated accurately by a power series in  $h$ . Algorithms are available to compute the coefficients  $C_n$  for arbitrary order in  $h$  [83].

With the kinetic energy operator expanded as in Eq. (§.90), one can set up a one-electron Schrödinger equation over a grid. One may assume a uniform grid, but this is not a necessary requirement.  $\psi(x_i, y_j, z_k)$  is computed on the grid by solving the eigenvalue problem:

$$\begin{aligned}
& -\frac{\hbar^2}{2m} \left[ \sum_{n_1=-M}^M C_{n_1} \psi_n(x_i + n_1 h, y_j, z_k) + \sum_{n_2=-M}^M C_{n_2} \psi_n(x_i, y_j + n_2 h, z_k) \right. \\
& \left. + \sum_{n_3=-M}^M C_{n_3} \psi_n(x_i, y_j, z_k + n_3 h) \right] + [V_{ion}(x_i, y_j, z_k) + V_H(x_i, y_j, z_k) \\
& + V_{xc}(x_i, y_j, z_k)] \psi_n(x_i, y_j, z_k) = E_n \psi_n(x_i, y_j, z_k) \tag{§.91}
\end{aligned}$$

If we have  $L$  grid points, the size of the full matrix resulting from the above problem is  $L \times L$ .

The grid we use is based on points uniformly spaced in a three dimensional cube as shown in Fig. §.5, with each grid point corresponding to a row in the matrix. However, many points in the cube are far from any atoms in the system and the wave function on these points may be replaced by zero. Special data structures may be used to discard these points and keep only those having a nonzero value for the wave function. The size of the Hamiltonian matrix is usually reduced by a factor of two to three with this strategy, which is quite important considering the large number of eigenvectors which must be saved. Further, since the Laplacian can be represented by a simple stencil, and since all local potentials sum up to a simple diagonal matrix,

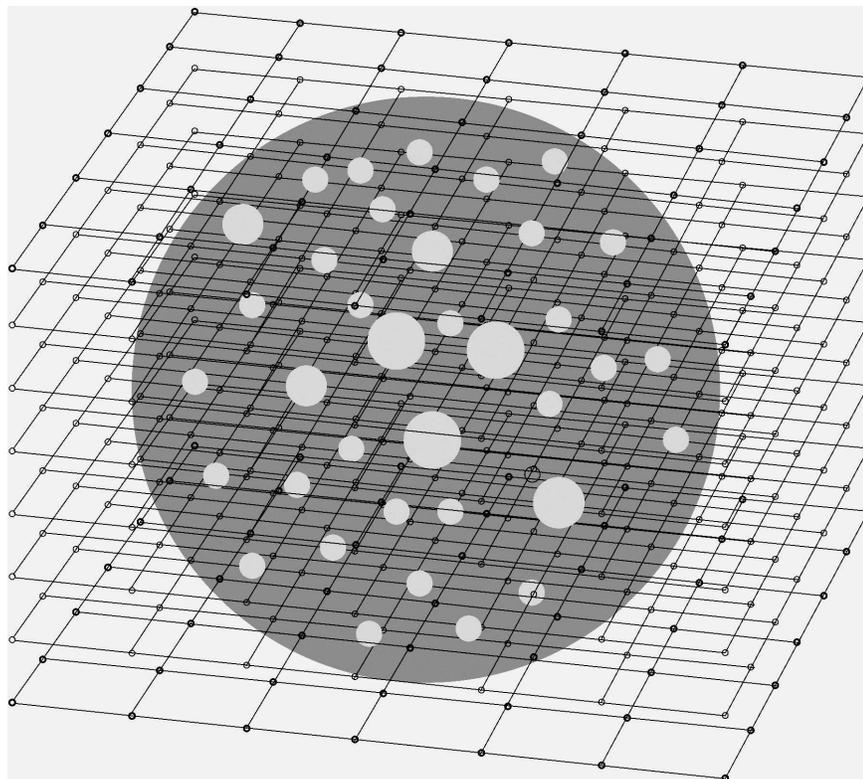


Figure §.5: Uniform grid illustrating a typical configuration for examining the electronic structure of a localized system. The dark gray sphere represents the domain where the wave functions are allowed to be nonzero. The light spheres within the domain are atoms.

the Hamiltonian need not be stored. Handling the ionic pseudopotential is complex as it consists of a local and a non-local term (Eqs. (§.84 )and (§.85)). In the discrete form, the nonlocal term becomes a sum over all atoms,  $a$ , and quantum numbers,  $(l, m)$  of rank-one updates:

$$V_{ion} = \sum_a V_{loc,a} + \sum_{a,l,m} c_{a,l,m} U_{a,l,m} U_{a,l,m}^T \quad (§.92)$$

where  $U_{a,l,m}$  are sparse vectors which are only non-zero in a localized region around each atom,  $c_{a,l,m}$  are normalization coefficients.

There are several difficulties in solving the eigen problems, in addition to the size of the matrices. First, the number of required eigenvectors is proportional to the atoms in the system, and can grow up to thousands, if not more. Besides storage, maintaining the orthogonality of these vectors can be a formidable task. Second, the relative separation of the eigenvalues becomes increasingly poor as the matrix size increases and this has an adverse effect on the rate of convergence of the eigenvalue solvers. Preconditioning techniques attempt to alleviate this problem. On the positive side, the matrix need not be stored as was mentioned earlier and this reduces storage requirement. In addition, good initial eigenvector estimates are available at each iteration from the previous self-consistency loop.

A popular form of extracting the eigenpairs is based on the generalized Davidson [84] method, in which the preconditioner is not restricted to be a diagonal matrix as in the Davidson method. (A detailed description can

be found in [85].) Preconditioning techniques in this approach are typically based on filtering ideas and the fact that the Laplacian is an elliptic operator [86]. The eigenvectors corresponding to the few lowest eigenvalues of  $\nabla^2$  are smooth functions and so are the corresponding wave functions. When an approximate eigenvector is known at the points of the grid, a smoother eigenvector can be obtained by averaging the value at every point with the values of its neighboring points. Assuming a cartesian  $(x, y, z)$  coordinate system, the low frequency filter acting on the value of the wave function at the point  $(i, j, k)$ , which represents one element of the eigenvector, is described by:

$$\left[ \frac{\psi_{i-1,j,k} + \psi_{i,j-1,k} + \psi_{i,j,k-1} + \psi_{i+1,j,k} + \psi_{i,j+1,k} + \psi_{i,j,k+1}}{12} \right] + \frac{\psi_{i,j,k}}{2} \rightarrow (\psi_{i,j,k})_{Filtered} \quad (\S.93)$$

It is worth mentioning that other preconditioners that have been tried have resulted in mixed success. The use of shift-and-invert [87] involves solving linear systems with  $A - \sigma I$ , where  $A$  is the original matrix and the shift  $\sigma$  is close to the desired eigenvalue. These methods would be prohibitively expensive in most situations, given the size of the matrix and the number of times that  $A - \sigma I$  must be factored. Alternatives based on an approximate factorization such as ILUT [88] are ineffective beyond the first few eigenvalues. Methods based on approximate inverse techniques have been somewhat more successful, performing better than filtering at additional preprocessing

and storage cost. Preconditioning ‘interior’ eigenvalues, i.e., eigenvalues located well inside the interval containing the spectrum, is still a very hard problem. Current solutions only attempt to dampen the effect of eigenvalues which are far away from the ones being computed. This is in effect what is achieved by filtering and sparse approximate inverse preconditioning. These techniques do not reduce the number of steps required for convergence in the same way that shift-and-invert techniques do. However, filtering techniques are inexpensive to apply and result in fairly substantial savings in iterations.

## §.6 Properties of confined systems: clusters

The electronic and structural properties of atomic clusters stand as one of the outstanding problems in materials physics. Clusters possess properties that are characteristic of neither the atomic nor solid state. For example, the energy levels in atoms may be discrete and well-separated in energy relative to  $kT$ . In contrast, solids have continuum of states (energy bands). Clusters may reside between these limits, *i.e.*, the energy levels may be discrete, but with a separation much less than  $kT$ .

Real space methods are ideally suited for investigating these systems. In contrast to plane wave methods, real space methods can examine non-periodic systems without introducing artifacts such as supercells. Also, one can easily examine charged clusters. In supercell configurations, unless a compensating background charge is added, the Coulomb energy diverges for

charged clusters.

A closely related issue concerns electronic excitations. In periodic systems, it is nontrivial to consider localized excitation, *e.g.*, with band theory exciting an atom in one cell, excites all atoms in all the equivalent cells. Density functional formalisms often avoid these issues by considering localized or non-periodic systems.

### §.6.1 Structure

Perhaps the most fundamental issue in dealing with clusters is determining their structure. Before any accurate theoretical calculations can be performed for a cluster, the atomic geometry of a system must be defined. However, this can be a formidable exercise. Serious problems arise from the existence of multiple local minima in the potential-energy-surface of these systems; many similar structures can exist with vanishingly small energy differences.

A convenient method to determine the structure of small or moderate sized clusters is *simulated annealing* [89]. Within this technique, atoms are randomly placed within a large cell and allowed to interact at a high (usually fictive) temperature. Within this temperature regime, atoms will sample a large number of configurations. As the system is cooled, the number of high energy configurations sampled is restricted. If the anneal is done slowly enough, the procedure should quench out structural candidates for the ground state structures.

Langevin molecular dynamics is well suited for simulated annealing meth-

ods. In Langevin dynamics, the ionic positions,  $\mathbf{R}_j$ , evolve according to

$$M_j \ddot{\mathbf{R}}_j = \mathbf{F}(\{\mathbf{R}_j\}) - \gamma M_j \dot{\mathbf{R}}_j + \mathbf{G}_j \quad (\S.94)$$

where  $\mathbf{F}(\{\mathbf{R}_j\})$  is the interatomic force on the  $j$ -th particle, and  $\{M_j\}$  are the ionic masses. The last two terms on the right hand side of Eq. (§.94) are the dissipation and fluctuation forces, respectively. The dissipative forces are defined by the friction coefficient,  $\gamma$ . The fluctuation forces are defined by random Gaussian variables,  $\{\mathbf{G}_j\}$ , with a white noise spectrum:

$$\langle G_i^\alpha(t) \rangle = 0 \quad \text{and} \quad \langle G_i^\alpha(t) G_j^\alpha(t') \rangle = 2\gamma M_i k_B T \delta_{ij} \delta(t - t') \quad (\S.95)$$

The angular brackets denote ensemble or time averages, and  $\alpha$  stands for the Cartesian component. The coefficient of  $T$  on the right hand side of Eq. (§.95) insures that the fluctuation-dissipation theorem is obeyed, *i.e.*, the work done on the system is dissipated by the viscous medium ([90,91]). The interatomic forces can be obtained from the Hellmann-Feynman theorem using the pseudopotential wave functions.

Langevin simulations using quantum forces can be contrasted with other techniques such as the Car-Parrinello method [92,93]. Langevin simulations as outlined above do not employ fictitious electron dynamics; at each time step the system is quenched to the Born-Oppenheimer surface and the quantum forces are determined. This approach requires a fully-self consistent treatment of the electronic structure problem; however, because the inter-

atomic forces are true quantum forces, the resulting molecular dynamics simulation can be performed with much larger time steps. Typically, it is possible to use steps an order of magnitude larger than in the Car-Parrinello method [94]. It should be emphasized that neither of these methods is particularly efficacious without the pseudopotential approximation.

To illustrate the simulated annealing procedure, we consider a silicon cluster of seven atoms. With respect to the technical details for this example, the initial temperature of the simulation was taken to be about 3000 K; the final temperature was taken to be 300 K. The annealing schedule lowered the temperature 500 K each 50 time steps. The time step was taken to be 5 fs. The friction coefficient in the Langevin equation was taken to be  $6 \times 10^{-4}$  a.u.<sup>1</sup> After the clusters reached a temperature of 300 K, they were quenched to 0 K. The ground state structure was found through a direct minimization by a steepest descent procedure.

Choosing an initial atomic configuration for the simulation takes some care. If the atoms are too far apart, they will exhibit Brownian motion, which is appropriate for Langevin dynamics with the interatomic forces zeroed. In this case, the atoms may not form a stable cluster as the simulation proceeds. Conversely, if the atoms are too close together, they may form a metastable cluster from which the ground state may be kinetically inaccessible even at the initial high temperature. Often the initial cluster is formed by a random

---

<sup>1</sup>1 atomic unit (a.u.) is defined by  $\hbar = m = e = 1$ . The unit of energy is the hartree (1 hartree = 27.2 eV); the unit of length is the bohr radius (1 bohr = 0.529 Å).

placement of the atoms with a constraint that any given atom must reside within 1.05 and 1.3 times the bond length from at least one atom where the bond length is defined by the crystalline environment. The cluster in question is placed in a spherical domain. Outside of this domain, the wave function is required to vanish. The radius of the sphere is such that the outmost atom is at least 6 a.u. from the boundary. Initially, the grid spacing was 0.8 a.u. For the final quench to a ground state structure, the grid spacing was reduced to 0.5 a.u. As a rough estimate, one can compare this grid spacing with a plane wave cutoff of  $(\pi/h)^2$  or about 40 Ry for  $h=0.5$  a.u.

In Fig. §.6, we illustrate the simulated anneal for the  $\text{Si}_7$  cluster. The initial cluster contains several incipient bonds, but the structure is far removed from the ground state by approximately 1 eV/atom. In this simulation, at about 100 time steps a tetramer and a trimer form. These units come together and precipitate a large drop in the binding energy. After another  $\sim 100$  time steps, the ground state structure is essentially formed. The ground state of  $\text{Si}_7$  is a bicapped pentagon, as is the corresponding structure for the  $\text{Ge}_7$  cluster. The binding energy shown in Fig. §.6 is relative to that of an isolated Si atom. Gradient corrections [18, 45], or spin polarization [95] have not been included in this example.. Therefore, the binding energies indicated in the Figure are likely to be overestimated by  $\sim 20\%$  or so.

In Fig. §.7, the ground state structures for  $\text{Si}_n$  are presented for  $n \leq 7$ . The structures for  $\text{Ge}_n$  are very similar to  $\text{Si}_n$ . The primary difference resides

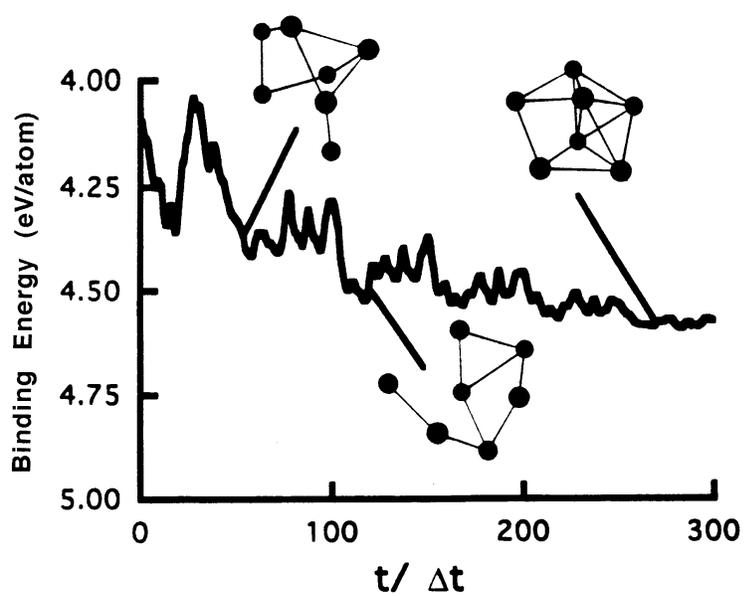


Figure §.6: Binding energy of  $\text{Si}_7$  during a Langevin simulation. The initial temperature is 3000 K; the final temperature is 300 K. Bonds are drawn for interatomic distances of less than  $2.5\text{\AA}$ . The time step is 5 fs.

in the bond lengths. The Si bond length in the crystal is 2.35 Å, whereas in Ge the bond length is 2.44 Å. This difference is reflected in the bond lengths for the corresponding clusters; Si<sub>n</sub> bond lengths are typically a few percent shorter than the corresponding Ge<sub>n</sub> clusters.

It should be emphasized that this annealing simulation is an optimization procedure. As such, other optimization procedures may be used to extract the minimum energy structures. Recently, a genetic algorithm has been used to examine carbon clusters [99]. In this algorithm, an initial set of clusters is “mated” with the lowest energy offspring “surviving”. By examining several thousand generations, it is possible to extract a reasonable structure for the ground state. The genetic algorithm has some advantages over a simulated anneal, especially for clusters which contain more than ~20 atoms. One of these advantages is that kinetic barriers are more easily overcome. However, the implementation of the genetic algorithm is more involved than an annealing simulation, *e.g.*, in some cases “mutations,” or *ad hoc* structural rearrangements, must be introduced to obtain the correct ground state [99].

### §.6.2 Photoemission spectra

A very useful probe of condensed matter involves the photoemission process. Incident photons are used to eject electrons from a solid. If the energy and spatial distributions of the electrons are known, then information can be obtained about the electronic structure of the materials of interest. For crystalline matter, the photoemission spectra can be related to the electronic

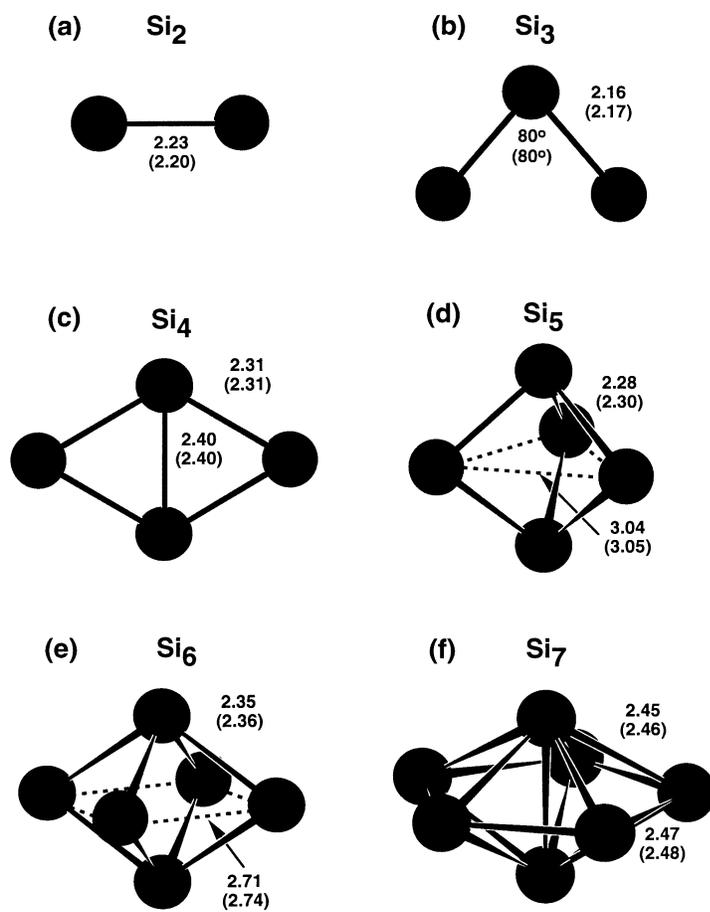


Figure §.7: Ground state geometries and some low-energy isomers of  $\text{Si}_n$  ( $n \leq 7$ ) clusters. Interatomic distances are in Å. The values in parentheses are from Ref. [96–98].

density of states. For confined systems, the interpretation is not as straightforward. One of the earliest experiments performed to examine the electronic structures of small semiconductor clusters examined negatively charged  $\text{Si}_n$  and  $\text{Ge}_n$  ( $n \leq 12$ ) clusters [100]. The photoemission spectra obtained in this work were used to gauge the energy gap between the highest occupied state and the lowest unoccupied state. Large gaps were assigned to the “magic number” clusters, while other clusters appeared to have vanishing gaps. Unfortunately, the first theoretical estimates [101] for these gaps showed substantial disagreements with the measured values. It was proposed by Cheshnovsky *et al.* [100], that sophisticated calculations including transition cross sections and final states were necessary to identify the cluster geometry from the photoemission data. The data were first interpreted in terms of the gaps obtained for *neutral* clusters; it was later demonstrated that *atomic relaxations* within the *charged* cluster are important in analyzing the photoemission data [102, 103]. In particular, atomic relaxations as a result of charging may change dramatically the electronic spectra of certain clusters. These charge induced changes in the gap were found to yield very good agreement with the experiment.

The photoemission spectrum of  $\text{Ge}_{10}^-$  illustrates some of the key issues. Unlike  $\text{Si}_{10}^-$ , the experimental spectra for  $\text{Ge}_{10}^-$  does not exhibit a gap. Cheshnovsky *et al.* interpreted this to mean that  $\text{Ge}_{10}^-$  does not exist in the same structure as  $\text{Si}_{10}^-$ . This is a strange result. Si and Ge are chemically similar and the calculated structures for both *neutral* structures are similar. The

lowest energy structure for both ten atom clusters is the tetracapped trigonal prism (labeled by *I* in Fig. §.8). The photoemission spectra for these clusters can be simulated by using Langevin dynamics. The clusters are immersed in a fictive heat bath, and subjected to stochastic forces. If one maintains the temperature of the heat bath and averages over the eigenvalue spectra, a density of states for the cluster can be obtained. The heat bath resembles a buffer gas as in the experimental setup, but the time intervals for collisions are not similar to the true collision processes in the atomic beam. The simulated photoemission spectrum for  $\text{Si}_{10}^-$  is in very good agreement with the experimental results, reproducing both the threshold peak and other features in the spectrum. If a simulation is repeated for  $\text{Ge}_{10}^-$  using the tetracapped trigonal prism structure, the resulting photoemission spectrum is *not* in good agreement with experiment. Moreover, the calculated electron affinity is 2.0 eV in contrast to the experimental value of 2.6 eV. Moreover, there is no reason to believe that the tetracapped trigonal prism structure is correct for  $\text{Ge}_{10}$  when charged. In fact, we find that the bicapped antiprism structure is lower in energy for  $\text{Ge}_{10}^-$ . The resulting spectra using both structures (*I* and *II* in Fig. §.8) are presented in Fig. §.9, and compared to the photoemission experiment. The calculated spectrum using the bicapped antiprism structure is in very good agreement with the photoemission. The presence of a gap is indicated by a small peak removed from the density of states [Fig. §.9(a)]. This feature is absent in the bicapped antiprism structure [Fig. §.9(b)] and consistent with experiment. For  $\text{Ge}_{10}$ , charging the

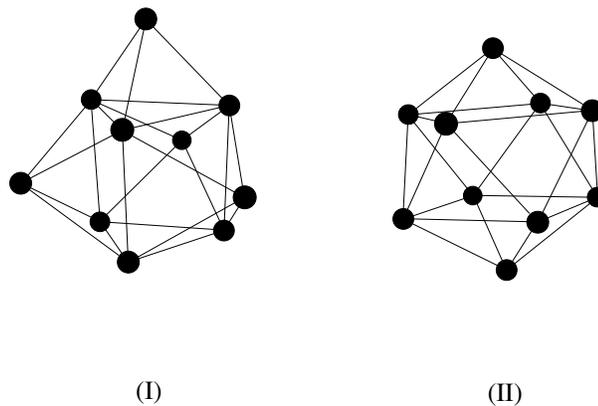


Figure §.8: Two possible isomers for  $\text{Si}_{10}$  or  $\text{Ge}_{10}$  clusters. (I) is a tri-capped trigonal prism cluster and (II) is a bi-capped anti-prism cluster .

structure reverses the relative stability of the two structures. This accounts for the major differences between the photoemission spectra.

It is difficult to assign a physical origin to a particular structure owing to the smaller energy differences involved . However, the bicapped structure has a higher coordination. Most chemical theories of bonding suggest that Ge is more metallic than Si, and as such, would prefer a more highly coordinated structure. The addition of an extra electron may induce a structure reflecting the metallic characteristic of Ge relative to Si.

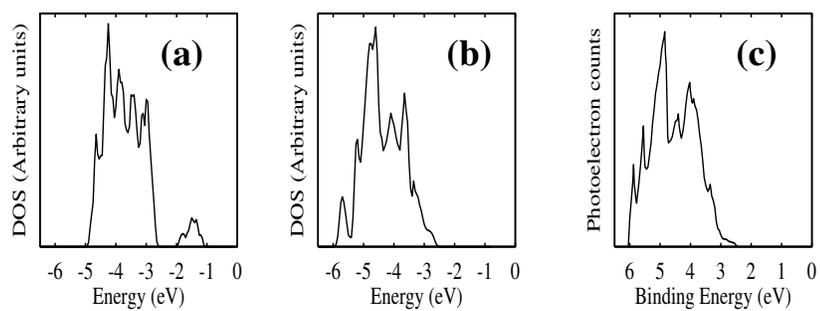


Figure §.9: Calculated density of states for Ge<sub>10</sub><sup>-</sup> in the tetracapped trigonal prism structure (a) and the bicapped antiprism structure (b). (c) Experimental photoemission spectra from Ref. [100].

### §.6.3 Vibrational modes

Experiments on the vibrational spectra of clusters can provide us with very important information about their physical properties. Recently, Raman experiments have been performed on clusters which have been deposited on inert substrates [104]. Since different structural configurations of a given cluster can possess different vibrational spectra, it is possible to compare the vibrational modes calculated for a particular structure with the Raman experiment in a manner similar to the previous example with photoemission. If the agreement between experiment and theory is good, this is a necessary condition for the validity of the theoretically predicted structure.

There are two common approaches for determining the vibrational spectra of clusters. One approach is to calculate the dynamical matrix for the ground state structure of the cluster:

$$M_{i\alpha,j\beta} = \frac{1}{m} \frac{\partial^2 E}{\partial R_i^\alpha \partial R_j^\beta} = -\frac{1}{m} \frac{\partial F_i^\beta}{\partial R_j^\alpha} \quad (\S.96)$$

where  $m$  is the mass of the atom,  $E$  is the total energy of the system,  $F_i^\alpha$  is the force on atom  $i$  in the direction  $\alpha$ ,  $R_i^\alpha$  is the  $\alpha$  component of coordinate for atom  $i$ . One can calculate the dynamical matrix elements by calculating the first order derivative of force versus atom displacement numerically. From the eigenvalues and eigenmodes of the dynamical matrix, one can obtain the vibrational frequencies and modes for the cluster of interest [105].

The other approach to determine the vibrational modes is to perform a

molecular dynamics simulation. The cluster in question is excited by small random displacements. By recording the kinetic (or binding) energy of the cluster as a function of the simulation time, it is possible to extract the power spectrum of the cluster and determine the vibrational modes. This approach has an advantage for large clusters in that one never has to do a mode analysis explicitly. Another advantage is that anaharmonic mode couplings can be examined. It has the disadvantage in that the simulation must be performed over a long time to extract accurate values for all the modes.

As an example, consider the vibrational modes for a small silicon cluster:  $\text{Si}_4$ . The starting geometry was taken to be a planar structure for this cluster as established from a simulated annealing calculation [105].

It is straightforward to determine the dynamical matrix and eigenmodes for this cluster. In Fig. §.10, the fundamental vibrational modes are illustrated. In Table §.1, the frequency of these modes are presented. One can also determine the modes via a simulation. To initiate the simulation, one can perform a Langevin simulation [102] with a fixed temperature at 300K. After a few dozen time steps, the Langevin simulation is turned off, and the simulation proceeds following Newtonian dynamics with “quantum” forces. This procedure allows a stochastic element to be introduced and establish initial conditions for the simulation without bias toward a particular mode. For this example, time step in the molecular dynamics simulation was taken to be 3.7 fs. The simulation was allowed to proceed for 1000 time steps or roughly 4 ps. The variation of the kinetic and binding energies is given in

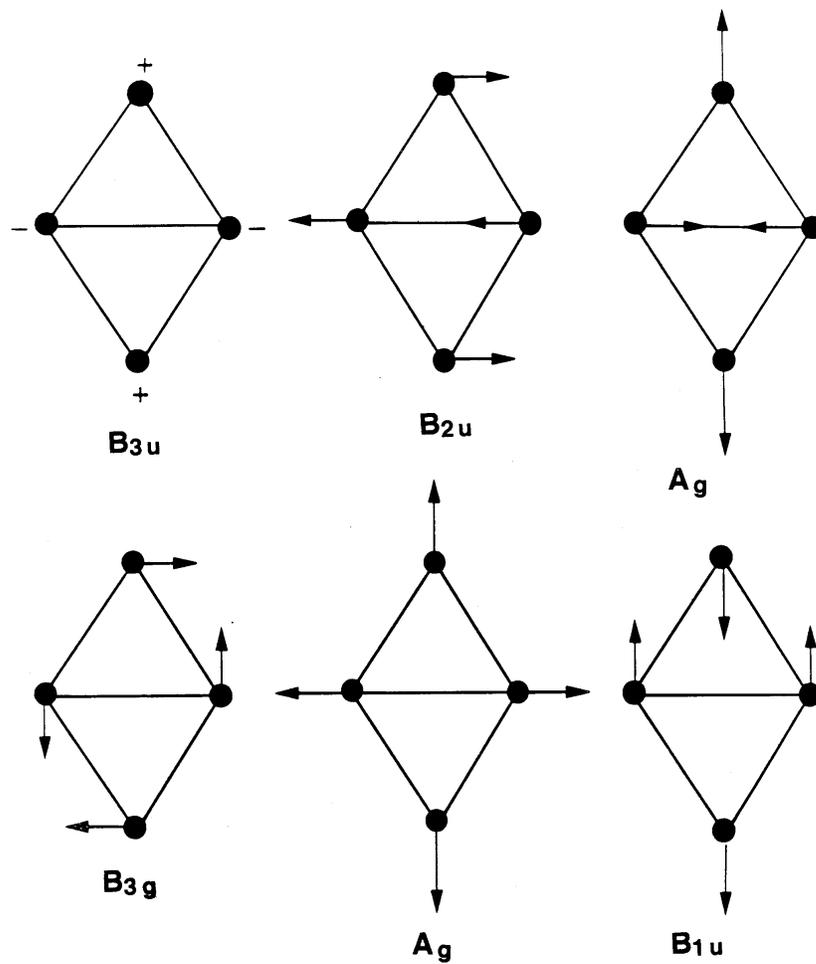


Figure §.10: Normal modes for a  $\text{Si}_4$  cluster. The + and - signs indicate motion in and out of the plane, respectively.

Table §.1: Calculated and experimental vibrational frequencies in a  $\text{Si}_4$  cluster. See Figure §.10 for an illustration of the normal modes. The frequencies are given in  $\text{cm}^{-1}$ .

	$B_{3u}$	$B_{2u}$	$A_g$	$B_{3g}$	$A_g$	$B_{1u}$
Experiment [104]			345		470	
Dynamical Matrix [105]	160	280	340	460	480	500
MD simulation [105]	150	250	340	440	490	500
HF [96]	117	305	357	465	489	529
LCAO [106]	55	248	348	436	464	495

Fig. §.11 as a function of the simulation time. Although some fluctuations of the total energy occur, these fluctuations are relatively small, *i.e.*, less than  $\sim 1$  meV, and there is no noticeable drift of the total energy. Such fluctuations arise, in part, because of discretization errors. As the grid size is reduced, such errors are minimized [105]. Similar errors can occur in plane wave descriptions using supercells, *i.e.*, the artificial periodicity of the supercell can introduce erroneous forces on the cluster. By taking the power spectrum of either the kinetic energy (KE) or binding energy (BE) over this simulation time, the vibrational modes can be determined. These modes can be identified with the observed peaks in the power spectrum as illustrated in Fig. §.12.

A comparison of the calculated vibrational modes from the molecular dynamics simulation and from a dynamical matrix calculation are listed in Table §.1. Overall, the agreement between the simulation and the dynamical matrix analysis is quite satisfactory. In particular, the softest mode, *i.e.*, the

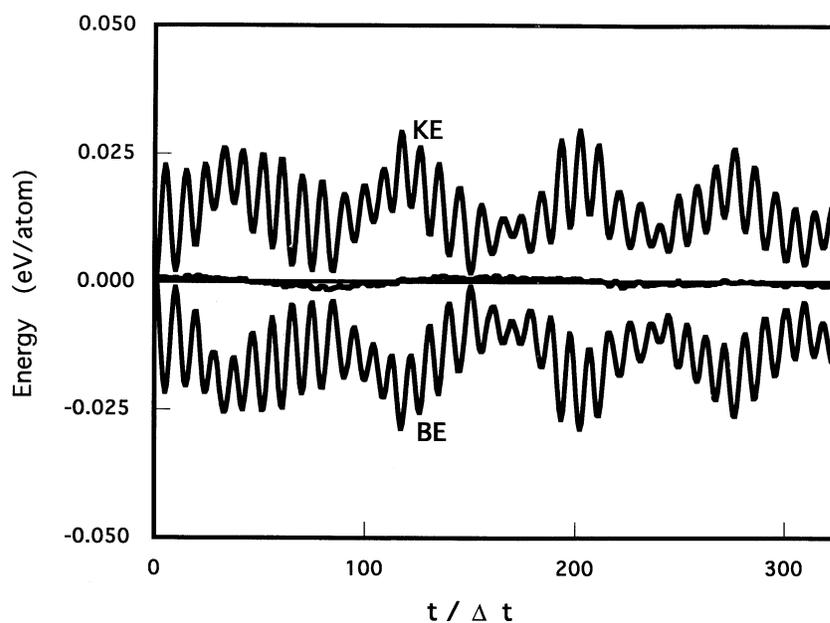


Figure §.11: Simulation for a  $\text{Si}_4$  cluster. The kinetic energy (KE) and binding energy (BE) are shown as a function of simulation time. The total energy (KE+BE) is also shown with the zero of energy taken as the average of the total energy. The time step,  $\Delta t$ , is 3.7fs.

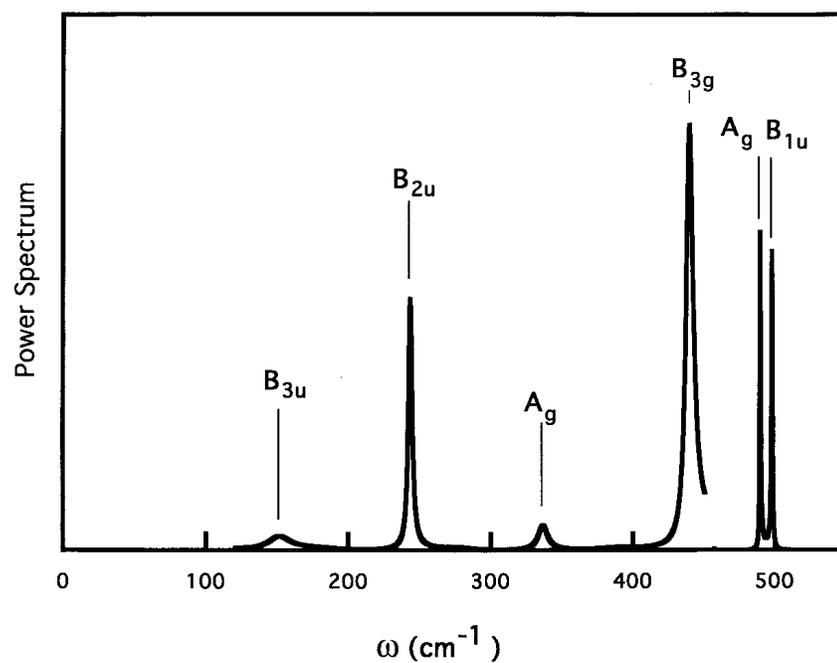


Figure §.12: Power spectrum of the vibrational modes of the Si<sub>4</sub> cluster. The simulation time was taken to be 4 ps. The intensity of the B<sub>3g</sub> and (A<sub>g</sub>, B<sub>1u</sub>) peaks has been scaled by 10<sup>-2</sup>.

$B_{3u}$  mode, and the splitting between the ( $A_g$ ,  $B_{1u}$ ) modes are well replicated in the power spectrum. The splitting of the ( $A_g$ ,  $B_{1u}$ ) modes is less than  $10 \text{ cm}^{-1}$ , or about 1 meV, which is at the resolution limit of any *ab initio* method.

The theoretical values are also compared to experiment. The predicted frequencies for the two  $A_g$  modes are surprisingly close to Raman experiments on silicon clusters [104]. The other allowed Raman line of mode  $B_{3g}$  is expected to have a lower intensity and has not been observed experimentally [104].

The theoretical modes using the formalism outlined here are in good accord (except the lowest mode) with other theoretical calculations given in Table §.1: an LCAO calculation [106] and a Hartree-Fock (HF) calculation [96]. The calculated frequency of the lowest mode, *i.e.*, the  $B_{3u}$  mode, is problematic. The general agreement of the  $B_{3u}$  mode as calculated by the simulation and from the dynamical matrix is reassuring. Moreover, the real space calculations agree with the HF value to within  $\sim 20\text{-}30 \text{ cm}^{-1}$ . On the other hand, the LCAO method yields a value which is 50 – 70% smaller than either the real space or HF calculations. The origin of this difference is not apparent. For a poorly converged basis, vibrational frequencies are often overestimated as opposed to the LCAO result which underestimates the value, at least when compared to other theoretical techniques. Setting aside the issue of the  $B_{3u}$  mode, the agreement between the measured Raman modes and theory for  $\text{Si}_4$  suggests that Raman spectroscopy can provide a

key test for the structures predicted by theory.

#### §.6.4 Polarizabilities

Recently polarizability measurements [107] have been performed for small semiconductor clusters. The polarizability tensor,  $\alpha_{ij}$ , is defined as the second derivative of the energy with respect to electric field components. For a noninteracting quantum mechanical system, the expression for the polarizability can be easily obtained by using second order perturbation theory where the external electric field,  $\mathcal{E}$ , is treated as a weak perturbation.

Within the density functional theory, since the total energy is not the sum of individual eigenvalues, the calculation of polarizability becomes a nontrivial task. One approach is to use density functional perturbation theory which has been developed recently in Green's function and variational formulations [108, 109].

Another approach, which is very convenient for handling the problem for *confined* systems, like clusters, is to solve the full problem exactly within the one electron approximation. In this approach, the external ionic potential  $V_{\text{ion}}(\mathbf{r})$  experienced by the electrons is modified to have an additional term given by  $-e\mathcal{E} \cdot \mathbf{r}$ . The Kohn-Sham equations are solved with the full external potential  $V_{\text{ion}}(\mathbf{r}) - e\mathcal{E} \cdot \mathbf{r}$ . For quantities like polarizability, which are derivatives of the total energy, one can compute the energy at a few field values, and differentiate numerically. Real space methods are suitable for such calculations on confined systems as the position operator  $\mathbf{r}$  is not ill-defined, as

is the case for supercell geometries.

There is another point that should be emphasized. It is difficult to determine the polarizability of a cluster or molecule owing to the need for a complete basis in the presence of an electric field. Often polarization functions are added to complete a basis and the response of the system to the field can be sensitive to the basis required. In both real space and plane wave methods, the lack of a “prejudice” with respect to the basis are considerable assets. The real space method implemented with a uniform grid possesses a nearly “isotropic” environment with respect to the applied field. The response can be easily checked with respect to the grid size by varying the grid spacing. Typically, the calculated electronic response of a cluster is not sensitive to the magnitude of the field over several orders of magnitude. In Fig. §.13, we illustrate the calculated polarizability as a function of the finite electric fields. For very small fields, the polarizability calculated by the change in dipole or energy is not reliable because of numerical inaccuracies such as roundoff errors. For very large fields, the cluster can be ionized by the field and again the accuracy suffers. However, for a wide range of values of the electric field, the calculated values are stable.

In Table §.2, we present some recent calculations for the polarizability of small Si and Ge clusters. It is interesting to note that some of these clusters have permanent dipoles. For example,  $\text{Si}_6$  and  $\text{Ge}_6$  both have nearly degenerate isomers. One of these isomers possesses a permanent dipole, the other does not. Hence, in principle, one might be able to separate the one

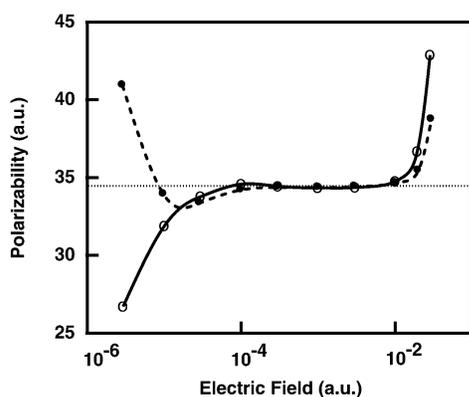


Figure §.13: One component of the polarizability tensor of  $\text{Si}_7$  as a function of the electric field. The dashed curve is the polarizability component from the second derivative of the energy with respect to the field; the solid curve is from the dipole derivative. For very small fields, the values are not accurate owing to the strict convergence criteria required in the wave functions to get accurate values. For high fields, the cluster is ionized. The dashed value is the predicted value for the cluster using a small, but finite field.

Table §.2: Static dipole moments and average polarizabilities of small silicon and germanium clusters.

Silicon			Germanium		
cluster	$ \mu $ (D)	$\langle\alpha\rangle$ ( $\text{\AA}^3/atom$ )	cluster	$ \mu $ (D)	$\langle\alpha\rangle$ ( $\text{\AA}^3/atom$ )
Si <sub>2</sub>	0	6.29	Ge <sub>2</sub>	0	6.67
Si <sub>3</sub>	0.33	5.22	Ge <sub>3</sub>	0.43	5.89
Si <sub>4</sub>	0	5.07	Ge <sub>4</sub>	0	5.45
Si <sub>5</sub>	0	4.81	Ge <sub>5</sub>	0	5.15
Si <sub>6</sub> (I)	0	4.46	Ge <sub>6</sub> (I)	0	4.87
Si <sub>6</sub> (II)	0.19	4.48	Ge <sub>6</sub> (II)	0.14	4.88
Si <sub>7</sub>	0	4.37	Ge <sub>7</sub>	0	4.70

isomer from the other via an inhomogeneous electric field.

### §.6.5 Optical spectra

The time dependent density functional formalism (Section §.3.4) is easy to implement in real space within the higher-order finite difference pseudopotential method [72, 73, 81]. The time dependent local density approximation (TDLDA) technique will be illustrated by considering the absorption spectra of sodium and hydrogenated silicon clusters. The ground-state structures of the clusters were determined by simulated annealing [102]. In all cases the obtained cluster geometries agreed well with the structures reported in other works [110, 111]. Since the wave functions for the unoccupied electron states are very sensitive to the boundary conditions, these calculations need to be performed within a relatively large boundary domain.

The calculated absorption spectrum for Na<sub>4</sub> is shown in Fig. §.14 along

with experiment. In addition, the spectrum generated by considering transitions between the LDA eigenvalues is shown. The agreement between TDLDA and experiment is remarkable, especially when contrasted with the LDA spectrum. TDLDA correctly reproduces the experimental spectral shape, and the calculated peak positions agree with experiment within 0.1 – 0.2 eV. The comparison with other theoretical work demonstrates that our TDLDA absorption spectrum is as accurate as the available CI spectra [112,113]. Furthermore, the TDLDA spectrum for the  $\text{Na}_4$  cluster appears to be in better agreement with experiment than the GW absorption spectrum calculated in Ref. [114].

The study of optical excitations in hydrogen terminated silicon clusters is essential for understanding absorption and emission of visible light in porous silicon [1]. Over the last decade,  $\text{Si}_n\text{H}_m$  clusters have been the subject of intensive experimental [116–120] and theoretical [121–127] research. However, disagreements among different theoretical models used for describing electronic excitations in these systems remain a subject of significant controversy. For the most part, the disagreements arise from the formulation of the optical gap in confined systems and the calculation of different components comprising the optical gap [128–131].

A common approach to hydrogenated silicon clusters is to consider  $\text{Si}_n\text{H}_m$  clusters where the arrangement of the silicon atoms corresponds to bulk silicon fragments. This is illustrated in Fig. §.15. The calculated absorption spectra of  $\text{Si}_n\text{H}_m$  clusters are shown in Fig. §.16. For the larger clusters

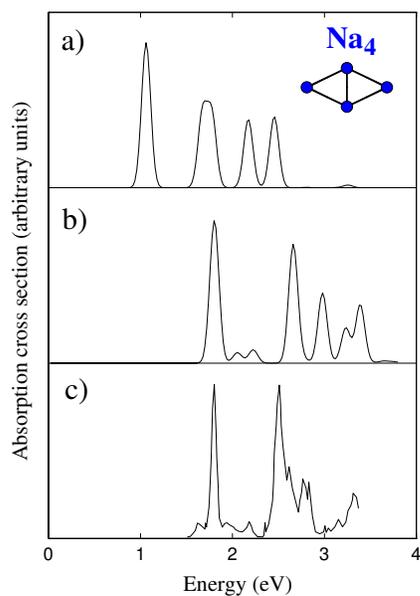


Figure §.14: The calculated and experimental absorption spectrum for  $\text{Na}_4$ . (a) shows a local density approximation to the spectrum using Kohn-Sham eigenvalues. (b) shows a TD-LDA calculation. Technical details of the calculation can be found in [32]. (c) panel is experiment from [115].

shown, only electronic transitions below a chosen energy threshold are displayed owing to computational constraints. The spectra of time-independent Kohn-Sham LDA eigenvalues will be illustrated here. As in the case of metallic and semiconductor clusters with free surfaces [32, 132, 133], the TDLDA spectra of  $\text{Si}_n\text{H}_m$  clusters are blue-shifted with respect to the Kohn-Sham eigenvalue spectra. Unlike optical spectra of “bare” semiconductor clusters considered in previous section, the spectra of hydrogenated silicon clusters do not display low energy transitions associated with the surface states. Photoabsorption gaps for  $\text{Si}_n\text{H}_m$  clusters are much larger than those of  $\text{Si}_n$  clusters with open surfaces.

In Table §.3, TDLDA values for the excitation energies of the first three  $\text{Si}_n\text{H}_m$  clusters are compared with experimental data [116, 126] as well as with the values calculated using the Bethe-Salpeter technique [121]. The last column in Table §.3 shows the Kohn-Sham LDA “ionization” energies of the clusters,  $-\epsilon_{\text{HOMO}}^{\text{LDA}}$ , given by the negative values of the energies for the highest occupied LDA electronic orbitals. Table §.3 demonstrates that the calculated TDLDA excitation energies for the transitions below, or close to  $-\epsilon_{\text{HOMO}}^{\text{LDA}}$  agree well with the experimental data and the Bethe-Salpeter values. The agreement, however, deteriorates for higher excitations, which lie above  $-\epsilon_{\text{HOMO}}^{\text{LDA}}$ . As the size of clusters increases, the energy of the first allowed excitation moves further down from the LDA “ionization” energy, and agreement with experiment improves.

For large  $\text{Si}_n\text{H}_m$  clusters, the first allowed optical transitions are always

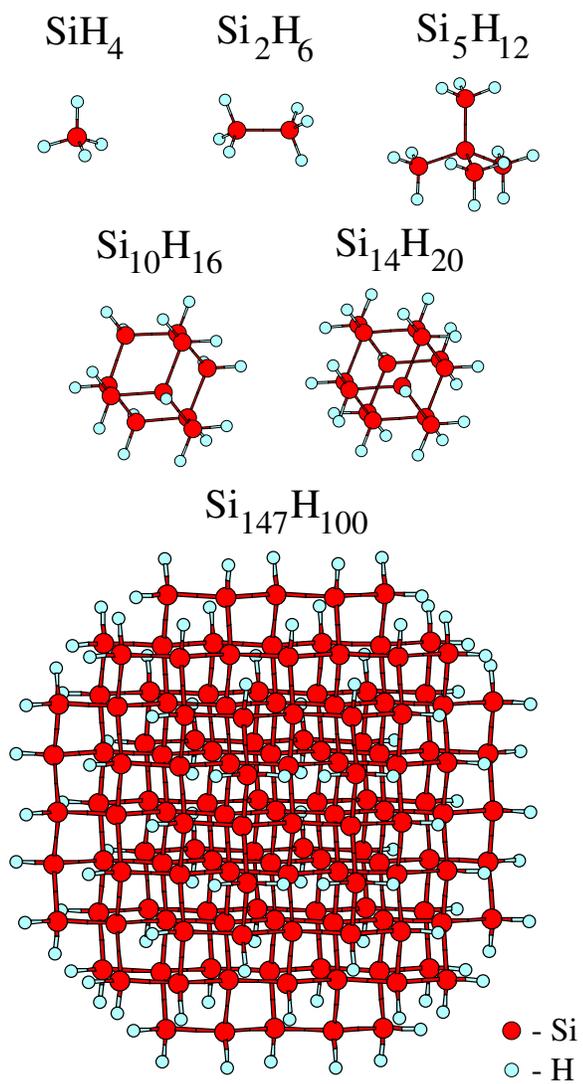


Figure §.15: Ball and stick models for hydrogenated silicon clusters.

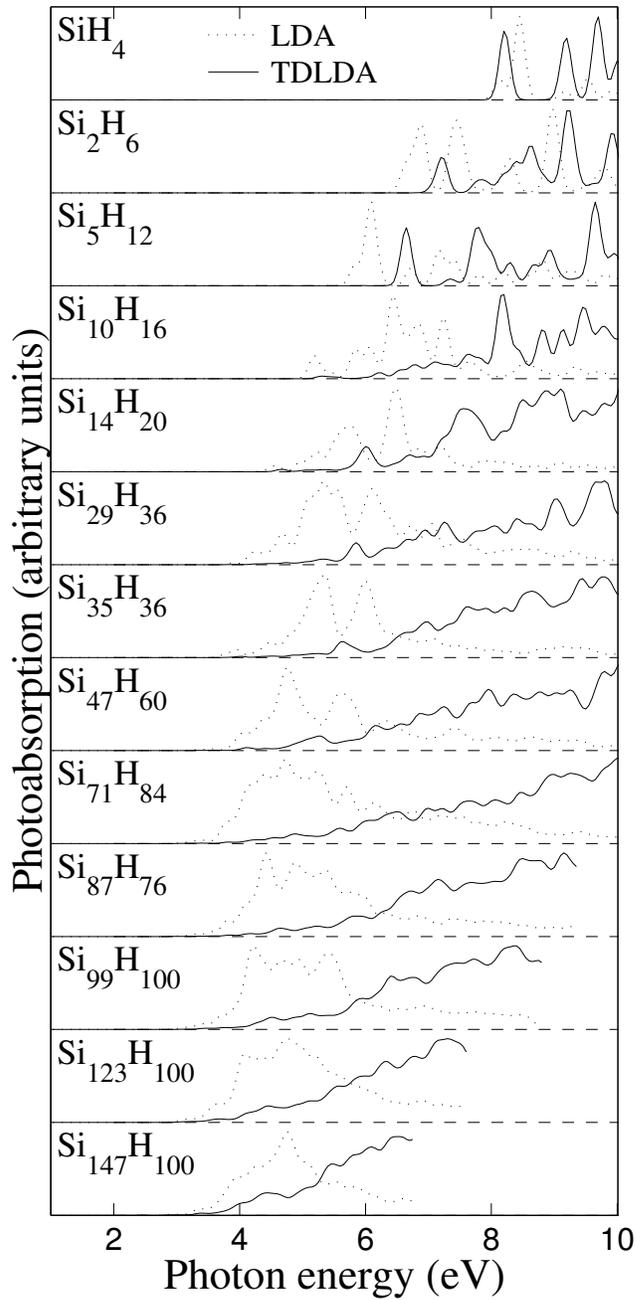


Figure §.16: Calculated TDLDA absorption spectra of  $\text{Si}_n\text{H}_m$  clusters (solid lines). Spectra of time-independent Kohn-Sham LDA eigenvalues (dotted lines) are shown for comparison. All spectra are broadened by 0.1 eV using a Gaussian convolution.

located below  $-\epsilon_{\text{HOMO}}^{\text{LDA}}$ . On this basis, one can argue that TDLDA should provide an accurate description for the photoabsorption gaps and the low energy optical transitions in larger  $\text{Si}_n\text{H}_m$  clusters.

Table §.3: Excitation energies of hydrogenated silicon clusters. The experimental optical absorption energies are taken from Ref. [116] (silane and disilane), and Ref. [126] (neopentasilane). The assignment of electronic excitations for silane and disilane corresponds to the Rydberg transitions. The Bethe-Salpeter (BS) excitation energies are adapted from Ref. [121].  $-\epsilon_{\text{HOMO}}^{\text{LDA}}$  is the time-independent LDA “ionization” energy. All values are in eV.

Cluster	Transition	Experiment	BS	TDLDA	$-\epsilon_{\text{HOMO}}^{\text{LDA}}$
$\text{SiH}_4$	$4s$	8.8	9.0	8.2	8.6
	$4p$	9.7	10.2	9.2	
	$4d$	10.7	11.2	9.7	
$\text{Si}_2\text{H}_6$	$4s$	7.6	7.6	7.3	7.5
	$4p$	8.4	9.0	7.8	
$\text{Si}_5\text{H}_{12}$	–	6.5	7.2	6.6	7.3

## §.7 Quantum confinement in nanocrystals

*Nanocrystals* are assemblages of atoms at the nanoscale where the atomic positions are characteristic of the crystalline state. Spherical nanocrystals are sometimes called “quantum dots.” Understanding the role of quantum confinement in altering optical properties of nanocrystals made from semiconductor materials is a problem of both technological and fundamental interest. In particular, the discovery of visible luminescence from porous Si [1] has focused attention on optical properties of confined systems. Although there is still debate on the exact mechanism of photoluminescence in porous

Si, there is a great deal of experimental and theoretical evidence that supports the important role played by quantum confinement in producing this phenomenon [119, 134].

Excitations in confined systems, like porous as the building blocks of porous Si, differ from those in extended systems due to quantum confinement. In particular, the components that comprise the excitation energies, such as quasiparticle and exciton binding energies change significantly with the physical extent of the system. So far, most calculations that model semiconductor nanocrystals have been of an empirical nature owing to major challenges to simulate these systems from first principles [120, 122–124, 135–139]. While empirical studies have shed some light on the physics of optical excitations in semiconductor nanocrystal, one often has to make assumptions and approximations that may not be justified. Efficient and accurate *ab initio* studies are necessary to achieve a better microscopic understanding of the size dependence of optical processes in semiconductor quantum dots. A major goal of our work is to develop such *ab initio* methods that handle systems from atoms to dots to crystals on an equal footing.

A problem using empirical approaches for semiconductor quantum dots centers on the transferability of the bulk interaction parameters to the nanocrystalline environment. The validity of this assumption, which postulates the use of fitted bulk parameters in a size regime of a few nanometers, is not clear, and has been questioned in recent studies [138, 139]. More specifically, quantum confinement-induced changes in the self-energy corrections, which

may affect the magnitude of the optical gaps significantly, are neglected in empirical approaches by implicitly assuming a “size-independent” correction that corresponds to that of the bulk. It follows that a reliable way to investigate optical properties of quantum dots would be to model them from first principles with no uncontrolled approximations or empirical data. However, there have been two major obstacles for the application of *ab initio* studies to these systems. First, due to large computational demand, accurate first principles calculations have been limited to small system sizes, which do not correspond to the nanoparticle sizes for which experimental data are available. Second, *ab initio* calculations performed within the local density approximation suffer from the underestimate of the band gap [29].

Recent advances in electronic structure algorithms using pseudopotentials (as outlined earlier) [66,71–82] and computational platforms, and alternative formulations of the optical gaps suitable for confined systems, the above-mentioned challenges for *ab initio* studies of quantum dots can be overcome. In particular, new electronic structure methods [66,71–82], implemented on massively parallel computational platforms allows one to model a cluster of more than 1,000 atoms in a straightforward fashion [140]. Such approaches can be illustrated by focusing on silicon quantum dots such as  $\text{Si}_{705}\text{H}_{300}$ . This system corresponds to spherically bulk-terminated Si clusters passivated by hydrogens at the boundaries, Fig. §.17. Computational details can be found in the literature [140].

For an  $n$ -electron system, the quasiparticle gap  $\varepsilon_g^{\text{qp}}$  can be expressed in

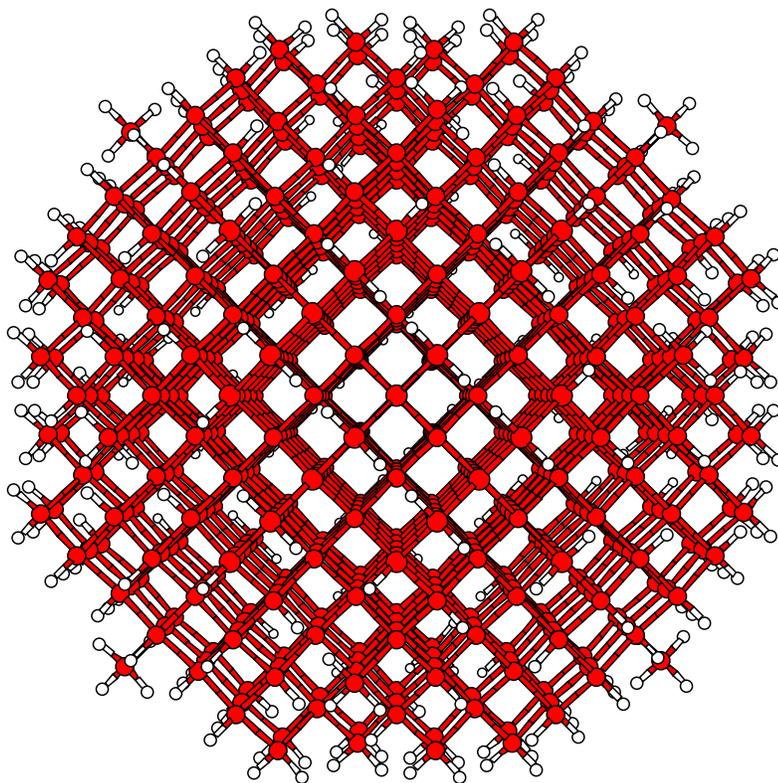


Figure §.17: Atomic structure of a Si quantum dot with composition  $\text{Si}_{525}\text{H}_{276}$ . The gray and white balls represent Si and H atoms, respectively. This bulk-truncated Si quantum dot contains 25 shells of Si atoms and is 27.2 Å in diameter.

terms of the ground state total energies  $E$  of the  $(n + 1)-$ ,  $(n - 1)-$ , and  $n$ -electron systems as

$$\varepsilon_g^{\text{qp}} = E(n + 1) + E(n - 1) - 2E(n) = \varepsilon_g^{\text{HL}} + \Sigma \quad (\S.97)$$

where  $\Sigma$  is the self-energy correction to the HOMO-LUMO gap  $\varepsilon_g^{\text{HL}}$  obtained within LDA. This definition is quite convenient for the calculation of the quasiparticle gap, as it is possible to excite individual electrons or holes from the ground state electronic configuration of a confined system. The calculation of  $\varepsilon_g^{\text{qp}}$  requires the self-consistent solutions of three different charge configurations of each quantum dot. The computational demand of this approach can be reduced significantly by using the wave functions of the neutral cluster to extract very good initial charge densities for the self-consistent solutions of the charged systems. Total energies for charged  $(n + 1)-$  and  $(n - 1)-$ electron systems can be calculated in a straightforward fashion [140].

Eq. (§.97) yields the correct quasiparticle gap  $\varepsilon_g^{\text{qp}}$ , if the *exact* exchange-correlation functional is used. Within the local density approximation, in the limit of very large systems ( $n \rightarrow \infty$ ), the gaps calculated using Eq. (§.97) approach the HOMO-LUMO gap  $\varepsilon_g^{\text{HL}}$  [130]. However, for small systems, Eq. (§.97) captures the correction to the LDA HOMO-LUMO gap quite accurately (Table §.4 ) when compared with available GW calculations [121]. Small deviations appear as the system size reaches approximately 1,000 atoms.

Table §.4: HOMO-LUMO and quasiparticle gaps  $\varepsilon_g^{\text{qp}}$  (Eq. §.97) calculated for hydrogenated Si clusters compared to quasiparticle gaps calculated within the GW approximation [121]. All energies are in eV.

	$\varepsilon_g^{\text{HL}}$	$\varepsilon_g^{\text{qp}}$	GW
SiH <sub>4</sub>	7.9	12.3	12.7
Si <sub>2</sub> H <sub>6</sub>	6.7	10.7	10.6
Si <sub>5</sub> H <sub>12</sub>	6.0	9.5	9.8
Si <sub>14</sub> H <sub>20</sub>	4.4	7.6	8.0

The size dependence of the quasiparticle and LDA HOMO-LUMO gaps, and self-energy corrections are shown in Fig. §.18. Both gap values and self-energy corrections are enhanced substantially with respect to bulk values, and are inversely proportional to the quantum dot diameter  $d$  as a result of quantum confinement. Specifically,  $\varepsilon_g^{\text{qp}}(d) - \varepsilon_{g,\text{bulk}}^{\text{qp}}$ ,  $\varepsilon_g^{\text{band}}(d) - \varepsilon_{g,\text{bulk}}^{\text{band}}$ , and  $\Sigma(d) - \Sigma_{\text{bulk}}$  scale as  $d^{-1.2}$ ,  $d^{-1.1}$ , and  $d^{-1.5}$ , respectively. The quasiparticle gaps shown in the figure are significantly higher compared to the gap values obtained in earlier semi-empirical calculations, although it is problematic in terms of any comparisons. The empirical gaps are obtained from potentials obtained from crystalline environments and scaled to dot sizes. The nature of these gaps is problematic in that they do not strictly correspond to quasiparticle gaps owing to their empirical roots. The main reason for the size difference in the gaps is the significant enhancement of electron self-energies due to quantum confinement, which cannot be properly taken into account in semi-empirical approaches.

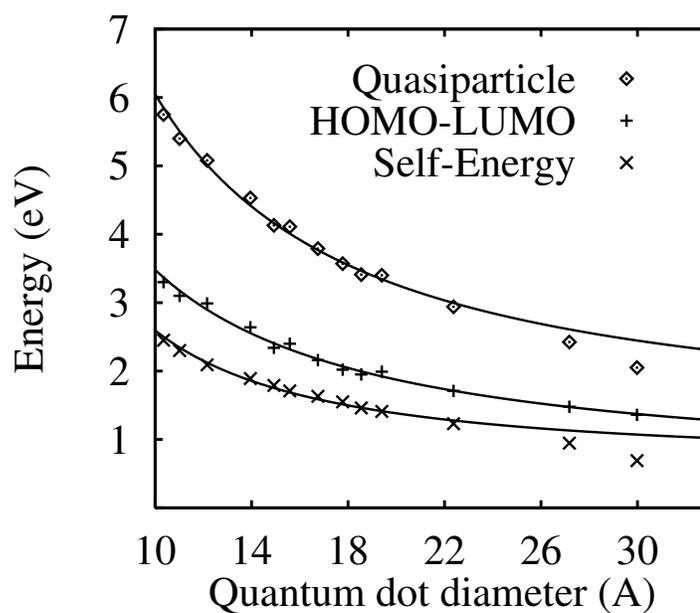


Figure §.18: Calculated quasiparticle ( $\diamond$ ) and HOMO-LUMO gaps (+), and self-energy corrections ( $\times$ ) as a function of the quantum dot diameter  $d$  (in Å). The solid lines are power-law fits to the calculated data approaching the corresponding bulk limits. For small deviations from the fits for large system sizes, see the text.

Since the *quasiparticle* gap refers to the energy to create a *non-interacting* electron-hole ( $e - h$ ) pair, one cannot compare these gaps directly to measurements of the *optical* gap. This issue is especially important for quantum dots in which the exciton radius becomes comparable to the size of the dot. quantum confinement in nanostructures enhances the bare exciton Coulomb interaction, and also reduces the electronic screening so that the exciton Coulomb energy  $E_{\text{Coul}}$  becomes comparable to the quasiparticle gap. In order to extract the optical gaps:

$$\varepsilon_g^{\text{opt}} = \varepsilon_g^{\text{qp}} - E_{\text{Coul}} \quad (\S.98)$$

the exciton Coulomb energy needs to be calculated accurately. Compared to  $E_{\text{Coul}}$ , exciton exchange-correlation energies are much smaller for the quantum dots studied in this work, and will therefore be neglected. Eq.(§.98) is a rigorous expression, provided the Coulomb energy can be properly computed.

A crude, yet commonly used, approximation to  $E_{\text{Coul}}$  comes from the effective mass approximation [141, 142]. Within the effective mass approximation, one assumes (i) an infinite potential barrier at the boundary of the quantum dot, and (ii) envelope wave functions of the form  $\psi(\mathbf{r}) \sim \frac{1}{r} \sin(2\pi r)/d$  for a noninteracting  $e - h$  pair. This yields (in a.u.)  $E_{\text{Coul}} = 3.572/\epsilon d$ . The effective mass approximation, although commonly used, cannot be expected to yield accurate exciton Coulomb energies, since in this approximation the microscopic features of the electron-hole wave functions inside the quantum

dot are neglected, and the wave functions are constrained to vanish abruptly outside the quantum dots, instead of decaying relatively slowly into the vacuum.

Thus,  $E_{\text{Coul}}$  is better calculated directly using *ab initio* pseudo wave functions. The exciton Coulomb energy can be written as

$$\begin{aligned} E_{\text{Coul}} &= \int d\mathbf{r}_1 |\psi_e(\mathbf{r}_1)|^2 V_{\text{scr}}^h(\mathbf{r}_1) \\ &= \int d\mathbf{r}_1 |\psi_e(\mathbf{r}_1)|^2 \int d\mathbf{r} \epsilon^{-1}(\mathbf{r}_1, \mathbf{r}) V_{\text{unscr}}^h(\mathbf{r}) \\ &= \iiint \epsilon^{-1}(\mathbf{r}_1, \mathbf{r}) \frac{|\psi_e(\mathbf{r}_1)|^2 |\psi_h(\mathbf{r}_2)|^2}{|\mathbf{r} - \mathbf{r}_2|} d\mathbf{r} d\mathbf{r}_1 d\mathbf{r}_2. \end{aligned} \quad (\S.99)$$

In this expression,  $V_{\text{scr}}^h$  and  $V_{\text{unscr}}^h$  are screened and unscreened potentials due to the hole,  $\psi_e$  and  $\psi_h$  are the electron and hole wave functions, and  $\epsilon^{-1}$  is the inverse of the microscopic dielectric matrix. One can define  $\tilde{\epsilon}^{-1}$  as

$$\int \epsilon^{-1}(\mathbf{r}_1, \mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}_2|} d\mathbf{r} \equiv \tilde{\epsilon}^{-1}(\mathbf{r}_1, \mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}, \quad (\S.100)$$

then the exciton Coulomb energy can be written as

$$E_{\text{Coul}} = \iint \tilde{\epsilon}^{-1}(\mathbf{r}_1, \mathbf{r}_2) \frac{|\psi_e(\mathbf{r}_1)|^2 |\psi_h(\mathbf{r}_2)|^2}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2. \quad (\S.101)$$

If  $\tilde{\epsilon}$  is taken to be unity, the unscreened  $E_{\text{Coul}}$  can be determined. The results are shown in Fig. §.19 along with the predictions of the EMA and recent empirical calculations [143]. The *ab initio* and empirical calculations

for the unscreened Coulomb energy are in quite good agreement with each other, both predicting smaller Coulomb energies and a softer power-law decay compared to the EMA. In particular, fitting the calculated data to a power law of the diameter as  $d^{-\beta}$ , we find  $\beta = 0.7$ .

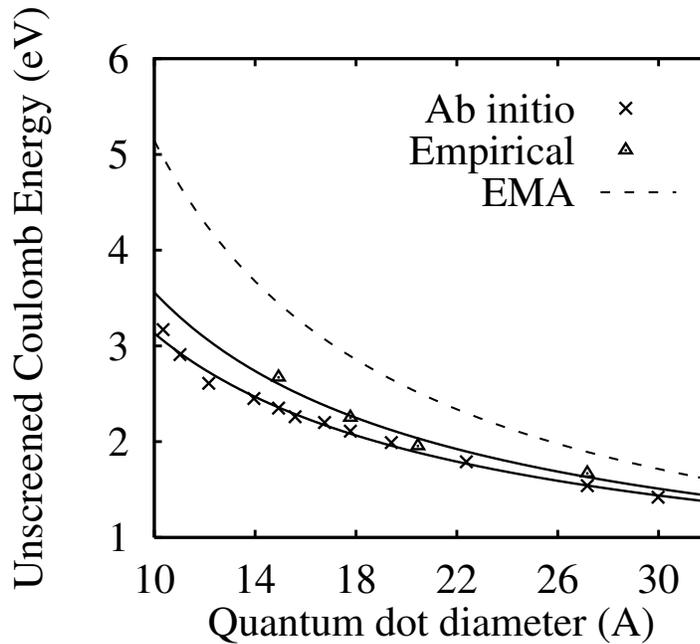


Figure §.19: Unscreened exciton Coulomb energies as a function of the quantum dot diameter  $d$  (in Å) calculated by (i) effective mass approximation (dashed line), (ii) direct empirical pseudopotential calculations ( $\Delta$  from Ref. [143]), (iii) direct *ab initio* pseudopotential calculations ( $\times$ ) as explained in the text. The solid lines are power-law fits to the calculated data.

An accurate calculation of  $E_{\text{Coul}}$  requires the inverse dielectric matrix  $\tilde{\epsilon}^{-1}(\mathbf{r}_1, \mathbf{r}_2)$  in Eq. (§.101). An *ab initio* calculation of  $\tilde{\epsilon}^{-1}(\mathbf{r}_1, \mathbf{r}_2)$  is computationally very demanding, although recent progress has been made on this

problem [144]. Earlier calculations used either the bulk dielectric constant or the reduced dielectric constant of the quantum dot for all  $e-h$  distances. These are rather simple approximations, since the screening is different at different length scales owing to the wave vector dependence of  $\tilde{\epsilon}$ . For example, when  $\mathbf{r}_1$  and  $\mathbf{r}_2$  in Eq. (§.101) are very close to each other, there will be practically no screening, and  $\tilde{\epsilon} \approx 1$ . Typically, both the hole and electron wave functions are well localized towards the center of the quantum dot, the screening will be reduced significantly, resulting in larger Coulomb energies compared to the case of using a single dielectric constant for all distances. Also, extracting the appropriate dielectric *constant* for these small systems is problematic.

One can improve on these approximations by explicitly using the wave functions as calculated by our pseudopotential approach and a realistic dielectric function that takes spatial variations of  $\tilde{\epsilon}$  into account. To calculate the screening dielectric functions  $\tilde{\epsilon}(\mathbf{r}_1, \mathbf{r}_2)$  of a particular quantum dot, one can proceed as follows: First, one applies a spatially modulated electric fields at several wave vectors to calculate the  $q$ -dependent polarizability  $\alpha(\mathbf{q})$  using a finite-field method. The  $q$ -dependent dielectric function  $\tilde{\epsilon}(\mathbf{q})$  can be obtained using a dielectric sphere model [145]. The results for  $\tilde{\epsilon}(\mathbf{q})$  of the  $\text{Si}_{87}\text{H}_{76}$  quantum dot are shown in Fig. §.20. After fitting the calculated  $\tilde{\epsilon}(\mathbf{q})$  to a rational polynomial function of  $q$  and Fourier-transforming to real space [146], we obtained the dielectric function  $\tilde{\epsilon}(r = |\mathbf{r}_1 - \mathbf{r}_2|)$ . Implicitly, we are assuming spatial isotropy in writing  $\tilde{\epsilon}(\mathbf{r}_1, \mathbf{r}_2) \approx \tilde{\epsilon}(r = |\mathbf{r}_1 - \mathbf{r}_2|)$ . As

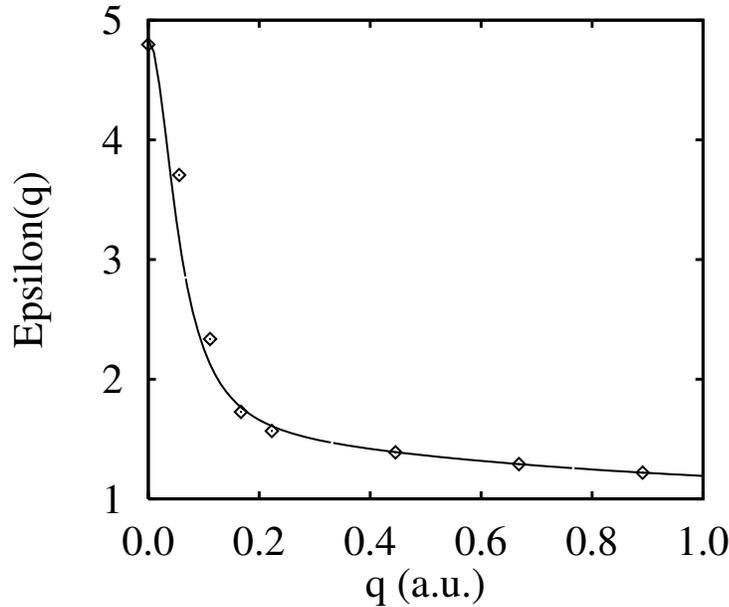


Figure §.20: Wavevector dependence of the dielectric function for the  $\text{Si}_{87}\text{H}_{76}$  quantum dot.

shown in Fig. §.20, the calculated  $\tilde{\epsilon}(\mathbf{q})$  has a very sharp drop to  $\approx 1$  beyond  $q = 0.2$  a.u., which corresponds roughly to the wavevector set by the linear dimension (or diameter) of this quantum dot. This sharp drop is typical for all quantum dots studied. In real space, this implies that the  $e-h$  interaction is very inefficiently screened inside the dot resulting in substantial excitonic Coulomb energies.

The resulting optical gaps  $\epsilon_g^{\text{opt}} = \epsilon_g^{\text{qp}} - E_{\text{Coul}}$  along with the quasiparticle gaps and experimental absorption data [117] from Si:H nanocrystals are shown in Fig. §.21. Although the calculated quasiparticle gaps are  $\sim 0.6$  to  $1.0$  eV larger than the experimental absorption data, the calculated optical gaps are in very good agreement with experiment.

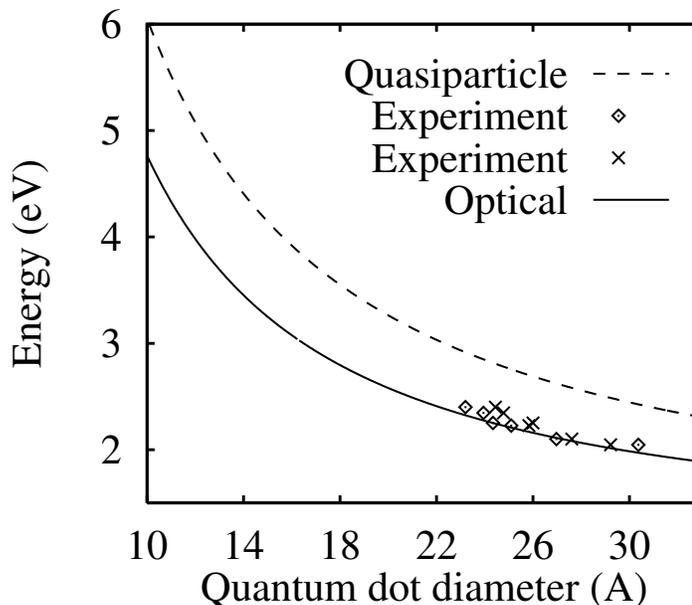


Figure 5.21: Calculated quasiparticle gaps (dotted line), optical gaps, and experimental absorption data from Si:H nanocrystals ( $\times$  and  $\diamond$  from Ref. [117]) as a function of the quantum dot diameter  $d$  (in Å). The two sets of experimental data ( $\times$  and  $\diamond$ ) differ by the method to estimate the nanocrystal size.

At this point, an interesting observation can be made about the good agreement of previous semi-empirical calculations with experiment [134–136]. In the above semi-empirical approaches, it is the underestimate of *both* the excitation gap *and* the exciton Coulomb energies (through the use of a static dielectric constant of either the bulk or the quantum dot), that results in calculated values in agreement with experiment. As a matter of fact, the bare gaps of Refs. [135] and [136] *without* the exciton Coulomb energies are in better agreement with the experiment. The *ab initio* pseudopotential results demonstrate that (i) the quasiparticle gaps in Si quantum dots are

actually higher than previously thought, and (ii) the exciton Coulomb energies, because of the wavevector dependence of the dielectric response function  $\epsilon(\mathbf{r}_1, \mathbf{r}_2)$ , are higher than previously calculated, resulting in optical gap values that are in good agreement with the experimental absorption data.

TDLDA methods can be used to examine hydrogenated silicon quantum dots. For large clusters or quantum dots, the absorption spectra become essentially quasi-continuous and it is incumbent to use care in defining the optical gap. In particular, a large number of low intensity transitions exist near the absorption edge. Taken individually, the oscillator strengths of these transitions would be located far below the experimentally detectable limit. As a result, identifying the first allowed optical transition in the case of large clusters is not a trivial task.

As the size of clusters increases, the absorption gaps gradually decrease, and the discrete spectra for small clusters evolve into quasi-continuous spectra for silicon nanocrystals. Fig. §.22 demonstrates that oscillator strength of dipole-allowed transitions near the absorption edge decreases with increasing cluster size. This fact is consistent with the formation of an *indirect* band gap in the limit of bulk silicon [126].

Rather than associating the optical gaps with the individual transitions, one can define a procedure for fixing the optical gap,  $E_{gap}^{opt}$ , via an integral of the oscillator strength. In particular, the following prescription has been

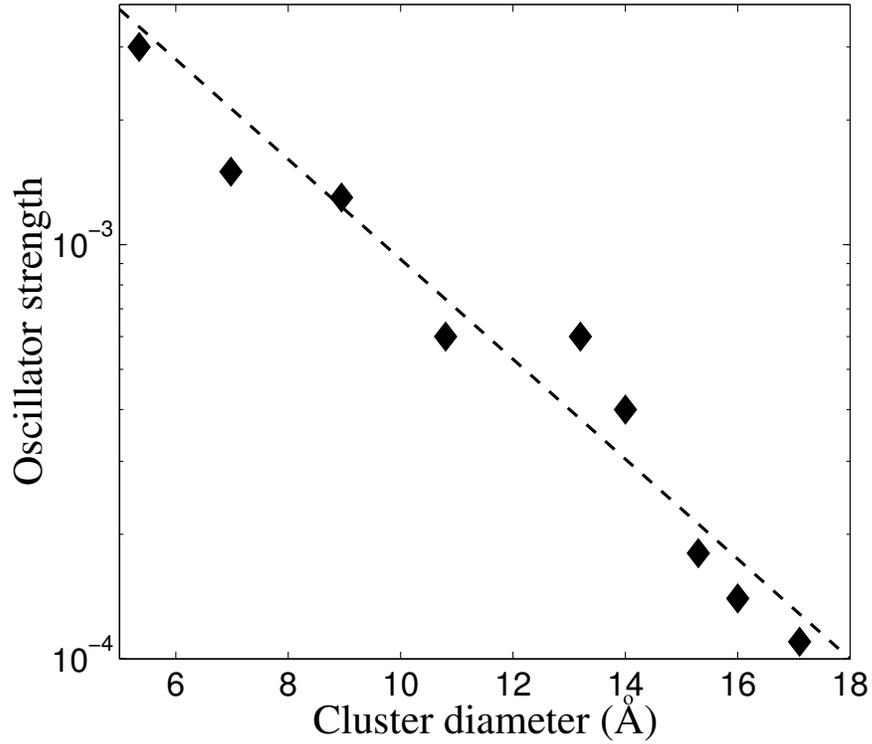


Figure §.22: Oscillator strength for optical transitions in hydrogenated silicon clusters as a function of cluster size. The strength is determined by considering transitions near the gap. The dashed line is a linear fit.

suggested [147] to define the gap:

$$pF = \int_0^{E_{gap}^{opt}} \sigma(\omega) d\omega \quad (\S.102)$$

where  $F$  is the total optical cross section,  $\sigma(\omega)$  is the optical cross section for a given frequency,  $\omega$  and  $p$  is some prescribed fraction of the total cross-section for the fixing the gap. For the photoabsorption gaps, a typical value

of  $p$  might be  $10^{-4}$ . This definition for the absorption gap does not affect the values of the optical gaps for small  $\text{Si}_n\text{H}_m$  clusters, since the intensity of their first allowed transitions is much higher than the selected threshold. An order of magnitude change in  $p$ , does not typically change the gap size by more than  $\pm 10$  meV. At the same time, Eq. §.102 offers a convenient way for the evaluation of optical gaps in large clusters.

The variation of the optical absorption gaps as a function of cluster size is shown in Fig. §.23. Along with the TDLDA values, we include optical gaps calculated by the Bethe-Salpeter (BS) technique [121]. For very small clusters,  $\text{SiH}_4$ ,  $\text{Si}_2\text{H}_6$ , and  $\text{Si}_5\text{H}_{12}$ , the gaps computed by the TDLDA method are close to the Bethe-Salpeter values, although for  $\text{Si}_{10}\text{H}_{16}$  and  $\text{Si}_{14}\text{H}_{20}$  our gaps are considerably smaller than the gaps calculated using the Bethe-Salpeter equation. At the same time, the TDLDA gaps for clusters in the size range from 5 to 71 silicon atoms are larger by  $\sim 1$  eV than the gaps calculated by the Hartree-Fock technique with the correlation correction included through the configuration-interaction approximation (HF-CI) [127].

These differences are consistent with the fact that the BS calculations systematically overestimate and the HF-CI calculations of Ref. [127] underestimate the experimental absorption gaps. For example, for the optical absorption gap of  $\text{Si}_5\text{H}_{12}$  the BS, TDLDA, and HF-CI methods predict the values of 7.2, 6.6, and 5.3 eV, respectively, compared to the experimental value of 6.5 eV. However, it is not clear whether the gaps of Ref. [127] refer to the optically-allowed or optically-forbidden transitions, which may of-

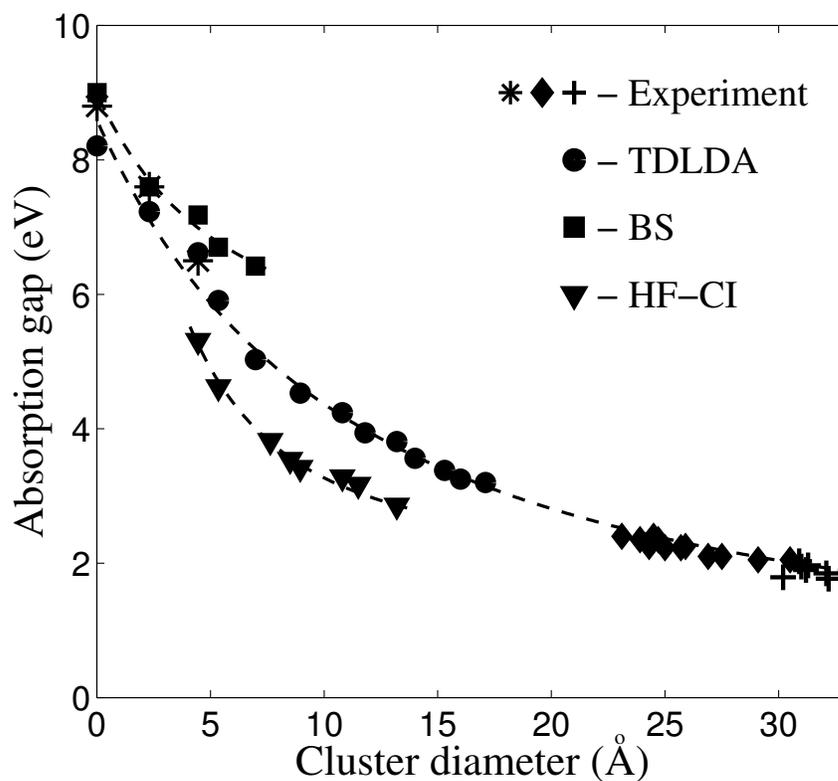


Figure 5.23: Variation of optical absorption gaps as a function of cluster diameter. Theoretical values shown in the plot include the gaps calculated by the TDLDA method (this work), by the Bethe-Salpeter technique (BS) [121] and by the Hartree-Fock method with the correlation included through the configuration-interaction approximation (HF-CI) [127]. Experimental values are taken from Refs. [116–118, 126]. The dashed lines are a guide to the eye.

fer a possible explanation for the observed discrepancy. For large clusters, we find the TDLDA optical gaps to be in generally good agreement with the photoabsorption gaps evaluated by the majority of self-energy corrected LDA [103, 126] and empirical techniques [124, 125, 136]. At present, the full TDLDA calculations for clusters larger than a few nm can exceed the capabilities of most computational platforms. Nevertheless, the extrapolation of the TDLDA curve in the limit of large clusters comes very close to the experimental values for the photoabsorption gaps. Software and hardware advances should make a direct verification of this possible in the near future.

In determining the optical gaps within a linear response approach, only excitations with an induced dipole are incorporated. In real time methods, the induced dipole term is calculated directly [36–38]. Within our frequency domain description, two factors enter in the ascertaining the existence of an induced dipole: the existence of a transition energy and the corresponding oscillator strength. Within TDLDA, these terms can be obtained from Eq. §.58 as  $\Omega_I$  and  $\mathbf{F}_I$ . These terms must always be considered together when predicting optical properties, although sometimes this is not done [26].

In Fig. §.24, we illustrate the lowest transitions without regard to oscillator strength for both LDA and TDLDA calculations. Transitions as defined by Eq. §.102 are shown. For these transitions, the gap is defined when the oscillator strength assumes a value of at least  $10^{-4}$  of the total optical cross-section. The main difference between LDA and TDLDA for these system is a strong blue-shift of the oscillator strength.

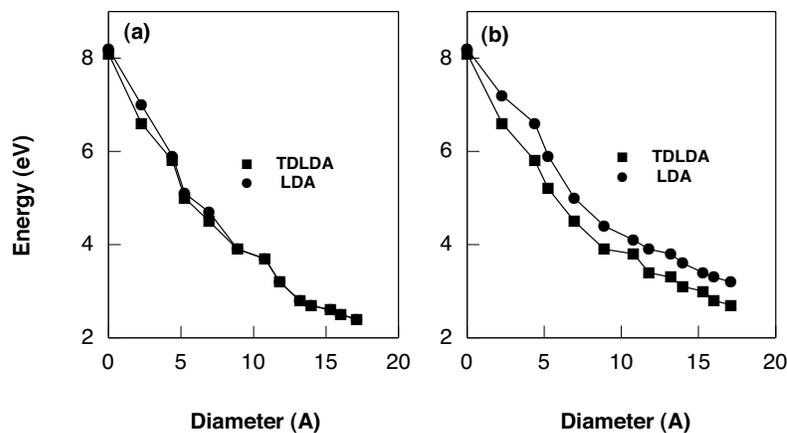


Figure 8.24: Gaps determined from LDA and TDLDA for hydrogenated silicon clusters. (a) The gaps plotted without regard to the oscillator strength. (b) The gaps determined using the criteria from Eq. 8.102.

It should be noted that *real time* methods for TDLDA do not involve unoccupied eigenvalues [36–38]. In this formalism, the absorption spectrum evolves from taking the power spectrum of the instantaneous induced dipole. The resolution of an optical transition is determined by the length of the time integration [36–38]. Since the frequency domain method and the real time method should yield the same spectrum, “virtual transitions,” *i.e.*, transitions that do not couple to the dipole, are not physically meaningful within frequency domain implementation of TDLDA.

### §.7.1 The role of oxygen in silicon quantum dots

Porous and nanocrystalline silicon studied in experiments are prepared under a variety of surface conditions determined by the etching technique and external chemical environments employed. Only a fraction of published experimental data refers to "pure" hydrogenated silicon dots [117]. Other measurements are performed on partially oxidized nanocrystals [148, 149]. For many cases, a precise chemical composition of nanocrystalline surfaces is not known [118, 150, 151]

However, most calculations for optical absorption and emission in silicon dots do not take into account differences in structure and chemical composition of the dot surface. This creates an ambiguity in the interpretation of experimental data. Almost all *ab initio* and empirical simulations available in literature use silicon dots passivated with hydrogen [103, 121, 122, 125–127, 136, 152], although some notable exceptions exist [153, 154]. This limitation also is true for structural issues, where only a few systems have been examined for reconstructed surfaces [155, 156].

Theoretical calculations [103, 121, 122, 125–127, 136, 152] based on a quantum confinement model show general agreement with experimental measurements [117] for optical absorption in hydrogen-passivated silicon clusters. In contrast, experiments performed on oxidized samples often display photoluminescence with energies significantly below the values of optical gaps predicted by the confinement model for clusters in the same size range [148, 149]. This disagreement could be greater than 1 eV. It has been suggested that the

onset of photoluminescence in silicon nanocrystals may be associated with the optical Stokes shift [157] and excitonic exchange splitting [158]. While these effects could be significant in small silicon dots, it appears that neither the Stokes shift, nor the excitonic exchange splitting alone could explain such a large disagreement between experiment and theory.

Recent experimental data present strong evidence that surface effects produce a very substantial impact on the electronic and optical properties of nanocrystalline silicon. Specifically, Wolkin *et al.* observed a large redshift of photoluminescence in porous silicon after exposure to open air [119]. The study reported a shift of photoluminescence of the order of 1 eV for samples composed of crystallites smaller than 2 nm in size. The observed redshift has been attributed to surface oxidation of silicon nanocrystals. According to the interpretation proposed in Ref. [119], oxygen creates trapped electron and hole states on nanocrystalline surfaces. The trapped surface states reduce the effective size of the optical gap. This mechanism can explain the difference between the energy of the measured photoluminescence and theoretical predictions based on the quantum confinement model.

Owing to a very large number of possible configurations for oxidized silicon clusters, current studies are often limited to the case of a single oxygen atom attached to the cluster surface. Oxidized clusters were prepared from regular hydrogen-terminated spherical dots by replacing two hydrogen atoms on the surface with a single atom of oxygen, followed by relaxation of all interatomic forces. The model geometries for oxidized clusters are illustrated

in Fig. §.25.

The calculated absorption spectra of oxidized silicon dots are shown in Figs. §.27 and §.28. In Fig. §.27 the spectra of small oxidized and nonoxidized clusters are illustrated. The addition of oxygen creates new absorption bands in the region of lower transition energies. Optical excitations with higher energies are also affected by oxidation, although some intense absorption peaks observed in nonoxidized clusters (such as the peaks at 6.6 and 7.8 eV for  $\text{Si}_5\text{H}_{12}$ ) appear to be only slightly shifted. Fig. §.28 shows the calculated spectra of the  $\text{Si}_{29}\text{OH}_{34}$  and  $\text{Si}_{35}\text{OH}_{34}$  clusters.

The overall change in optical absorption caused by the addition of a single oxygen atom is less apparent larger clusters. To make the effect of oxidation more evident, we plot in Fig. §.28 the differential spectra calculated as the difference in optical absorption of the same cluster before and after oxidation. Positive values of differential photoabsorption correspond to the new absorption peaks that appear only after oxidation. The differential absorption spectra for  $\text{Si}_{29}\text{OH}_{34}$  and  $\text{Si}_{15}\text{OH}_{34}$  clearly show the presence of low-energy optical transitions associated with surface oxygen. The calculated optical absorption gaps in oxidized and nonoxidized silicon dots are compared in Fig. §.26. The TDLDA gaps for nonoxidized  $\text{Si}_7\text{H}_7$  clusters are adapted from our previous work [152]. The spectra are essentially quasi-continuous and exhibit a large number of low-intensity transitions near the absorption edge. As such, the effective optical gaps were evaluated at a very small but nonzero fraction of the complete electronic oscillator strength as

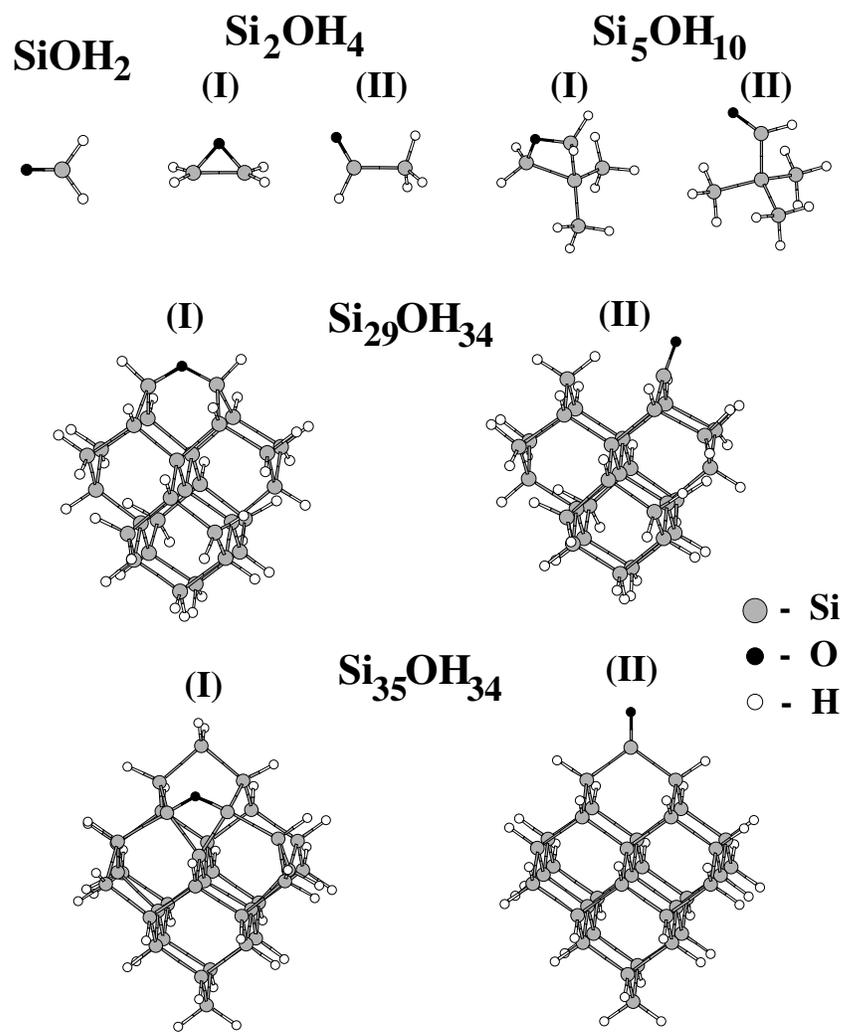


Figure §.25: Model geometries for hydrogenated silicon with oxygen.

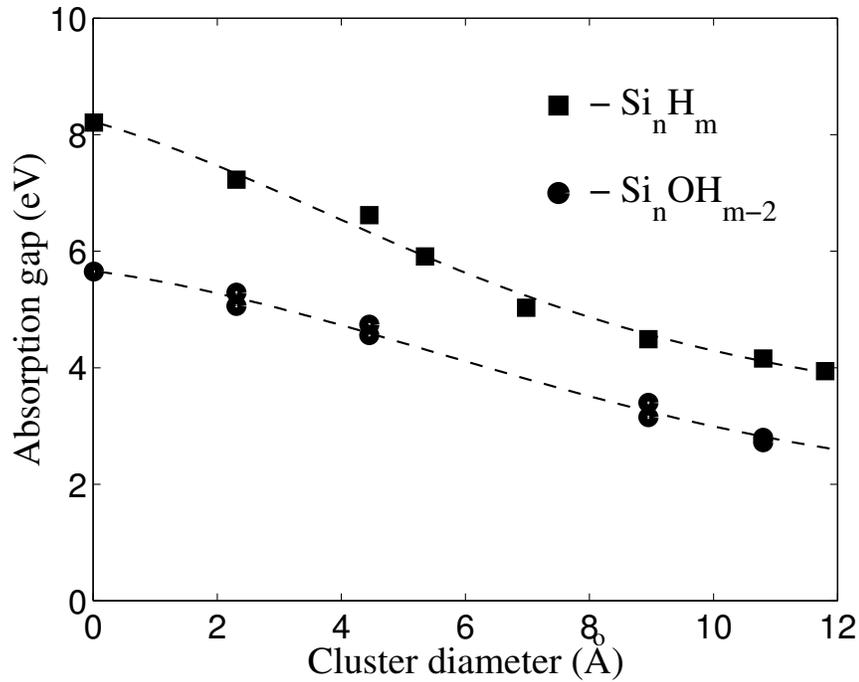


Figure §.26: Comparison between the optical absorption gaps of regular and oxidized hydrogen-terminated silicon clusters. The gaps for  $\text{Si}_n \text{H}_m$  clusters are adapted from Ref. [152]. The dashed lines are a guide to the eye.

in Eq. §.102. The same criterion in defining the gap for silicon quantum dots was used for the oxidized silicon clusters. Fig. §.26 demonstrates that surface oxidation reduces optical gaps in hydrogenated silicon clusters by as much as 1-2 eV. The change in the size of optical gaps is consistent with the redshift of photoluminescence observed in Ref. [119] and is likely responsible for the disagreement between experimental photoluminescence from oxidized silicon nanocrystals and theoretical estimates based on the quantum confinement model.

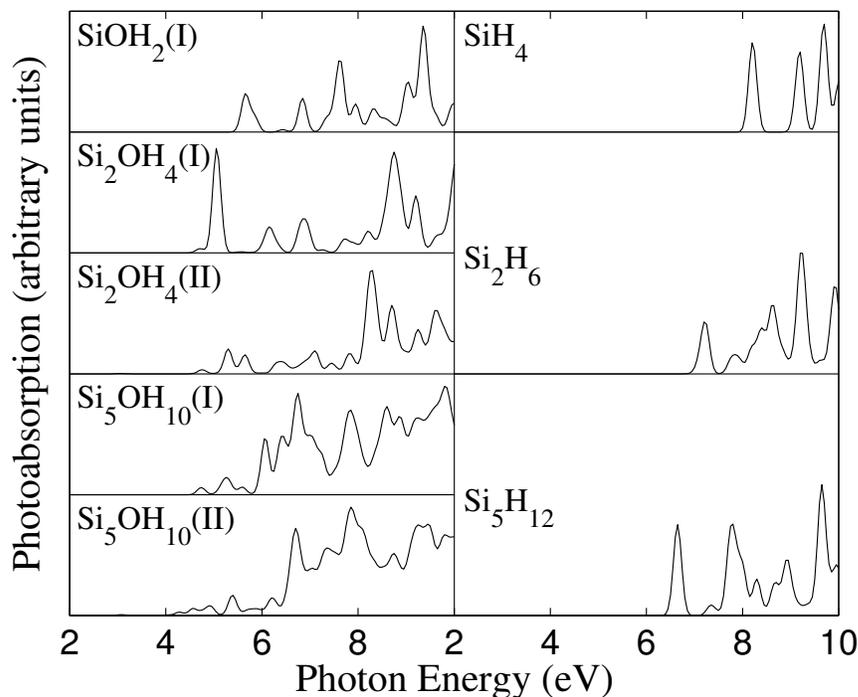


Figure 3.27: Left: calculated TDLDA absorption spectra of oxidized hydrogen-terminated silicon clusters. Right: TDLDA spectra of nonoxidized clusters. All spectra were broadened by 0.1 eV using a Gaussian convolution.

A surprising result of oxygen absorption is the small difference observed in the optical gaps between cluster isomers with Si=O and Si-O-Si bonds on the surface. At the same time, Figs. 3.27 and 3.28 reveal substantial differences in the shape of optical spectra for these clusters. One can understand this difference by examining the mechanism of the gap formation in two selected clusters:  $\text{Si}_{35}\text{OH}_{34}$  (I) and (II). The order of electronic levels near the gap is illustrated for both isomers in Figs. 3.29 and 3.30, respectively.

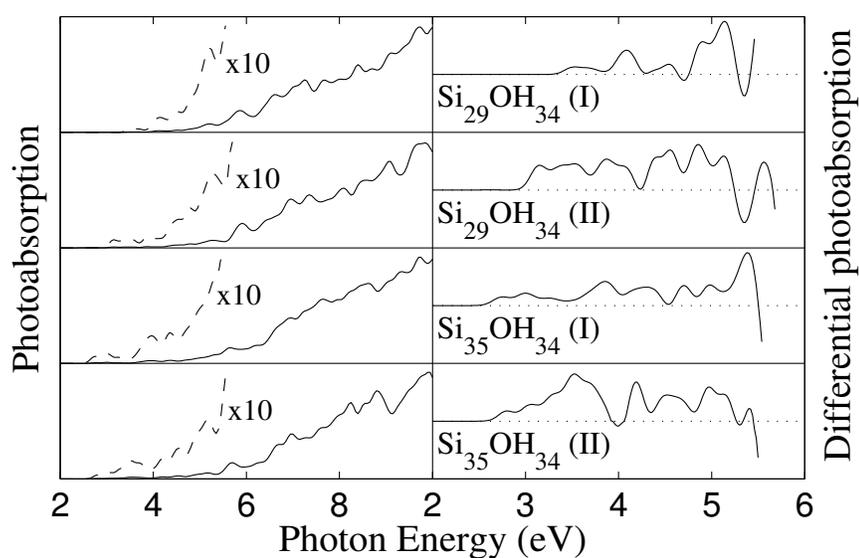


Figure §.28: Left: calculated TDLDA absorption spectra of  $\text{Si}_{29}\text{OH}_{34}$  and  $\text{Si}_{35}\text{OH}_{34}$  clusters. Right: difference in optical absorption between clusters with and without oxygen on the surface. All spectra are broadened by 0.1 eV.

These diagrams represent simplified schemes that show only the dominant single-electron Kohn-Sham transitions within the TDLDA description and do not account for correlations among individual excitations. The energies of optical transitions shown in these figures correspond to one-electron singlet TDLDA excitations [32]. They differ from transition energies of the TDLDA optical spectra shown in Figs. §.27 and §.28, which correspond to collective electronic excitations. Nevertheless, the single-electron diagrams are useful for the qualitative analysis of optical transitions in oxidized silicon dots.

The authors of Ref. [119] proposed that photoluminescence in small oxidized silicon clusters occurs between the trapped electron and hole states, both of which are associated with the double Si=O bond on the cluster surface. Specifically, the trapped electron state is a p-state localized on silicon and the trapped hole state is a p-state localized on oxygen. Spatial distributions of electron densities for the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO) of the  $\text{Si}_{35}\text{OH}_{34}$  (II) cluster plotted in Fig. §.27 confirm that these states are indeed represented by p- states mainly localized on the silicon and oxygen atoms. However, the distributions of HOMO and LUMO electron densities for the  $\text{Si}_{35}\text{OH}_{34}$  (I) cluster shown in Fig. §.28 reveal a different picture. The LUMO state is, for the most part, localized on two silicon atoms that form the Si-O-Si bonds. At the same time, the HOMO state is not localized on the oxygen atom. Instead, this electronic state is spread among the layers of silicon atoms surrounding the Si-O-Si fragment. In both cases, the direct

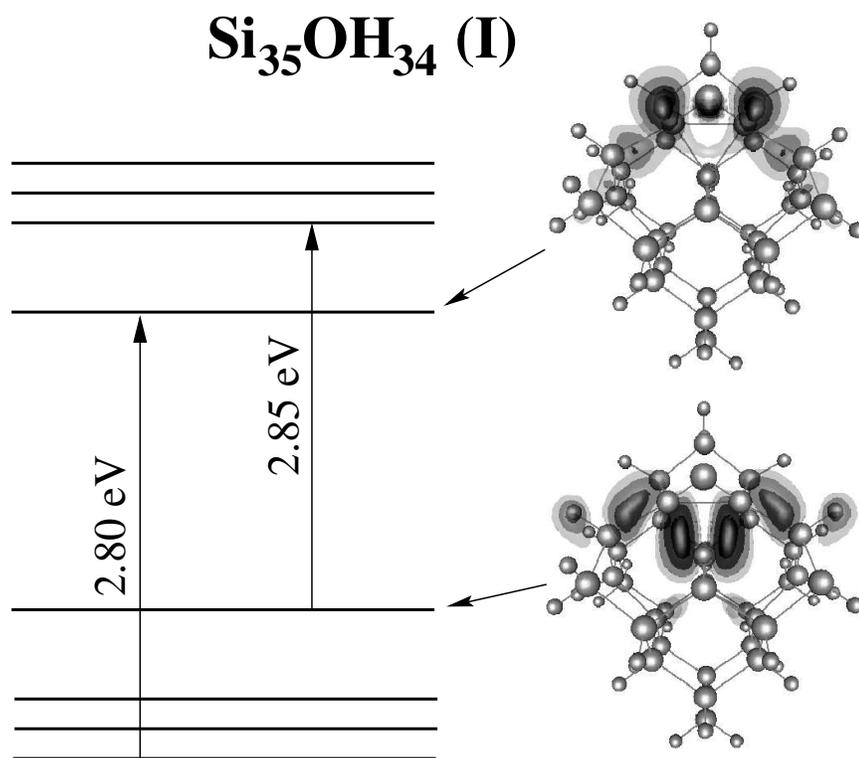


Figure §.29: Schematic representation of electronic levels in the vicinity of the gap for Si<sub>35</sub>OH<sub>34</sub> (I) clusters. Spatial distributions of electron densities are shown for the HOMO and LUMO states.

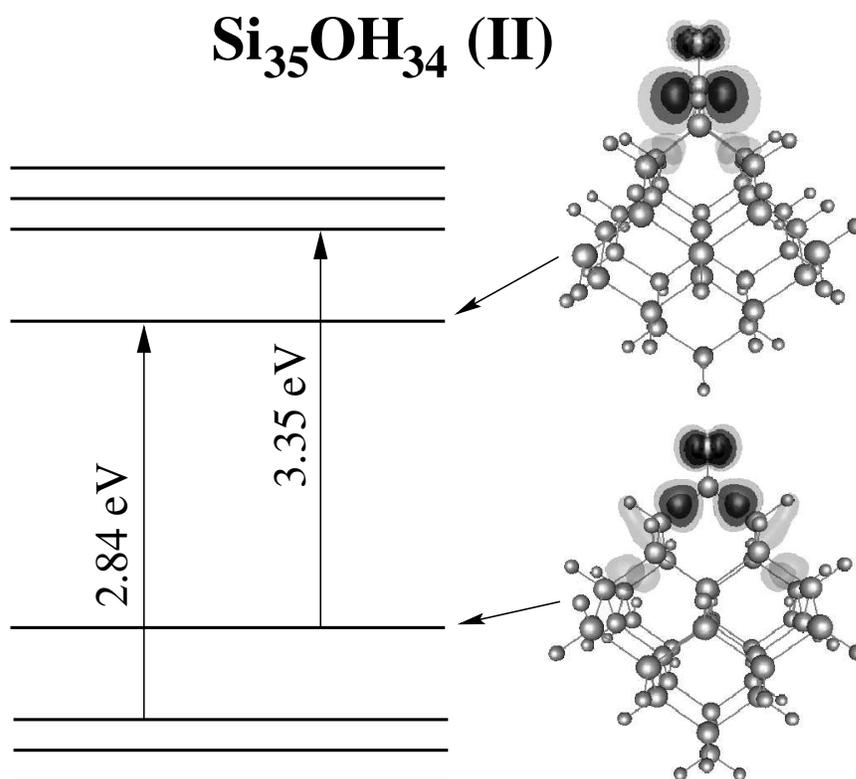


Figure §.30: Schematic representation of electronic levels in the vicinity of the gap for Si<sub>35</sub>OH<sub>34</sub> (II) clusters. Spatial distributions of electron densities are shown for the HOMO and LUMO states.

dipole transitions between the HOMO and LUMO states are forbidden. The absorption edge for  $\text{Si}_{35}\text{OH}_{34}$  (II) is formed mainly by transitions from lower occupied orbitals to the LUMO state. For this cluster, transitions from the HOMO state to higher unoccupied orbitals do not contribute to optical absorption near the gap. For  $\text{Si}_{35}\text{OH}_{34}$  (I), however, both of these types of electronic transitions are involved in the formation of the absorption edge.

Such calculations show that even a low concentration of oxygen on the surface can substantially alter the optical properties of silicon nanoclusters. However, experimental studies are not always limited to clusters with low oxygen content. Some limited studies have been performed on dots with a higher concentration of oxygen, e.g,  $\text{Si}_{35}\text{O}_6\text{H}_{24}$ . This cluster was prepared from the hydrogen-terminated dot  $\text{Si}_{35}\text{H}_{36}$  by replacing 12 outer-shell hydrogen atoms with oxygen to form six double Si=O bonds at the positions symmetrically equivalent to that shown in Fig. §.25 for  $\text{Si}_{35}\text{OH}_{34}$  (II). The increase in oxygen coverage caused a further reduction of the absorption gap to 2.4 eV. This value was approximately 0.4 eV lower than the absorption gap for  $\text{Si}_{35}\text{OH}_{34}$  (II), and almost 1.6 eV lower than the gap for the nonoxidized cluster  $\text{Si}_{35}\text{H}_{36}$ . The principal mechanism of gap formation for  $\text{Si}_{35}\text{O}_6\text{H}_{24}$  appears to be similar to that for  $\text{Si}_{35}\text{OH}_{34}$ . The additional reduction of the absorption gap in case of  $\text{Si}_{35}\text{O}_6\text{H}_{24}$  could be explained by interactions among oxygen-induced electronic states. The absorption gap for  $\text{Si}_{35}\text{OH}_{34}$  is reduced by the presence of localized oxygen-induced levels. In the limit of large clusters, the positions of these levels should be essentially independent of the

cluster size. Since the gaps in silicon dots decrease with increasing cluster size as a result of diminishing quantum confinement, at some point the oxygen-induced states are expected to cross over the electronic levels from the body of the cluster. After this point, the oxygen-induced states would no longer be located inside the gap. Calculations suggest that depending on the fraction of oxygen coverage, the oxygen-induced states should not cross over the levels from the body of the cluster for silicon dots up to approximately 20-25 Å in diameter [153]. For larger dots, the overall effect of surface oxidation on the optical properties is likely to be less important.

### §.7.2 Doping quantum dots

Electronic and optical properties of semiconductor nanostructures are strongly affected by quantum confinement due to the reduced dimensionality of these systems [159]. In nanocrystals or quantum dots, where motion of electrons (or holes) is limited in all three dimensions, quantum confinement results in a strong increase of the optical excitation energies when compared to the bulk. One expects that other electronic and optical properties such as the dielectric properties will be affected as well.

In bulk semiconductors, shallow donors (or acceptors) are crucial in determining the transport properties required to construct electronic devices. However, these properties are expected to be significantly altered in highly confined systems such as quantum dots. As a consequence, important questions exist as to whether dopants will continue to play a role similar to that

in bulk semiconductors and on whether new applications such as quantum computation [160, 161] will become possible in the nano regime.

Experimental studies of shallow impurities in quantum dots, such as P in Si nanocrystals, have been slow to address such issues. In part, this is due to difficulties in preparation of samples in a controllable manner, *e.g.*, it is hard to ensure that a quantum dot contains only one impurity. For such reasons, only a few experimental studies have focused on doping of quantum dots. These studies have utilized photoluminescence and electron spin resonance measurements most of which have been performed on silicon quantum dots. Increasing the dopant concentration results in distinct changes in its photoluminescence properties such as suppression of the signal [162] and a blue-shift of photoluminescence maxima with decreasing particle size in heavily *p*-doped porous silicon [163]. It is also not clear whether or not the doping of Si nanocrystals provides a generation of free charge carriers in these systems [163, 164].

Electron spin resonance measurements are a popular tool for examining impurities in semiconductors and have recently been applied to these systems. Spin resonance experiments determine the hyperfine splitting (HFS) of the defect electron levels, which are directly related to localization of the dopant electron density on the impurity site [165]. In Si nanocrystals with radii of 10 nm doped with P, a hyperfine splitting of 110 G has been observed [166]. This splitting is in sharp contrast to the bulk value of 42 G. A size dependence of the HFS also exists in Si dots with radii around 50 nm [167], although in

this case it is likely influenced by an asymmetrical shape of Si crystalites. Recently, a strong size dependence of the HFS of P atoms was observed in much smaller nanocrystallites with radii of 2 - 3 nm [168].

Theoretical studies of shallow impurities in quantum dots have also lagged relative to calculations for macroscopic systems. The large number of atoms and low symmetry hinder such studies. Some empirical studies have been performed for impurities in quantum dots [138, 169, 170]. These calculations involve various parameters, which are commonly assumed to have bulk-like values. A common drawback in these studies is the use of a generic hydrogen-like potential to model the impurity atom.

The real-space, *ab initio* pseudopotential density functional method [171] has been applied to the electronic properties of a single phosphorus impurity in a hydrogenated Si quantum dot containing hundreds of atoms [172]. The nanocrystals were modeled as spherical, bulk-terminated Si clusters whose surface is passivated by hydrogen atoms. One silicon atom is substituted by a phosphorus atom.

P, Si and H atoms were modeled using *ab initio* Troullier-Martins pseudopotentials [63, 65]. Parameters for the Si and H pseudopotentials and other technical details are given elsewhere [171, 172].

Several substitutional geometries for the P atom have been explored, *e.g.*, a P placed at the center of the dot, off center, and on the surface. No significant relaxations of the Si atoms were found in the vicinity of the P atom. The largest change in position occurred when the P was positioned off

center. In this case, the P atom was shifted by about 0.1 a.u. in the outward direction.

In contrast to supercell approach, real space method allows one to examine charged clusters in a straightforward manner [171]. Ionization energies  $I_d$  for P-doped nanocrystals and affinity energies  $A_p$  for pure Si nanocrystals have been determined using charged systems:

$$I_d = E(n - 1) - E(n), \quad (\S.103)$$

$$A_p = E(n) - E(n + 1), \quad (\S.104)$$

where  $E$  is the ground state total energies of the  $n-$ ,  $(n + 1)-$  and  $(n - 1)-$ electron systems. The binding energy  $E_B$  for the donor atom can be calculated as a difference between these two quantities:

$$E_B = I_d - A_p. \quad (\S.105)$$

This definition of the binding energy  $E_B$  corresponds to two separate processes: The doped dot is ionized, *i.e.* the electron is physically removed from the nanocrystal. The affinity energy may be calculated by considering an isolated neutral dot of equal size and adding an electron. A similar approach has been utilized in the tight-binding calculations [138], where the binding energy was calculated as a difference between the lowest conduction levels of the same crystallite with one excess electron with and without impurity.

This definition of the binding energy for the donor atom can be contrasted with that for a bulk system, where this quantity is defined as the difference between the dopant electron level and conduction band continuum, *i.e.*, the binding energy is equal to the ionization energy of the defect atom. In nanocrystals or quantum dots, such a definition is problematic since an electron being excited into an unoccupied state (below the vacuum level) will be confined by the physical size of the dot and will continue to interact strongly with the impurity atom.

The calculated ionization, affinity and binding energy as a function of quantum dot radius  $R$  are shown in Fig. §.31. The ionization energies for pure hydrogenated Si nanocrystals are also given for comparison. The ionization and affinity energies for pure Si quantum dots have values close to those calculated recently for hydrogenated Ge nanocrystals. A surprising feature in Fig. §.31 a is that the ionization energy  $I_d$  shows virtually no dependence on the size of the dot.

The dependence of  $I_d(R)$  is different from the behavior of the ionization energy in Si quantum dots where this quantity is very large at small radii and gradually decreases, scaling as  $R^{-1.1}$ , to its bulk value. Although this dependence of the ionization energy on radius is weaker than  $R^{-2}$  law predicted by effective mass theory [173,174], it is, nevertheless, a consequence of spatial confinement of electrons (holes) in quantum dots. It is surprising that this behavior is absent in the functional dependence of  $I_d(R)$ . The binding energy  $E_B$ , which scales as  $R^{-0.8}$ , is shown in Fig. §.31 b. These

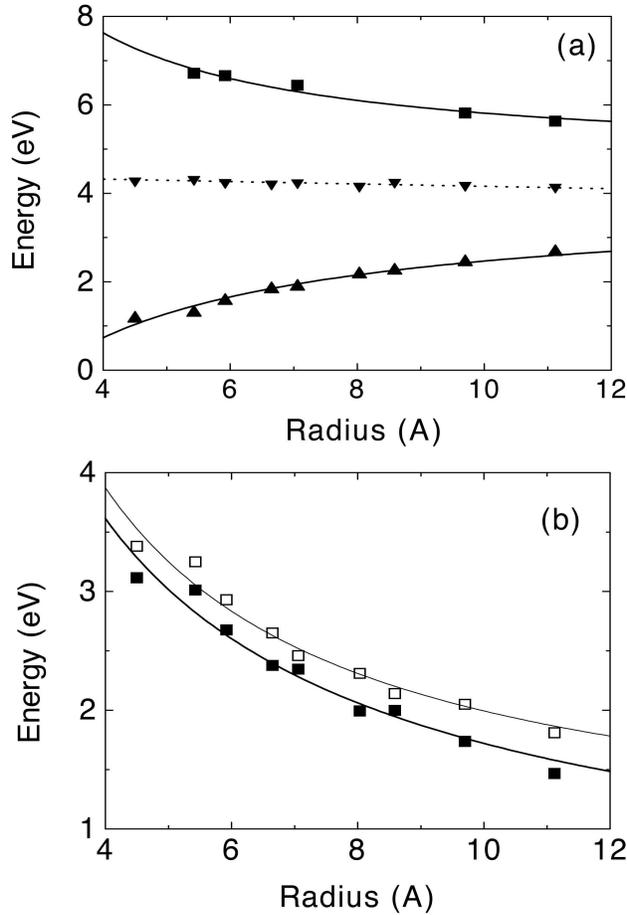


Figure §.31: (a) Ionization energy  $I_d$  for phosphorus-doped nanocrystal (▼) and electron affinity  $A_p$  (▲) as a function of nanocrystal's radius  $R$ . Ionization energy for pure hydrogenated Si nanocrystals (■) is also shown. Solid lines are the best fits to calculated points, dotted line is a guide to an eye. (b) Binding energy  $E_B$  (■) and energy difference between defect level with single occupancy and highest occupied state with double occupancy (□) as a function of the dot's radius.

values are close to results of the tight-binding method [138] even though the ionization energy has a constant value in this range of sizes. Also plotted in Fig. §.31 b is the dependence of the “band gap”, *i.e.* the difference between the lowest level with single occupancy and the highest doubly occupied level, in P-doped systems. This quantity is blue-shifted with respect to the bulk values where it should be approximately equal to Si band gap. Comparison with results for pure hydrogenated Si dots [140] of the same radius shows that it is smaller by about 10% than values of the HOMO-LUMO gaps. The large values of the binding energy suggest that for dots in this size regime, the donors cannot be considered as shallow. This is largely due to the weak screening present in quantum dots and the physical confinement of the donor electron within the dot.

The nature of the Si-P bond can be clarified by examining the charge density of the dopant electron  $|\Psi(r)|^2$  for several dot sizes. In Fig. §.32, we illustrate the charge profile for the case when the impurity is at the dot center. The density is plotted along [100] direction; results in other directions are similar. At all dots radii, the dopant wave function is strongly localized around the impurity site, *i.e.*, the majority of the charge is within the P-Si bond length. From effective mass calculations [173, 174], it follows that the envelope wave function of the dopant electron is given by  $j_0(\pi r/R) \propto \sin(\pi r/R)/r$ . The calculated charge profile in Fig. §.32 is at variance with this description. This difference in the spatial distributions can be attributed to the weaker screening in quantum dots. At these sizes,

the dielectric constant is several times smaller than the bulk value [140,144], giving rise to the increase of the effective electron-impurity potential and stronger localization of the electron around the defect atom.

Given the charge distribution of the dopant electron, one can evaluate the isotropic hyperfine parameter which determines the contact interaction between the electron and defect nuclei. The method of Van de Walle and Blöchl [175] is used to extract the hyperfine parameters from pseudo-charge densities. The hyperfine parameters for a P atom positioned in the dot center are given in Fig. §.33. At small sizes, the hyperfine parameter is very large owing to strong localization of the electron around impurity. As the radius increases, the value of  $A$  decreases. Our calculated results scale with radius  $R$  of the dot as  $R^{-1.5}$  (effective mass theory gives  $R^{-3}$ ). In Figure 3, we also present the experimental data of Ref. [168]. The measured values of the hyperfine parameter falls on the best fit to calculated results; computational limitations prevent us from comparing directly to measured values.

The hyperfine values are not strongly dependent on the choice of the P site. Other sites have been tested by replacing one of Si atoms in each shell with a P atom while retaining the passivating hydrogen atoms. The ionization and binding energies were unchanged to within  $\sim 5\%$ , independent of the impurity atom position [172].

The value of the isotropic hyperfine parameter also remains largely unchanged, save for the outermost layers of the dot. This behavior is demonstrated in Fig. §.34, where hyperfine parameter is plotted as a function of

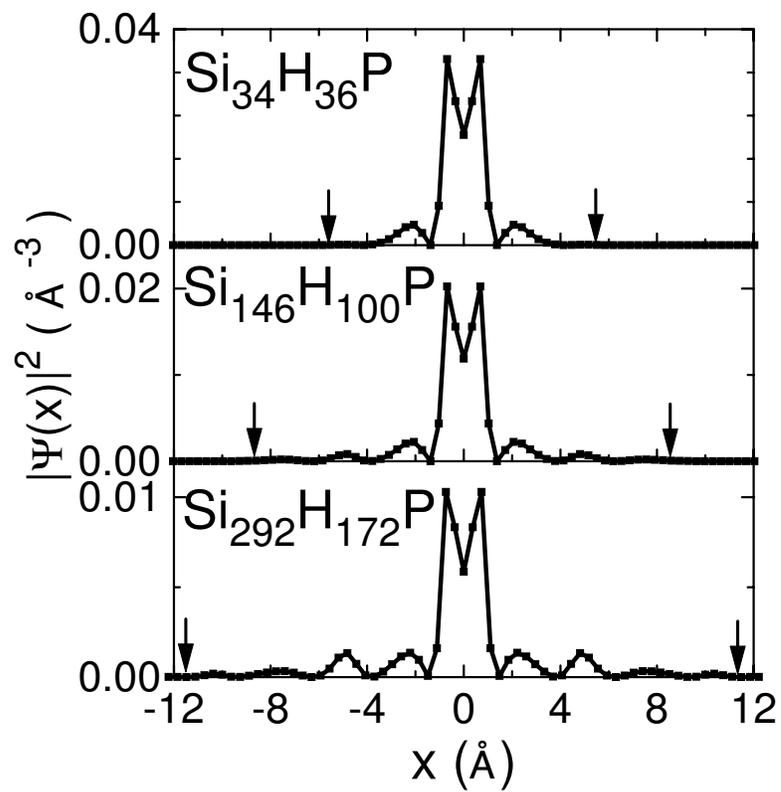


Figure §.32: Charge density for the dopant electron along  $[100]$  direction.  $x/R$  is the  $x$ -coordinate normalized on the dot's radius  $R$ .

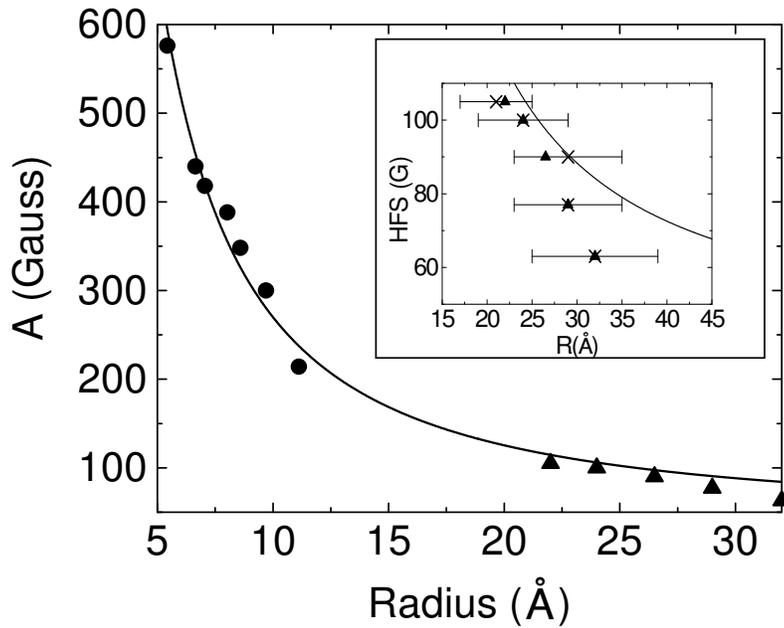


Figure §.33: Calculated ( $\bullet$ ) and experimental ( $\blacktriangle$ ) isotropic hyperfine parameter  $A$  vs. dot's radius  $R$ . The solid line is the best fit to calculations (bulk value of hyperfine parameter 42 G was used to obtain this fit). The Inset shows experimental data of Ref. [162] together with the fit to results of calculations. Two sets of experimental points correspond to the average size of nanocrystals ( $\times$ ) and the size of nanocrystals ( $\blacktriangle$ ) estimated from comparison of photoluminescence energies for doped and undoped samples (for more details, see Ref. [162]).

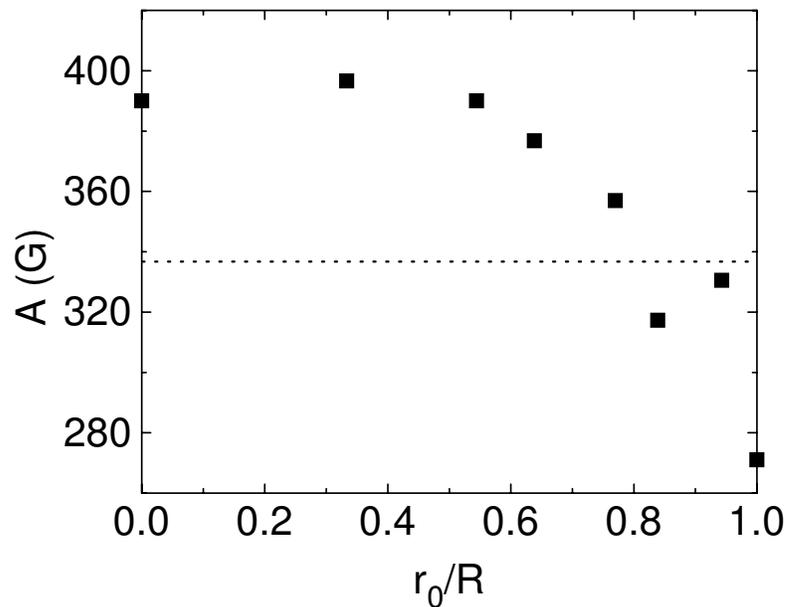


Figure 3.34: Isotropic hyperfine parameter  $A$  in  $\text{Si}_{86}\text{H}_{76}\text{P}$  quantum dot as a function of the normalized position  $r_0/R$  of the P atom in the quantum dot. The dotted line is the average value of the hyperfine parameter.

defect position in a representative dot  $\text{Si}_{86}\text{H}_{76}\text{P}$ . Near the surface, the P atom density becomes more delocalized and the hyperfine parameter shows a notable decrease in value from the value when P resides in the dot center. However, the average value of the hyperfine parameter over all sites is only about 15 % lower than the value obtained when the P atom is at the center of the nanocrystal which further increases agreement with experimental data.

## §.8 **Acknowledgments**

We would like to acknowledge support from the National Science Foundation, the United States Department of Energy and the Minnesota Supercomputing Institute.



# Bibliography

- [1] L. T. Canham, *Appl. Phys. Lett.* **57**, 1046 (1990).
- [2] J. M. Ziman, *Electrons and Phonons*, Oxford University Press, 1960.
- [3] A. Haug, *Theoretical Solid State Physics*, Pergamon Press, 1972.
- [4] O. Madelung, *Introduction to Solid State Theory*, Springer-Verlag, 1996.
- [5] M. Head-Gordon, *J. Phys. Chem.* **100**, 13213 (1996).
- [6] L. H. Thomas, *Proc. Camb. Phil. Soc.* **23**, 542 (1926).
- [7] E. Fermi, *Z. Phys.* **48**, 73 (1928).
- [8] S. Lundqvist and N. H. March, editors, *Theory of the Inhomogeneous Electron Gas*, Plenum, 1983.
- [9] C. Kittel, *Quantum Theory of Solids*, Wiley, 1987.
- [10] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Holt Rinehart, 1976.

- [11] J. C. Slater, Phys. Rev. **81**, 385 (1951).
- [12] J. C. Slater, *Quantum Theory of Molecules and Solids, Vol. 1-4*, McGraw Hill, 1964-1974.
- [13] J. C. Slater, *Quantum Theory of Matter*, McGraw Hill, 1968.
- [14] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
- [15] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
- [16] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
- [17] J. M. Ziman, *Principles of the Theory of Solids*, Cambridge University Press, 2nd edition, 1986.
- [18] J. P. Perdew, K. Burke, and Y. Wang, Phys. Rev. B **54**, 16533 (1996).
- [19] J. R. Chelikowsky and S. G. Louie, editors, *Quantum Theory of Materials*, Kluwer, 1996.
- [20] E. K. U. Gross, J. F. Dobson, and M. Petersilka, in *Density Functional Theory*, edited by R. F. Nalewajski, page 81, Springer-Verlag, 1996.
- [21] M. Petersilka, U. J. Gossmann, and E. K. U. Gross, Phys. Rev. Lett. **76**, 1212 (1996).
- [22] V. R. Saunders and J. H. van Lenthe, Mol. Phys. **48**, 923 (1983).
- [23] R. J. Buenker, S. D. Peyerimhoff, and W. Butscher, Mol. Phys. **35**, 771 (1978).

- [24] B. Bernu, D. M. Ceperley, and W. A. Lester Jr., *J. Chem. Phys.* **93**, 552 (1990).
- [25] B. Bernu, D. M. Ceperley, and W. A. Lester Jr., *J. Chem. Phys.* **95**, 7782 (1991).
- [26] A. J. Williamson, J. C. Grossman, R. Q. Hood, A. Puzder, and G. Galli, *Phys. Rev. Lett.* **89**, 196803 (2002).
- [27] L. J. Sham and T. M. Rice, *Phys. Rev.* **144**, 708 (1966).
- [28] L. Hedin, *Phys. Rev.* **139**, A796 (1965).
- [29] M. S. Hybertsen and S. G. Louie, *Phys. Rev. B* **34**, 5390 (1986).
- [30] M. Casida, in *Recent Advances in Density-Functional Methods, Part I*, edited by D. Chong, page 155, World Scientific, Singapore, 1995.
- [31] M. Casida, in *Recent Developments and Applications of Modern Density Functional Theory*, edited by J. Seminario, page 391, Elsevier, Amsterdam, 1996.
- [32] I. Vasiliev, S. Ögüt, and J. R. Chelikowsky, *Phys. Rev. Lett.* **82**, 1919 (1999).
- [33] K. Raghavachari, D. Ricci, and G. Pacchioni, *J. Chem. Phys.* **116**, 825 (2002).
- [34] J. Jaramillo and G. E. Scuseria, *Theor. Chem. Acc.* **105**, 62 (2000).

- [35] S. Hirata and M. Head-Gordon, *Chem. Phys. Lett.* **314**, 291 (1999).
- [36] K. Yabana and G. F. Bertsch, *Phys. Rev. B* **54**, 4484 (1996).
- [37] K. Yabana and G. F. Bertsch, *Z. Phys. D* **42**, 219 (1997).
- [38] K. Yabana and G. F. Bertsch, *Int. J. Quant. Chem.* **75**, 55 (1999).
- [39] E. K. U. Gross and W. Kohn, *Phys. Rev. Lett.* **55**, 2850 (1985).
- [40] E. K. U. Gross and W. Kohn, *Adv. Quant. Chem.* **21**, 255 (1990).
- [41] G. Onida, L. Reining, and A. Rubio, *Rev. Mod. Phys.* **74**, 601 (2002).
- [42] L. Reining, V. Olevano, A. Rubio, and G. Onida, *Phys. Rev. Lett.* **88**, 066404 (2002).
- [43] Y. H. Kim and A. Görling, *Phys. Rev. Lett.* **89**, 096402 (2002).
- [44] J. P. Perdew, K. Burke, and Y. Wang, *Phys. Rev. B* **54**, 16533 (1996).
- [45] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- [46] M. E. Casida and D. R. Salahub, *J. Chem. Phys.* **113**, 8918 (2000).
- [47] R. van Leeuwen and E. J. Baerends, *Phys. Rev. A* **49**, 2421 (1994).
- [48] I. Vasiliev, *Phys. Stat. Sol.* (submitted).
- [49] G. D. Mahan, *Many-Particle Physics*, Plenum Press, New York, 1981.
- [50] D. M. Ceperley, *Phys. Rev. B* **18**, 3126 (1978).

- [51] J. P. Perdew and A. Zunger, *Phys. Rev. B* **23**, 5048 (1981).
- [52] D. E. Beck, *Phys. Rev. B* **43**, 7301 (1991).
- [53] I. Vasiliev, S. Ögüt, and J. R. Chelikowsky, *Phys. Rev. B* **65**, 115416 (2002).
- [54] W. R. Burdick, Y. Saad, L. Kronik, I. V. M. Jain, and J. Chelikowsky, *Comp. Phys. Comm.* , (in press) (2003).
- [55] J. R. Chelikowsky and M. L. Cohen, in *Handbook of Semiconductors*, edited by T. S. Moss and P. T. Landsberg, Elsevier, Amsterdam, 2nd edition, 1992.
- [56] E. Fermi, *Nuovo Cimento* **11**, 157 (1934).
- [57] D. R. Hamann, M. Schlüter, and C. Chiang, *Phys. Rev. Lett.* **43**, 1494 (1979).
- [58] A. Zunger and M. L. Cohen, *Phys. Rev. B* **18**, 5449 (1978).
- [59] G. P. Kerker, *J. Phys. C* **13**, L189 (1980).
- [60] G. Bachelet, D. R. Hamann, and M. Schlüter, *Phys. Rev. B* **26**, 4199 (1982).
- [61] H. S. Greenside and M. Schlüter, *Phys. Rev. B* **26**, 4199 (1983).
- [62] D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990).

- [63] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
- [64] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
- [65] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).
- [66] C. Y. Fong, *Topics in Computational Materials Science*, World Scientific, 1998.
- [67] A. Briley, M. R. Pederson, K. A. Jackson, D. C. Patton, and D. V. Porezag, Phys. Rev. B **58**, 1786 (1997).
- [68] K. A. Jackson, M. R. Pederson, D. V. Porezag, Z. Hajnal, and T. Fraunheim, Phys. Rev. B **55**, 2549 (1997).
- [69] J. R. Chelikowsky and S. G. Louie, Phys. Rev. B **29**, 3470 (1984).
- [70] R. W. Jansen and O. F. Sankey, Phys. Rev. B **36**, 6520 (1987).
- [71] S. R. White, J. W. Wilkins, and M. P. Teter, Phys. Rev. B **39**, 5819 (1989).
- [72] J. R. Chelikowsky, N. Troullier, and Y. Saad, Phys. Rev. Lett. **72**, 1240 (1994).
- [73] J. R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad, Phys. Rev. B **50**, 11135 (1994).
- [74] J. R. Chelikowsky et al., Comp. Phys. Comm. **85**, 325 (1995).

- [75] E. L. Briggs, D. J. Sullivan, and J. Bernholc, Phys. Rev. B **52**, R5471 (1995).
- [76] F. Gygi and G. Galli, Phys. Rev. B **52**, R2229 (1995).
- [77] G. Zumbach, N. A. Modine, and E. Kaxiras, Solid State Commun. **99**, 57 (1996).
- [78] T. Ono and K. Hirose, Phys. Rev. Lett. **82**, 5016 (1999).
- [79] Y. H. Kim, I. H. Lee, and R. M. Martin, Comp. Phys. Comm. **131**, 10 (2000).
- [80] I. H. Lee, Y. H. Kim, and R. M. Martin, Phys. Rev. B **61**, 4397 (2000).
- [81] T. L. Beck, Rev. Mod. Phys. **74**, 1041 (2000).
- [82] J.-L. Fattebert and J. Bernholc, Phys. Rev. B **62**, 1713 (2000).
- [83] B. Fornberg and D. M. Sloan, Acta Numerica **94**, 203 (1994).
- [84] R. B. Morgan and D. S. Scott, SIAM J. Sci. Comput. **7**, 817 (1986).
- [85] Y. Saad, A. Stathopoulos, J. R. Chelikowsky, K. Wu, and S. Ogut, BIT **36**, 563 (1996).
- [86] C. H. Tong, T. F. Chan, and C. C. J. Kuo, SIAM J. Sci. Stat. Comput. **13**, 227 (1992).
- [87] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice Hall, Englewood Cliffs, 1980.

- [88] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing, 1996.
- [89] R. Biswas and D. R. Hamann, Phys. Rev. B **34**, 895 (1986).
- [90] R. Kubo, Rep. Prog. Theor. Phys. **29**, 255 (1966).
- [91] H. Risken, *The Fokker-Planck Equation*, Springer-Verlag, 1984.
- [92] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
- [93] R. Car and M. Parrinello, Phys. Rev. Lett. **60**, 204 (1988).
- [94] N. Binggeli, J. L. Martins, and J. R. Chelikowsky, Phys. Rev. Lett. **68**, 2956 (1992).
- [95] F. W. Kutzler and G. S. Painter, Phys. Rev. B **45**, 288 (1992).
- [96] C. Rohlfing and K. Raghavachari, J. Chem. Phys. **96**, 2114 (1992).
- [97] K. Raghavachari and V. Logovinsky, Phys. Rev. Lett. **55**, 2853 (1985).
- [98] K. Raghavachari and C. Rohlfing, J. Chem. Phys. **94**, 3670 (1990).
- [99] D. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
- [100] O. Cheshnovsky et al., Chem. Phys. Lett. **138**, 119 (1987).
- [101] D. Tomanek and M. Schluter, Phys. Rev. Lett. **56**, 1055 (1986).
- [102] N. Binggeli and J. R. Chelikowsky, Phys. Rev. B **50**, 11764 (1994).

- [103] S. Ogut and J. R. Chelikowsky, *Phys. Rev. B* **55**, R4914 (1997).
- [104] E. C. Honea et al., *Nature* **366**, 42 (1993).
- [105] X. Jing, N. Troullier, J. R. Chelikowsky, K. Wu, and Y. Saad, *Solid State Comm.* **96**, 231 (1995).
- [106] R. Fournier, S. B. Sinnott, and A. E. DePristo, *J. Chem. Phys.* **97**, 4149 (1992).
- [107] R. Schäfer, S. Schlect, J. Woenckhaus, and J. A. Becker, *Phys. Rev. Lett.* **76**, 471 (1996).
- [108] S. Baroni, P. Gianozzi, and A. Testa, *Phys. Rev. Lett.* **58**, 1861 (1987).
- [109] X. Gonze, D. C. Allan, and M. P. Teter, **68**, 3603 (1992).
- [110] J. L. Martins, J. Buttet, and R. Car, *Phys. Rev. B* **31**, 1804 (1985).
- [111] I. Moullet, J. L. Martins, F. Reuse, and J. B. J. Buttet, *Phys. Rev. B* **42**, 11598 (1990).
- [112] V. Bonacic-Koutecky, P. Fantucci, and J. K. Koutecky, *J. Chem. Phys.* **93**, 3802 (1990).
- [113] V. Bonacic-Koutecky, P. Fantucci, , and J. Koutecky, *Chem. Phys. Lett.* **166**, 32 (1990).
- [114] G. Onida, L. Reining, R. W. Godby, R. del Sole, and W. Andreoni, *Phys. Rev. Lett.* **75**, 181 (1995).

- [115] C. R. C. Wang, S. Pollack, D. Cameron, and M. M. Kappes, *Chem. Phys. Lett.* **93**, 3787 (1990).
- [116] U. Itoh, Y. Toyoshima, H. Onuki, N. Washida, and T. Ibuki, *J. Chem. Phys.* **85**, 4867 (1986).
- [117] S. Furukawa and T. Miyasato, *Phys. Rev. B* **38**, 5726 (1988).
- [118] D. J. Lockwood, A. Wang, and B. Bryskiewicz, *Solid State Commun.* **89**, 587 (1994).
- [119] M. V. Wolkin, J. Jorne, P. M. Fauchet, G. Allan, and C. Delerue, *Phys. Rev. Lett.* **82**, 197 (1999).
- [120] T. Takagahara and K. Takeda, *Phys. Rev. B* **46**, 15578 (1992).
- [121] M. Rohlfing and S. G. Louie, *Phys. Rev. Lett.* **80**, 3320 (1998).
- [122] N. A. Hill and K. B. Whaley, *Phys. Rev. Lett.* **75**, 1130 (1995).
- [123] C. Delerue, M. Lannoo, and G. Allan, *Phys. Rev. Lett.* **76**, 3038 (1996).
- [124] C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* **48**, 11024 (1993).
- [125] L. W. Wang and A. Zunger, *J. Phys. Chem* **100**, 2394 (1994).
- [126] B. Delley and E. F. Steigmeier, *Phys. Rev. B* **47**, 1397 (1993).
- [127] R. J. Baierle, M. J. Caldas, E. Molinari, and S. Ossicini, *Solid State Commun.* **102**, 545 (1997).

- [128] A. Franceschetti, L. W. Wang, and A. Zunger, *Phys. Rev. Lett.* **83**, 1269 (1999).
- [129] S. Ögüt, J. R. Chelikowsky, and S. G. Louie, *Phys. Rev. Lett.* **1999**, 1270 (1999).
- [130] R. W. Godby and I. D. White, *Phys. Rev. Lett.* **80**, 3161 (1998).
- [131] S. Ögüt, J. R. Chelikowsky, and S. G. Louie, *Phys. Rev. Lett.* **80**, 3162 (1998).
- [132] I. Vasiliev, S. Ogut, and J. R. Chelikowsky, *Phys. Rev. B* **60**, 8477 (1999).
- [133] C. Troparevsky, L. Kronik, and J. R. Chelikowsky, *Phys. Rev. B* **65**, 33311 (2002).
- [134] D. J. Lockwood, *Solid State Comm.* **92**, 101 (1994).
- [135] J. P. Proot, C. Delerue, and G. Allan, *Appl. Phys. Lett.* **61**, 1948 (1992).
- [136] L. W. Wang and A. Zunger, *J. Phys. Chem.* **98**, 2158 (1994).
- [137] N. A. Hill and K. B. Whaley, *Phys. Rev. Lett.* **76**, 3039 (1996).
- [138] M. Lannoo, C. Delerue, and G. Allan, *Phys. Rev. Lett.* **74**, 3415 (1995).
- [139] G. Allan, C. Delerue, M. Lannoo, and E. Martin, *Phys. Rev. B* **52**, 11982 (1995).

- [140] S. Ogut, J. R. Chelikowsky, and S. G. Louie, Phys. Rev. Lett. **79**, 1770 (1997).
- [141] L. Brus, J. Phys. Chem. **90**, 2555 (1986).
- [142] Y. Kayanuma, Phys. Rev. B **38**, 9797 (1988).
- [143] A. Franceschetti and A. Zunger, Phys. Rev. Lett. **78**, 915 (1997).
- [144] S. Ogut, R. Burdick, Y. Saad, and J. R. Chelikowsky, Phys. Rev. Lett. **90**, 127401 (2003).
- [145] I. Vasiliev, S. Ogut, and J. R. Chelikowsky, Phys. Rev. Lett. **78**, 4805 (1997).
- [146] G. Srinivasan, Phys. Rev. B **178**, 1244 (1969).
- [147] R. Schäfer and J. A. Becker, Phys. Rev. B **54**, 10296 (1996).
- [148] S. Schuppler et al., Phys. Rev. B **52**, 4910 (1995).
- [149] J. von Behren, T. van Buuren, M. Zacharias, E. H. Chimowitz, and P. M. Fauchet, Solid State Comm. **105**, 317 (1998).
- [150] Y. Kanemitsu et al., Phys. Rev. B **48**, 2827 (1993).
- [151] V. Lehmann and U. Gösele, Appl. Phys. Lett. **58**, 856 (1991).
- [152] I. Vasiliev, S. Ögüt, and J. R. Chelikowsky, Phys. Rev. Lett. **86**, 1813 (2001).

- [153] I. Vasiliev, J. R. Chelikowsky, and R. M. Martin, *Phys. Rev. B* **65**, 121302 (2002).
- [154] A. Puzder, A. J. Williamson, J. C. Grossman, and G. Galli, *Phys. Rev. Lett.* **88**, 097401 (2002).
- [155] L. Mitas, J. Therrien, R. Twisten, G. Belomoin, and M. H. Nayfeh, *Appl. Phys. Lett.* **78**, 1918 (2001).
- [156] I. Vasiliev and R. M. Martin, *Phys. Stat. Sol. (b)* **233**, 5 (2002).
- [157] E. Martin, C. Delerue, G. Allan, and M. Lannoo, *Phys. Rev. B* **50**, 18258 (1994).
- [158] T. Takagahara and K. Takeda, *Phys. Rev. B* **53**, R4205 (1996).
- [159] A. D. Yoffe, *Adv. Phys.* **50**, 1 (2001).
- [160] B. E. Kane, *Nature* **393**, 133 (1998).
- [161] D. P. DiVincenzo, *Nature* **393**, 113 (1998).
- [162] A. Mimura, M. Fujii, S. Hayashi, D. Kovalev, and F. Koch, *Phys. Rev. B* **62**, 12625 (2000).
- [163] G. Mauckner, W. Rebitzer, K. Thonke, and R. Sauer, *Solid State Comm.* **91**, 717 (1994).
- [164] D. Kovalev, H. Heckler, G. Polisski, and F. Koch, *Phys. Status Solidi B* **215**, 871 (1999).

- [165] G. Feher, Phys. Rev. **114**, 1219 (1959).
- [166] J. Müller, F. Finger, R. Carius, and H. Wagner, Phys. Rev. B **60**, 11666 (1999).
- [167] B. J. Pawlak, T. Gregorkiewicz, C. A. J. Ammerlaan, and P. F. A. Alkemade, Phys. Rev. B **64**, 115308 (2001).
- [168] M. Fujii, A. Mimura, S. Hayashhi, Y. Yamamoto, and K. Murakami, Phys. Rev. Lett. **89**, 206805 (2002).
- [169] C. Y. Fong, H. Zhong, B. M. Klein, and J. S. Nelson, Phys. Rev. B **49**, 7466 (1994).
- [170] I. H. Lee, K. H. Ahn, Y. H. Kim, R. M. Martin, and J. P. Leburton, Phys. Rev. B **60**, 13720 (1999).
- [171] J. R. Chelikowsky, J. Phys. D: Appl. Phys. **33**, R33 (2000).
- [172] D. Melnikov and J. R. Chelikowsky, Phys. Rev. Lett. , in press (2004).
- [173] L. E. Brus, J. Chem. Phys. **79**, 4403 (1983).
- [174] L. E. Brus, J. Chem. Phys. **79**, 5566 (1983).
- [175] P. E. B. C. G. Van de Walle, Phys. Rev. B **47**, 4244 (1993).