

Ab Initio Energy Techniques

Gerbrand Ceder, MIT (1999)

Using quantum mechanics, it is now possible to compute the energy of a material with no more input knowledge than the nuclear charge and number of electrons on all the atoms and the value of a few physical constants. Although quantum mechanics allows one, in principle, to determine physical properties with arbitrary accuracy, its application in condensed matter is usually accompanied by several approximations. These approximations have led to the hundreds of different acronyms, indicating differences in methods of solution (Table 1). Before we discuss these approximations, it may be useful to evaluate under what conditions we may want to use quantum mechanics.

Some reasons to use quantum mechanics:

Control and Accuracy: Since one starts with a form for the energy that is fundamentally correct (the Schrödinger equation, or its relativistic counterpart, the Dirac equation), it is easier to control and evaluate the approximations that go into the results than when using simpler models, such as empirical potentials. Also, quantum mechanical methods usually offer a hierarchy of approximations making it possible, at least in principle, to push the envelope and check your results by doing calculations that are less and less approximate.

Novel Materials: In some cases one simply does not have experimental data on materials. In that case, first-principles techniques can be extremely useful. Especially in the design of novel materials, the capability to determine some properties of materials before they are actually synthesized can be extremely useful. Later on we will discuss which "properties" one can trust from modern first-principles methods.

Electronic Effects: The physics of some effects can simply not be captured with simple energy models. In that case, quantum mechanics, is indispensable. An example is the segregation of As to grain boundaries in Ge. In contrast to the usual driving forces for grain boundary segregation, such as size and surface energy, As segregates to a Ge grain boundary for purely electronic effects: The "extra" valence electron on As (as compared to Ge) can occupy a lower band near the boundary.

Table 1: Selection of Acronyms in First-Principles Methods

DFT	Density Functional Theory
LDA	Local Density Approximation
LSD	Local Spin Density approximation
LMTO-ASA	Linear-Muffin-Tin-Orbital method in the Atomic Sphere Approximation
LMTO-FP	Linear-Muffin-Tin-Orbital method in the Full Potential Approximation
(F)LAPW	(Full Potential) Linearized Augmented-Plane-Wave method
ASW	Augmented Spherical Wave method
TB	Tight Binding
PP	Pseudopotential
LCAO	Linear Combination of Atomic Orbitals
SSCAD	Spherical-Self-Consistent Atomic Deformation Method
PIB	Potential Induced Breathing
HF	Hartree Fock
CI	Configuration Interaction
MP(2)	Moller-Pleset perturbation theory to second order

The Many-Particle Schrödinger Equation

The "real" quantum mechanical task is really a beast of a problem. It consist of finding the eigenvalues (E) and eigenfunctions (Ψ) of the following problem* :

$$H\Psi = E\Psi \quad (1)$$

This is an *eigenvalue* equation. It basically means that given the definition of H we need to find all the functions Ψ so that when H operates on Ψ it returns a function proportional to Ψ . If we knew those functions Ψ we would have a complete specification of the state of our system (electrons and nuclei). The proportionality constant in (1) would be the energy of the state. The Ψ with the lowest energy therefore gives the ground state whereas the other solution Ψ 's correspond to excited states. By itself equation (1) is rather innocent looking. But wait until you realize what H is. H is the Hamiltonian operator that describes how to get all the contributions to the total energy of the systems. Most of these are classical (electrostatic) except for the kinetic energy:

$$H = \sum_{i=1}^{N_e} \frac{-\hbar^2}{2m_e} \nabla_i^2 + \sum_{I=1}^{N_n} \frac{-\hbar^2}{2m_I} \nabla_I^2 - \sum_i \sum_I \frac{Z_I e^2}{|r_i - R_I|} + \frac{1}{2} \sum_I \sum_{J \neq I} \frac{Z_I Z_J e^2}{|R_J - R_I|} + \frac{1}{2} \sum_i \sum_{j \neq i} \frac{e^2}{|r_j - r_i|} \quad (2)$$

In equation (2) lowercase i is used to label electronic coordinates and uppercase I and J for nuclear coordinates. To get rid of the fundamental constants in this equation one usually works in atomic units when doing quantum mechanics. A list of the common atomic units is given below. The two you should be most familiar with are the *Bohr* for length and the *Hartree* for energy.

* This is the time-independent Schrödinger equation. To see how it is derived from the more fundamental time-dependent equation you can consult any introductory textbook on quantum mechanics.

Table: Definition of atomic units

Quantity	Symbol	Value
length	a_0 (bohr)	$5.2918 \times 10^{-11} \text{m}$
mass	m_e	$9.1095 \times 10^{-31} \text{kg}$
time	t	$2.4189 \times 10^{-17} \text{s}$
energy	E_h (hartree)	$4.3598 \times 10^{-18} \text{J}$
charge	e	$1.6022 \times 10^{-19} \text{C}$
angular momentum	$k = h/2$	$1.0546 \times 10^{-34} \text{J s}$
electric field	E	$5.1423 \times 10^{-11} \text{V m}^{-1}$
electric field gradient	$-V_{zz}$	$9.7174 \times 10^{21} \text{V m}^{-2}$
magnetic induction	B	$2.3505 \times 10^5 \text{T}$
electric dipole	P_e	$8.4784 \times 10^{-30} \text{C m}$
electric quadrupole	Θ_e	$4.4866 \times 10^{-40} \text{C m}^2$
magnetic moment	ρ_M	$1.8548 \times 10^{-23} \text{J T}^{-1}$
polarizability	α	$1.6488 \times 10^{-41} \text{C}^2 \text{m}^2 \text{J}^{-1}$
magnetizability	χ	$7.8910 \times 10^{-29} \text{J T}^{-2}$

The first approximation that one makes is to decouple the nuclear and electronic degrees of freedom. Electrons moves much faster than nuclei so that it can be assumed that the electrons are always in the lowest energy state for the given nuclear configuration. You can think of this as the electrons moving through a landscape formed by the ions. On the time scale of the electrons, this landscape does not change, even although on a longer time scale the ions may be vibrating. The Hamiltonian now looks somewhat simpler (in reduced units):

$$H = E_{nuclei}(\{\mathbf{R}_I\}) - \sum_{i=1}^{N_e} \nabla_i^2 + \sum_{i=1}^{N_e} V_{nuclei}(\mathbf{r}_i) + \frac{1}{2} \sum_j^{N_e} \sum_{i \neq j}^{N_e} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (3)$$

This simplification, which is pretty good, goes under the name *Born-Oppenheimer** approximation, or *adiabatic* approximation. Practically, this means that for each set of nuclear coordinates we have a Hamiltonian and a

* Yes, this is the same Oppenheimer that worked on the bomb.

corresponding set of wave functions and energies. The lowest energy as a function of the nuclear coordinates is often referred to as the *Born-Oppenheimer* surface. You can think of it as the function that gives the energy at each configuration of the nuclei. Obviously, knowing this would allow for great materials simulations ! But unfortunately, we are still not out the woods. Let us investigate carefully each of the four terms.

$E_{nuclei}(\{\mathbf{R}_I\})$ This term is simply the electrostatic and kinetic (vibrations) energy of the nuclei. At zero temperature, we have no kinetic energy (apart from the zero-point motion, so that this term reduces to a simple electrostatic sum over point charges. This can be obtained with the usual trick of Ewald summation.

$\sum_{i=1}^{N_e} \nabla_i^2$ This is the kinetic energy of the electrons. Note that the ∇^2 operator basically measures the curvature of the wave function. Rapidly varying wave functions will have high kinetic energy.

$\sum_i V_{nuclei}(\mathbf{r}_i)$ This is the energy of the electronic charge in the electrostatic potential from the nuclei. V_{nuclei} is basically a sum of $1/r$ potentials from all the nuclear charges.

$\frac{1}{2} \sum_j^{N_e} \sum_{i \neq j}^{N_e} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$ This is the beast ! It is the electrostatic interaction between all

the electrons. Given of the order of 10^{23} electrons in a material, this term couples all of them. See now why they call it a "many"-body Schrödinger equation ?

The last term is truly a stumbling block in solving the Schrödinger equation. If it were not present things would be much simpler. The first three terms

only contain the coordinate of one electron at a time*. If we only had these terms we could solve the quantum mechanics problem one electron at a time. That is instead of 10^{23} electrons at a time!

There is hope however. In the next section we will discuss how, by solving one-electron problems iteratively, one can get close to the true solution of the many-particle Schrödinger equation. There are basically two popular approaches. In the Hartree Fock (HF) methods one tries to construct an approximate many-electron wave function from one-electron wave functions determined in some appropriate effective potential. As we will see, Hartree Fock methods are computationally demanding and usually only applied to small molecules. In Density Functional Methods (DFT) the many-electron problem is exactly transformed to a one-electron problem in an unknown effective potential. Approximations come from approximations to this effective potential. Before we discuss both methods in greater detail it is instructive to look at the general concept of *self-consistent* one-electron methods.

Self-Consistent One-Electron Methods

Since the part that makes equation (3) unsolvable is the electron-electron interaction, it is obvious we need to tinker with this term. Suppose an electron experienced, instead of the instantaneous interaction from all the other electrons, an "averaged" potential from the presence of the other electrons. Then we could add that potential to the first three terms in equation (3) and we would have a Schrödinger equation with only single-electron terms. This is the concept of *mean field* approximations. What should this mean-field potential be? It is obviously the electrostatic interaction with charge density formed by all the other electrons in the

* Actually, the first term contains no electronic coordinate. Since it is the part of the energy arising only from the nuclei, it does not depend on the electronic state and enters simply as a constant in the Schrödinger equation.

system. This charge density can be computed from the wave functions for each electron* :

$$\rho(\mathbf{r}) = \sum_i^{N_e} |\varphi_i(\mathbf{r})|^2 \quad (4)$$

The potential from this charge density is the mean-field potential, also called the *Hartree* potential:

$$V_{MF}(\mathbf{r}) = \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' = \sum_j \int \frac{|\varphi_j(\mathbf{r}')|^2}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}' \quad (5)$$

With this mean-field potential we now have a single one-electron Schrödinger equation:

$$H_{MF}\varphi = -\nabla^2\varphi + V_{nuclei}(\mathbf{r})\varphi + V_{MF}(\mathbf{r})\varphi = E\varphi \quad (6)$$

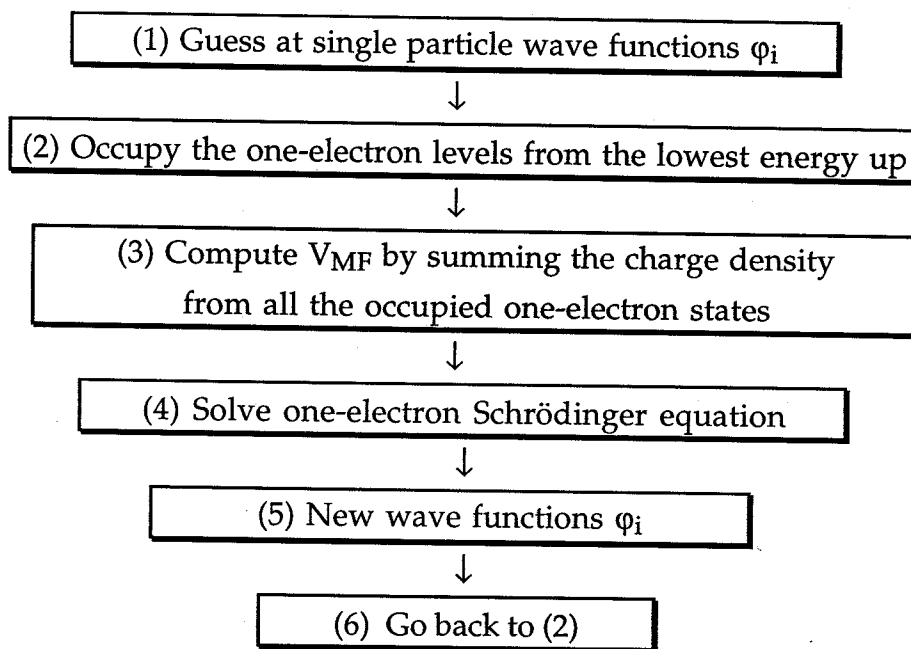
For an infinite system, $V_{MF}(\mathbf{r}_i)$ is identical for all electrons so that we have a single Schrödinger equation to solve: The same one for each electron. Now that is an improvement ! If we solve for the eigenfunctions and eigenvalues of this Hamiltonian we obtain the one-electron states and energies. Note however that there is a slight problem. The unknown one-electron wave functions appear in V_{MF} . So, we need the unknown wave functions to determine V_{MF} . The solution to this problem is to solve equation (5) and (6) iteratively: One makes a trial guess for the wave functions of all electrons and calculates its V_{MF} through equation (5). With this effective potential the Schrödinger equations is solved to get new wave functions. From these, a new V_{MF} is calculated, and so on. The procedure is terminated when the charge density does not vary much anymore. At this point *self consistency* is reached, meaning that the wave functions and effective potential are self-

* The electron for which the field is calculated should be left out of this sum. If one works on macroscopic systems with 10^{23} atoms, including it in the charge density causes a negligible error.

consistent with each other, i.e. the wave functions correspond to those one would get from solving the Schrödinger equation with that potential. Typically, self-consistency is reached in less than a dozen iterations, although in some cases oscillations around the optimal solutions can slow down convergence. For this reason, the new solution in each iteration is gradually "mixed" with the old solution. It has been found that this greatly improves stability of the iteration cycle.

The Hartree approximation is never used by itself as there are two fundamental problems with it: The lack of *exchange* interaction and electron *screening* (correlation). The procedure of solving a one-electron Hamiltonian in a self-consistent manner is however key to most ab-initio quantum mechanical methods.

Self-Consistent One-Electron Methods



The Exchange Problem: Hartree-Fock

Before we discuss the Hartree Fock approach we need to say a few words about the electron "coordinates". An electron not only has a positional coordinate but also a spin coordinate ("up" or "down"). This is the same spin that is responsible for magnetism. From now on we will include this spin degree of freedom and write the coordinate as $r_i s_i$, where s_i indicates the spin part of electron i .

In the Hartree procedure we did not say anything about the many-electron wave functions. Even for the same set of single-electron wave functions there are many ways to construct a many-electron wave function. One can quite easily prove that the Hartree approximation corresponds to taking the many-electron wave function as a product of the one-electron wave functions:

$$\Psi(r_1 s_1, r_2 s_2, \dots, r_i s_i, \dots, r_N s_N) = \varphi_1(r_1 s_1) \varphi_2(r_2 s_2) \dots \varphi_i(r_i s_i) \dots \varphi_N(r_N s_N) \quad (7)$$

This should make sense to you: Remember that the square of the wave function is the probability to find the system in a given state. If all electrons move independently of each other, the combined wave function should be the product of the single-particle wave functions, similar to the factoring of probabilities in statistics. The most serious problem with this product wave function is that it violates the Pauli principle which states that the wave function of a many particle system has to be anti-symmetric* with respect to an interchange of two particles. Equation (7) is neither symmetric nor anti symmetric in the particle coordinates. The way to anti-symmetrize the all-electron wave function with respect to any interchange of two electrons is to construct it as a determinant of the single electron wave functions:

* Anti-symmetric means that the wave function has to change sign when the coordinates of two particles are interchanged. It should be clear that no physical observables can change with such a coordinate exchange since the system before and after the change are indistinguishable. This is no problem for the wave function as it is not an observable.

$$\Psi(\mathbf{r}_1s_1, \mathbf{r}_2s_2, \dots, \mathbf{r}_is_i, \dots, \mathbf{r}_Ns_N) = \begin{vmatrix} \varphi_1(\mathbf{r}_1s_1) & \varphi_1(\mathbf{r}_2s_2) & \cdots & \varphi_1(\mathbf{r}_Ns_N) \\ \varphi_2(\mathbf{r}_1s_1) & \varphi_2(\mathbf{r}_2s_2) & \cdots & \varphi_2(\mathbf{r}_Ns_N) \\ \vdots & \vdots & & \vdots \\ \varphi_N(\mathbf{r}_1s_1) & \varphi_N(\mathbf{r}_2s_2) & \cdots & \varphi_N(\mathbf{r}_Ns_N) \end{vmatrix} \quad (8)$$

This determinant is called a *Slater* determinant. Interchanging the coordinates of two electrons corresponds to the interchange of two columns which, as we know from linear algebra, changes the sign of the determinant. Note that the Pauli principle will reduce the Coulombic interaction between electrons as it keeps electrons apart. When two electrons (with the same spin) occupy exactly the same coordinate the determinant vanishes. Exchange is therefore an "interaction" caused by the Pauli principle, but with no potential energy that represents it in the Hamiltonian. This is the true free lunch ! It can be shown that the one-electron functions φ in the Hartree Fock approximation can be obtained by solving a Schrödinger equation with an additional term:

$$H_{HF}\varphi_i = -\nabla_i^2\varphi_i + V_{nuclei}(\mathbf{r})\varphi_i + V_{MF}(\mathbf{r})\varphi_i - \sum_j \int d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \varphi_j^*(\mathbf{r}')\varphi_i(\mathbf{r}')\varphi_j(\mathbf{r})\delta_{s_i,s_j} = E\varphi_i \quad (9)$$

This is the same as the mean field equations except for the last term which is the result of the exchange interaction. This term will create multiple two-center integrals in the practical solution which makes the Hartree Fock approximation very time-consuming. For example, a fairly accurate calculation for crystalline Be in the hcp structure requires 16 million two-electron integrals !

Practically

The best known package for Hartree Fock calculations is GaussianXX (XX refers to the year of issue). As of this writing we are up to Gaussian98. It can do Hartree Fock and much, much more. If you want a really nice interface to the code you can by the Gaussian GUI interface from Molecular Simulation

Inc (MSI)* These programs use the nice properties of Gaussian functions to make the integration work a lot simpler. They use terminology such as STO-3G to indicate a type of calculation. This label indicates the type of *basis* that is used to perform the calculations. We will say a bit more about a *basis* later in these notes. You can find a more detailed description of the nomenclature in this field in the chapter from Grant and Richards.

Hartree Fock methods are mainly applied in chemistry to study (small) molecules. Its poor scaling (around N^4) limits its application to large systems. For the purists among you it does however offer a systematic way to approach the correct solution. Hartree Fock with a large enough basis set can give you the most accurate wave function with no approximations for the exchange interactions. Including correlation effects in a systematic way can lead to very accurate results. This is however only reserved for those among you with their own personal supercomputer account.

Correlation

In the Hartree approach an electron interacts with the average charge density produced by the other electrons. So there is no instantaneous interaction between electrons, only an averaged-out interaction. In reality, electronic movement is *correlated*. Electrons can "avoid" each other. Correlation can be clarified with a simple example: Think of two electrons hopping between localized states on a set of identical nuclei. On average, the electron density will be the same on all nuclei, and in the Hartree approximation, each electron will interact with this average density. In the true system, the electrons will probably never sit together on the same nucleus. So, although the average charge density formed by the two electrons is identical on all nuclei, one electron will stay away from a nucleus when the other electron sits there. You can see that the Hartree approximation will overestimate the electrostatic energy between the electrons as it will contain some contribution from the electronic densities at the same nucleus. Correlation is present in the true all-electron wave function but not in an

* Recently (1996) MSI merged with Biosym, another company that markets software for molecular modeling and ab-initio calculations.

approach at the Hartree level. It is therefore completely absent in Hartree Fock methods. Several methods have been devised to include correlation approximately into Hartree Fock. You will see schemes with names such as CI (*Configuration Interaction*), MP2 or MP3 (*Moller-Pleset Perturbation Theory* to second or third order) or CC (*Coupled Cluster* method). In general all these methods take lots of computer time and scale anywhere from N^5 to N^7 ! So clearly, only for small molecules !