

# Density Functional Theory

## Contents

- *DFT Origins and Overview*
- *The Kohn-Sham Equations*
- *Interpretation of One-Particle Energies*
- *Spin-Polarization*
- *Generalized Gradient Approximation*
- *Atomic Partial Charges and Bader Charge Analysis*
- *Implicit Solvation Model*

## 1 DFT Origins and Overview

In the mid-1960's, Hohenberg and Kohn [1] proved a rather remarkable theorem, which states that from knowledge of the electron density of a physical system such as a solid, a surface, or a molecule everywhere in space the total internal electronic energy (henceforth, the “energy”) of that system is uniquely determined. One can therefore express the energy of an atomistic system as an exclusive functional of its electron density. Furthermore, in most situations of practical interest, to an excellent approximation the total electronic energy of the system does not depend on the motions of the atomic nuclei. This is the Born-Oppenheimer Approximation, which is widely valid, owing to the disparate masses of electrons and nuclei. Thus the atomic nuclei usually can be treated as fixed point charges when calculating the electronic energy of the system. Therefore, the variable part of  $E[\rho]$  can be expressed as a functional exclusively of the electron density everywhere in space,

$$E = E[\rho]. \quad (1)$$

The manner in which  $E[\rho]$  is determined for particular physical systems defines the subject of density functional theory (DFT). The idea of using the electron density as the fundamental entity in a quantum mechanical theory of matter originated in the early days of quantum mechanics in the 1920's, especially with the work of Thomas [2] and of Fermi [3]. However, in subsequent decades, it was the Hartree-Fock (HF) approach [4], rather than DFT, which was first developed and applied to small molecular systems. Calculations on realistic solid state systems were then out of reach. In 1951, Slater [5] used ideas from the electron gas with the intention to simplify Hartree-Fock theory to a point where electronic structure calculations on solids became feasible. Slater's work, which led to the so-called  $X_\alpha$ -method [6], has contributed tremendously to the development of electronic structure calculations.

[1] Pierre Hohenberg and Walter Kohn, “Inhomogeneous Electron Gas,” *Physical Review B* 136, no. 3 (1964): 864-871.

[2] L H Thomas, “The Calculation of Atomic Fields,” *Proc. Cambridge Philos. Soc.* 23 (1927): 542-548.

[3] E Fermi, “Eine Statistische Methode Zur Bestimmung Einiger Eigenschaften Des Atoms Und Ihre Anwendung Auf Die Theorie Des Periodischen Systems,” *Zeitschrift Fur Physik* 48 (1928): 73-79.

[4] D R Hartree, “The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods,” *Proc. Cambridge Philos. Soc.* 24 (1928): 89-110; D R Hartree, “The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part II. Some Results and Discussion,” *Proc. Cambridge Philos. Soc.* 24 (1928): 111-132; Vladimir Fock, “Näherungsmethode Zur Lösung Des Quantenmechanischen Mehrkörperproblems,” *Zeitschrift Fur Physik* 61, no. 1 (January 1930): 126-148; Vladimir Fock, “‘Selfconsistent Field’ Mit Austausch Fur Natrium,” *Zeitschrift Fur Physik* 62, no. 11 (November 1930): 795-805.

[5] J C Slater, “A Simplification of the Hartree-Fock Method,” *Physical Review* 81, no. 3 (February 1951): 385-390.

[6] J C Slater, *Quantum Theory of Molecules and Solids*, McGraw-Hill, 1963.

Hartree-Fock theory and DFT are both mean-field, ground state theories, in which each electron in the system is assumed to move in a time-averaged, smeared out field produced by all the other electrons in the system. Both methods then employ variational principles to obtain a set of coupled equations collectively describing the ground state of the many-electron system, which are, respectively, called the Hartree-Fock equations or the Kohn-Sham equations [7]. In their respective methods, these equation sets function as the effective single-particle Schrödinger equations of the electronic system. Either set of equations defines a pseudoeigenvalue problem in which the eigenvalues depend on the eigenfunctions. Thus, these equations must be solved iteratively until self-consistency is achieved between the input and output eigenfunctions and the total energy.

In HF Theory, the occupied single-particle energy eigenvalues may be interpreted directly as a physical observables, based on Koopman's Theorem [8]. Modern-day extensions of HF theory include Moeller-Plesset [9] and Coupled Cluster [10] methods, but these are primarily applied to molecular systems. DFT and its refinements (hybrid functionals, GW methods, etc.) are the modern computational tools of choice for the electronic structure of solids. In DFT, a theorem by Janak [11], discussed later, provides a physical interpretation of DFT eigenvalues analogous to that of Koopman's Theorem for HF eigenvalues. It is noted, however, that in neither case is the total system energy merely a sum of the eigenvalues of the occupied single-electron states multiplied by their occupancies.

HF Theory and DFT differ most fundamentally in the manner, in which the antisymmetry requirement on many-body wavefunctions of groups of indistinguishable fermions, such as electrons, is enforced. This requirement states that when the labels of any two indistinguishable fermions in the list of particles described by the wavefunction are interchanged, or "exchanged," the value of the wavefunction must change sign. The most important consequence of antisymmetry is exclusion, which prohibits any two particles described by the same indistinguishable fermion wavefunction from having the same exact list of quantum numbers. This principle forces the electrons of a many-electron system to populate a set of states of increasingly higher energy, in accordance with the so called Aufbau principle. This effect dictates the basic structure of many-electron ground states. Separately, antisymmetry mediates the Coulomb interaction between electrons by keeping pairs of electrons in similar states and with parallel spins, farther apart. This effect is known as the exchange interaction.

HF Theory enforces antisymmetry by expressing a manybody wave function as the determinant of a matrix whose  $ij$ 'th element is the  $[i]$ 'th single-particle wavefunction of the  $[j]$ 'th particle in the system. Here  $[i]$  represents the  $i$ 'th complete set of single-particle quantum numbers defining the occupied single-particle state "i," and  $[j]$  represents the three spatial coordinates and the spin of the  $j$ 'th particle. (Interchanging any two particle labels in a determinantal wavefunction is the equivalent of interchanging two rows or columns of the matrix whose determinant is formed. Thus, determinantal wavefunctions are manifestly antisymmetric.) In DFT, the electron density is formed as a simple sum over the square moduli of the occupied states, so the effects of antisymmetry do not appear naturally in the energy operator, that is, the Hamiltonian. Therefore, terms accounting for antisymmetry effects, principally the exchange term, must be manually inserted. In either theory, state occupancies are manually constrained in accordance with exclusion, and orthonormality of the eigenfunctions is enforced by means of a set of Lagrange multipliers. However, the two differing means of imposing or incorporating antisymmetry have far-reaching consequences.

Because of exclusion, a many-electron wavefunction will tend toward zero amplitude whenever any two electrons with parallel spins are close together. But even electrons with opposing spins repel one another through the Coulomb interaction, which becomes increasingly strong as two electrons approach one another. Thus, regardless of their spins, the instantaneous position of any individual electron in a many-electron system is correlated with the instantaneous positions of all other electrons in the system. Mean-field theories like HF and DFT can only capture these "correlation effects" in an approximate manner.

For historical reasons, the "correlation energy" is defined as the difference between the exact non-relativistic electronic energy and the Hartree-Fock energy of the system. In HF-based approaches, the correlation energy may be estimated by combining multiple determinantal wave functions, or elements thereof in the same

[7] Walter Kohn and L J Sham, "Self Consistent Equations Including Exchange and Correlation Effects," *Physical Review A* 140, no. 4 (1965): 1133-1138.

[8] T. Koopmans, "Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms.," *Physica (Amsterdam)* 1, (1934): 104

[9] C. Moeller and M. S. Plesset, "Note on an Approximation Treatment for Many-Electron Systems", *Physical Review* 46, (1934): 618

[10] J. Cizek, "On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods.," *The Journal of Chemical Physics* 45, (1966): 4256

[11] J Janak, "Proof That  $\delta E/\delta N_i = E$  in Density-Functional Theory," *Physical Review B* 18, no. 12 (December 15, 1978): 7165.

wavefunction. In DFT-based approaches, the correlation energy is expressed as a spatial integral over a function of the local electron density, often with gradient corrections. This local density function, called the correlation potential, can be estimated for an “inhomogeneous” electron gas, such as that existing within any real material, based on first principles theoretical results for the “homogeneous,” or uniform electron gas. In the Local Density Approximation (LDA) [7], the correlation potential of a homogeneous electron gas at a particular point is set equal to the correlation potential of a homogeneous electron gas of the same density as exists at that point in the inhomogeneous system. In the Generalized Gradient Approximation (GGA), a density gradient correction term is included to account for inhomogeneity, and in addition a number of so-called sum rules are enforced to be obeyed. Generally, the biggest improvements offered by the GGA relative to the LDA are realized for purposes of geometry optimization.

The correlation energy for an inhomogeneous electron gas converges to the uniform electron gas result of the same local density at both extremes of density: infinite electron density, at which the Thomas-Fermi screening length goes to zero; and vanishingly small density, at which correlation effects go to zero. In real materials with electron densities between the extremes, expressions for the correlation potential are necessarily imprecise. However, they must satisfy various additional conditions.

Antisymmetry also leads to an “exchange” term in the total energy expression, resulting from integrals of products of single-particle wavefunctions with interchanged, or “exchanged,” state and particle labels, multiplied by the Coulomb operator. In HF Theory, this term can produce lists of computationally arduous integrals over products of functions localized at up to four different nuclear centers. The root of the difficulty is that the exchange potential is intrinsically “nonlocal,” in that it depends on the values of wavefunctions at all points within the computational domain. In DFT, this term is replaced by a much more computationally tractable integral involving only point functions of the local charge density and the Coulomb operator.

The exchange and correlation potentials, both originating from antisymmetry, usually are combined in DFT functionals into a joint exchange-correlation (XC) potential. In essence, traditional DFT accepts a less precise local description of the Coulomb exchange term, which is intrinsically nonlocal, in return for more serviceable approximations of electron correlation, combined with greater computational facility. This greater computational efficiency enables one to analyze more complex, and often more realistic material models with larger numbers of atoms. DFT also has less difficulty with open-shell electronic configurations, as are encountered, for example, in transition metals, which cannot be described by a single-determinant wavefunction.

Traditional DFT, that is, the LDA and GGA, are recommended for properties of the electronic ground state of systems, such as total energies, forces, and phonon spectra. In several modern extensions of traditional DFT, a degree of non-local exchange, combined with a screening factor, or even exact HF exchange, is mixed into the XC potential. These hybrid functionals can significantly improve the description of excited states, bandgap energies, and optical and Fermi surface properties, but at a greater computational cost. *MedeA VASP* offers these advanced functionalities in addition to LDA and GGA options.

## 2 The Kohn-Sham Equations

In solid-state systems, molecules, and atoms, the electron density is a scalar function defined at each point  $\mathbf{r}$  in real space,

$$\rho = \rho(\mathbf{r}) \quad (2)$$

The electron density and the total energy depend on the type and arrangements of the atomic nuclei. Therefore, one can write

$$E = E[\rho(\mathbf{r}), \{R_\alpha\}] \quad (3)$$

The set  $\{R_\alpha\}$  denotes the positions of all atoms  $\alpha$  in the system under consideration. Eq. (3) is the key to the atomic-scale understanding of electronic, structural, and dynamic properties of matter. If one has a way of evaluating this expression one can, for example, predict the equilibrium structure of a solid, the reconstruction of free surfaces, and the equilibrium geometries of molecules adsorbed on surfaces.

Furthermore, the derivative of the total energy Eq. (3) with respect to a displacement of an atomic nucleus from equilibrium yields the force acting on that atom corresponding to that displacement. This fact enables efficient searches for stable structures and, perhaps more importantly, the study of dynamical processes

such as diffusion and the reactions between atoms and molecules on surfaces. As already mentioned, the computational methods described here exploit the Born-Oppenheimer approximation, in which it is assumed that the motions of the electrons are infinitely faster than those of the nuclei. This fact justifies calculating the electronic structure for a fixed atomic arrangement, and then moving the atoms according to classical mechanics. Born-Oppenheimer is an especially good approximation for ground state properties of heavy atoms like tungsten, but may introduce some errors for light atoms such as hydrogen or lithium. This approximation also breaks down in certain special cases, such as in superconductors below their critical temperatures.

In DFT, the total energy is decomposed into three parts: a part describing the electronic kinetic energy, an electrostatic or Coulomb energy part, and the so-called exchange-correlation energy already introduced,

$$E = T_0 + U + E_{xc} \quad (4)$$

The most straightforward term in the total energy is the Coulomb energy  $U$ . It is purely classical, and consists of the electrostatic energy arising from the Coulomb attraction between electrons and nuclei, the repulsion between all electronic charges, and the repulsion between nuclei,

$$U = U_{en} + U_{ee} + U_{nn} \quad (5)$$

With

$$U_{en} = -e^2 \sum Z_\alpha \int \frac{\rho(r)}{|r - R_\alpha|} dr \quad (6)$$

$$U_{ee} = e^2 \int \int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr' \quad (7)$$

$$U_{nn} = e^2 \sum \frac{Z_\alpha Z_{\alpha'}}{|R_\alpha - R_{\alpha'}|} \quad (8)$$

where  $e$  is the elementary charge of a proton and  $Z_\alpha$  is the atomic number of an atom  $\alpha$ . The summations extend over all atoms, and the spatial integrations are over the entire computational domain. Once the electron density at all points, and the atomic numbers and positions of all atomic nuclei are known, the above expression can be evaluated using the techniques of classical electrostatics.

The kinetic energy term,  $T_0$ , is more subtle and is discussed in detail here. In DFT, the “real” electrons of a system are replaced by “effective” electrons, often called “electron quasiparticles”, with the same charge as real electrons. The physical justification for the independent particle picture, which actually refers to electron quasiparticles, is that the wavefunction defining the state of a many-particle quantum system can be represented as a linear combination of products of single-particle wavefunctions. (Note: unless a spin-polarized version of DFT is considered, these so called single-particle states can actually hold a pair of electrons with identical charge densities, but opposite spin.) In principle, all the physically observable characteristics of the many-body system can be reproduced in this way. In its ground state, individual particles may exist only in one of a finite set of allowed single-particle states. Particles in different single-particle states can only interact (i.e., scatter) with one another if energy is added to the system to lift one or more particles into previously unoccupied states of higher energy. It follows that in the ground state, the electron quasiparticles occupying the single-particle states of the system, which are unable to scatter with one another, act in much the same way as if they were truly independent particles.

Now as previously discussed, the instantaneous motion of a “real” electron is correlated with the instantaneous motions of all other electrons in the system. Hence, because effective electrons, by definition, see only a constant, time-averaged effective potential created by all other electrons in the system, terms must be added to the effective single-particle energy operator, or Hamiltonian, to capture dynamic correlation effects. The total density of all effective electrons is the same as the real total electron density. However, the effective masses of effective electrons, or electron quasiparticles, are such that a version of Newton’s Second Law remains applicable, but in a generalized form: where the scalar electron mass is replaced by the effective mass tensor  $M$ , reflecting the fact that electrons in a crystal subject to an external applied force are simultaneously subject to the crystal field of the material.

$T_0$  is the sum of the kinetic energies of all effective electrons moving as independent particles. (Usually, it is not necessary to explicitly maintain this distinction between real and effective electrons.) If the state of each effective electron is described by a single-electron wave function,  $\Psi_i$ , then the kinetic energy of all effective electrons in the system is given by,

$$T_0 = \sum_i n_i \int \Psi_i^*(r) \left[ \frac{-\hbar^2}{2m} \nabla^2 \right] \Psi_i(r) dr \quad (9)$$

This expression is the sum of the expectation values of single-electron kinetic energies.  $n_i$  denotes the number of electrons in the state  $i$ . By construction, the effects of the crystal field, and dynamical correlations between the electrons are excluded from  $T_0$ , so that the free electron masses, rather than the effective masses, are relevant here. (More on this point, below.)

The third term in Eq. (4), the exchange-correlation energy,  $E_{xc}$ , includes all remaining complicated electronic contributions to the total energy. The most important of these contributions is the exchange term, which is a direct consequence of the exclusion principle. As previously mentioned, when the energy of a many-electron system is evaluated, the antisymmetry requirement gives rise to terms involving products of pairs of single-electron wave functions in which the state and particle coordinate labels are “exchanged” relative to one another. This term in the energy enters with the opposite sign from that of the direct Coulomb term, and therefore reduces the net average Coulomb repulsion between electrons of parallel spin.

The exchange term in the energy is purely quantum in nature with no classical analog, and as mentioned, it is only nonzero for pairs of electrons with parallel spin. Forcing two electrons with parallel spin to be in the same place requires the antisymmetric many-electron wave function to equal its own negative, meaning that the wavefunction must have vanishing amplitude whenever two electrons with parallel spin approach one another. (This would be true, even if electrons were uncharged and experienced no Coulomb repulsion!) This fact means that each electron carries with it an “exchange hole,” which can be thought of as a small surrounding volume of space in which the probability of finding another electron with parallel spin is dynamically reduced relative to the probability expected from mean field theories. Slater showed that the total charge integrated over the entire exchange hole is  $+e^{162}$ .

In consequence of exchange effects, the antisymmetric nature of many-electron wavefunctions naturally incorporates a measure of dynamic correlation into mean field approximations to the system energy, but only for pairs of electrons with parallel spin. This is the case with DFT. However, the effects of dynamic correlation between pairs of electrons with antiparallel spins are as yet completely unaccounted for in the current treatment. The remaining effects of dynamic correlations between all electrons, regardless of spin, must be captured by introducing additional terms to the expression for  $E[\rho]$ .

As an illustration of the relative importance of the various contributions to the total electronic energy, the total energy of a single carbon atom is approximately -1,019 eV, that of a silicon atom is -7,859 eV, and that of a tungsten atom is -439,634 eV. In all three cases, the kinetic energy and the Coulomb energy contributions are of similar magnitude but of opposite sign (a result guaranteed by the Virial Theorem). The magnitude of the exchange-correlation term is about 10% of that of the Coulomb term, and is attractive for electrons (because the exchange-hole is positive). The correlation energy is smaller in magnitude than the exchange energy, but it nevertheless plays an important role in determining the details of the length and strength of interatomic bonds, as well as of band gap energies, energy densities of states and Fermi Surface properties in crystalline solids.

Compared with the total energy per atom in a solid material, the binding energy of an atom in a solid or on a surface is quite small, in the range from about 1 to 8 eV. Energy differences influencing changes in the equilibrium positions of atoms, as for example on a reconstructed surface, can be even smaller. For example, only about 0.03 eV are required to flip an asymmetric Si-dimer on a reconstructed Si(001) surface from one configuration into another, reversing the role of the upper and lower Si atoms. It is a tremendous challenge for any theory to cope with a range of energies of nearly five order of magnitude, from tenths to thousands of eV. Density functional theory meets this challenge remarkably well in a surprisingly wide range of cases.

The Hohenberg-Kohn Theorem, which is central to DFT, states that the ground state total energy of an electronic system, which is uniquely determined by the corresponding ground state charge density, is at its minimum value and stationary with respect to first-order variations in that charge density. That is,

$$\frac{\partial E[\rho]}{\partial \rho} \Big|_{\rho=\rho_0} = 0 \quad (10)$$

In conjunction with the kinetic energy, we have introduced single-electron wave functions  $\Psi_i$ , from which the total electron density can be generated.

$$\rho(r) = \sum_i n_i |\Psi_i(r)|^2 \quad (11)$$

As in the expression for the kinetic energy,  $n_i$  denotes the occupation number of the eigenstate  $i$ , which is represented by the single-electron wave function  $\Psi_i$ . To this point in its development, DFT is formally exact in the sense that no approximations yet have been made in describing the many-electron interactions. By construction,  $\rho$  in Eq. (11) is the exact many-electron density.

Equations that can be used for practical density functional calculations next must be derived, which necessarily involve approximations. Through equations (9) and (11) we have introduced single-electron wave functions. A change in these wave functions corresponds to a variation in the electron density. Therefore, the variational condition (10) can be used to derive the conditions for the one-particle wave functions that lead to the ground state electron density. To this end, one substitutes Eq. (11) into expression (10) and varies the total energy with respect to each wave function. This procedure leads to the following equations:

$$\left[ \frac{-\hbar^2}{2m} \nabla^2 + V_{eff}(r) \right] \Psi_i(r) = \epsilon_i \Psi_i(r) \quad (12)$$

with

$$V_{eff}(r) = V_c(r) + \mu_{xc}[\rho(r)] \quad (13)$$

Equations (12) are called the Kohn-Sham equations. The electron density, which is constructed from the single-electron wave functions that are the solutions of Equations (12), is the ground state density that corresponds to the minimum total energy. The solutions of the Kohn-Sham equations form an orthonormal set, *i.e.*

$$\int \Psi_i^*(r) \Psi_j(r) dr = \delta_{ij} \quad (14)$$

This additional constraint was implemented in Equations (12) by introducing “Lagrange multipliers,”  $\epsilon_i$ , into Equation (10). These “Lagrange multipliers” are effective single-electron energy eigenvalues and their interpretation will be discussed more later. These eigenvalues are used to determine the occupation numbers  $n_i$  of the single-electron states by applying the Aufbau principle. According to this principle, eigenstates are filled to their maximum occupancy allowed by i Exclusion in order of increasing eigenvalues, starting with the state with the lowest (*i.e.*, most negative) eigenvalue. For non-spin polarized systems, each state may be occupied by up to two electrons until all electrons are accommodated. In spin polarized systems, each state is occupied by at most one electron. In either case, the highest occupied state or states is/are permitted to have partial occupancy.

As a consequence of the partitioning of the total energy in Eq. (4), the Hamiltonian operator in the Kohn-Sham equations (12) contains three terms, one for the kinetic energy, the second for the Coulomb potential, and the third for the exchange-correlation potential. The kinetic energy term is the standard second-order differential operator familiar from single-particle Schrödinger equations and its construction does not require knowledge of the specific system being modeled. In contrast, the Coulomb potential operator,  $V_c$ , and the exchange-correlation potential operator,  $\mu_{xc}$ , both depend on the electron distribution in the specific system under consideration.

The Coulomb, or electrostatic potential  $V_c$  at point  $r$  is generated from the electric charges of all nuclei and electrons in the system. It can be evaluated directly in real space via,

$$V_c(r) = -e^2 \sum_{\alpha} \frac{Z_{\alpha}}{|r - R_{\alpha}|} + e^2 \int \frac{\rho(r')}{|r - r'|} dr' \quad (15)$$

In condensed matter systems it is more convenient to use Poisson’s equation, given below, to evaluate the Coulomb potential:

$$\nabla^2 V_c(r) = -4\pi e^2 q(r) \quad (16)$$

Here,  $q$  denotes both the electronic charge distribution  $\rho$  and the positive point charges of the nuclei at positions  $R_{\alpha}$ .

The exchange-correlation potential is related to the exchange-correlation energy by

$$\mu_{xc} = \frac{\partial E_{xc}[\rho]}{\partial \rho} \quad (17)$$

Equation (17) is formally exact in the sense that it does not contain any approximations to the complete many-body interactions. In practice, however, the exchange-correlation energy (and thus the exchange-correlation potential) is not known for a spatially varying electron gas. Therefore, one needs to develop approximate expressions for the XC potential in terms of the electron density (since, as a consequence of the Kohn-Sham theorem, the exchange-correlation energy depends only on the electron density). A simple and, as it turns out, surprisingly good approximation is to assume that the exchange-correlation energy depends only on the local electron density around each volume element  $dr$ . This is called the local density approximation (LDA)

$$E_{xc}[\rho] = \int \rho(r) \epsilon_{xc}^0(\rho(r)) dr \quad (18)$$

Below is an illustration of the basic idea. In any atomic arrangement such as a crystal, a surface, or a molecule, there is a certain electron density  $\rho$  at each point  $r$  in space. The LDA then rests on two basic assumptions:

1. the exchange and correlation effects come predominantly from the immediate vicinity of a point  $r$  and
2. these exchange and correlation effects do not depend strongly on variations of the electron density in the vicinity of  $r$ .

If conditions (i) and (ii) are reasonably well fulfilled, then the contribution from volume element  $dr$  would be the same as if this volume element were surrounded by a constant electron density of the same value as within  $dr$ . This is an excellent approximation for many metallic systems, but results in more significant errors in systems with strongly varying electron densities.

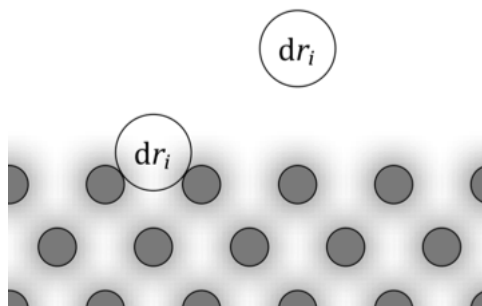


Illustration of the local density approximation: For the purpose of computing the exchange-correlation energy in a volume element  $dr_i$ , the electron density  $\rho_i$  around point  $r_i$  is assumed to be constant in the near surrounding volume. Note that the value of  $\rho_i$  is different in each volume element.

A system of interacting electrons with a constant density is called a homogeneous electron gas. Extensive theoretical efforts have been made to understand and characterize such an idealized system. In particular, the exchange-correlation energy per electron of a homogeneous electron gas,  $\epsilon_{xc}^0(\rho)$  has been determined by several approaches such as many-body perturbation theory by Hedin and Lundqvist [12], and with quantum Monte-Carlo methods by Ceperley and Alder [13]. As a result,  $\epsilon_{xc}(\rho)$ , is quite accurately known for a range of densities. For practical calculations,  $\epsilon_{xc}(\rho)$  is expressed as an analytical function of the electron density. There are various analytical forms with different coefficients in their representation of the exchange-correlation terms. These coefficients are not adjustable parameters, but rather they are determined from first principles. Hence, the LDA is a first-principles approach, which is distinct in meaning from an “exact” approach, in the sense that the quantum mechanical problem is solved without any adjustable, arbitrary, or system-dependent parameters.

Below, an example for analytical expressions used in LDA computer programs is shown. Explicit forms for the local density exchange were given originally by Gàspar [14] and Kohn & Sham [7]. Correlation terms are according to Hedin & Lundqvist [12]. Exchange and correlation energies per electron are denoted by  $\epsilon$  and

[12] L Hedin and B I Lundqvist, “Explicit Local Exchange-Correlation Potentials,” *Journal of Physics C: Solid State Physics* 4, (1971): 2064-2083. U von Barth and L Hedin, “A Local Exchange-Correlation Potential for the Spin Polarized Case. I,” *Journal of Physics C: Solid State Physics*, 1972.

[13] D M Ceperley, “Ground State of the Electron Gas by a Stochastic Method,” *Physical Review Letters* 45, no. 7 (August 1980): 566-569.

[14] Reszö Gàspar, “Über Eine Approximation Des Hartree-Fockschen Potentials Durch Eine Universelle Potentialfunktion,” *Acta Physica Academiae Scientiarum Hungaricae* 3, no. 3 (April 1954): 263-286.

the corresponding potentials by  $\mu$ . Both quantities are given in Hartree atomic units (1 Hartree = 2 Rydberg = 27.21165 eV). The units for the electron density are number of electrons/(Bohrradius)<sup>3</sup>.

	Energy	Potential
	$\epsilon_{xc} = \epsilon_x + \epsilon_c$	$\mu = \frac{\partial \rho \epsilon}{\partial \rho}$
Exchange	$\epsilon_x = -\frac{3}{2} \left( \frac{3}{\pi} \rho \right)^{\frac{1}{3}}$	$\mu_x = -2 \left( \frac{3}{\pi} \rho \right)^{\frac{1}{3}}$
Correlation	$\epsilon_c = -c \left[ \left( 1 + x^3 \right) \ln \left( 1 + \frac{1}{x} \right) + \frac{x}{2} - x^3 - \frac{1}{3} \right]$	$\mu_c = -c \ln \left( 1 + \frac{1}{x} \right)$
	$c = 0.0225 \quad x = \frac{r_s}{21} \quad r_s = \left( \frac{3}{4\pi\rho} \right)^{\frac{1}{3}}$	

Note that there are two types of exchange-correlation terms, one type for the energy and the other for the potential. The energy,  $\epsilon_{xc}(\rho)$ , is needed to evaluate the total energy and the potential term,  $\mu_{xc}(\rho)$ , is required for the Kohn-Sham equations. The two terms are, following (17) and (18), related.

$$\mu_{xc} = \frac{\partial (\rho \epsilon_{xc}(\rho))}{\partial \rho} \quad (19)$$

Using the explicit formulas given above, one can evaluate the exchange-correlation potential within the local density approximation for any electron density distribution  $\rho$ . Thus, all terms of the effective single-particle operator in the Kohn-Sham equations are now defined, and one can proceed with a computational implementation.

### 3 Interpretation of One-Particle Energies

The fundamental quantities in density functional theory are the electron density and the corresponding total energy, but not the one-particle eigenvalues. However, the one-electron picture is so useful that one seeks to exploit the Kohn-Sham eigenvalues and one-particle wave functions as much as possible. The one-particle energies of effective electrons have been introduced in the derivation of the Kohn-Sham equations as Lagrange multipliers. The Kohn-Sham equations have the form of an eigenvalue problem in which each wave function has an associated eigenvalue  $\epsilon_i$  with an occupation number of  $n_i$ . Janak's theorem [11] provides a relationship between the total energy and these eigenvalues.

$$\epsilon_i = \frac{\partial E}{\partial n_i} \quad (20)$$

The eigenvalue  $\epsilon_i$  equals the change of the total energy with respect to the change in the occupation number of level  $i$ . However, it is desirable to seek a more direct physical interpretation of these eigenvalues. The independent particle picture, in which many-electron states are viewed as phase-coherent superpositions of one-electron states, was firmly established in solid-state and molecular physics long before the advent of DFT. For example, the distinction between a metal and an insulator is explained by differences in their energy band structures. (Energy bands are one-electron energies of allowed states of a periodic potential plotted in reciprocal space as a function of different directions.) These same band structures inform the electronic and optical characteristics of semiconductors and semiconductor/metal junctions. Photoemission experiments also are conveniently interpreted in a one-electron picture, often offering quite reasonable quantitative agreement between theory and experiment. Likewise, the analysis of the *s*, *p*, and *d* character of partial electron energy densities of states (i.e., the number of allowed states per unit energy) has become an extremely useful tool in the understanding of chemical bonding in alloys and compounds. Electronic band structures and partial densities of states are easily obtained from *MedeA VASP*.

The LDA and GGA levels of theory often can provide qualitatively reasonable optical spectra. For this purpose, it often suffices to evaluate only the "direct" optical transitions, which are those for which the wave



vectors (or crystal momenta) of the initial and final electronic states remain the same. For each of these transitions, absorption can occur for photons whose frequency corresponds to the energy difference between the initially occupied and unoccupied states. Integration over all nonequivalent wave vectors in the reciprocal space of the periodic potential gives the joint energy density of states. The intensity of each transition is then computed from the wave functions for each pair of occupied and unoccupied states for all nonequivalent wave vectors.

The computed electric dipole matrix elements combine with the joint density of states, yielding the imaginary part of the dielectric function  $\epsilon_2$ . From  $\epsilon_2$  one obtains the real part of the dielectric function  $\epsilon_1$  from the Kramers-Kronig Relation. Knowledge of  $\epsilon_1$  and  $\epsilon_2$ , i.e. the complex dielectric function, is sufficient to derive the absorption coefficient, reflectivity, and the refractive index. In *MedeA VASP*, the relevant matrix elements are directly computed by VASP. The Kramers-Kronig transformation is performed by VASP, which provides the complex dielectric function as a function of energy. From this primary information the other optical functions are computed and displayed in *MedeA*. The dielectric function as computed by VASP is accessible in the OUTCAR file. If desired, the individual dipole transition matrix elements can be accessed there as well.

The direct interpretation of the Kohn-Sham eigenvalues to derive excitation energies often gives quantitative agreement with experimental photoemission spectra and reasonable qualitative agreement with optical spectra (if excitonic effects are minor). However, significant discrepancies with experiment often arise regarding quantities such as energy band gaps and Fermi surface properties of semiconductors and insulators. Discrepancies greater than a factor of two can be found between measured band gap values and the LDA or GGA eigenvalues corresponding to the highest occupied and lowest unoccupied single-electron states. The LDA/GGA even predicts some narrow band gap semiconductors such as InSb to be semimetals, with overlapping valence and conduction bands. This is because the LDA and GGA were developed to study system ground states, whereas optical properties and energy band gaps necessarily involve the virtual, or excited states of a system.

Thus, the discrepancies mentioned above are not necessarily a “failure” of the LDA and GGA, inasmuch as they represent an inaccurate application of LDA/GGA results beyond their range of validity. However, as such, these difficulties do reveal an important limitation of the LDA, and one which is not substantively improved by the GGA. In the derivation of the Kohn-Sham equations, the effective single-particle eigenvalues were never shown or postulated to be physically observable excitation energies! Only the total electron density and the corresponding total system energies predicted by the LDA and GGA have rigorous physical interpretations. It was to be expected, therefore, that the LDA and GGA would have difficulties predicting properties of the system requiring accurate knowledge of excited states.

Nevertheless, DFT eigenvalues and eigenfunctions offer a very useful foundation for more rigorous theories that can predict excited state properties of materials, especially those of semiconductors and insulators, with far greater accuracy than the LDA/GGA itself. One of the most rigorous, and computationally demanding approaches is that of the Green’s Function - Screened Coulomb Exchange method, commonly known as the “GW” method, which was pioneered by Hybertsen & Louie [15]. More computationally tractable approaches add a measure of nonlocal exchange to the LDA/GGA local exchange-correlation potential, usually with an exponential screening factor. These so called hybrid functional and screened exchange methods often approach the same level of quantitative accuracy as the GW method (other than for excitonic properties), at a fraction of the computational cost. The greater computational facility of hybrid functional methods sometimes translates into results with greater overall reliability, owing to the ability to perform more thorough k-space sampling, or to analyze more realistic geometric models than would be possible in GW computations. VASP offers a full menu of GW, screened exchange and hybrid functional options.

The highest occupied electronic level in a metallic system is called the Fermi energy or Fermi level,  $E_F$ . (In a semiconductor or an insulator, the Fermi Energy lies in an energy band gap.) The nature of the electronic states at  $E_F$  play a crucial role in determining materials properties such as electrical and thermal conductivity, magnetism, and superconductivity. The *MedeA Electronics* module enables visualization of the Fermi surface and provides access to the transport properties. On surfaces, the energy difference between  $E_F$  and the electrostatic potential in the vacuum region,  $V_0$ , above the surface is the work function,  $\phi$ . While in general the Kohn-Sham eigenvalues are not excitation energies, it was shown by Schulte [16] that for a metallic system the highest occupied Kohn-Sham eigenvalue can be directly interpreted as the work function. Thus, the agreement between experimental and calculated work functions provides a good test for the quality of

[15] Mark Hybertsen and Steven Louie, “First-Principles Theory of Quasiparticles: Calculation of Band Gaps in Semiconductors and Insulators,” *Physical Review Letters* 55, no. 13 (September 23, 1985): 1418.

[16] F Schulte, “On the Theory of the Work Function,” *Zeitschrift Fur Physik B Condensed Matter* 27, no. 4 (1977): 303-307.

actual calculations. With present LDA approaches, the calculated values are typically within 0.1-0.2 eV of the experimental results.

## 4 Spin-Polarization

So far, the discussion of density functional theory was restricted to non-spin-polarized cases. However, many systems such as magnetic transition metals or molecules such as O<sub>2</sub> involve unpaired electrons or molecular radicals and thus require a spin-polarized method. In such systems, the number of electrons with “spin-up” can be different from that with “spin-down”. In the early 1970’s, von Barth & Hedin [17] and Gunnarson, Lundqvist & Lundqvist [18] generalized density functional theory to accommodate spin-polarized systems. This resulted in a spin density functional theory known as the local spin density (LSD) approximation.

The explicit form of local spin density exchange-correlation terms given by von Barth & Hedin is,

$$\begin{aligned} \epsilon_{xc}(\rho, \zeta) &= \epsilon_{xc}^p(\rho, \zeta) + [\epsilon_{xc}^f(\rho, \zeta) - \epsilon_{xc}^p(\rho, \zeta)] f(\zeta) \\ \epsilon_{xc}^p &= \epsilon_x^p + \epsilon_c^p \quad \epsilon_x^p = \frac{3}{2} \left( \frac{3}{\pi} \rho \right)^{\frac{1}{3}} \quad \epsilon_c^p = -0.0225 F \left( \frac{r_s}{21} \right) \\ \epsilon_{xc}^f &= \epsilon_x^f + \epsilon_c^f \quad \epsilon_x^f = 2^{\frac{1}{3}} \epsilon_x^p \quad \epsilon_c^f = -0.01125 F \left( \frac{r_s}{53} \right) \\ \zeta &= \frac{\rho_\uparrow - \rho_\downarrow}{\rho} \quad f(\zeta) = \frac{(1+\zeta)^{\frac{4}{3}} + (1-\zeta)^{\frac{4}{3}} - 2}{2^{\frac{4}{3}}} \\ F(x) &= (1+x^3) \ln \left( 1 + \frac{1}{x} \right) + \frac{x}{2} - x^3 - \frac{1}{3} \end{aligned}$$

### Exchange-correlation potential

$$\begin{aligned} \mu_{xc}^\sigma &= A(\rho) \left( \frac{2\rho_\sigma}{\rho} \right)^{\frac{1}{3}} + B(\rho) \quad \text{with } \sigma = \uparrow \text{ or } \downarrow \\ A(\rho) &= \mu_{xc}^\sigma + v_c(r_s) \quad B(\rho) = \mu_c^p(r_s) \\ \mu_c^p &= -0.0225 \ln \left( 1 + \frac{21}{r_s} \right) \quad r_s = \left( \frac{3}{4\pi\rho} \right)^{\frac{1}{3}} \\ v_c &= -\frac{4}{3} \frac{1}{2^{\frac{1}{3}} - 1} \left[ 0.01125 F \left( \frac{r_s}{53} \right) - 0.0225 F \left( \frac{r_s}{21} \right) \right] \end{aligned}$$

Energies and potentials are given in Hartree atomic units; the units for electron and spin densities are number of electrons/(Bohrradius)<sup>3</sup>.

In the local spin density functional (LSDF) theory, the fundamental quantities are both the electron density,  $\rho$ , and the spin density,  $\sigma$ . The spin density is defined as the difference between the density of the spin-up

[17] U von Barth and L Hedin, “A Local Exchange-Correlation Potential for the Spin Polarized Case. I,” Journal of Physics C: Solid State Physics, 1972.

[18] O Gunnarson, BI Lundqvist, and S Lundqvist, “Screening in a Spin-Polarized Electron Liquid,” Solid State Communications 11, no. 1 (1972): 149-153.

electrons and the density of the spin-down electrons

$$\sigma(r) = \rho_{\uparrow}(r) - \rho_{\downarrow}(r) \quad (21)$$

with the total electron density,

$$\rho(r) = \rho_{\uparrow}(r) + \rho_{\downarrow}(r) \quad (22)$$

In LSDF theory, the exchange-correlation potential for spin-up electrons is in general different from that for spin-down electrons. Consequently, the effective potential (13) becomes dependent on the spin. Thus, the Kohn-Sham equations (12) in their spin-polarized form are

$$\left[ \frac{-\hbar^2}{2m} \nabla^2 + V_{eff}^{\sigma}(\epsilon_i^{\sigma}) \right] \Psi_i^{\sigma}(r) = \epsilon_i^{\sigma} \Psi_i^{\sigma}(r) \quad (23)$$

$\sigma = \uparrow$  or  $\downarrow$

with

$$V_{eff}^{\sigma} = V_c + \mu_{xc}^{\sigma}[\rho(r), \sigma(r)] \quad (24)$$

The exchange-correlation potential in LSDF theory depends on both the electron density and the spin density, as written in equation (24). There are two sets of single-particle wave functions, one for spin-up electrons and one for spin-down electrons, each with their corresponding one-electron eigenvalues. For the case of equal spin-up and spin-down densities, the spin density is zero throughout space and LSDF theory becomes identical with the LDF approach. Notice that in spin-polarized calculations, the occupation of single-particle states is 1 or 0, but there is still only one Fermi energy. In magnetic systems, the spin-up and spin-down electrons are often referred to as “majority” and “minority” spin systems. The table in the next section gives an example of the local spin density exchange-correlation formula by von Barth & Hedin [17].

## 5 Generalized Gradient Approximation

A large number of total energy calculations have shown that the LDA gives interatomic bond lengths within  $\pm 0.05$  of experiment or better for a great variety of solids, surfaces, and molecules. However, the following systematic errors, all associated with “overbinding,” are found:

- most lattice parameters predicted with LDA are too short.
- weak bonds are noticeably too short, for example the Ni-C bond in the Ni carbonyl Ni(CO)<sub>4</sub>, the bond between two magnesium atoms (which are closed shell systems), and the length of hydrogen bonds such as that in the water dimer HOH-OH<sub>2</sub>;
- the binding energies calculated with the LDA are typically too large, sometimes by as much as 50% in strongly bound systems and even more in weakly bound materials.

Gradient-corrected density functionals as suggested by Perdew [19], Becke [20], Perdew & Wang [21] and Perdew, Burke & Ernzerhof [22] offer a remedy. The basic idea in these schemes is the inclusion of terms in the exchange-correlation expressions that depend on the gradient of the electron density and not only on its value at each point in space. Therefore, these corrections are also sometimes referred to as “non-local” (or “semi-local”) potentials, as the gradient term implicitly involves the value of the density at more than one point. However, this kind of term is distinct from globally non-local potentials such as exact exchange. The table below gives the form suggested by Becke [20] for the exchange part and by Perdew [19] for the correlation part.

[19] John P Perdew, “Density Functional Approximation for the Correlation Energy of the Inhomogeneous Electron Gas,” *Physical Review B* 33, no. 12 (1986): 8822-8824.

[20] A D Becke, “Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior,” *Physical Review A* 38, no. 6 (1988): 3098.

[21] John P Perdew and Yue Wang, “Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy,” *Physical Review B* 45, no. 23 (June 1992): 13244-13249.

[22] John P Perdew, Kieron Burke, and Matthias Ernzerhof, “Generalized Gradient Approximation Made Simple,” *Physical Review Letters* 77, no. 18 (October 1996): 3865-3868.

While dissociation energies calculated with these corrections rival the best post-Hartree-Fock quantum chemistry methods in accuracy, gradient-corrected density functional calculations are computationally much less demanding and more widely applicable. Gradient-corrected density functionals have been studied extensively for molecular systems, for example by Andzelm & Wimmer [23]. The results are very encouraging and this approach could turn out to be of great value in providing quantitative thermochemical data.

The one-particle eigenvalues obtained with gradient-corrected exchange-correlation potentials are not significantly different from the LDA eigenvalues. Therefore, these potentials do not (and are not intended to) remove the discrepancy between calculated and measured energy band gaps. Their primary value is in better describing ground state properties, whereas bandgap energies involve unoccupied excitation states as well as occupied states.

### Gradient-correction to the total energy for exchange by Becke and correlation by Perdew

$$\begin{aligned}
 E_{GGA} &= E_{LSD} + E_X^G + E_C^G \\
 E_X^G &= b \sum \int \frac{\rho_\sigma \chi_\sigma^2}{1 + 6b\chi_\sigma \sinh^{-1} \chi_\sigma} dr \\
 \chi_\sigma &= \frac{|\nabla \rho|}{\rho^{3/4}} \quad \sigma = \uparrow \text{ or } \downarrow \\
 E_C^G &= \int f(\rho_\uparrow, \rho_\downarrow)^{-g(\rho)\nabla \rho} |\nabla \rho|^2 dr
 \end{aligned} \tag{25}$$

Above, energies are given in Hartree atomic units; the units for the electron and spin densities are number of electrons / (Bohr radius)<sup>3</sup>. The constant  $b$  in Becke's formula is a parameter fitted to the exchange energy of inert gases. The explicit form of the functions  $f$  and  $g$  in Perdew's expression for the correlation energy is given in the original paper.

## 5.1 Relativistic Effects

Electrons near an atomic nucleus achieve such high kinetic energies that relativistic effects become noticeable even for light atoms near the beginning of the periodic table. For elements with an atomic number above about 54 (Xe) these relativistic effects become quite important and should be included in electronic structure calculations. The relativistic mass enhancements of electrons in states concentrated near a nucleus lead to a contraction of the electronic charge distribution compared with that which would be predicted from a non-relativistic treatment. For atoms with about  $Z > 54$  non-relativistic calculations therefore can overestimate bond lengths by 0.1 Å and more. Furthermore, relativistic effects change the relative energy of  $s$ ,  $p$ ,  $d$ , and  $f$ -states, which can have a significant impact on bonding mechanisms and energies, and energy band topology.

Relativistic effects lead to a spin-orbit splitting between states of differing orbital and total (i.e., orbital plus spin) angular momentum that would otherwise be degenerate (have the same energy). So, for example, a non-relativistic set of single-electron states  $(n, l)$  with principle quantum number  $n$  and orbital angular momentum  $l$  would comprise a  $2(2l+1)$ -fold degenerate multiplet of states with differing azimuthal orbital angular momentum quantum number  $m_l$  and z-component spin quantum number  $m_s$ . When spin-orbit coupling is turned on, this multiplet is split into a  $(2l+2)$ -fold multiplet with  $j = l + 1/2$ , and a  $2l$ -fold multiplet with  $j = l - 1/2$ , where  $j$  is the total angular momentum quantum number. The magnitude of spin-orbit splitting in the  $f$ -shell of atomic Ce is about 0.3 eV between the  $4f_{5/2}$  and  $4f_{7/2}$  states. For core atomic states, the spin-orbit splitting can be very large. For example, the  $2p_{1/2}$  and  $2p_{3/2}$  core states in W are split by 1351 eV. Spin-orbit effects substantively change bandgap values in heavy atom III-V semiconductors, and can lift or change Fermi surface degeneracies, changing optoelectronic and thermoelectric properties of semiconductors. Important effects on surfaces such as Kerr rotation in magneto-optical devices, or x-ray dichroism involve spin-orbit splitting. A relativistic electronic structure theory is necessary to capture these effects. This is accomplished by solving the Dirac equations, as discussed, for example, in the textbooks by Bjorken & Drell [24] and by Messiah [25]. Within a spherically symmetric potential, the Dirac equations, like the non-

[23] J Andzelm and Erich Wimmer, "Density Functional Gaussian-Type-Orbital Approach to Molecular Geometries, Vibrations, and Reaction Energies," *Journal of Physical Chemistry* 96, no. 2 (1992): 1280.

[24] JD Bjorken and SD Drell, *Relativistic Quantum Mechanics*, McGraw-Hill, 1964. JD Bjorken and SD Drell, *Relativistic Quantum Fields*, McGraw-Hill, 1965.

[25]

relativistic Schrödinger equation, can be decomposed into radial and angular parts. As an illustration, we show the radial equations.

$$\frac{-dF_{nlj}(r)}{dr} + \frac{\kappa}{r}F_{nlj}(r) = [E - m + V_{eff}(r)] \quad (26)$$

$$\frac{dG_{nlj}(r)}{dr} + \frac{\kappa}{r}G_{nlj}(r) = [E - m + V_{eff}(r)] \quad (27)$$

$$\begin{array}{cccc}
 & s & p & d & f \\
 l = & 0 & 1 & 2 & 3 \\
 k = & -1 & 1, -2 & 2, -3 & 3, -4 \\
 j = & \frac{1}{2} & \frac{1}{2}, \frac{3}{2} & \frac{3}{2}, \frac{5}{2} & \frac{5}{2}, \frac{7}{2} \\
 \kappa = & \begin{cases} -(l+1) \\ l \end{cases} & \text{and } j = & \begin{cases} l + \frac{1}{2} \\ l - \frac{1}{2} \end{cases}
 \end{array} \quad (28)$$

$$\kappa = \begin{cases} -(l+1) \\ l \end{cases} \quad \text{and } j = \begin{cases} l + \frac{1}{2} \\ l - \frac{1}{2} \end{cases} \quad (29)$$

$F$  and  $G$  are called the “large” and “small” components of the radial wave function. The quantum numbers  $n$  and  $l$  are the principal energy quantum number and the orbital angular momentum quantum number, respectively, and are unchanged from the non-relativistic case.  $j$  is the total angular momentum quantum number, which along with the orbital angular momentum quantum number, determines the spin-orbit-split energy levels, and is used as a subscript to label states such as  $2p_{\frac{1}{2}}$ ,  $2p_{\frac{3}{2}}$  and the  $4f_{\frac{5}{2}}$ . The quantum number  $\kappa$  is a convenient quantity used within relativistic computer programs. The radial part of the charge density is constructed from the large and small components by,

$$\rho(r) = \sum [|F_{nlj}(r)|^2 + |G_{nlj}(r)|^2] \quad (30)$$

By neglecting the small part of the wavefunction, one can derive an approximate treatment of relativistic effects in terms of two-component spinors. Expanding the resulting expression in powers of the fine structure constant and keeping only terms to zero'th and first order yields a sum of two “scalar-relativistic” energy terms and the spin-orbit term. The spin-orbit term represents the coupling of the magnetic field created by an electron's orbital motion to its own intrinsic magnetic moment. Relativistic effects on exchange and correlation aren't as easily analyzed.

The correct treatment of exchange and correlation in a fully relativistic density functional theory is a difficult problem that has not yet been completely resolved. However, reasonable approximations are available.

Koelling & Harmon [26] proposed a semi-relativistic (or scalar-relativistic) treatment of these effects. This approach involves an averaging over the spin-orbit splitting, but retains the kinematic effects. This restores most of the simplicity of a non-relativistic method, but still gives an excellent representation of the core electron distribution and the appropriate (spin-orbit averaged) energies of the valence electrons.

## 6 Atomic Partial Charges and Bader Charge Analysis

The concepts of atomic charges, partial charges, angular momentum and crystal field-split charge projections, and charge transfer are described in this section. These concepts are heavily used for describing and understanding chemical bonding and reactions, not only for molecular systems, but also for surface and bulk systems. The approaches available in *MedeA VASP* for using these concepts are briefly outlined here.

Atomic charges are not precisely defined, owing to the continuous nature of the electron density distribution, the quantum-mechanical uncertainty of electron positions, and the somewhat arbitrary nature of space partitioning among neighboring atoms. Therefore, a quantitative analysis of atomic charges is usually only possible with respect to a suitable reference system, such as bulk vs. surface, or bound vs. isolated atom, etc.

In particular, the method used to partition space into volumes assigned to specific atoms and the interstitial volume strongly affects the quantitative results of a charge analysis. Most straightforwardly, space is partitioned into spheres of different radii around each atom position and the space in between. The VASP code

1. Messiah, “Quantum Mechanics”, North-Holland Publishing Company, Amsterdam, (1974)

[26] D D Koelling and B N Harmon, “A Technique for Relativistic Spin-Polarised Calculations,” *Journal of Physics C: Solid State Physics*, 1977.

considers such a partitioning, and applies two different algorithms to compute atomic charges and charge contributions from different angular momentum channels ( $s$ ,  $p$ ,  $d$ ,  $f$  partial charges) inside the spheres:

i) Projections  $P$  of the wave functions  $f_{nk}$  onto a set of spherical harmonics  $Y_{lm}^N$  centered at each atomic site  $N$  (which represents a natural basis set for atomic-like wave functions)

$$P_{Nlmnk} = \langle Y_{lm}^N | \Phi_{nk} \rangle \quad (31)$$

and the corresponding augmentation part. For this projection scheme the atomic sphere radii can be freely chosen. *MedeA* applies covalent radii for this purpose as a default, which could be modified by setting the RWIGS parameter in the **Add to Input** Tab of the *MedeA VASP GUI*.

ii) Alternatively, a fast projection scheme onto the PAW spheres is implemented, for which the radii cannot be freely chosen but are fixed at the values of the PAW sphere radii for each atom.

Further details for applying these projection schemes are provided in the section on the DOS/Optic/Tensors Tab of the *MedeA VASP GUI* and in the context sensitive Help panel.

The above projection schemes are based on necessarily somewhat arbitrary sphere radii, which strongly affect the charge values. A number of schemes for charge analysis have been suggested that attempt to overcome this limitation. Some of these schemes, such as Mulliken charges and Coulson charges, are based on a population analysis of the wave functions applicable when basis functions centered on the atoms are used. Other schemes, such as Bader charges, Hirshfeld charges, and Voronoi deformation density, are based on a partitioning of the electron density distribution, and are therefore independent of the type of basis functions employed.

The *MedeA VASP GUI* offers automated access to a Bader decomposition of space into areas attributed to each atom [27]. The rule for how to bound the space assigned to each atom is based purely on the total charge density as computed by VASP. The boundaries between atom volumes are so-called zero flux surfaces  $S$ , i.e. surfaces on which the electron density  $\rho$  satisfies the zero-flux boundary condition

$$\nabla\rho(r_s) \cdot n(r_s) = 0 \quad (32)$$

for each point  $r_s$  of the surface, where  $n(r_s)$  is the unit vector normal to the surface at  $r_s$ . At each point of a dividing surface, the gradient of the electron density has no component normal to the surface, i.e. the charge density exhibits a local minimum perpendicular to the surface. A surface of local minimum charge density is an intuitive and natural boundary with which to define the volume assigned to an enclosed atom. The regions bounded by these dividing surfaces are called Bader regions. In most cases a Bader region will contain a nucleus, but it is also possible that a Bader region may contain no nucleus. Algorithmically, the Bader regions are identified by evaluating paths of steepest ascent confined to each grid point used for representing the charge density. Those grid points that have paths leading to the same terminal maximum charge belong to the same Bader region. The total charge within a Bader region is evaluated by the sum over all grid points contained therein, and this value provides a very good measure of the total electronic charge of the enclosed atom, the Bader charge.

The spatial extent of atoms may be defined by Bader regions, or arbitrarily by custom spheres, or by PAW spheres. These values are used not only to attribute total charges,  $s$ ,  $p$ ,  $d$ , and  $f$  charge characters, and magnetic moments, but also for calculating and displaying atom- and angular momentum-projected energy densities of states.

## 7 Implicit Solvation Model

*MedeA VASP* enables users to simulate the effects of solvation in calculations for molecules, surfaces, adsorption processes and surface reactions. Solvation effects can be included for single point total energies, geometry optimizations and molecular dynamics simulations. Furthermore, the *MedeA* Phonon computational module for vibrational properties enables users to analyze surface processes under solvation. Using

[27] W. Tang, E. Sanville, and G. Henkelman, "A grid-based Bader analysis algorithm without lattice bias", *Journal of Physics: Condensed Matter* 21, (2009): 084204. E. Sanville, S. D. Kenny, R. Smith, and G. Henkelman, "An improved grid-based algorithm for Bader charge allocation", *Journal of Computational Chemistry* 28, (2007): 899-908. G. Henkelman, A. Arnaldsson, and H. Jónsson, "A fast and robust algorithm for Bader decomposition of charge density", *Computational Materials Science* 36, (2006): 254-360

the *MedeA Transition State Search* module, one can generate energy profiles and reaction barriers for solvated surface processes. The method called VASPsol was implemented in the group of Richard Hennig at University of Florida [28].

There are two main approaches for studying solvation effects by means of quantum theory. One approach treats the entire solute/solvent system in an atomistic and quantum mechanical manner, which extends the *ab initio* approach explicitly to all solvent molecules. Due to the large number of solvent molecules often required to model physical processes of interest, and the very large number of configurations needed to extract converged equilibrium properties, such an explicit approach is computationally demanding and often prohibitive. The alternative approach, as pursued by VASPsol, treats only the solute portion of the system by *ab initio* methods, whereas these solute atoms interact with a solvent whose properties are simulated using a continuum approach. The average over the solvent degrees of freedom is implicitly represented by the properties of the solvent bath. Such an implicit solvation model is computationally much more tractable and can be quite accurate, provided that all interactions between solute and solvent are considered with appropriate detail.

There are three contributions to solvation effects taken into account by the VASPsol approach: The most important contribution, especially for polar or ionic solutes and surfaces and polar solvents, is the electrostatic interaction, which is modeled by the dielectric constant as an input parameter. If both the solute/surface and the solvent are non-polar, the dispersive (Van der Waals) interactions might become more important than electrostatics. Finally, if the solvent molecules are quite large, the energy for creating a cavity within the solvent may become the dominant term, which is modeled by a surface tension parameter.

For the details of the derivation of the solvation model the user is referred to the publication of the VASPsol implementation. The basic assumptions and approximations are outlined here. As a starting point, the free energy  $A$  of the combined solute/solvent system is written as the sum of two terms: 1) a functional, denoted  $F$ , of the total electron density and the thermodynamically averaged atomic densities of the solvent molecules, and 2) a term describing the electrostatic energy of the system.

$$A = F[n_{tot}, \{N_i(r)\}] + \int d^3r V_{ext}(r) \left( \sum_i Z_i N_i(r) - n_{tot}(r) \right) \quad (33)$$

Here the total electron density is the sum of the electron density of the solute and the solvent  $n_{tot}(r) = n_{solute}(r) + n_{solvent}(r)$ ,  $N_i(r)$  is the thermodynamically averaged atomic density of all the solvent species  $i$ , and  $V_{ext}(r)$  is the external potential of the solute nuclei. The ground state free energy  $A_0$  is determined by a stepwise minimization of this functional, first over the solvent electron density, and then over the solute electron density. The usual Kohn-Sham density functional is separated into a solute term  $A_{KS}$ , and a remaining term  $A_{diel}$  capturing all the interactions of the solute with the solvent as well as the internal energy of the solvent.  $A_{diel}$  is minimized with respect to the average atomic densities of the solvent  $N_i(r)$ , finally yielding the ground state free energy as,

$$A_0 = \min_{n_{solute}(r)} \{A_{KS}[n_{solute}(r), V_{ext}(r)] - \int d^3r V_{ext}(r) n_{solute}(r) + \tilde{A}_{diel}[n_{solute}(r), V_{ext}(r)]\} \quad (34)$$

$A_0$  is a functional only of the electron density and the external potential of the solute. All solvent effects are cast into the functional  $\tilde{A}_{diel}$ , which is obtained from minimization over the solvent charge density and the average atomic densities of the solvent. As such,  $A_0$  represents a continuum model for the solvent, the ground state of which is determined by the solute electronic structure.

The expression (34) is exact. However, approximations for  $\tilde{A}_{diel}$  are needed to enable practical computations. One approximation employed for this purpose treats electrostatic solute-solvent interactions. It is introduced by a term included into the functional  $\tilde{A}_{diel}$  to account for the electrostatic interaction between the solute electronic structure and the charge distribution induced in the solvent. Assuming a linear dependence of the solvent polarization on the electric field close to the solute, the solvent polarization can be described by the local relative permittivity, that is, the local dielectric function of the solvent,  $\epsilon(r)$ . In addition, to account for cavitation and dispersion, which is a dominant effect in the first solvation shell, an interface term proportional to the area accessible to the solvent is introduced into the functional  $\tilde{A}_{diel}$ . This is the free energy contribution,

$$A_{cav} = \tau \int d^3r |\nabla S| \quad (35)$$

[28] K. Mathew, R. Sundararaman, K. Letchworth-Weaver, T.A. Arias, R.G. Hennig, "Implicit solvation model for density-functional study of nanocrystal surfaces and reaction pathways", *Journal of Chemical Physics* 140, (2014): 084106

Here  $\tau$  is an effective surface tension parameter characterizing the effects of cavitation, dispersion and repulsion between solvent and solute not covered by electrostatics, and  $S(r)$  is a cavity shape function. Inside the dielectric cavity the relative permittivity is assumed to be the vacuum value  $\epsilon(r) = 1$ . Outside the cavity, the bulk value of the solvent is reached with the induced charges being placed at the surface. Furthermore, a diffuse cavity is assumed, i.e. the relative permittivity is smoothly varying as a functional of the electronic charge density of the solute, which ensures that the derivatives of the energy functional will be continuous. The following functional dependence of the relative permittivity of the solvent on the solute electronic charge density is assumed,

$$\epsilon(n_{solute}(r)) = 1 + (\epsilon_b - 1)S(n_{solute}(r)) \quad (36)$$

where  $\epsilon_b$  is the relative permittivity (dielectric constant) of the bulk solvent, and the cavity shape function is,

$$S(n_{solute}(r)) = \frac{1}{2}erfc\left\{\frac{\ln\frac{n_{solute}(r)}{n_c}}{\sigma\sqrt{2}}\right\} \quad (37)$$

The parameter  $n_c$  defines the value of the electron density at which the cavity forms and the parameter  $\sigma$  determines the width of the diffuse cavity. Within the range given by  $\sigma$  the relative permittivity is ramped up smoothly from the vacuum value of 1 inside the solute to  $\epsilon_b$  inside the bulk solvent, and for  $\epsilon_b \rightarrow 1$  the free energy becomes the vacuum value. This smooth variation mimics the dielectric response of the first solvation shell, in which the relative permittivity is known to be smaller than in the bulk, due to the higher electric field near the solute (dielectric saturation).

Based on these approximations to electrostatic, cavitation and dispersive interactions, the self-consistent solution of the Kohn-Sham equations for the solute/solvent system has become computationally practicable. From the solute charge density the combined electrostatic potential due to the electronic and nuclear charges of the solute in a polarizable medium is obtained by iteratively solving a generalized Poisson equation

$$\nabla \cdot [\epsilon(n_{solute}(r))\nabla\phi(r)] = -4\pi\{N_{solute}(r) - n_{solute}(r)\} \quad (38)$$

where  $N_{solute}(r)$  are the effective core charges approximated by Gaussians and  $n_{solute}(r)$  is the valence electronic charge density of the solute. The Kohn-Sham Hamiltonian includes two additional terms resulting from solvation, an electrostatic and a cavitation term

$$V_{el} = -\frac{d\epsilon(n_{solute}(r))}{dn_{solute}(r)}\frac{|\nabla\phi|^2}{8\pi} \quad V_{cav} = \tau\frac{d|\nabla S|}{dn_{solute}(r)} \quad (39)$$

The resulting energy expression has two additional related contributions accounting for electrostatic interactions and cavitation,

$$E_{el} = -\frac{1}{8\pi}\int d^3r\epsilon(n_{solute}(r))|\nabla\phi|^2 \quad E_{cav} = \tau\int d^3r|\nabla S| \quad (40)$$

In addition, the Hellman-Feynman force expression requires two correction terms due to solvation.

The parameters of this implicit solvation model can be set in the *INCAR* file as input parameters to the VASP code. The most prominent parameter, which is usually known from experimental data, is the bulk relative permittivity (or dielectric constant) of the solvent,  $\epsilon_b$ , which correspond to the *EB\_K* keyword of VASP, and can be set directly in the *MedeA VASP GUI*. The other parameters, i.e.  $n_c$  corresponding to *NC\_K*,  $S$  corresponding to *SIGMA\_K*, and  $\tau$  corresponding to the *TAU* keyword, can be set from the **Add to Input** tab. Unfortunately, these parameters are not directly informed by experimental data. The default parameters are suitable for water: the dielectric constant  $\epsilon = 78.4$  is set from experiment, whereas the shape function parameters  $n_c$  and  $\sigma$  and the effective surface tension  $\tau$  are obtained from a fit to experimental data on solvation energies of molecules in water:  $n_c = 0.0025 \text{ \AA}^{-3}$ ,  $\sigma = 0.6$ , and  $\tau = 0.525 \text{ meV \AA}^{-2}$ .