6 Density Functional Theory

The basis for Density Functional Theory (DFT) is the proof by Hohenberg and Kohn\(^1\) that the ground-state electronic energy is determined completely by the electron density \(\rho\), see Appendix B for details. In other words, there exists a one-to-one correspondence between the electron density of a system and the energy. The significance of this is perhaps best illustrated by comparing to the wave function approach. A wave function for an \(N\)-electron system contains \(3N\) coordinates, three for each electron (four if spin is included). The electron density is the square of the wave function, integrated over \(N-1\) electron coordinates, this only depends on three coordinates, independently of the number of electrons. While the complexity of a wave function increases with the number of electrons, the electron density has the same number of variables, independently of the system size. The "only" problem is that although it has been proven that each different density yields a different ground-state energy, the functional connecting these two quantities is not known. The goal of DFT methods is to design functionals connecting the electron density with the energy.\(^2\)

A note on semantics: a function is a prescription for producing a number from a set of variables (coordinates). A functional is similarly a prescription for producing a number from a function, which in turn depends on variables. A wave function and the electron density are thus functions, while an energy depending on a wave function or an electron density is a functional. We will denote a function depending on a set of variables with parentheses, \(f(x)\), while a functional depending on a function is denoted with brackets, \(F[f]\).

Comparing this with the wave mechanics approach, it seems clear that the energy functional may be divided into three parts, kinetic energy, \(T[\rho]\), attraction between the nuclei and electrons, \(E_{ne}[\rho]\), and electron-electron repulsion, \(E_{ee}[\rho]\) (the nuclear-nuclear repulsion is a constant in the Born–Oppenheimer approximation). Furthermore, with reference to Hartree–Fock theory (eq. (3.33)), the \(E_{ee}[\rho]\) term may be divided into a Coulomb and an Exchange part, \(J[\rho]\) and \(K[\rho]\), implicitly including correlation energy in all terms. The \(E_{ne}[\rho]\) and \(J[\rho]\) functionals are given by their classical expressions, where the factor of \(1/2\) in \(J[\rho]\) allows the integration to run over all space for both
variables.

$$E_{nc}[\rho] = \sum_n \int \frac{Z_n \phi(r)}{|R_n - r|} \, dr$$

$$J[\rho] = \frac{1}{2} \int \int \rho(r) \rho(r') \, dr \, dr'$$

(6.1)

This approach is known as Thomas-Fermi (TF) theory, including the exchange energy for non-interacting uniform electron gas. For such a system it may be shown that $T[\rho]$ and

$$K[\rho] = -\frac{1}{2} \int \int \rho(r) \rho(r') \, dr \, dr'$$

(6.2)

The energy functional $E_{TF}[\rho] = T_{TF}[\rho] + E_{nc}[\rho] + J[\rho]$ is known as Thomas-Fermi-Dirac (TFD) model, constituting the Thomas-Fermi-Dirac (TFD) model.

The assumption of a non-interacting uniform electron gas does not hold very well for atomic and molecular systems. Total energies are in error by 15-50%, but more serious is it that neither TF nor TFD theories predict bonding, molecules simply do not exist. The $T$ and $K$ functionals may be improved by addition of terms which depend not only on the density itself, but also on its derivative(s). This is equivalent of considering a non-uniform electron gas, and performing a Taylor-like expansion of the density. Addition of such gradient correction terms improves the results, for example bonding is now allowed, but in general it has been found that this is not a viable approach for constructing DFT models capable of yielding results comparable to those obtained by wave mechanics methods.

The foundation for the use of DFT methods in computational chemistry was the introduction of orbitals by Kohn and Sham. The main problem in Thomas-Fermi models is that the kinetic energy is represented poorly. The basic idea in the Kohn and Sham (KS) formalism is splitting the kinetic energy functional into two parts, one of which can be calculated exactly, and a small correction term.

Assume for the moment a Hamilton operator of the following form with $0 \leq \lambda \leq 1$.

$$H_\lambda = T + V_{ext}(\lambda) + \lambda V_{xc}$$

(6.3)

The $V_{ext}$ operator is equal to $V_{nc}$ for $\lambda = 1$, for intermediate $\lambda$ values, however, it is assumed that the external potential $V_{ext}(\lambda)$ is adjusted so that the same density is obtained for both $\lambda = 1$ (the real system) and $\lambda = 0$ (a hypothetical system with non-interacting electrons). For the $\lambda = 0$ case the exact solution to the Schrödinger equation is given as a Slater determinant composed of (molecular) orbitals, $\phi_j$, for which the
exact kinetic energy functional is given as

\[ T_S = \sum_{i=1}^{N} \langle \phi_i | -\frac{1}{2} \nabla^2 | \phi_i \rangle \]  

(6.1)

The subscript \( S \) denotes that it is the kinetic energy calculated from a Slater determinant. The \( \lambda = 1 \) case corresponds to interacting electrons, and eq. (6.4) is therefore only an approximation to the real kinetic energy, but a substantial improvement over the TF formula (eq. (6.2)).

Another way of justifying the use of eq. (6.4) for calculating the kinetic energy is by reference to natural orbitals (eigenvectors of the density matrix, Section 9.5). The exact kinetic energy can be calculated from the natural orbitals (NO) arising from the exact density matrix.

\[ T[\rho_{\text{exact}}] = \sum_{i=1}^{\infty} n_i \langle \phi_i^{\text{NO}} | -\frac{1}{2} \nabla^2 | \phi_i^{\text{NO}} \rangle \]

\[ \rho_{\text{exact}} = \sum_{i=1}^{\infty} n_i |\phi_i^{\text{NO}}|^2 \]

(6.5)

\[ N = \sum_{i=1}^{\infty} n_i \]

The orbital occupation numbers \( n_i \) (eigenvalues of the density matrix) will be between 0 and 1, corresponding to the number of electrons in the orbital. Note that the representation of the exact density normally will require an infinite number of natural orbitals. The first \( N \) occupation numbers (\( N \) being the total number of electrons in the system) will normally be close to 1, and the remaining close to 0.

Since the exact density matrix is not known, the (approximate) density is written in terms of a set of auxiliary one-electron functions, orbitals, as

\[ \rho(r) = \sum_{i=1}^{N} |\phi_i(r)|^2 \]

(6.6)

This corresponds to eq. (6.5) with occupation numbers of exactly 1 or 0. The “missing” kinetic energy from eq. (6.4) is thus due to the occupation numbers deviating from being exactly 1 or 0.

The key to Kohn–Sham theory is thus the calculation of the kinetic energy under the assumption of non-interacting electrons (in the same sense as HF orbitals in wave mechanics describe non-interacting electrons) from eq. (6.4). In reality the electrons are interacting, and eq. (6.4) does not provide the total kinetic energy. However, just as HF theory provides \( \sim 99\% \) of the correct answer, the difference between the exact kinetic energy and that calculated by assuming non-interacting orbitals is small. The remaining kinetic energy is absorbed into an exchange–correlation term, and a general DFT energy expression can be written as

\[ E_{\text{DFT}}[\rho] = T[\rho] + E_{\text{ex}}[\rho] + J[\rho] + E_{\text{xc}}[\rho] \]

(6.7)

By equating \( E_{\text{DFT}} \) to the exact energy, this expression may be taken as the definition of
$E_{xc}$, it is the part which remains after subtraction of the non-interacting kinetic energy, and the $E_{ne}$ and $J$ potential energy terms.

$$E_{xc}[\rho] = (T[\rho] - T_{S}[\rho]) + (E_{xc}[\rho] - J[\rho])$$  \hspace{1cm} (6.8)

The first parenthesis in eq. (6.8) may be considered the kinetic correlation energy, while the second contains both exchange and potential correlation energy.

The exchange energy is by far the largest contributor to $E_{xc}$. For the neon atom, for example, the exchange energy is $-12.11$ a.u., while the correlation energy is $-0.39$ a.u. (as calculated by wave mechanics methods). Since the exchange energy dominates $E_{xc}$, one may reasonably ask why we do not calculate this term "exactly" from orbitals (analogously to the kinetic energy), by the formula known from wave mechanics (eq. (3.33)), and only calculate the computationally difficult part, the correlation energy, by DFT. Although this has been tried, it gives poor results. The basic problem is that the DFT definitions of exchange and correlation energies are not completely equivalent to their wave mechanics counterparts. The correlation energy in wave mechanics is defined as the difference between the exact energy and the corresponding Hartree--Fock value, and the exchange energy is the total electron--electron repulsion minus the Coulomb energy. These energies have both a short- and long-range part (in terms of the distance between two electrons). The long-range correlation is essentially the "static" correlation energy (i.e. the "multi-reference" part, see Section 4.6) while the short-range part is the "dynamical" correlation. The long-range part of the exchange energy in wave mechanics effectively cancels the long-range part of the correlation energy. The definitions of exchange and correlation in DFT, however, are local (short range), they only depend on the density at a given point and in the immediate vicinity (via derivatives of the density). The cancellation at long range is (or should be) implicitly built into the exchange--correlation functional. Calculating the exchange energy by wave mechanics and the correlation by DFT thus destroys the cancellation.

The strength of DFT is that only the total density needs to be considered. In order to calculate the kinetic energy with sufficient accuracy, however, orbitals have to be reintroduced. Nevertheless, as discussed in Section 6.5, DFT has a computational cost which is similar to HF theory, with the possibility of providing more accurate (exact, in principle) results.

The major problem in DFT is deriving suitable formulas for the exchange--correlation term. Assume for the moment that such a functional is available, the problem is then similar to that encountered in wave mechanics HF theory: determine a set of orthogonal orbitals which minimize the energy (the requirement of orthogonal orbitals basically enforces the Pauli principle). Since the $J[\rho]$ and $E_{xc}[\rho]$ functional depends on the total density, a determination of the orbitals involves an iterative sequence. The orbital orthogonality constraint may be enforced by the Lagrange method (section 14.6), again in analogy with wave mechanics HF methods (eq. (3.34)).

$$L[\rho] = E_{DFT}[\rho] - \sum_{\beta} \lambda_{\gamma} [\langle \phi_{\gamma} | \phi_{\beta} \rangle - \delta_{\gamma \beta}]$$  \hspace{1cm} (6.9)

Requiring the variation of $L$ to vanish provides a set of equations involving an effective one-electron operator ($h_{KS}$), similar to the Fock operator in wave mechanics.
The Lagrange multipliers can again be associated with molecular orbital energies, and the highest occupied orbital energy is the ionization potential (Koopmans' theorem), but only if the exact exchange–correlation functional is employed. Since this is not the case in actual calculations, the orbital energies in practice do not carry quite the same significance as in HF theory. The unknown KS orbitals may be determined by numerical methods, or expanded in a set of basis functions, analogously to the HF method (Section 3.5).

Although it is clear that there are many similarities between wave mechanics HF theory and DFT, there is an important difference. If the exact \( E_{xc} [\rho] \) was known, DFT would provide the exact total energy, including electron correlation. DFT methods therefore have the potential of including the computationally difficult part in wave mechanics, the correlation energy, at a computational effort similar to that for determining the uncorrelated HF energy. Although this is certainly the case for approximations to \( E_{xc} [\rho] \) (as illustrated below), this is not necessarily true for the exact \( E_{xc} [\rho] \). It may well be that the exact \( E_{xc} [\rho] \) functional is so complicated that the computational effort for solving the KS equations will be similar to that required for solving the Schrödinger equation (exactly) with a wave mechanics approach. Indeed, unless one believes that the Schrödinger equation contains superfluous information, this is likely to be the case. Since exact solutions are generally not available in either approach, the important question is instead: what is the computational cost for generating a solution of a certain accuracy. In this respect DFT methods have very favourable characteristics.

It is possible to prove that the exchange–correlation potential is a unique functional, valid for all systems, but an explicit functional form of this potential has been elusive. The difference between DFT methods is the choice of the functional form of the exchange–correlation energy. There is little guidance from theory how such functionals should be chosen, and consequently many different potentials have been proposed. Functional forms are often designed to have a certain limiting behaviour (for example including the uniform electron gas limit), and fitting parameters to known accurate data. Which functional is the better will have to be settled by comparing the performance with experiments or high-level wave mechanics calculations.

One of the more recent approaches for designing \( E_{xc} \) functionals is based on "inverting" eq. (6.7). An accurate electron density may be calculated by advanced wave
mechanics methods (e.g., coupled cluster), and a set of KS orbitals which yield this density can be determined by a "constrained search" method involving a minimization of the kinetic energy. The idea was originally introduced by Levy and Perdew and has since been elaborated on by Parr and others. Recent implementations and applications have been reported by Baerends et al. and Handy et al. This enables a calculation of the exchange-correlation energy (by subtracting $T$, $V_{xc}$, and $J$ from the total DFT energy), thereby giving the direct dependence of $V_{xc}$ on the density $\rho$. Since the DFT area is fairly new in computational chemistry, there are at present no clear "standard" methods, like MP2 and CISD in traditional $ab$ initio theory, and the calibration of different methods is much less developed. Some examples of functionals which currently are used quite extensively are given below, but since DFT is an active area of research, new and improved potentials are likely to emerge.

It is customary to separate $E_{xc}$ into two parts, a pure exchange $E_x$ and a correlation part $E_c$, although it is not clear that this is a valid assumption (cf. the above discussion of the definition of exchange and correlation). Each of these energies is often written in terms of the energy per particle (energy density), $\varepsilon_x$ and $\varepsilon_c$.

$$E_{xc}[\rho] = E_x[\rho] + E_c[\rho] = \int \rho(r) \varepsilon_x[\rho(r)] \, dr + \int \rho(r) \varepsilon_c[\rho(r)] \, dr$$  \hspace{1cm} (6.12)

The corresponding potential required in eq. (6.10) is given as the derivative of the energy with respect to the density.

$$V_{xc}(r) = \frac{\partial E_{xc}[\rho]}{\partial \rho(r)} = \varepsilon_{xc}[\rho(r)] + \rho(r) \frac{\partial \varepsilon_{xc}(r)}{\partial \rho}$$  \hspace{1cm} (6.13)

As mentioned in the start of Chapter 4, the correlation between electrons of parallel spin is different from that between electrons of opposite spin. The exchange energy is "by definition" given as a sum of contributions from the $\alpha$ and $\beta$ spin densities, as exchange energy only involves electrons of the same spin. The kinetic energy, the nuclear-electron attraction and Coulomb terms are trivially separable.

$$E_x[\rho] = E_x^\alpha[\rho_\alpha] + E_x^\beta[\rho_\beta]$$

$$E_c[\rho] = E_c^{\alpha\alpha}[\rho_\alpha] + E_c^{\beta\beta}[\rho_\beta] + E_c^{\alpha\beta}[\rho_\alpha, \rho_\beta]$$  \hspace{1cm} (6.14)

The total density is the sum of the $\alpha$ and $\beta$ contributions, $\rho = \rho_\alpha + \rho_\beta$, and for a closed-shell singlet these are identical ($\rho_\alpha = \rho_\beta$). Functionals for the exchange and correlation energies may be formulated in terms of separate spin-densities; however, they are often given instead as functions of the spin polarization $\zeta$ (normalized difference between $\rho_\alpha$ and $\rho_\beta$), and the radius of the effective volume containing one electron, $r_S$.

$$\zeta = \frac{\rho_\alpha - \rho_\beta}{\rho_\alpha + \rho_\beta} = \frac{4}{3} \pi r_S^3 = \rho^{-1}$$  \hspace{1cm} (6.15)

In the formulas below it is implicitly assumed that the exchange energy is a sum over both the $\alpha$ and $\beta$ densities.

6.1 Local Density Methods

In the Local Density Approximation (LDA) it is assumed that the density locally can be treated as a uniform electron gas, or equivalently that the density is a slowly varying
6.1 LOCAL DENSITY METHODS

function. The exchange energy for a uniform electron gas is given by the Dirac formula (eq. (6.2)).

\[
E_x^{\text{LDA}}[\rho] = -C_x \int \rho^{1/3}(r) dr \\
\varepsilon_x^{\text{LDA}}[\rho] = -C_x \rho^{1/3}
\]  

(6.16)

In the more general case, where the \( \alpha \) and \( \beta \) densities are not equal, LDA (where the sum of the \( \alpha \) and \( \beta \) densities is raised to the 4/3 power) has been virtually abandoned and replaced by the Local Spin Density Approximation (LSDA) (which is given as the sum of the individual densities raised to the 4/3 power, eq. (6.17)).

\[
E_x^{\text{LSDA}}[\rho] = -2^{1/3} C_x \int [\rho_\alpha^{1/3} + \rho_\beta^{1/3}] dr \\
\varepsilon_x^{\text{LSDA}}[\rho] = -2^{1/3} C_x [\rho_\alpha^{1/3} + \rho_\beta^{1/3}]
\]  

(6.17)

LSDA may also be written in terms of the total density and the spin polarization.

\[
\varepsilon_x^{\text{LSDA}}[\rho] = -\frac{3}{2} C_x \rho^{1/3} [(1 + \zeta) \rho^{1/3} + (1 - \zeta) \rho^{1/3}]
\]  

(6.18)

For closed-shell systems LSDA is equal to LDA, and since this is the most common case, LDA is often used interchangeably with LSDA, although this is not true in the general case (eqs. (6.16) and (6.17)). The \( X_\alpha \) method proposed by Slater in 1951 can be considered as an LSDA method where the correlation energy is neglected and the exchange term is given as

\[
\varepsilon_{X_\alpha}[\rho] = -\frac{3}{2} \alpha C_x \rho^{1/3}
\]  

(6.19)

With \( \alpha = 2/3 \) this is identical to the Dirac expression. The original \( X_\alpha \) method used \( \alpha = 1 \), but a value of \( 3/4 \) has been shown to give better agreement for atomic and molecular systems. The name Slater is often used as a synonym for the L(SD)A exchange energy involving the electron density raised to the 4/3 power (1/3 power for the energy density).

The correlation energy of a uniform electron gas has been determined by Monte Carlo methods for a number of different densities. In order to use these results in DFT calculations, it is desirable to have a suitable analytic interpolation formula. This has been constructed by Vosko, Wilk and Nusair (VWN) and is in general considered to be a very accurate fit. It interpolates between the unpolarized (\( \zeta = 0 \)) and spin polarized (\( \zeta = 1 \)) limits by the following functional.

\[
\varepsilon_{\text{VWN}}(r_\beta, \zeta) = \varepsilon_c(r_\beta, 0) + \varepsilon_c(r_\beta) \left[ \frac{f(\zeta)}{f(0)} \right] \left[ \left| 1 - \zeta \right|^4 + [\varepsilon_c(r_\beta, 1) - \varepsilon_c(r_\beta, 0)] f(\zeta) \zeta^4 \right]
\]

\[
f(\zeta) = \frac{(1 + \zeta)^{4/3} + (1 - \zeta)^{4/3} - 2}{2(2^{1/3} - 1)}
\]  

(6.20)
The $\varepsilon_{c}(r_S, \zeta)$ and $\varepsilon_{\sigma}(r_S)$ functionals are parameterized as in eq. (6.21):

$$
\varepsilon_{c/\sigma}(x) = A \left( \frac{\ln \frac{x^2}{X(x)} + \frac{2}{Q} \tan^{-1} \left( \frac{Q}{Q + x} \right)}{\left( \frac{x - x_0}{X(x)} \right)^{2} + \frac{2(x + 2x_0)}{Q} \tan^{-1} \left( \frac{Q}{2x + x} \right)} \right)
$$

(6.21)

$$
X(x) = x^2 + \ell \cdot x + c
Q = \sqrt{4c - \ell^2}
$$

The parameters $A, x_0, \ell$ and $c$ are fitting constants, different for $\varepsilon_{c}(r_S, 0)$, $\varepsilon_{c}(r_S, 1)$, and $\varepsilon_{\sigma}(r_S)$.

A modified form for $\varepsilon_{c/\sigma}(r_S)$ has been given by Perdew and Wang, and is used in connection with the PW91 functional described in Section 6.2.

$$
\varepsilon_{c/\sigma}^{PW91}(x) = -2\alpha \rho(1 + \alpha x^2) \ln \left( \frac{1 + \frac{1}{2\alpha(\beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4)}}{\frac{1}{\rho}} \right)
$$

(6.22)

Here $\alpha, \beta_1, \beta_2, \beta_3$ and $\beta_4$ are suitable constants.

The LSDA approximation in general underestimates the exchange energy by $\sim 10\%$, thereby creating errors which are larger than the whole correlation energy. Electron correlation is furthermore overestimated, often by a factor close to 2, and bond strengths are as a consequence underestimated. Despite the simplicity of the fundamental assumptions, LSDA methods are often found to provide results with an accuracy similar to that obtained by wave mechanics IFF methods.

### 6.2 Gradient Corrected Methods

Improvements over the LSDA approach have to consider a non-uniform electron gas. A step in this direction is to make the exchange and correlation energies dependent not only on the electron density, but also on derivatives of the density. Such methods are known as Gradient Corrected or Generalized Gradient Approximation (GGA) methods (a straightforward Taylor expansion does not lead to an improvement over LSDA, it actually makes things worse, thus the name generalized gradient approximation). GGA methods are also sometimes referred to as non-local methods, although this is somewhat misleading since the functionals depend only on the density (and derivatives) at a given point, not on a space volume as for example the Hartree–Fock exchange energy.

Perdew and Wang (PW86) proposed modifying the LSDA exchange expression to that shown in eq. (6.23), where $x$ is a dimensionless gradient variable, and $\alpha, b$ and $c$ being suitable constants (summation over equivalent expressions for the $\alpha$ and $\beta$ densities is implicitly assumed).

$$
\varepsilon_{\sigma}^{PW86} = \varepsilon_{\sigma}^{LDA}(1 + \alpha x^2 + bx^4 + cx^6)^{1/15}
$$

(6.23)

$$
x = \frac{[\nabla \rho]}{\rho^{\gamma/3}}
$$

where

The $\beta$ parameter in PW86 (6.23)

And also has the form

This is an improvement over the density functional theory (DFT) results:

Perdew and Wang's approach has come to be known as the Perdew-Wang (PW86) exchange functional.

The correlation energy functional $\varepsilon_{\sigma}$, also known as the correlation potential, is given by the sum of the exchange potential and the correlation potential.

Parr (1986) provides a detailed discussion of the correlation energy functional and its role in DFT calculations.

where
6.2 GRADIENT CORRECTED METHODS

Becke\(^{15}\) proposed a widely used correction (B or B88) to the LSDA exchange energy, which has the correct \(-r^{-1}\) asymptotic behaviour for the energy density (but not for the exchange potential),\(^{16}\)

\[
\varepsilon_{\text{B88}} = \varepsilon_{\text{LDA}} + \Delta_{\text{B88}}^{\text{B88}}
\]

\[
\Delta_{\text{B88}}^{\text{B88}} = -\beta p^{1/3} \frac{x^2}{1 + 6/x \sinh^{-1}x}
\]  

(6.24)

The \(\beta\) parameter is determined by fitting to known atomic data and \(x\) is defined in eq. (6.23).

Another functional form (not a correction) proposed by Becke and Roussel (BR)\(^{17}\) has the form

\[
\varepsilon_{\text{BR}}^{\text{BR}} = -\frac{2 - 2e^{-\alpha \rho}}{4b} - \alpha \rho e^{-\alpha \rho}
\]

\[
\alpha^2 e^{-\alpha \rho} = 8\pi \rho
\]

\[
\alpha (\alpha b - 2) = b \frac{\nabla^2 \rho - 2D}{\rho}
\]

\[
D = \sum_{i} |\nabla \phi_i|^2 - \frac{(\nabla \rho)^2}{4\rho}
\]

(6.25)

This functional contains derivatives of the orbitals, not just the gradient of the total density, and is computationally slightly more expensive. Despite the apparent difference in functional form, exchange expressions (6.24) and (6.25) have been found to provide similar results.

Perdew and Wang have proposed an exchange functional similar to B88 to be used in connection with the PW91 correlation functional given below (eq. (6.30)).

\[
\varepsilon_{\text{PW91}}^{\text{PW91}} = \varepsilon_{\text{LDA}} \left(1 + x_1 \sinh^{-1}(x_2) + (\alpha_2 + \alpha_3 e^{-\beta_2 x^2}) x^2\right)
\]  

(6.26)

where \(\alpha_1-5\) and \(b\) again are suitable constants and \(x\) is defined in eq. (6.23).

There have been various gradient corrected functional forms proposed for the correlation energy. One popular functional (not a correction) is due to Lee, Yang and Parr (LYP)\(^{18}\) and has the form

\[
\varepsilon_{\text{LYP}}^{\text{LYP}} = -\frac{a}{(1 + d \rho^{-1/3})} - \frac{\gamma e^{-\rho^{-1/3}}}{9(1 + d \rho^{-1/3}) \rho^{8/3}}
\]

\[
\times \left[18(2^{2/3})C_F(p_i^{3/3} + \rho_i^{3/3}) - 18 \rho w + \rho_\alpha (2\nabla^2 \rho_\alpha) + \rho_{\beta}(2\nabla^2 \rho_{\beta})\right]
\]

\[
\gamma = 2 \left[1 - \frac{\mu^2_{\alpha} + \rho^2_{\alpha}}{\rho^2_{\alpha}}\right]
\]

\[
\mu_w = \frac{1}{8} \left(\frac{\nabla^2 \rho_\alpha}{\rho_\alpha} - \nabla^2 \rho_\alpha\right)
\]

(6.27)

where the \(a, \beta, \gamma, \rho\) and \(\mu\) parameters are determined by fitting to data for the helium
atom. The \( \epsilon_{v} \) functional is known as the local Weizsäcker kinetic energy density. Note that the \( \gamma \)-factor becomes zero when all the spins are aligned (\( \rho = \rho_{\uparrow} = \rho_{\downarrow} = 0 \)), i.e., the LYP functional does not predict any parallel spin correlation in such a case (e.g., the LYP correlation energy in triplet \( \text{He} \) is zero). The appearance of the second derivative of the density can be removed by partial integration\(^{19} \) to give eq. (6.28).

\[
\epsilon_{v}^{\text{LYP}} = -4\alpha \frac{\rho_{\uparrow}\rho_{\downarrow}}{\rho^{3}(1 + d\rho^{1/3})}
\left\{ \frac{144(2^{2/3})C_{\gamma}(\rho_{\uparrow}^{5/3} + \rho_{\downarrow}^{5/3}) + (47 - 7\delta)|\nabla \rho|^{2}}{18} \right. \\
- \left. \frac{(45 - \delta)(|\nabla \rho_{\uparrow}|^{2} + |\nabla \rho_{\downarrow}|^{2})}{\rho^{}\frac{2}{3}} \right. \\
+ \left. \frac{2\rho^{\frac{4}{3}}(|\nabla \rho_{\uparrow}|^{2} + |\nabla \rho_{\downarrow}|^{2} - |\nabla \rho|^{2})}{\rho^{\frac{2}{3}}} \right. \\
\left. - \frac{(\rho_{\uparrow}^{2} |\nabla \rho_{\uparrow}|^{2} + \rho_{\downarrow}^{2} |\nabla \rho_{\downarrow}|^{2})}{\rho^{2}} \right. \\
\omega = \frac{e^{-\phi r^{1/3}}}{(1 + d\rho^{1/3})\rho^{1/3}}
\delta = c\rho^{1/3} + \frac{d\rho^{1/3}}{(1 + d\rho^{1/3})}
\varepsilon_{v}^{\text{P86}} = \varepsilon_{v}^{\text{LDA}} + \Delta\varepsilon_{v}^{\text{P86}}
\Delta\varepsilon_{v}^{\text{P86}} = \frac{\Phi C(\rho)|\nabla \rho|^{2}}{\tilde{f}(\xi)\rho^{2/3}}
\tilde{f}(\xi) = 2^{1/3} \frac{\left\{ \frac{1 + \xi}{2} \right\}^{5/3}}{\left\{ \frac{1 - \xi}{2} \right\}^{5/3}}
\Phi = \frac{\alpha C(\infty)|\nabla \rho|}{\tilde{f}(\xi)\rho^{2/3}}
C(\rho) = \gamma_{1} + \frac{\gamma_{2} + \gamma_{3}\rho_{S} + \gamma_{4}\rho_{S}^{2}}{1 + \gamma_{5}\rho_{S} + \gamma_{6}\rho_{S}^{2} + \gamma_{7}\rho_{S}^{3}}
\varepsilon_{v}^{\text{PW91}} = \varepsilon_{v}^{\text{LDA}} + \Delta\varepsilon_{v}^{\text{PW91}}[\rho] = \rho \left[ H_{0}(t, r_{S}, \xi) + H_{1}(t, r_{S}, \xi) \right]
H_{0}(t, r_{S}, \xi) = \tilde{f}(\xi) \frac{3}{2} \frac{t^{2} + At^{4}}{1 + At^{2} + A^{2}t^{4}}
H_{1}(t, r_{S}, \xi) = \Phi \frac{16}{\pi^{2}} \frac{1}{(3\pi^{2})^{1/3}} |\nabla \rho|^{2} e^{-\Delta s^{2}f(\xi)^{2}}
f(\xi) = \frac{1}{2} \left[ \left( 1 + \xi \right)^{2/3} + \left( 1 - \xi \right)^{2/3} \right]
t = \frac{192}{\pi^{2}} \frac{1}{2f(\xi)\rho^{7/6}}
A = \alpha e^{-\epsilon_{v}(r_{S}^{2}\xi)/f(\xi)} - 1
\]
where }{\varepsilon(r_s, c)}{\text{ is the PW92 parameterization of the LSDA correlation energy functional (eq. (6.22)), } x} and }{C(p)}{\text{ are as defined in eqs. (6.23) and (6.29), and } \alpha, \beta, \epsilon \text{ and } \delta}{\text{ are suitable constants.}}

It should be noted that several of the proposed functionals violate fundamental restrictions, such as predicting correlation energies for one-electron systems (for example P86 and PW91) or failing to have the exchange energy cancel the Coulomb self-repulsion (Section 3.3, eq. (3.32)). One of the more recent functionals which does not have these problems is due to Becke (B95), \cite{becke1993} which has the form

\[
\begin{align*}
\varepsilon_{c}^{\text{B95}} & = \varepsilon_{c}^{\beta} + \varepsilon_{c}^{\alpha} + \varepsilon_{c}^{\delta} \\
\varepsilon_{c}^{\alpha} & = \left[ \frac{1}{2} \alpha \left( \psi_{a}^{2} + \psi_{b}^{2} \right) \right]^{-1} \varepsilon_{c}^{\text{PW91,0,0}} \\
\varepsilon_{c}^{\beta} & = \left[ 1 + \beta \psi_{a}^{2} \right]^{-2} \frac{D_{\alpha}}{D_{\alpha}^{\text{LDA}}} \varepsilon_{c}^{\text{PW91,0,g}} \\
D_{\alpha}^{\text{LDA}} & = 2^{-3/2} \frac{C_{F} \rho_{a}^{1/3}}{N} \varepsilon_{c}^{\beta}
\end{align*}
\] (6.31)

Here \( \alpha \) runs over \( \alpha \) and \( \beta \) spins, \( \psi_{a} \) and \( D_{\alpha} \) have been defined in eqs. (6.23) and (6.25), \( \alpha \) and \( \beta \) are fitting parameters, and \( \varepsilon_{c}^{\text{PW91}} \) is the Perdew–Wang parameterization of the LSDA correlation functional (eq. (6.22)).

### 6.3 Hybrid Methods

From the Hamiltonian in eq. (6.3) and the definition of the exchange–correlation energy in eq. (6.8), an exact connection can be made between the exchange–correlation energy and the corresponding potential connecting the non-interacting reference and the actual system, see appendix B for details. The resulting equation is called the Adiabatic Connection Formula (ACF) \cite{perdew1981} and involves an integration over the parameter \( \lambda \) which “turns on” the electron–electron interaction.

\[
E_{xc} = \int_{0}^{1} \langle \Psi_{\lambda} | \nabla_{xc} (\lambda) | \Psi_{\lambda} \rangle d\lambda
\] (6.32)

In the crudest approximation (taking \( \nabla_{xc} \) to be linear in \( \lambda \)) the integral is given as the average of the values at the two end-points,

\[
E_{xc} \approx \frac{1}{2} \langle \Psi_{0} | \nabla_{xc} (0) | \Psi_{0} \rangle + \frac{1}{2} \langle \Psi_{1} | \nabla_{xc} (1) | \Psi_{1} \rangle
\] (6.33)

In the \( \lambda = 0 \) limit, the electrons are non-interacting and there is consequently no correlation energy, only exchange energy. Furthermore, since the exact wave function in this case is a single Slater determinant composed of KS orbitals, the exchange energy is exactly that given by Hartree–Fock theory (eq. (3.33)). If the KS orbitals are identical to the HF orbitals, the “exact” exchange is precisely the exchange energy calculated by HF wave mechanics methods. The last term in eq. (6.33) is still unknown. Approximating it by the LSDA result defines the Half-and-Half (H+H) method. \cite{perdew1986}

\[
E_{xc}^{\text{H+H}} = \frac{1}{2} E_{xc}^{\text{exact}} + \frac{1}{2} (E_{xc}^{\text{LSDA}} + E_{xc}^{\text{LSDA}})
\] (6.34)

Since the GGA methods give a substantial improvement over LDA, a generalized version of the Half-and-Half method may be defined by writing the exchange energy as
a suitable combination of LSDA, exact exchange and a gradient correction term. The correlation energy may similarly be taken as the LSDA formula plus a gradient correction term.

\[ E_{xc}^{B3} = (1 - \alpha)E_{xc}^{LSDA} + \alpha E_{xc}^{exact} + \beta \Delta E_x^{B88} + E_{xc}^{LSDA} + \epsilon \Delta E_{xc}^{GGA} \]  

(6.35)

Models which include exact exchange are often called hybrid methods, the names Adiabatic Connection Model (ACM) and Becke 3 parameter functional (B3) are examples of such hybrid models defined by eq. (6.35). The \( \alpha, \beta \) and \( \epsilon \) parameters are determined by fitting to experimental data and depend on the form chosen for \( E_{xc}^{GGA} \). Typical values are \( \alpha \sim 0.2, \beta \sim 0.7 \) and \( \epsilon \sim 0.8 \). Owing to the substantially better performance of such parameterized functionals the Half-and-Half model is rarely used anymore. The B3 procedure has been generalized to include more fitting parameters, however, the improvement is rather small.

### 6.4 Performance

The specification of a DFT method requires selection of a suitable form for the exchange and correlation energies. Although one may in principle select an LSDA form for one of them and a gradient form for the other, this is not really consistent. Within the LSDA approximation, the exchange is given explicitly by the Dirac–Slater expression (6.17), and the only difference is the interpolation function used for reproducing the (very good) Monte Carlo results for the correlation energy. Since the VWN formula (6.20)/(6.21) generally is considered a good interpolating function, the term LSDA has become almost synonymous with the acronym SVWN. Gradient corrected methods have typically used the B88 exchange (6.24), or the B3/ACM hybrid (6.35), and either the LYP, P86 or PW91 correlation (6.27–6.30). Associated acronyms are BLYP, BP86, BPW91, B3LYP, B3P86 and B3PW91.

Gradient corrected methods usually perform much better than LSDA. For the G2-1 data set (see Section 5.5), omitting electron affinities, the mean absolute deviations shown in Table 6.1 are obtained. The improvement achieved by adding gradient terms is impressive, and hybrid methods (like B3PW91) perform almost as well as the elaborate G2 model for these test cases. For a somewhat larger set of reference data, called the G2-2 set, the data shown in Table 6.2 are obtained.

In general it is found that GGA methods often give geometries and vibrational frequencies for stable molecules of the same or better quality than MP2, at a computational cost similar to HF. For systems containing multi-reference character, where MP2 usually fails badly, DFT methods are often found to generate results of a

Table 6.1 Comparison of the performance of DFT methods by mean absolute deviations (kcal/mol)

<table>
<thead>
<tr>
<th>Method</th>
<th>G2</th>
<th>LSDA</th>
<th>B88</th>
<th>BPW91</th>
<th>B3PW91</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atomization Energies</td>
<td>1.2</td>
<td>35.7</td>
<td>3.9</td>
<td>5.7</td>
<td>2.4</td>
</tr>
<tr>
<td>Ionization Potentials</td>
<td>1.4</td>
<td>6.3</td>
<td>11.2</td>
<td>4.1</td>
<td>3.8</td>
</tr>
<tr>
<td>Proton Affinities</td>
<td>1.0</td>
<td>5.6</td>
<td>2.4</td>
<td>1.5</td>
<td>1.2</td>
</tr>
</tbody>
</table>
6.4 PERFORMANCE

Table 6.2 Comparison of the performance of DFT methods (kcal/mol)

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean absolute deviation</th>
<th>Maximum absolute deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>G2</td>
<td>1.6</td>
<td>8.2</td>
</tr>
<tr>
<td>G2(MP2)</td>
<td>2.0</td>
<td>10.1</td>
</tr>
<tr>
<td>G2(MP2, SVP)</td>
<td>1.9</td>
<td>12.5</td>
</tr>
<tr>
<td>SVWN</td>
<td>90.9</td>
<td>228.7</td>
</tr>
<tr>
<td>BLYP</td>
<td>7.1</td>
<td>28.4</td>
</tr>
<tr>
<td>BPW91</td>
<td>7.9</td>
<td>32.2</td>
</tr>
<tr>
<td>B3LYP</td>
<td>3.1</td>
<td>20.1</td>
</tr>
<tr>
<td>B3PW91</td>
<td>3.5</td>
<td>21.8</td>
</tr>
</tbody>
</table>

Another significant advantage is that DFT methods based on unrestricted determinants (analogously to UHF, Section 3.7) for open-shell systems are not very prone to "spin contamination", i.e., \(S^2\), is normally close to \(S_z(S_z+1)\) (see also Sections 4.4 and 11.5.3). This is basically a consequence of electron correlation being included in the single determinantal wave function (by means of \(E_x\)). Actually, it has been argued that "spin contamination" is not well defined in DFT methods\(^{28}\) and that \(S^2\), should not be equal to \(S_z(S_z+1)\). The argument is basically that real systems display "spin polarization", e.g., there are points in space where \(\rho_{\beta}\) is larger than \(\rho_{\alpha}\) (assuming that the number of \(\alpha\)-electrons is larger than the number of \(\beta\)-electrons). This effect cannot be achieved by a restricted open-shell type determinant (analogous to ROHF), only by an unrestricted treatment which allows the \(\alpha\)- and \(\beta\)-orbitals to be different. Another consequence of the presence of \(E_x\) is that restricted type determinants are much more stable towards symmetry breaking to an unrestricted determinant (Section 3.8.3) than Hartree–Fock wave functions. For ozone (Section 4.4), for example, it is not possible to find a lower energy solution corresponding to UHF for "pure" DFT methods (LSDA, BLYP, BPW91), although those including exact exchange (B3LYP, B3PW91) display a triplet instability.

Weak interactions due to dispersion (van der Waals type interactions) are poorly described by current functionals.\(^{30}\) Owing to the general overestimation of bond strengths, LDA does predict an attraction between for example rare gas atoms, although not very accurately, while essentially all gradient corrected methods predict a purely repulsive interaction (at least when corrected for basis set superposition error). Hydrogen bonding, however, is mainly electrostatic, which is reasonably well accounted for by DFT methods. There are indications that relative energies are predicted less accurately by DFT methods, and that transition structures are sometimes poorly described, but as already mentioned, the number of systems for which DFT methods have been calibrated is still fairly small.

Finally, DFT methods are at present not well suited for excited states of the same symmetry as the ground state. The absence of a wave function makes it difficult to ensure orthogonality between the ground and excited states.
6.5 Computational Considerations

The KS orbitals can be determined by a numerical procedure, analogous to numerical HF methods. In practice such procedures are limited to small systems, and essentially all calculations employ an expansion of the KS orbitals in an atomic basis set.

\[ \phi_l = \sum_{n} c_{nl} \chi_n \]  

(6.36)

The basis functions are normally the same as used in wave mechanics for expanding the HF orbitals, see Chapter 5 for details. Although there is no guarantee that the exponents and contraction coefficients determined by the variational procedure for wave functions are also optimum for DFT orbitals, the difference is presumably small since the electron densities derived by both methods are very similar.31

The variational procedure again leads to a matrix equation in the atomic orbital basis which can be written in the form (compare eq. (3.50)).

\[ \mathbf{h}_{KS} \mathbf{C} = \mathbf{S} \mathbf{C} \varepsilon \]
\[ h_{\alpha\beta} = \langle \chi_\alpha | \mathbf{h}_{KS} | \chi_\beta \rangle \]
\[ S_{\alpha\beta} = \langle \chi_\alpha | \chi_\beta \rangle \]

(6.37)

The \( \mathbf{h}_{KS} \) matrix is analogous to the Fock matrix in wave mechanics, and the one-electron and Coulomb parts are identical to the corresponding Fock matrix elements. The exchange–correlation part, however, is given in terms of the electron density, and possibly also involves derivatives of the density (or orbitals, as in the BR functional, eq. (6.25)).

\[ \int \chi_\alpha(r) V_{xc}[\rho(r), \nabla \rho(r)] \chi_\beta(r) dr \]  

(6.38)

Since the \( V_{xc} \) functional depends on the integration variables implicitly via the electron density, these integrals cannot be evaluated analytically, but must be generated by a numerical integration.

\[ \int \chi_\alpha(r) V_{xc}[\rho(r), \nabla \rho(r)] \chi_\beta(r) dr \approx \sum_{k=1}^{G} V_{xc}[\rho(r_k), \nabla \rho(r_k)] \chi_\alpha(r_k) \chi_\beta(r_k) \nabla v_k \]  

(6.39)

As the number of grid points \( G \) goes to infinity, the approximation becomes exact. In practice the number of points is selected based on the desired accuracy in the final results, i.e. if the energy is only required with an accuracy of \( 10^{-3} \), the number of integration points can be smaller than if the energy is required with an accuracy of \( 10^{-5} \).32 There are also some technical skills involved in selecting the optimum distribution of a given number of points to yield the best accuracy, i.e. the points should be dense where the function \( V_{xc} \) varies most. The grid is usually selected as being spherical around each nucleus, making it dense in the radial direction near the nucleus, and dense in the angular part in the valence space. For typical applications 1 000–10 000 points are used for each atom.33 It should be noted that only the larger of such
grids approach saturation, i.e. in general the energy will depend on the number (and location) of grid points. In order to compare energies for different systems, the same grid must therefore be used. The grid plays the same role for $E_{xc}$ as the basis set for the other terms. Just as it is improper to compare energies calculated with different basis sets, it is not justified to compare DFT energies calculated with different grid sizes. Furthermore, an incomplete grid will lead to “grid superposition errors” analogous to basis set superposition errors (Section 5.8).

With an expansion of the orbitals in basis functions, the number of integrals necessary for solving the KS equations increases as $M^4$, owing to the Coulomb integrals in the $J$ functional (and possibly also “exact” exchange in the hybrid methods). The number of grid points for the numerical $E_{xc}$ integration (eq. (6.39)) increases linearly with system size, and the computational effort for the exchange–correlation term increases as $GM^2$, i.e. a cubic dependence on system size. When the Coulomb (and possibly “exact” exchange) term is evaluated directly from integrals over basis functions, DFT methods scale formally as $M^4$. However, as discussed in Section 3.8.6, the Coulomb (and exchange) part can be calculated with an effort which scales only as $M^3$ for large systems with for example fast multipole methods. The numerical integration required for the exchange and correlation parts may also be reduced to a computation cost which scales linearly with system size, i.e. with modern techniques DFT is a true linear scaling method. This opens up the possibilities of performing accurate calculations on systems containing thousands of atoms, which is likely to have impacts on many areas outside traditional computational chemistry.

Nevertheless, the formal $M^4$ scaling has spawned approaches which reduce the dependence to $M^3$. This may be achieved by fitting the electron density to a linear combination of functions, and using the fitted density in evaluating the $J$ integrals in the Coulomb term.

$$\rho \approx \sum_{\alpha} \alpha_{\alpha}^f \chi_{\alpha}^f$$

(6.40)

The density fitting functions may or may not be the same as those used in expanding the orbitals. The fitting constants $\alpha_{\alpha}^f$ are chosen so that the Coulomb energy arising from the difference between the exact and fitted densities is minimized, subject to the constraint of charge conservation. The $J$ integrals then become

$$\int \chi_{\alpha}^f(1) \chi_{\alpha}^f(1) \frac{1}{|r_1 - r_2|} \chi_{\alpha}^f(2) dr_1 dr_2$$

(6.41)

which only involves three basis functions, thereby reducing the computational effort to $M^3$.

In some cases the exchange–correlation potential $V_{xc}$ is also fitted to a set of functions, similarly to the fitting of the density.

$$V_{xc} \approx \sum_{\alpha} b_{\alpha}^f \chi_{\alpha}^f$$

(6.42)

Again the set of fitting functions may or may not be the same as the orbital and/or the density basis functions. Once the potential has been fitted, the exchange–correlation energy may be evaluated from integrals involving three functions, analogously to eq.
(6.41). This fitting does not reduce the formal scaling, since the exchange–correlation term already is of order $M^3$.

Many of these fitting schemes were derived before linear scaling techniques (Section 3.8.6) were fully developed, and it is not clear whether they have any advantages. For calculation of energy derivatives, they actually seem counterproductive, since the fitting procedures seriously complicate the computational expressions.\textsuperscript{36}

The use of grid-based techniques for the numerical integration of the exchange–correlation contribution has some disadvantages when derivatives of the energy are desired. For this reason there is interest in developing grid-free DFT methods where the exchange–correlation potential is expressed completely in terms of analytical integrals.\textsuperscript{37}

In practice a DFT calculation involves an effort similar to that required for an HF calculation. Furthermore, DFT methods are one-dimensional just as HF methods are: increasing the size of the basis set allows a better and better description of the KS orbitals. Since the DFT energy depends directly on the electron density, it is expected that it has basis set requirements similar to those for HF methods, i.e. close to converged with a TZ(2df) type basis.

Should DFT methods be considered \textit{ab initio} or semi-empirical? If \textit{ab initio} is taken to mean the absence of fitting parameters, LSDA methods are \textit{ab initio}, but gradient corrected methods may or may not be. The LSDA exchange energy contains no parameters, and the correlation functional is known accurately as a tabulated function of the density. The VWN correlation functional (eqs. (6.20) and (6.21)) is merely a suitable interpolation formula necessarily for practical calculations; the constants do not represent fitting parameters chosen to improve the performance for atomic and molecular systems. Some gradient corrected methods (e.g. the B88 exchange and the LYP correlation), however, contain parameters which are fitted to give the best agreement with experimental atomic data, but the number of parameters is significantly smaller than for semi-empirical methods. The semi-empirical PM3 method (Section 3.10.5), for example, has 18 parameters for each atom, while the B88 exchange functional only has one fitting constant, valid for the whole periodic table. Other functionals are derived entirely from theory, and can consequently be considered “pure” \textit{ab initio}.

If \textit{ab initio} is taken to mean that the method is based on theory which in principle is able to produce the exact results, DFT methods are \textit{ab initio}. The only caveat is that current methods cannot yield the exact results, even in the limit of a complete basis set, since the functional form of the exact exchange–correlation energy is not known. Wave mechanics employ the exact Hamilton operator and make approximations for the wave function, while DFT makes approximations in the Hamilton operator, and it is easier to improve on the wave function description than to add correction to the operator. It is perhaps a little disturbing that seemingly very different functionals give results of similar quality.\textsuperscript{38} Although gradient corrected DFT methods have been shown to give impressive results, even for theoretically difficult problems, the lack of a systematic way of extending a series of calculations to approach the exact result is a major drawback of DFT. The results converge towards a certain value as the basis set is increased, but theory does not allow an evaluation of the errors inherent in this limit (like the systematic overestimation of vibrational frequencies with wave mechanics HF methods). Furthermore, although a progression of methods such as LSDA, BPW91
and B3PW91 provides successively lower errors for a suitable set of reference data (such as used for calibrating the Gaussian-2 model), there is no guarantee that the same progression will provide better and better results for a specific property of a given system. Indeed, LSDA methods may in some cases provide better results, even in the limit of a large basis set, than either of the more "complete" gradient corrected models. The quality of a given result can therefore only be determined by comparing the performance for similar systems where experimental or high quality wave mechanics results are available. In this respect DFT resembles semi-empirical methods. Nevertheless, DFT methods, especially those involving gradient corrections and hybrid methods, are significantly more accurate (and the errors are much more uniform) than those of for example the MNDO family, and DFT is consequently a valuable tool for systems where a (very) high accuracy is not needed.

References

3. F. Block, Z. Physik, 57 (1929), 545.

7.1 Classical Equations

A single-electron basis of...