Physica Scripta. Vol. T109, 9–39, 2004

Basic Density-Functional Theory—an Overview

U. von Barth*

Departamento de Fisica Teorica, Facultad de Ciencias, Universidad de Valladolid, 47011 Valladolid, Spain

Received June 13, 2003; accepted February 6, 2004

PACS Ref: 71.15.Mb, 71.15.Nc, 71.45.Gm

Abstract

In these notes I have given a personally flavored exposé of static density-functional theory (DFT). I have started from standard many-body physics at a very elementary level and then gradually introduced the basic concepts of DFT. Successively more advanced topics are added and at the end I even discuss a few not yet published theories.

The discussion represents many of the personal views of the author and there is no attempt at being comprehensive. I fully realize that I am often 'unfair' in treating the achievements of other researchers.

Many topics of standard DFT are deliberately left out like, e.g., time-dependence, excitations, and magnetic or relativistic effects.

These notes represent a compilation of a series of lectures given at at the EXC!TING Summer School *DFT beyond the ground state* at Riksgränsen, Sweden in June of 2003.

1. Background

Many reviews and articles on the topic of density-functional theory (DFT) start by the proclamation that, over the past so and so years, DFT has become, by far, the most prominent tool for the calculation of the ground-state properties of electronic systems. Although somewhat vacuous the statement is definitely true. In fact, DFT calculations of the electronic properties of real materials have nowadays turned into an extensive industrial endeavor.

The idea of using the density as the basic variable for the description of the energies of electronic systems goes back almost to the advent of quantum mechanics and the realization that the solution of the full equation of Schrödinger was beyond reach in most cases. The statistical atom of Gombás [1] and the approximations by Thomas [2] and Fermi [3] were early attempts in this direction. Then, of course, came the Hohenberg and Kohn [4] theorems in the mid sixties followed by the work by Kohn and Sham [5]. They demonstrated that the electron density of a fully interacting system could actually, in a rigorous way, be obtained from simple one-electron theory. At that time, most researchers involved with the calculation of the electronic properties of atoms, molecules, and solids where strongly influenced by the school of J. C. Slater. Properties were calculated from one-electron theory using a statistical approximation but only for the effect of exchange and correlation. The latter was obtained in 1951 by Slater [6] as an average over the Fermi sea of the self-energy of the homogeneous electron gas treated within the Hartree–Fock approximation. As a matter of fact, from a numerical point of view, that theory was not very different from modern DFT within the local-density approximation (LDA).

In the beginning of the seventies, Slater's so called $X\alpha$ method still dominated the scene and few people realized the immense conceptual importance of the new DFT. A younger scientist at IBM Research, A. Williams, was, however, one of those propagating the idea that if there is an, in principle, exact theory, why not try to use it. Eventually, Williams and collaborators produced a compilation of LDA results for almost the entire periodic table of elements—a very useful piece of work which is often used as a reference even today [7]. DFT, and then within the LDA, slowly gathered momentum until it, in the early eighties, by many people, was thought to be the answer to almost all questions. The journals were full of LDA calculations where the agreement with experiment was almost perfect and where remaining discrepancies were blamed on numerical flaws. Gradually came a realization that there were severe problems with the LDA. Perhaps mainly from solid-state physicist dealing also with more localized systems like molecules. Not really by the theoretical chemists who, at the time, were quite immune to the ideas of DFT. It was realized that the LDA always resulted in a certain overbinding depending on the degree of localization of the constituent electrons.

Because the LDA is based on the homogeneous but interacting electron gas, the inclusion of gradient corrections was the natural way to proceed in order to go beyond the LDA. These kind of corrections were actually worked out already in the original paper by Hohenberg and Kohn [4]. The next important steps were taken by Langreth, Perdew, and Mehl [8,9] in the late seventies and early eighties and John Perdew and collaborators have ever since been perusing the painstaking work which has carried those corrections to the level of sophistication where we are today [10].

But it was not John Perdew who turned the quantum chemists around. That achievement was signed Axel Becke [11]. One of his ideas is that molecules show no resemblance to the electron gas and gradient corrections should be designed to reproduce accurate results for prototype localized systems. With Becke's gradient corrections, and those of others, the atomization energies of a large number of well known molecules improved to the point where most established and commercially available codes for molecular calculations today can be run also in DFT mode. In fact, after the beginning of the nineties the popularity of DFT within the community of quantum chemists grew to the point where Walter Kohn was awarded the Nobel Prize in Chemistry, in 1998. Professor Kohn shared the prize with the chemist John Pople [12]

---

* Permanent address: Lund University, SE-22362 Lund, Sweden

who is the major player in the development of computer codes for electronic structure calculations in molecules.

I am inclined to hand out a mark of "not passed" to the Nobel Committee for physics and perhaps to the entire physics community for allowing themselves to be beaten by chemists in realizing the importance of DFT. After all, DFT is more accurate and perhaps more indispensable in solids as compared to the case of molecules.

These notes could rather easily be mistaken for a review article on DFT—and then a comparatively inferior and incomplete one. But they are definitely not intended as one. Instead, my aim is to guide a beginner through the basic principles of DFT up to a level where he or she will have acquired a basic understanding of what is involved, which are the major methods, the main obstacles, the predominant fields of applications, and what kind of accuracy one can expect. It would be great if a follower of these notes would also end up having some idea about in what direction the field will develop in the near future. There are many topics which will not be discussed here at all because they are covered by other contributors to the present meeting. Typical such topics are the the spin-dependent generalization [13,14], the pseudizing of the theory [15–17], excited states and the time-dependent version of the theory (TDDFT) [18,19], current-density functional theory [20], relativistic DFT [21,22], DFT at higher temperatures [23], and many-component DFT [24].

Looking through the list of references you will find that the majority of them are relatively old. This is connected, in part, to my opinion that there has not been so much novel theoretical development over the past ten years. Instead we have seen an explosion in the production of results for a very large number of different physical systems of interest to material science, magnetic properties, bio-active molecules and drugs, nano-systems and devices, etc. etc. Since my interest has been in the development of new theory, these advances are only mentioned here.

The list of references given here is in no way complete or 'fair' to many workers in the field. The list reflects my personal preferences and to some extent also my lack of knowledge caused by my absence from the field in later years. I have basically referred only to work which has had some impact on my own work within the field. For a more comprehensive coverage of DFT and a better description of other peoples work, I refer the interested reader to a number of excellent review articles in the field [25–39].

2. Many-electron theory

2.1. *The total energy*

Most presentations of DFT follow the historical path of the founding fathers. Unfortunately, many beginners find this formulation rather abstract and difficult to grasp. In a beautiful but very short paper from 1979 [40], Mel Levy gave a much less abstract presentation of the basic functionals of the theory. For pedagogical reasons, we prefer the Levy approach and will later make extensive use of this approach. But we will begin by investigating how far we can proceed without any use of DFT. We believe that our approach has certain pedagogical advantages in so far as it builds up expectations to the point where DFT becomes the most natural next step.

The problem we like to address is that of finding the ground-state energy of a many-electron system. As usual in quantum mechanics, we then solve the eigenvalue problem,

$$H\Psi = E\Psi$$

for the Hamiltonian H and look for the lowest eigenvalue. Unfortunately, in the case of a solid or a large molecule the Hamiltonian has a dreadful appearance,

$$H = \sum_{i=1}^{N} -\frac{1}{2}\nabla_i^2 + \sum_{i=1}^{N} w(\boldsymbol{r}_i) + \frac{1}{2}\sum_{i\neq j}^{N} v(\boldsymbol{r}_i - \boldsymbol{r}_j).$$

Here, the first term represents the kinetic energy of all the N electrons in the system. The second term represents the interaction with the external (to the electrons) charges and for a molecule or a solid it would, e.g., look like

$$w(\boldsymbol{r}) = -\sum_{\mu=1}^{M} Z_\mu v(\boldsymbol{r} - \boldsymbol{R}_\mu)$$

where Z_μ are the different number of protons on the different constituent M atoms and $v(\boldsymbol{r}) = 1/r$ is the ordinary Coulomb interaction. Notice that we have here transformed ourselves to the world of atoms by putting $m = \hbar = e^2/(4\pi\varepsilon_o) = 1$. In this world the unit of length is one Bohr radius equal to 0.529 Å and the unit of energy is one Hartree or 27.21 eV. Finally, the last term is the very strong and important repulsion energy between all the electrons (the factor of 1/2 makes sure that every interaction is included only once).

For any system of practical interest the number of electrons would range from ten to 10^{23} and the number of degrees of freedom in the Hamiltonian is prohibitively large. Still, in the absence of the last term in the Hamiltonian, it would actually have been feasible to find the total energy of the system from simple one-electron theory. This follows from the fact that, without the electron-electron interaction terms, the Hamiltonian would have been a sum of identical one-electron Hamiltonians—one for each electron. And they must all be identical because all electrons in nature are indistinguishable. The eigenstates of such a Hamiltonian can be shown to be Slater determinants consisting of one-electron orbitals which are solutions to the mentioned one-electron Hamiltonian and the energies of the many-body states are just the sum of the eigenvalues of the constituent orbitals. We are, however, not that fortunate but we might consider the possibility of approximating the inter-electron part of the Hamiltonian by a sum of terms corresponding to each electron moving in some average field of all the other electrons. This was exactly what was done in the so called $X\alpha$ approach mentioned in the introductory section and also Hartree–Fock theory has this form. And, as we shall see later, also DFT can be cast in this mean-field form although the resulting equations are no approximations but give exact answers to certain well defined questions.

For the time being, we notice that the total ground-state energy can be obtained as the expectation value of the

Hamiltonian with respect to some exact or approximate wave-function Ψ according to

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

But in the process of evaluating this expression we will find that most degrees of freedom integrate out and only some partial knowledge of the full wave function is needed in order to obtain the energy. Let us first consider the simplest term in the Hamiltonian, i.e., the energy arising from the interaction with the external potential. We have, in obvious notation,

$$W = \langle\Psi|\hat{W}|\Psi\rangle = \sum_{i=1}^{N} \langle\Psi|w(\boldsymbol{r}_i)|\Psi\rangle$$

$$= \int d^3r\, w(\boldsymbol{r}) \sum_{i=1}^{N} \langle\Psi|\delta(\boldsymbol{r}-\boldsymbol{r}_i)|\Psi\rangle = \int d^3r\, w(\boldsymbol{r})\langle\Psi|\hat{n}(\boldsymbol{r})|\Psi\rangle$$

where we have defined the density operator $\hat{n}(\boldsymbol{r})$ through the relation

$$\hat{n}(\boldsymbol{r}) = \sum_{i=1}^{N} \delta(\boldsymbol{r}-\boldsymbol{r}_i).$$

Thus, defining the ground-state density $n(\boldsymbol{r})$ of the system by

$$n(\boldsymbol{r}) = \langle\Psi|\hat{n}(\boldsymbol{r})|\Psi\rangle$$

we have the very intuitive and very classical relation

$$W = \int d^3r\, w(\boldsymbol{r})\, n(\boldsymbol{r}).$$

The important message displayed by this relation is that, for this part W of the energy, only a very limited piece of information from the full wave function is needed, in this case, the density given by

$$n(\boldsymbol{r}) = N\int d^3r_2 \ldots d^3r_N d\sigma_2 \ldots d\sigma_N \\ \times |\Psi(\boldsymbol{r},\sigma;\boldsymbol{r}_2,\sigma_2;\ldots;\boldsymbol{r}_N,\sigma_N)|^2.$$

Here, the factor N, i.e. the total number of electrons, enters because the symmetry of the wave function makes all terms contribute an equal amount in the sum over electrons above. (We have written the discrete sum of the two values of the spin variables σ_i as a generalized integral.)

Turning next to second most complicated term in the total energy, i.e. the kinetic energy T, we obviously obtain

$$T=\langle\Psi|\hat{T}|\Psi\rangle = \sum_{i=1}^{N} \langle\Psi| -\frac{1}{2}\nabla_i^2|\Psi\rangle$$

$$=-\frac{N}{2}\int \Psi^*(x_1,x_2,\ldots,x_N)\nabla_1^2\Psi(x_1,x_2,\ldots,x_N)dx_1\, dx_2\ldots dx_N.$$

In the latter integral we have used the short-hand notation x_i for the set of variables $(\boldsymbol{r}_i,\sigma_i)$. Thus, defining what in the trade is known as the one-particle density matrix according to

$$\Gamma^{(1)}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') \\ = \int \Psi^*(\boldsymbol{r},\sigma;x_2,\ldots,x_N)\Psi(\boldsymbol{r}',\sigma';x_2,\ldots,x_N)dx_2\ldots dx_N$$

we have

$$T = -\frac{1}{2}\sum_{\sigma,\sigma'}\int\left[\nabla^2\Gamma^{(1)}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}')\right]_{\boldsymbol{r}'=\boldsymbol{r}} d^3r.$$

We see again that a rather limited information from the full wave function is needed in order to calculate the total kinetic energy of the system. For future reference and for dealing also with magnetic systems, we notice that the diagonal components of the one-particle density matrix represent the spin densities $n_\sigma(\boldsymbol{r})$,

$$n_\sigma(\boldsymbol{r}) = \Gamma^{(1)}_{\sigma,\sigma}(\boldsymbol{r},\boldsymbol{r})$$

in terms of which the density is just

$$n(\boldsymbol{r}) = \sum_\sigma n_\sigma(\boldsymbol{r}).$$

Finally, we turn to the most complicated part of the energy, i.e., the interaction energy U which, in terms of the full wave function, is given by

$$U = \frac{1}{2}\sum_{i\neq j}^{N}\langle\Psi|v(\boldsymbol{r}_i-\boldsymbol{r}_j)|\Psi\rangle$$

$$= \frac{1}{2}N(N-1) \\ \times\int|\Psi(\boldsymbol{r}_1,\sigma_1;\boldsymbol{r}_2,\sigma_2;\ldots;\boldsymbol{r}_N,\sigma_N)|^2 v(\boldsymbol{r}_1-\boldsymbol{r}_2)dx_1\, dx_2\ldots dx_N.$$

Also in this case, the particular contraction of the total wave function giving the interaction energy, has a special name. It is called the diagonal of the two-particle density matrix and it is given by

$$\Gamma^{(2)}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') = N(N-1) \\ \times\int|\Psi(\boldsymbol{r},\sigma;\boldsymbol{r}',\sigma';x_3,\ldots,x_N)|^2\, dx_3\ldots dx_N.$$

This quantity has a very physical interpretation. The operator representing the density of electron i at the point $\boldsymbol{r}$ in space with spin σ is obviously $\delta(\boldsymbol{r}-\boldsymbol{r}_i)\delta_{\sigma,\sigma_i}$. And there is a similar expression giving the density of electron j at another point $\boldsymbol{r}'$ with spin σ'. We can then construct the density corresponding to any electron at the point $\boldsymbol{r}$ with spin σ given the fact that there is another one at the point $\boldsymbol{r}'$ with spin σ'. This quantity must obviously have contributions from all pairs of electrons and we write it

$$\hat{n}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') = \sum_{i\neq j}^{N}\delta(\boldsymbol{r}-\boldsymbol{r}_i)\delta_{\sigma,\sigma_i}\delta(\boldsymbol{r}'-\boldsymbol{r}_j)\delta_{\sigma',\sigma_j}.$$

We have here excluded the term with $i \neq j$ because no electron can be at two different points in space at the same time. Calculating the expectation value of this two-electron density in the ground-state state gives,

$$\Gamma^{(2)}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') = \langle \Psi | \hat{n}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') | \Psi \rangle.$$

The quantity $\Gamma^{(2)}$ has a number of intuitively understandable properties. According to the Pauli exclusion principle, two electrons can never be at the same place if they have the same spin. Thus,

$$\Gamma^{(2)}_{\sigma,\sigma}(\boldsymbol{r},\boldsymbol{r}) = 0$$

which follows directly from the antisymmetry of the wave function. This symmetry also shows that $\Gamma^{(2)}$ is symmetric in its arguments

$$\Gamma^{(2)}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') = \Gamma^{(2)}_{\sigma',\sigma}(\boldsymbol{r}',\boldsymbol{r}).$$

The operator $\hat{n}(\boldsymbol{r},\boldsymbol{r}')$ is easily seen to be a positive definite operator so that

$$\Gamma^{(2)}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') \geq 0.$$

Let us then integrate one of the spatial variables over all space to obtain

$$\begin{aligned}\int \hat{n}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}')\mathrm{d}^3r' &= \sum_{i \neq j}^{N} \delta(\boldsymbol{r}-\boldsymbol{r}_i)\delta_{\sigma,\sigma_i}\delta_{\sigma',\sigma_j} \\ &= \sum_{i,j}^{N} \delta(\boldsymbol{r}-\boldsymbol{r}_i)\delta_{\sigma,\sigma_i}\delta_{\sigma',\sigma_j} - \sum_{i=1}^{N} \delta(\boldsymbol{r}-\boldsymbol{r}_i)\delta_{\sigma,\sigma_i}\delta_{\sigma',\sigma_i} \\ &= \hat{n}_\sigma(\boldsymbol{r})\hat{N}_{\sigma'} - \hat{n}_\sigma(\boldsymbol{r})\delta_{\sigma,\sigma'}\end{aligned}$$

where the operator $\hat{n}_\sigma(\boldsymbol{r})$, of course, corresponds to the spin density at $\boldsymbol{r}$,

$$\hat{n}_\sigma(\boldsymbol{r}) = \sum_{i=1}^{N} \delta(\boldsymbol{r}-\boldsymbol{r}_i)\delta_{\sigma,\sigma_i}$$

The operator $\hat{N}_\sigma$ counts the total number of electrons with spin σ,

$$\hat{N}_\sigma = \sum_{i=1}^{N} \delta_{\sigma,\sigma_i}.$$

The latter operator is just a constant $(2/\hbar)$ times the z-component of the total spin. Since the Hamiltonian does not contain any spin variables it commutes with the total spin and, in a finite system, one can then always consider the ground state as an eigenfunction of the total spin. Writing $(\hbar/2)N_\sigma$ for the corresponding eigenvalue and taking an expectation value with respect to the ground state of the above operator relation we obtain,

$$\int \Gamma^{(2)}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}')\ \mathrm{d}^3r' = n_\sigma(\boldsymbol{r})\{N_\sigma - \delta_{\sigma,\sigma'}\}.$$

Due to the symmetry of $\Gamma^{(2)}$ we would certainly have obtained a similar result by instead integrating over $\boldsymbol{r}$ and we realize that the spin densities $n_\sigma(\boldsymbol{r})$ can be considered to be factors in $\Gamma^{(2)}$. We thus define a new quantity g, called the pair-correlation function, according to the relation

$$\Gamma^{(2)}_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') = n_\sigma(\boldsymbol{r})n_{\sigma'}(\boldsymbol{r}')g_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}').$$

The function $\Gamma^{(2)}$ is an expectation value of a two-particle density and thus more related to densities than to probabilities. Dividing by the spin-densities the pair-correlation function g becomes more related to probabilities. In fact, if the electrons were completely independent the two-particle density would be just the product of the one-electron densities and the pair-correlation function would equal unity. Thus, the degree of correlation between the electrons is measured by how much smaller than unity g is. And, from the definition, it is certainly always positive as one would expect from a probability.

Using the new definition of g and the obvious relation

$$\int n_\sigma(\boldsymbol{r})\ \mathrm{d}^3r = N_\sigma,$$

our previous integral formula for the pair density $\Gamma^{(2)}$ becomes

$$\int n_{\sigma'}(\boldsymbol{r}')\{g_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}') - 1\}\ \mathrm{d}^3r' = -\delta_{\sigma,\sigma'}$$

after division by $n_\sigma(\boldsymbol{r})$. Consequently, the quantity $n[g-1]$ represents a negative particle density, i.e., a lack of electrons which integrates up to negative unity for electrons of the the same spin and to zero for those of opposite spin. This is the famous exchange-correlation hole surrounding every electron in the system. Using g, the formula for the interaction energy U becomes

$$U = \frac{1}{2}\sum_{\sigma,\sigma'} \int n_\sigma(\boldsymbol{r})n_{\sigma'}(\boldsymbol{r}')g_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}')v(\boldsymbol{r}-\boldsymbol{r}')\ \mathrm{d}^3r\mathrm{d}^3r'.$$

We see that this energy involves a sum over spin indices and we do not need the detailed spin structure of the pair-correlation function g in order to calculate U. It is, therefore, convenient to define the spin averaged pair-correlation function g by means of the relation

$$n(\boldsymbol{r})n(\boldsymbol{r}')g(\boldsymbol{r},\boldsymbol{r}') = \sum_{\sigma,\sigma'} n_\sigma(\boldsymbol{r})n_{\sigma'}(\boldsymbol{r}')g_{\sigma,\sigma'}(\boldsymbol{r},\boldsymbol{r}').$$

It should be remembered, however, that the full spin structure is of interest in magnetic system and is also important to attempts to approximate the pair-correlation function. This is due to the fact that the correlations between electrons of different spin is much stronger and different in character as compared to the case of the like spin electrons. The latter are kept apart by the Pauli exclusion principle thus reducing their mutual Coulomb repulsion.

The largest part of the interaction energy is the classical electrostatic Coulomb interaction U_0 between the electrons. In most electronic systems, this is a huge part of the total energy, much larger than the energies associated with exchange and correlation, and it is not very difficult to calculate. It is given by

$$U_0 = \frac{1}{2}\int n(\boldsymbol{r})n(\boldsymbol{r}')v(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r'.$$

It is thus customary to separate out this part of the interaction energy. In terms of our spin averaged pair-correlation function g the interaction energy U becomes

$$U = U_0 + \frac{1}{2}\int n(\boldsymbol{r})n(\boldsymbol{r}')\{g(\boldsymbol{r},\boldsymbol{r}')-1\}v(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r'.$$

The last term in this expression is referred to as the exchange-correlation part if the interaction energy often written U_{xc}. In terms of the spin averaged pair-correlation function g the famous sum rule for the exchange-correlation hole becomes

$$\int n(\boldsymbol{r}')\{g(\boldsymbol{r},\boldsymbol{r}')-1\}\,\mathrm{d}^3r' = -1.$$

We see from this that whereas the classical Coulomb energy, also sometimes called the Hartree energy, is an interaction between N electrons and N electrons, the exchange correlation part U_{xc} of the interaction energy is the classical Coulomb interaction between each of the N electrons with one negative unit of charge, i.e., with its exchange-correlation hole.

Collecting all pieces of the total energy we have

$$E = -\tfrac{1}{2}\mathrm{Tr}\,[\nabla^2\Gamma^{(1)}] + \int wn + \tfrac{1}{2}\int nvn' + \tfrac{1}{2}\int nn'\{g-1\}v,$$

where we have suppressed all integration variables and used the symbol Tr to signify the operations of making the arguments equal and integrating over all space, in analogy with the concept of the trace of an ordinary matrix.

2.2. *The Hellman-Feynman theorem*

As we shall see shortly, it is not difficult to find rather accurate approximations to the spin averaged pair-correlation function of many electronic systems. These approximations will give errors in the already rather small exchange-correlation energies of the order of 5% or less. It has, however, proved to be difficult to find approximations to the one-particle density matrix $\Gamma^{(1)}$ which are accurate enough for calculating the very large kinetic energy of the system. Fortunately, one can use the well known Hellman–Feynman theorem of textbook quantum mechanics to transform the calculation of the total kinetic energy into a calculation of the kinetic energy of non-interacting electrons plus a small correction to the correlation part of the interaction energy. It is actually rather simple to demonstrate how this is done. We imagine that we scale down the strength—now called λ—of the Coulomb interaction between the electrons. For instance, in an atom this would have a drastic effect on the density of the electrons which would then become much more compact and pile up close to the nucleus. This effect would result from the reduction in the screening by the other electrons of the attraction of one electron to the nucleus. We will therefore add an external λ-dependent one-body potential to the system in such a way that the density is not affected by the reduction in the inter-electron Coulomb interaction. It is perfectly legitimate to ask if this is actually possible to achieve. This is a very difficult question and is referred to as the w-representability problem of DFT. It is normally assumed to be possible but it is very difficult to prove in an exact way. But most physicists are content with the fact that, in practice, it has always been possible to achieve w-representability within the desired degree of accuracy.

The energy E_λ of the system will, of course, change as a result of the reduction in the Coulomb interaction. For obtaining the derivative of E_λ with respect to the parameter λ, the Hellman–Feynman theorem states that it is sufficient to differentiate the explicit λ-dependence in the Hamiltonian. Due to the stationary property of the ground-state energy E_λ, the λ-dependence of the ground-state wave function Ψ_λ does not enter. Thus, in obvious notation,

$$\begin{aligned}\frac{\partial E_\lambda}{\partial\lambda} &= \left\langle\Psi_\lambda\left|\frac{\partial H_\lambda}{\partial\lambda}\right|\Psi_\lambda\right\rangle = \left\langle\Psi_\lambda\left|\frac{\partial\hat{W}_\lambda}{\partial\lambda}\right|\Psi_\lambda\right\rangle + \left\langle\Psi_\lambda|\hat{U}|\Psi_\lambda\right\rangle \\ &= \int\frac{\partial w_\lambda(\boldsymbol{r})}{\partial\lambda}n_\lambda(\boldsymbol{r})\,\mathrm{d}^3r + \tfrac{1}{2}\int n_\lambda(\boldsymbol{r})n_\lambda(\boldsymbol{r}')g_\lambda(\boldsymbol{r},\boldsymbol{r}')v(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r'.\end{aligned}$$

But the λ-dependence of $w_\lambda(\boldsymbol{r})$ was defined to be such that the density remained constant and equal to the density $n(\boldsymbol{r})$ of the fully interacting system at all λ. Therefore, integrating again with respect to λ, we obtain

$$\begin{aligned}E - E_o &= \int[w(\boldsymbol{r}) - w_0(\boldsymbol{r})]n(\boldsymbol{r})\,\mathrm{d}^3r \\ &\quad + \tfrac{1}{2}\int n(\boldsymbol{r})n(\boldsymbol{r}')\tilde{g}(\boldsymbol{r},\boldsymbol{r}')v(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r'.\end{aligned}$$

Here, E_o is the total energy of a system of N non-interacting electrons moving in the external potential $w_o(\boldsymbol{r})$ and having the same density $n(\boldsymbol{r})$ as the fully interacting ($\lambda = 1$) system. This means that we find E_o by solving the one-electron Schrödinger equation

$$\{-\tfrac{1}{2}\nabla^2 + w_o(\boldsymbol{r})\}\phi_k(\boldsymbol{r}) = \varepsilon_k\phi_k(\boldsymbol{r})$$

to obtain

$$E_o = \sum_{k=1}^{N}\varepsilon_k = \sum_{k=1}^{N}\langle\phi_k|-\tfrac{1}{2}\nabla^2|\phi_k\rangle + \int w_o(\boldsymbol{r})n(\boldsymbol{r})\,\mathrm{d}^3r$$

where

$$n(\boldsymbol{r}) = \sum_{k=1}^{N}|\phi_k(\boldsymbol{r})|^2.$$

The quantity $\tilde{g}$ is just the normal pair-correlation function g but averaged over all values of the interaction strength λ. Thus,

$$\tilde{g}(\boldsymbol{r},\boldsymbol{r}') = \int_0^1 g_\lambda(\boldsymbol{r},\boldsymbol{r}')\,\mathrm{d}\lambda.$$

Again adding up the pieces of the total energy we find

$$E = \sum_{k=1}^{N} \langle \phi_k | -\tfrac{1}{2}\nabla^2 | \phi_k \rangle + \int wn + \tfrac{1}{2}\int nvn' + \tfrac{1}{2}\int nn'\{\tilde{g} - 1\}v,$$

where we suppressed integration variables as we did previously.

This is a truly remarkable results. It means that we can always find the total energy of an interacting many-electron system by doing one-electron theory. And we have not even mentioned density-functional theory (DFT) at this stage. What we have to do is to find a one-electron potential that gives the corresponding non-interacting system of N electrons a density which is identical to the true density of the interacting system. And we must find the exact average pair-correlation function $\tilde{g}$ giving rise to the comparatively small exchange-correlation energy.

As we shall see in the next section, is not very difficult to find rather accurate approximations to the latter quantity. The first task is, however, considerably more difficult in the sense that it requires the formal machinery of DFT. Fortunately, an accurate total energy actually does not require a very accurate one-electron potential. This is due to the variational property of the total energy as a functional of the density as will be explained in Section 3.

We will end this section by making a comment which we find both interesting and amusing. By means of the local-density approximation (LDA) of the following section we have an explicit expression for the last term, i.e., the exchange-correlation part of the total energy from the last equation. We could, of course, then do exactly as in Thomas–Fermi theory and minimize our, by now, explicit expression for the total energy by varying the density. This would certainly entail varying the one-electron potential w_o giving rise to that density and we would find the lowest energy when w_o is chosen equal to what is nowadays referred to as the total Kohn–Sham potential within the LDA. Consequently, we could have started to do local-density calculations without ever having heard about DFT. Of course, in order to proceed toward an exact theory allowing us, e.g., to go beyond the LDA we need DFT. And one should certainly not underestimate the tremendous conceptual impact of the theorems by Hohenberg, Kohn, and Sham.

2.3. *The local-density approximation*

We end this section by a short discussion on the well know local-density approximation (LDA). Of course, the LDA is almost always associated with DFT—and rightfully so. We will, however, introduce it already here in an attempt to demonstrate how far toward accurate total energies one can actually get without invoking DFT. The purpose of this approach is certainly not to rob the revered fathers of DFT of any glory. It is rather to build up understanding and expectations to the point where DFT becomes the one and only way to proceed.

Our task is here to approximate the average pair-correlation function $\tilde{g}$ of the inhomogeneous and interacting system in order to obtain a good approximation to what we from now and on will refer to as the exchange-correlation part of the total energy, E_{xc},

$$E_{xc} = \frac{1}{2}\int n(\boldsymbol{r})n(\boldsymbol{r}')\{\tilde{g}(\boldsymbol{r},\boldsymbol{r}') - 1\}v(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r'.$$

We begin by approximating the exact $\tilde{g}$ by that of the homogeneous but interacting electron gas which we call $\tilde{g}_h$. But this function is also a function of the density of the homogeneous gas and one is immediately faced with the problem of choosing some effective density $\bar{n}$ for the approximation. This is actually a standard problem when one wants to, as is often done, approximate two-point functions in inhomogeneous systems with the corresponding quantities from the homogeneous gas. The choice which leads to the simplest algebra is $\bar{n} = n(\boldsymbol{r})$, which is here the preferred choice. We notice that these approximations for $\tilde{g}$ will become exact when the density approaches a constant. We finally complete our approximation by also replacing the argument $\boldsymbol{r}'$ in the second density by $\boldsymbol{r}$. Again this approximation becomes exact when the density tends to a constant and we can also argue that the largest contribution to the integral is coming from regions where the density is large. It can be shown that the function $\tilde{g} - 1$ differs little from zero when the two arguments $\boldsymbol{r}$ and $\boldsymbol{r}'$ are further apart than the average distance between electrons. Thus, the integral has large contributions only from regions when $\boldsymbol{r}$ and $\boldsymbol{r}'$ are close which means that it is quite reasonable to replace the argument $\boldsymbol{r}'$ by $\boldsymbol{r}$ in the second density if the density does not vary appreciably over an inter electronic distance. We have completed the LDA and obtain (notice that, due to the complete translational symmetry of the homogeneous gas, the function $\tilde{g}_h$ can only be a function of $\boldsymbol{r}-\boldsymbol{r}'$)

$$\begin{aligned} E_{xc}^{LDA} &= \frac{1}{2}\int n(\boldsymbol{r})n(\boldsymbol{r})\{\tilde{g}_h(\boldsymbol{r}-\boldsymbol{r}'; n(\boldsymbol{r})) - 1\}v(\boldsymbol{r}-\boldsymbol{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r' \\ &= \frac{1}{2}\int \mathrm{d}^3r\; n(\boldsymbol{r})n(\boldsymbol{r})\{\tilde{g}_h(\boldsymbol{r}'; n(\boldsymbol{r})) - 1\}v(\boldsymbol{r}')\,\mathrm{d}^3r' \\ &= \int \mathrm{d}^3r\; n(\boldsymbol{r})\varepsilon_{xc}(n(\boldsymbol{r})). \end{aligned}$$

Here, the function $\varepsilon_{xc}(n)$ is just the exchange-correlation energy per electron of the homogeneous and interacting electron gas obtained by evaluating the exact expression for E_{xc} above at a constant density n and then dividing by the total number N of electrons. (Notice that $N = \Omega n$ where Ω is the total volume of the gas).

We could actually have arrived at this expression in a more direct way as follows. If the density of an inhomogeneous system does not vary that rapidly we can split the whole system into small boxes of volume d^3r and consider each box as containing a homogeneous electron gas with a density equal to the local density $n(\boldsymbol{r})$ at the coordinate $\boldsymbol{r}$ of the box. The contribution to the energy of each box is, of course, the exchange-correlation energy per particle $\varepsilon_{xc}(n(\boldsymbol{r}))$ of the gas in that box times the total number, $n(\boldsymbol{r})\,\mathrm{d}^3r$, of electrons in the box. Summing over all boxes gives us back the LDA above.

In order to use the LDA in practical calculations, we must certainly supply a table of accurate exchange-correlation energies of the homogeneous gas as a function of the density n. Such results are obtained from large-scale Monte Carlo simulations [41] and parameterized for convenient access in computer codes. Just one word of advice. There are many parameterizations in the literature and most of them are, for all practical purposes, next to equivalent [26,42,43]. But avoid those which are based on very complicated analytical expressions which use up an undue amount of computer time.

We will give a detailed account of the successes and failures of the LDA later on.

3. Density-functional theory

3.1. *The kinetic-energy functional*

As the name density-functional theory (DFT) suggests, DFT is a many-body theory based on the idea of using only the density as the basic variable for describing many-electron systems. And it is a formulation in terms of functionals of the density. As mentioned in the introduction, the rather abstract functionals of DFT were formulated in a rather hands-on way by Mel Levy in 1979 [40]. And we will here prefer this so called constrained search approach by Levy to the original formulation by Hohenberg, Kohn, and Sham. We will begin by defining the functional for the kinetic energy, T_o.

Given an arbitrary particle density $n(\boldsymbol{r})$, we think of all anti-symmetric many-body wave functions which yield that density n. In mathematical language this set M of states is written

$$M(n) = \{|\Psi\rangle \mid \langle\Psi|\hat{n}(\boldsymbol{r})|\Psi\rangle = n(\boldsymbol{r})\}.$$

Here, $\hat{n}(\boldsymbol{r})$ is the density operator for the N electrons as defined in Section 2.1. Although the density $n(\boldsymbol{r})$ could be chosen freely, it should obey certain conditions like being everywhere non-negative and integrating up to N electrons. It is appropriate to ask whether there is always an anti-symmetric state yielding the density $n(\boldsymbol{r})$ or, in other words, if the set $M(n)$ is not empty. The answer to this question is actually yes. This problem is referred to as the N-representability problem of N-electron densities and we will not further dwell on this rather complicated issue.

The operator $\hat{T}$ corresponds to the kinetic energy of the N electrons and we will now study its expectation values with respect to states in the set $M(n)$. And we search for the smallest possible such expectation value. These expectation values are bounded from below because the kinetic energy is a positive definite operator. This means that the set of such expectation values has an *infimum* meaning that there is a number, say T_o, such that all the expectation values are larger than or equal to T_o and that there is always at least one expectation value smaller than $T_o + \varepsilon$, no matter how small $\varepsilon > 0$. We write

$$T_o[n] = \underset{|\Psi>\in M(n)}{\text{infimum}} \langle\Psi|\hat{T}|\Psi\rangle.$$

From the point of view of a physicist it is enough to know that $T_o[n]$ is the smallest possible kinetic energy which can be achieved among many-body states all having the density $n(\boldsymbol{r})$. In other words, the value $T_o[n]$ is the minimum of all expectation values $\langle\Psi|\hat{T}|\Psi\rangle$ provided also $\langle\Psi|\hat{n}(\boldsymbol{r})|\Psi\rangle = n(\boldsymbol{r})$. Consequently, in order to find $T_0[n]$, we are faced with a minimization under constraints, and there is one constraint for each point in space since the expectation value of the density operator at each point $\boldsymbol{r}$ in space has to equal the prescribed density $n(\boldsymbol{r})$. Such a constrained minimization is most easily achieved through the use of Lagrangian multipliers. And we must have one Lagrangian parameter $V(\boldsymbol{r})$ for each point $\boldsymbol{r}$ in space and then minimize

$$\langle\Psi|\hat{T}|\Psi\rangle + \int V(\boldsymbol{r})n(\boldsymbol{r})\,\mathrm{d}^3r = \langle\Psi|\hat{T}|\Psi\rangle + \int V(\boldsymbol{r})\langle\Psi|\hat{n}(\boldsymbol{r})|\Psi\rangle\,\mathrm{d}^3r$$
$$= \langle\Psi|\hat{T} + \hat{V}|\Psi\rangle$$

with respect to arbitrary variations of the wave function or state $|\Psi\rangle$ (notice, however, that it still has to obey the anti-symmetrization postulate). Here, the 'one-body' quantum mechanical operator $\hat{V}$ is defined through the relation

$$\hat{V} = \sum_{i=1}^{N} V(\boldsymbol{r}_i)$$

But this is a common problem in quantum mechanics that we know how to solve. It is the problem of finding the ground-state energy of the Hamiltonian

$$H_o = \hat{T} + \hat{V}$$

which is a sum of one-electron Hamiltonians. Thus, we just have to construct all eigensolutions to the one-dimensional Schrödinger equation

$$\left\{-\tfrac{1}{2}\nabla^2 + V(\boldsymbol{r})\right\}\phi_k(\boldsymbol{r}) = \varepsilon_k\phi_k(\boldsymbol{r})$$

and then we know that the ground-state wave function is the normalized Slater determinant consisting of the N solutions with the lowest possible one-electron eigenvalues (N is the number of electrons in the system). The ground-state energy is just the sum of those lowest eigenvalues and

$$T_o[n] = \langle\Psi|\hat{T}|\Psi\rangle = \sum_{k=1}^{N}\left\langle\phi_k\middle|-\tfrac{1}{2}\nabla^2\middle|\phi_k\right\rangle$$

Then, of course, the infinite set of Lagrangian parameters $V(\boldsymbol{r})$, which, in this case, obviously acts as the external one-electron potential has to be adjusted until the density given by

$$n(\boldsymbol{r}) = \sum_{k=1}^{N}|\phi_k(\boldsymbol{r})|^2$$

becomes equal to the chosen one.

The whole procedure can now be summarized in words. In order to find the value of the functional $T_o[n]$ of the density n at a particular density $n(\boldsymbol{r})$, we just solve the one-electron Schrödinger problem in some potential $V(\boldsymbol{r})$ which is then adjusted until the sum of the squares of the corresponding one-electron orbitals agrees with the chosen density. The value of the functional for the kinetic energy is then the total kinetic energy of this non-interacting problem. Notice that the procedure relies on the possibility of always being able to find a potential (Lagrangian parameters) which reproduces the chosen density. This is very difficult to show in an exact way and requires very strict conditions on the chosen densities. We will not go in to this rather complicated issue here. This problem is referred to as the non-interacting w-representability problem in the density-functional literature. We will be content by the fact that, in practice, one has always been able to find to such a potential within some rather high degree of accuracy. And many such calculations have been carried out since I started the activity in 1982 (see e.g. Ref. [32]).

The next step in our formal development of DFT is to study how the functional T_o changes when the density n undergoes a small change to $n+\delta n$. The Lagrangian multipliers must obviously change to, say $V+\delta V$, in order to produce the new density $n+\delta n$. The first-order change in T_o becomes

$$\delta T_o = \delta\left\{\sum_{k=1}^{N}\varepsilon_k - \int V(\boldsymbol{r})n(\boldsymbol{r})\,\mathrm{d}^3r\right\}$$
$$= \sum_{k=1}^{N}\delta\varepsilon_k - \int \delta V(\boldsymbol{r})n(\boldsymbol{r})\,\mathrm{d}^3r - \int V(\boldsymbol{r})\delta n(\boldsymbol{r})\,\mathrm{d}^3r.$$

But, when the one-electron potential $V(\boldsymbol{r})$ changes, the one-electron eigenvalues also change and from first-order perturbation theory we know that

$$\delta\varepsilon_k = \langle\phi_k|\delta V|\phi_k\rangle$$

Hence,

$$\sum_{k=1}^{N}\delta\varepsilon_k = \int \delta V(\boldsymbol{r})n(\boldsymbol{r})\,\mathrm{d}^3r$$

and we obtain

$$\delta T_o = -\int V(\boldsymbol{r})\delta n(\boldsymbol{r})\,\mathrm{d}^3r,$$

an equation which is close to the defining equation to what is know as the functional derivative of the functional T_o with respect to the density. There is only one slight catch. When defining a functional derivative, one must allow for completely arbitrary variations in the density $n(\boldsymbol{r})$ but, here, we are only allowed to consider density changes which conserve the number of electrons. For such density changes, we see that adding a constant to the potential $V(\boldsymbol{r})$ will produce no change in T_o because

$$\int \delta n(\boldsymbol{r})\,\mathrm{d}^3r = 0.$$

We are thus only allowed to infer

$$\frac{\delta T_o}{\delta n(\boldsymbol{r})} = -V(\boldsymbol{r}) + \mu_0$$

with some unknown constant μ_0.

We can carry our formal analysis one step further by also studying second-order changes in the kinetic energy as we change the density. As we shall see later, this will be of importance when we do linear response theory and gradient expansions. If we change the potential $V(\boldsymbol{r})$ by the amount $\delta V(\boldsymbol{r})$ the resulting density will change to $n(\boldsymbol{r})+\delta n(\boldsymbol{r})$. From the above so called Euler equation we obtain

$$\int \frac{\delta^2 T_o}{\delta n(\boldsymbol{r})\delta n(\boldsymbol{r}')}\delta n(\boldsymbol{r}')\,\mathrm{d}^3r' = -\delta V(\boldsymbol{r}) + \delta\mu_0.$$

On the other hand, as a matter of definition, the first-order change $\delta n(\boldsymbol{r})$ in the particle density resulting from a change $\delta V(\boldsymbol{r})$ in the total one-electron potential is proportional to this δV with a 'constant' of proportionality called the non-interacting density response function $\chi_o(\boldsymbol{r},\boldsymbol{r}')$ of the system, i.e.,

$$\delta n(\boldsymbol{r}) = \int \chi_o(\boldsymbol{r},\boldsymbol{r}')\delta V(\boldsymbol{r}')\,\mathrm{d}^3r'.$$

The non-interacting response function χ_0 is often also called the Lindhard [44] function and is easily obtained from standard second-order perturbation theory as described in most text books on quantum mechanics. Combining the last two equations leads to the conclusion

$$\frac{\delta^2 T_o}{\delta n(\boldsymbol{r})\delta n(\boldsymbol{r}')} = -\chi_0^{-1}(\boldsymbol{r},\boldsymbol{r}'),$$

meaning that the second functional derivative of the kinetic energy with respect to the density is the negative of the inverse of the non-interacting static density response function of the system. Notice that, in order to draw this conclusion, we have assumed that

$$\int \chi_0(\boldsymbol{r},\boldsymbol{r}')\,\mathrm{d}^3r' = 0$$

which allows us to ignore the constant $\delta\mu_0$. This relation is, however, valid in all finite systems on account of particle conservation. (In a metallic system, χ_0 integrates to a finite value actually proportional to the density of states at the Fermi level and one has to be a little more careful in handling the constant $\delta\mu_0$. The conclusion with regard to the second derivative of T_o remains, however, unaltered.)

We end the present subsection with the remark that T_o is obviously the kinetic energy of non-interacting electrons—despite the fact we started out searching for the minimum kinetic energy among arbitrary many-body wave-functions of density n! The underlying reason for this somewhat limited result is the one-body character of the operator for the kinetic energy being a sum of terms each containing the variables of only one of the electrons. As we shall see in the next subsection, this will not be true about the *many-body* functional $F[n]$ defined there.

3.2. *The functional F[n]*

An important ingredient in defining the functional T_o for the kinetic energy was the fact that the corresponding operator $\hat{T}$ is positive definite and thus has a spectrum bounded from below. The operator $\hat{T}+\hat{U}$, where $\hat{U}$ is the operator

$$\hat{U}=\tfrac{1}{2}\sum_{i\neq j}^{N} v(\boldsymbol{r}_i-\boldsymbol{r}_j)$$

corresponding to the total interaction energy between the electrons, has the same property, i.e., it is positive definite. (In the definition of $\hat{U}, v$ is the Coulomb interaction $v(r)=1/r$, as discussed in subsection 2.1). The operator $\hat{T}+\hat{U}$ defines the functional $F[n]$ of the density n in a way completely analogous to the definition of T_o above. We define

$$F[n]=\underset{|\Psi>\in M(n)}{\text{infimum}}\,\langle\Psi|\hat{T}+\hat{U}|\Psi\rangle$$

We point out that the operator $\hat{T}+\hat{U}$ contains terms involving the coordinates of two electrons, i.e., it is a 'two-electron' operator. We find $F[n]$ by a minimization of

$$\langle\Psi|\hat{T}+\hat{U}+\hat{V}|\Psi\rangle$$

with respect to arbitrary variations of the wave function or state $|\Psi\rangle$. Here, as above, the operator $\hat{V}$ is the one-body operator acting as external potential but corresponding to the Lagrangian multipliers needed to keep the density equal to the chosen one during the minimization. As above, this problem is identical to the quantum mechanical problem of finding the ground-state energy of the Hamiltonian $H=\hat{T}+\hat{U}+\hat{V}$. And, then, just as above, the potential $V(\boldsymbol{r})$ (= the Lagrangian multipliers) has to be adjusted such that the ground-state density of H becomes equal to the chosen density $n(\boldsymbol{r})$. And we will here assume without proof that this is always possible to achieve. We have thus established that the value of the functional $F[n]$ is just the sum of the kinetic and the interaction energies of a system of interacting electrons in an external one-body potential which gives the system a ground-state density equal to the chosen density n. If the ground-state energy of the Hamiltonian H is E, this minimal value obviously has to be

$$F[n]=E-\int V(\boldsymbol{r})n(\boldsymbol{r})\,\mathrm{d}^3r.$$

As we did in the case of the functional T_o above, we can also discuss how the functional $F[n]$ changes due to small changes $\delta n(\boldsymbol{r})$ in the density n. In order to have a slightly different density, the Lagrangian parameters have to change from V to $V+\delta V$, and first-order perturbation theory then says that the ground-state energy E of H changes by δE where

$$\delta E=\int n(\boldsymbol{r})\delta V(\boldsymbol{r})\,\mathrm{d}^3r.$$

Consequently, the corresponding change δF in F is given by

$$\delta F[n]=-\int V(\boldsymbol{r})\delta n(\boldsymbol{r})\,\mathrm{d}^3r$$

from which we, in analogy with above, conclude that

$$\frac{\delta F}{\delta n(\boldsymbol{r})}=-V(\boldsymbol{r})+\mu$$

with some unknown constant μ.

For the purpose of later doing response theory and gradient expansions we can, in analogy with the case of the kinetic energy, also conclude that

$$\frac{\delta^2 F}{\delta n(\boldsymbol{r})\delta n(\boldsymbol{r}')}=-\chi^{-1}(\boldsymbol{r},\boldsymbol{r}'),$$

where χ^{-1} is the inverse of the full static density response function of the interacting system. (Static because there is no time dependence in the perturbation).

3.3. *The minimum of the energy functional*

Having defined the functional $F[n]$ of the density n, it appears very natural to define the functional $E[n,w]$,

$$E[n,w]=F[n]+\int n(\boldsymbol{r})w(\boldsymbol{r})\,\mathrm{d}^3r,$$

of the two independent variables n and w. At a fixed 'potential' w, this functional has a minimum at the ground-state density n of that many-electron system which has w as external potential. And the value at the minimum is the ground-state energy of that system. In order to demonstrate this minimal property we let $|\Psi>$ be the ground state of the Hamiltonian $H=\hat{T}+\hat{U}+\hat{W}$ with the ground-state density $n=\langle\Psi|\hat{n}|\Psi\rangle$. We also introduce the ground state $|\Psi_1\rangle$ of the Hamiltonian $H=\hat{T}+\hat{U}+\hat{W}_1$ having a different ground-state density n_1. The two potentials w and w_1 must, obviously, differ by more than a constant in order for the two densities n and n_1 to be different. According to our definitions in the previous subsections and to the variational property of the ground state $|\Psi\rangle$, we have

$$\begin{aligned}E[n,w]&=F[n]+\int nw=\langle\Psi|\hat{T}+\hat{U}|\Psi\rangle+\int nw\\&=\langle\Psi|\hat{T}+\hat{U}+\hat{W}|\Psi\rangle\leq\langle\,\Psi_1|\hat{T}+\hat{U}+\hat{W}|\Psi_1\rangle\\&=F[n_1]+\int n_1 w=E[n_1,w]\end{aligned}$$

where we, for simplicity, have suppressed the integration variables.

For the first functional derivative of the functional $E[n,w]$ with respect to the density n at the ground-state density corresponding to the external potential w we obtain from above,

$$\left(\frac{\delta E}{\delta n(\boldsymbol{r})}\right)_w=\frac{\delta F}{\delta n(\boldsymbol{r})}+w(\boldsymbol{r})=\mu.$$

This is the 'quasi' stationary property of the total energy as a function of the density at a fixed external potential w. (The quantity which is strictly stationary is, obviously, $E-\mu N$, where N is the total number of electrons—but this amounts to the same thing when N is kept fixed). Notice that we are free to study arbitrary variations of the energy functional $E[n, w]$ with respect to both variables n and w. By choosing that density which renders the functional $E[n, w]$ stationary with respect to the density at a fixed external potential w, the variables n and w become coupled such that the density n becomes *the* ground-state density corresponding to the external potential w. We can then write $E[n] = E[n, w[n]]$ and we can consider the ground-state energy of any many-electron system as a functional of only the density n. This latter functional has, of course, in general, no stationary point at some density n.

3.4. *Kohn–Sham theory*

We are now in the position to appreciate the impact of these functionals of DFT on our previous discussion of the relevance of one-electron equations to the total energies of many-electron systems. Assume that $n(\mathbf{r})$ is the ground-state density of a many-electron system subject to the external potential $w(\mathbf{r})$. And let again $w_o(\mathbf{r})$ be that potential which produces the same density $n(\mathbf{r})$ in a non-interacting system. According to the discussion above, we have the following two expressions for the total energy E of the many-body system,

$$E = T_o[n] + \int n(\mathbf{r})w(\mathbf{r})\,\mathrm{d}^3r + \tfrac{1}{2}\int n(\mathbf{r})n(\mathbf{r}')v(\mathbf{r}-\mathbf{r}')\,\mathrm{d}^3r\mathrm{d}^3r' + E_{xc}[n]$$

where the exchange-correlation energy E_{xc} is defined in Subsection 2.3, and

$$E = F[n] + \int n(\mathbf{r})w(\mathbf{r})\,\mathrm{d}^3r.$$

Consequently,

$$F[n] = T_0[n] + \tfrac{1}{2}\int n(\mathbf{r})n(\mathbf{r}')v(\mathbf{r}-\mathbf{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r' + E_{xc}[n].$$

Notice that this equation clearly demonstrates E_{xc} to be a functional of the density n because all other terms are. Furthermore, for the functional derivatives of T_o and F with respect to the density taken at the ground-state density $n(\mathbf{r})$ we have,

$$\frac{\delta T_o}{\delta n(\mathbf{r})} = -w_o(\mathbf{r}) + \mu_o$$

and

$$\frac{\delta F}{\delta n(\mathbf{r})} = -w(\mathbf{r}) + \mu$$

Inserting these two equations into that obtained by functionally differentiating the functional $F[n]$ above, we have

$$w_o(\mathbf{r}) = w(\mathbf{r}) + V_H(\mathbf{r}) + v_{xc}(\mathbf{r}) + \delta\mu.$$

Here, the so called Hartree potential V_H is just the functional derivative of the classical Coulomb interaction energy,

$$V_H(\mathbf{r}) = \int v(\mathbf{r}-\mathbf{r}')n(\mathbf{r}')\,\mathrm{d}^3r'$$

and the so called exchange-correlation potential v_{xc} is the functional derivative of the exchange-correlation energy E_{xc} with respect to the density,

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}}{\delta n(\mathbf{r})}.$$

The quantity $\delta\mu$ is some yet undetermined constant of little relevance since the potential floor is of no consequence in non-relativistic physics. We have thus arrived at our desired recipe for constructing the potential $w_o(\mathbf{r})$ which produces the correct ground-state density of the fully interacting system in a completely non-interacting fashion.

From our exact expression for the exchange-correlation energy E_{xc} we obtain the following exact expression for the exchange-correlation potential $v_{xc}(\mathbf{r})$,

$$v_{xc}(\mathbf{r}) = \int n(\mathbf{r}')\{\tilde{g}(\mathbf{r},\mathbf{r}') - 1\}v(\mathbf{r}-\mathbf{r}')\,\mathrm{d}^3r' + \tfrac{1}{2}\int n(\mathbf{r})n(\mathbf{r}')\frac{\delta\tilde{g}(\mathbf{r},\mathbf{r}')}{\delta n(\mathbf{r})}v(\mathbf{r}-\mathbf{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r'.$$

The predominant way of implementing DFT is to construct some reasonable approximation to the exchange-correlation energy E_{xc} like, e.g., the LDA and then to construct $v_{xc}(\mathbf{r})$ by functionally differentiating the approximate E_{xc} with respect to the density. In the case of the LDA this gives

$$v_{xc}(\mathbf{r}) = \mu_{xc}(n(\mathbf{r}))$$

where $\mu_{xc}(n)$ is the exchange-correlation part of the chemical potential of the homogeneous electron gas at density n. It actually has certain computational advantages to have the exchange-correlation energy and potential being consistent in this way. But there are also many drawbacks as will be discussed later. For instance, in the case of several accurate gradient approximations, the potential produced by functional differentiation is often singular and can actually drive the density away from the correct result. In such cases we have, for many years, advocated the practice of doing separate approximations for the potential v_{xc} directly from the exact expression above. This practice is strongly supported by the variational principle discussed in Section 3.3.

In modern work, however, a different exact equation, referred to as the Sham–Schlüter equation [45], forms the starting point for the construction of accurate exchange-correlation potentials v_{xc}.

3.5. *The gradient expansion*

Our previous discussion of the total energy of the many-electron systems has led us to the point where all remaining difficulties are deferred to finding reasonable approximations to the exchange-correlation energy E_{xc} of the system. And we have already introduced the LDA which goes a long way toward this end. By construction, the LDA is correct for the homogeneous electron gas and accurate for densities that vary little over distances smaller than an inverse Fermi wave vector. A natural step beyond the LDA would, therefore, be to allow the energy functional to depend not only on the local value of the density but also on its different local gradients. By adding successively higher-order gradients, the gradient expansion is supposed to correct the results of the LDA when the density varies more and more rapidly. One should, however, not have to assume that the deviations from a constant density are also small. But the gradient expansion should, of course, be valid also in that case and it is actually easier to determine the different gradient corrections in this regime. Thus, we will here apply response theory to the homogeneous electron gas in order to see how such gradient corrections can be obtained.

We expose the gas to a weak and slowly varying external perturbation and expand the exchange-correlation energy E_{xc} in a Taylor series in the deviation $\delta n(\boldsymbol{r})$ from the homogeneous density n_o,

$$E_{xc}[n] = E_{xc}[n_o] + \int v_{xc}(n_o)\delta n + \tfrac{1}{2}\int K_{xc}\ \delta n\ \delta n' + \tfrac{1}{6}\int L_{xc}\delta n\ \delta n'\delta n'' + \dots .$$

Here, we have suppressed integration variables, and

$$K_{xc}(\boldsymbol{r},\boldsymbol{r}') = \frac{\delta^2 E_{xc}}{\delta n(\boldsymbol{r})\ \delta n(\boldsymbol{r}')},$$

and

$$L_{xc}(\boldsymbol{r},\boldsymbol{r}',\boldsymbol{r}'') = \frac{\delta^3 E_{xc}}{\delta n(\boldsymbol{r})\ \delta n(\boldsymbol{r}')\ \delta n(\boldsymbol{r}'')}.$$

Due to the full translational symmetry of the electron gas the potential v_{xc} is a constant, K_{xc} can only depend on the distance between $\boldsymbol{r}$ and $\boldsymbol{r}'$, and L_{xc} must be invariant under the operation of displacing all arguments by the same amount. Since we are considering particle conserving perturbations the linear term in the expansion vanishes. Furthermore, we are interested in the corrections to the LDA and thus expand the LDA result in a similar manner. We obtain

$$E_{xc}^{LD}[n] = \int n_o\varepsilon_{xc}(n_o) + \int \frac{\partial(n\varepsilon_{xc})}{\partial n_o}\delta n + \tfrac{1}{2}\int \frac{\partial^2(n\varepsilon_{xc})}{\partial n_o^2}(\delta n)^2 + \tfrac{1}{6}\int \frac{\partial^3(n\varepsilon_{xc})}{\partial n_o^3}(\delta n)^3 + \cdots$$

Our purpose here, is to illustrate the origin of gradient terms rather than to develop a full fledged gradient expansion. Consequently, in what follows, we will, for simplicity, restrict the analysis to linear response, i.e., to terms of second order in the deviations from the homogeneous limit. Subtracting the LDA result from the exact one and noting that $\mu_{xc} = \partial(n\varepsilon_{xc})/\partial n$ gives the following second-order correction to the LDA,

$$E_{xc}[n] = E_{xc}^{LD}[n] + \tfrac{1}{2}\int\int \delta n\{K_{xc} - \mu'_{xc}\cdot\delta(\boldsymbol{r}-\boldsymbol{r}')\}\delta n'.$$

In Fourier space this becomes

$$E_{xc}[n] = E_{xc}^{LD}[n] + \tfrac{1}{2}\int \frac{\mathrm{d}^3q}{(2\pi)^3}\{K_{xc}(\boldsymbol{q}) - K_{xc}(0)\}|\delta n_{\boldsymbol{q}}|^2$$

where we have used the exact relation $K_{xc}(0) = \partial\mu_{xc}/\partial n$ which follows from the compressibility sum rule of the electron gas [46]. If we now make the crucial assumption that the perturbed density has appreciable Fourier components only at small q, we can approximate K_{xc} by its small-q Taylor series

$$K_{xc}(q) = K_{xc}(0) + 2B_{xc}(n_o)q^2 + \cdots$$

and going back to real space we obtain

$$E_{xc}[n] = E_{xc}^{LD}[n] + \int B_{xc}(n_o)|\nabla n|^2\ \mathrm{d}^3r.$$

This gives the lowest-order gradient correction to the LDA. Of course, in a strongly inhomogeneous system like an atom or a molecule, we might ask ourselves what number (n_o) to use for the homogeneous density. It can be shown that it is quite appropriate to again use the local density $n(\boldsymbol{r})$ rather than some average density for n_o. In fact, this procedure eliminates the necessity to include many of the higher-order gradients from higher-order response functions. But the validity of this procedure as well as of the entire gradient expansion depends in a crucial way on satisfying certain consistency relations between response functions of different order. For instance, one such requirement is the exact relation [47]

$$\frac{\partial K_{xc}(\boldsymbol{q})}{\partial n} = L_{xc}(\boldsymbol{q},\boldsymbol{q}).$$

We have so far said little about how to calculate the different so called exchange-correlation kernels like K_{xc} and L_{xc} on which the gradient expansion rests. But in previous sections we have laid down the machinery for determining these kernels. For instance, taking the second functional derivative with respect to the density of the relation connecting the functional $F[n]$ to the functional $E_{xc}[n]$ for the exchange-correlation energy and inserting our results for those derivatives gives

$$-\chi^{-1} = -\chi_o^{-1} + v + K_{xc}.$$

This very important relation can be used in either of two ways.

(*i*) From, e.g., many-body perturbation theory we might be able to calculate an accurate static density response function χ of the electron gas. Then, the above relation determines K_{xc} and its small-q Taylor expansion in reciprocal space and thus a whole set of gradient coefficients according to the above analysis. Such calculations were done as early as 1970 by Geldart and Taylor [48] and repeated more accurately by Engel and Vosko in 1990 [49] but including only effects of the inter-electronic Coulomb interaction to lowest order. Including also correlation effects to the level of the so called random-phase approximation (RPA) the gradient coefficient $B_{xc}(n)$ has been calculated by Ma and Brueckner [50] in the high-density limit and later by Geldart and Rasolt [51] at metallic densities. Rasolt calculated the same coefficient for spin-polarized systems in 1977 [52].

(*ii*) Having some accurate expression for the exchange-correlation energy $E_{xc}[n]$ of the inhomogeneous system, we can form the second derivative with respect to the density to obtain the kernel K_{xc} and thus the full density response function χ of the system. This way of using the equation for the response function demonstrates the important fact that the exact static density response function of any electronic system can, in principle, always be obtained from DFT.

With regard to the second-order response function L_{xc} which gives rise to a gradient correction of the form

$$\Delta E_{xc}[n] = \int C_{xc}(n(\boldsymbol{r})) \frac{|\nabla n(\boldsymbol{r})|^2 \nabla^2 n}{|n(\boldsymbol{r})|^3} \, \mathrm{d}^3 r,$$

much less work has been done. A student of mine calculated the coefficient C_x including only exchange effects [47] and, to my knowledge, nothing is known about the effects of correlations on this coefficient. Our coefficient was later incorporated in a *Meta-GGA* by Perdew and collaborators [53]. This particular GGA was more successful than others and it appeared to have a better ability to distinguish between single, double, and triple bonds in molecules (see Section 5.6).

In cases when the density varies more rapidly and does have appreciable amplitude at larger $\boldsymbol{q}$-vectors, one cannot use the Taylor expansions of the response functions around $q = 0$. The derivation above suggests, however, that it is not at all necessary to use a gradient approximation. Using the fact that

$$\int \{K_{xc}(\boldsymbol{r} - \boldsymbol{r}'; n_o) - \mu'_{xc} \cdot \delta(\boldsymbol{r} - \boldsymbol{r}')\} \, \mathrm{d}^3 r' = 0,$$

the second-order correction to E_{xc}^{LD} can readily be rearranged to read ([4])

$$E_{xc}[n] = E_{xc}^{LD}[n] - \frac{1}{4} \int K_{xc}(\boldsymbol{r} - \boldsymbol{r}'; n_o) \{n(\boldsymbol{r}) - n(\boldsymbol{r}')\}^2 \, \mathrm{d}^3 r \, \mathrm{d}^3 r'.$$

This correction to the LDA is obviously valid for arbitrarily rapid but small density variations. It represents an infinite summation of the gradient terms. When, however, we want to apply this correction to a strongly inhomogeneous system like, e.g., an atom, we are forced to decide what to use as an average density n_o in the kernel K_{xc}. In a solid we could, e.g., use the average density of the valence electrons. In an atom it becomes more natural to think of a local density but, since the correction involves two points in space ($\boldsymbol{r}$ and $\boldsymbol{r}'$), there are a number of possible choices, e.g., $n((\boldsymbol{r} + \boldsymbol{r}')/2)$ or $(n(\boldsymbol{r}) + n(\boldsymbol{r}'))/2$. Gunnarsson *et al.* [54] have shown that the first choice gives an infinite correction for an atom whereas the second choice gives a reasonable result. These conflicting results of two similar and seemingly sound approximations could be viewed as a complete breakdown of gradient expansions. The origin of the sensitivity to the choice of average density is the use of linear response theory underlying the entire discussion above. Thus, our conflicting results is merely a reflection of the fact that an atom or, for that matter, a solid is *not* a linear perturbation of the homogeneous electron gas. As we shall see shortly, things are not as bad as they seem. Gradient expansions are not convergent in the sense that successively more accurate results can be obtained by adding more terms. They represent asymptotic expansions, meaning that an approximation with a few gradient terms added will be more accurate than the LDA in cases with relatively slow density variations. When the density varies more rapidly, we might be better off without gradient terms. As we shall see, the criterion for the validity of the LDA is not as strict as the corresponding criterion for the gradient expansion. Over the past fifteen years, people have tried to circumvent this problem by applying different cut-off procedures designed to retain the advantages of gradient corrections for moderate density variations without displaying the breakdown characteristic of the straightforward gradient expansion in cases with rapid density variations.

4. The performance of the LDA

4.1. *The realm of the LDA*

The extreme usefulness of the LDA has been common knowledge for many years. We will here give a short resume of results obtained from the LDA for different physical properties and state some general conclusions concerning the quality of these results.

We will, however, first remind the reader that static DF theory, with a few exceptions, is strictly applicable only to ground-state properties and, secondly, we will say a few words about to what systems the theory should be applied.

As a test of the quality of new approximations to E_{xc}, DF theory is often applied to atoms and small molecules. With regard to this practice we offer the following remarks:

(*i*) These systems are, for their physical properties, best treated by other many-body techniques such as Many-Body-Perturbation Theory (MBPT) or Configuration-Interaction expansions (Cl) which offer a well defined route to successively more accurate answers.

(*ii*) The interior of atoms is extremely difficult to handle by DF techniques due to the rapid density variations in these regions. Fortunately, the physics of these regions have little to do with the physics of everyday life (chemical bonding etc.).

(*iii*) The almost non-vanishing densities in the exterior of atoms cause gradient corrections to diverge and approximate *xc*-holes to be misplaced and deformed but contain little energy. More importantly—such regions do exist only in very sparse solids. Consequently, atoms and small molecules represent very severe tests on the quality of different approximations and care must be exercised in order not to have the conclusions obscured by irrelevant difficulties.

The application of Cl expansions or MBPT to complicated systems of considerable interest such as solids with many atoms per unit cell, surfaces, atoms or molecules adsorbed on surfaces, and large molecules, is computationally prohibitive. This is the realm of DF theory provided it can be made accurate enough.

4.2. *Results*

The following pages with tables I–X demonstrate the successes of the Local-Density Approximation (LDA) as applied to a variety of physical properties of solids. These pages also indicate some of the deficiencies of the LDA. Some of the results have been taken from the literature and in most cases I have given the source. Other results have been produced in the course of my own research.

4.3. *Successes*

The most striking feature discernible in the data shown below, is the remarkable accuracy of the LDA as applied to a variety of physical properties in many different systems. We will here summarize the existing experience from numerous applications of the LDA as follows:

- Binding energies are often better than 1 eV but in some s–d bonded systems the error can be twice or even three times as large. There is a systematic over binding.
- Equilibrium distances are generally accurate to within 0.1 Å They are systematically too short.
- Vibrational frequencies are accurate to within 10–20%. There exist occasional cases with larger errors.
- Charge densities are better than 2%.
- Geometries are accurate.
- LDA results are nearly always much better than those of the Hartree–Fock (HF) approximation.
- Most importantly, physical trends are generally correct.

We end this subsection by stressing that whereas the errors in the binding energies of the LDA can be large, the binding distances are usually very accurate. Thus, the energy surfaces of the LDA are rather parallel to the true ones.

Table I. Total energies (in eV) of a few atoms.

Atom	LDA	exp
H	13.3	13.6
H^-	14.4	14.4
Al	6567	6592

4.4. *Deficiencies*

The results shown above, however, also indicate some cumbersome deficiencies. Most notable is the systematic overbinding predicted by the LDA, particularly for the s-d bonded systems. The overbinding is also, although to a much lesser extent, reflected in a small but relatively systematic underestimate of the bonding distances. We end this section with a small list of systems for which the LDA predicts the wrong ground states.

- The transition-metal oxides FeO and CoO are erroneously predicted to be metallic. But, MnO and NiO come out as anti-ferromagnetic insulators in accordance with experiment [71].
- Solid Fe is predicted to be an fcc paramagnet [72] but is a bcc ferromagnet at low temperatures.
- In many semiconductors, the LDA gives the metal-insulator transition at much too large volumes. [73]
- The LDA predicts the wrong dissociation limits for a large number of molecules. [29,35,37]
- The LDA predicts incorrect ground states for many atoms. [29,32,37]
- The LDA gives unstable negative atomic ions in many cases when these are stable. [32]
- Et cetera, et cetera..

4.5. *s-p Transfer energies*

Gunnarsson and Jones [74] suggest that the root of the problem is the inability of the LDA to properly account for so called s-to-p and s-to-d transfer energies. When a solid or molecule is formed even from rather "round" atoms the electronic density stretches out toward neighboring atoms to form bonds. In one-electron calculations the stretching is accomplished by transferring electrons from states of lower angular momentum to states with higher angular momentum (*e.g.* s-d) in the bonding process. The physical picture is illustrated, *e.g.*, by the case of Si. The Si atom has four valence electrons $3s^2, 3p^2$. In forming the solid we would, to a first approximation, promote one of the s electrons to a p orbital and then form the tetrahedrally arranged hybrids. Thus, the atom in the solid has the approximate configuration $3s, 3p^3$ and bonding has resulted in a transfer of one electron from an s- to a p-

Table II. Ground-state properties of the molecules H_2O, NH_3, and CO_2, as obtained from the LDA and from experiment. The results are taken from the work by Müller, Jones, and Harris [55]. We assume that the numerical errors involved in obtaining the LDA results are negligible in comparison to the deviation between theory and experiment. The equilibrium distances in Table II are probably exceptions to our assumption because they do not conform to the general expectations of bond distances being too short within the LDA.

	H_2O		NH_3		CO_2	
	LDA	exp	LDA	exp	LDA	exp
d	1.84	1.81	1.94	1.91	2.21	2.20
θ	106	105	108	107	180	180
ω_s	3680	3657	3335	3337	1420	1388
ω_b	1590	1595	820	950	730	667
μ	0.732	0.730	0.564	0.583	0	0

d is the equilibrium distance in atomic units.
θ is the equilibrium bond angle in degrees.
ω_s is the stretching frequency in cm^{-1}.
ω_b is the bending frequency in cm^{-1}.
μ is the dipole moment in atomic units.

Table III. Ionization potentials (in eV) of a few atoms. LPM designates results from the gradient corrected scheme by Langreth, Perdew and Mehl [8,9] and Hu and Langreth [56]. In the cases of Mn and Fe we refer to the removal energy of a 4s electron.

Atom	exp	LDA	LPM
He	24.6	24.2	24.8
Li	5.4	5.6	5.6
Be	9.3	8.4	9.0
Na	5.1	5.3	5.1
Ca	6.1	6.3	6.0
Mn	7.4	7.9	7.1
Fe	7.9	8.4	7.8

Table IV. Cohesive energies (in eV) of a few solids as obtained from the LDA and from experiment. The Si result is from Ref. [57]. The other results are taken from the book by Moruzzi et al. *[7] and from Refs. [58–60].*

	Na	Mg	Si	Ti	Zr	Ni
exp	1.1	1.5	4.6	4.9	6.3	4.4
LDA	1.1	1.6	5.1	6.1	6.8	5.7

Table V. Lattice parameters in atomic units for a few solids as obtained from the LDA and from experiment. With one exception, Si, the data are taken from the book by Moruzzi et al. *[7]. The Si result is from Ref. [57].*

	Na	Mg	Si	Ti	Zr	Ni
exp	8.0	8.5	10.3	7.8	8.2	6.7
LDA	7.7	8.4	10.2	7.6	8.2	6.6

Table VI. The binding energies (in eV) of the first-row dimers as obtained from the LDA, from the LPM scheme [8,9], from the GGA according to PBE [39,61,62], and from experiment. All results (except PBE) are taken from the work by Becke [63,64].

	H_2	Li_2	B_2	C_2	N_2	O_2	F_2
exp	4.8	1.1	3.0	6.3	9.9	5.2	1.7
LDA	4.9	1.0	3.9	7.3	11.6	7.6	3.4
LPM	5.0	0.6	3.3	6.1	10.2	6.4	2.4
PBE	4.5	0.9	—	—	10.5	6.2	2.3

Table VII. Equilibrium distances in Bohr of the first-row dimers. The LDA results are from the work by Becke [63].

	H_2	Li_2	B_2	C_2	N_2	O_2	F_2
exp	1.40	5.05	3.00	2.35	2.07	2.28	2.68
LDA	1.45	5.12	3.03	2.35	2.07	2.27	2.61

orbital. This costs energy but, in the LDA, we pay too small a price for the transfer resulting, on an absolute scale, in too low an energy for the solid. Consequently, the LDA overestimates the cohesive energy of Si and in this case the actual error is 0.5 eV or 11% (See Table IV). A similar situation exists in the transition metals although the transfer in that case is from s, p to d [7]. According to Gunnarsson and Jones [74] the errors in the s-p or s-d transfer energies are due to the insensitivity of the LDA to the nodal structure of the one-electron wave functions. Consider *e.g.* the fluorine atom which has a ^{2}P ground state constructed from the configuration $2s_{\Uparrow}2s_{\downarrow}2p^3_{\Uparrow}2p^2_{\downarrow}$. An s-p transfer is achieved by promoting one of the 2s electrons to the empty hole in the 2p-shell resulting in a ^{2}S excited state. From HF theory, the accompanying change in the exchange energy of the L-shell is easily found to be

$$\tfrac{2}{3}G_1(s,p) - \tfrac{9}{25}F_2(p,p) + \tfrac{1}{2}[F_o(s,s) - F_o(p,p)]$$

in terms of the usual Slater [75] integrals. Inserting the actual orbitals of F gives a 6.7 eV increase in the exchange energy. This rather large value is associated with the extra angular node of the 2p- relative to the 2s-orbital. In the LDA the exchange energy is given by

$$E_x^{LD}[n_{\Uparrow}, n_{\downarrow}] = A_x \sum_{\sigma} \int [n_\sigma(\mathbf{r})]^{4/3}\, d^3r$$

where $n_\uparrow$ and $n_\downarrow$ are the spin-up and spin-down densities and where A_x is equal to $-(3/4)(6/\pi)^{1/3}$ in Hartrees. Thus, a substantial change in the exchange energy associated with the s-p transfer requires the square of the s- and p-orbitals to be rather different. This is not the case in the F atom because both orbitals belong to the same principal-quantum number shell (the L-shell). Consequently, the cost of the transfer is much too small in the LDA (~0.6 eV). Although the total error is considerably reduced [32] (to ~2.6 eV) by correlation effects and self-consistency, it is still large enough to explain a large part of the error (1.7 eV from Table VI) in the binding energy of the F_2 molecule.

Clearly, the LDA in which the exchange-correlation energy only depends on the local density cannot account for changes in the nodal structure of the wave functions. One could, however, hope that xc-functionals based on gradients would be capable of picking up such nodal dependencies. Unfortunately, this does not seem to be the case as can be seen in Table XI showing results for the aforementioned s-p transfer treated by different methods. The Generalized Gradient Approximation (GGA) by Langreth, Perdew, Mehl and Hu (LPM) [8,9,56] in conjunction with spherically averaged densities perform almost as badly as the LDA.

One could guess that this disappointing result might be a consequence of the spherical averaging, a procedure which certainly softens the nodal structure closely connected to the angular dependence. Therefore, it is even more disappointing to see the results marked LPM-NS in Table XI. They are obtained from the LPM scheme by taking full account of the asphericities of the spin-densities. The results are marginally better than those of the spherical approximation and we conclude that a functional constructed from the density and its first spatial derivatives is not able to respond to a change in the nodal structure of the wave functions. Indeed, s-d transfer energies have been calculated in transition-metal atoms by Kutzler and Painter [76] with the same disappointing outcome. They also tested an improved generalized-gradient approximation by Perdew and Wang [77,78] to be discussed in Section 5.4 and

Table VIII. Vibrational frequencies in cm^{-1} of the first- and second-row dimers as obtained from the LDA and from experiment. The results are taken from the work by Becke [63].

	H_2	Li_2	B_2	C_2	N_2	O_2	F_2
exp	4400	350	1050	1860	2360	1580	890
LDA	4190	330	1030	1880	2380	1620	1060

Table IX. Heat of formation (in eV) of a few compounds as obtained from the LDA and from experiment. The data are taken from the work by Williams, Kübler, and Gelatt [65,66].

	MgAg	AlZr	SiZr	NiAl	CuZn
exp	0.19	0.44	0.81	0.61	0.12
LDA	0.24	0.44	0.73	0.74	0.14

Table X. a The saturated magnetic moment (Bohr magnetons), b the hyperfine field (kiloGauss), and c the spin-susceptibility enhancement factor for a few metals. The data are taken from the work by Janak et al. *[67–70].*

	X.a Magnetic moment Fe	Co	Ni
exp	2.22	1.56	0.61 0.59
LDA	2.15	1.56	
	X.b Hyperfine field Fe	Co	Ni
exp	339	217	75 80
LDA	260	220	
	X.c Susceptibility enhancement Li	Na	K
exp	2.50	1.65	1.70
LDA	2.25	1.71	1.95

Table XI. Exchange energies ($-E_x$, in eV) in two different configurations of the fluorine atom as obtained from Hartree–Fock (HF), from the LDA, and from the LPM scheme (Ref. [9]) evaluated at spherically averaged densities. LPM-NS designates LPM exchange energies evaluated at the correct non-spherical densities. The numbers illustrate the failure of all DF approximations in describing the correct (HF) s-to-p transfer energy.

	$2s_{\Uparrow}2s_{\downarrow}2p^3_{\Uparrow}2p^2_{\downarrow}$	$2s_{\Uparrow}2p^3_{\Uparrow}2p^3_{\downarrow}$	ΔE_x
HF	272.5	264.4	8.1
LDA	246.4	244.1	2.2
LPM	267.4	264.6	2.8
LPM-NS	267.6	264.6	3.0

there referred to as the PW86 scheme. Although this scheme is somewhat better than both the LDA and the LPM scheme, the improvement is marginal. (In this case the LDA is actually better than the LPM scheme.)

The s-d transfer errors in atoms is one possible mechanism responsible for the cohesive energy errors of extended systems. Correlations, however, strongly reduce the effect already in atoms and screening as well as smoother densities in the solids will reduce the effect even further. As we saw above, two proposed generalized-gradient approximations which are known to be superior to the LDA with regard to binding energies of molecules and solids (see Sections 5.3 to 5.6) fail to improve the s-d transfer energies. From this we conclude that the s-p or s-d transfer mechanism is probably not the most important mechanism behind the failure of the LDA with regard to cohesive energies.

4.6. *The near-degeneracy problem*

Cl and MBPT dominate the world of Quantum Chemistry (QC), i.e. the physics of small to large molecules. These methods are reliable and can often provide the 1 kcal/mol (1 eV = 23 kcal/mol) accuracy of interest to the physics of chemical reactions [12]. They are, however, enormously time-consuming and there would be a breakthrough in QC if DF methods could be brought to chemical accuracy. It is therefore rather disappointing to see the LD results for the binding energies of the first row dimers in Table VI. The much larger binding-energy errors in these finite systems can be understood in terms of their more rapidly varying densities. Clearly, the explanations based on the s-p transfer mechanism becomes more relevant but also other effects come into play. One such effect can be referred to as the near-degeneracy problem in the DF theory. Consider the H_2^+-molecule at large internuclear separation R. The correct wave function is close to

$$\psi(\mathbf{r}) = \sqrt{\tfrac{1}{2}}\{\phi(\mathbf{r}) + \phi(\mathbf{r}-\mathbf{R})\} = \sqrt{\tfrac{1}{2}}\{\phi_a + \phi_b\}$$

where $\phi(\mathbf{r})$ is the wave function of hydrogen: $\phi = e^{-r}/\sqrt{\pi}$. The expectation value of the full Hamiltonian

$$H = -\frac{1}{2}\nabla^2 - \frac{1}{r_a} - \frac{1}{r_b} + \frac{1}{R}$$

with respect to this wave function is, to exponential accuracy, equal to $\langle\phi|H|\phi\rangle = -1\,\mathrm{Ry}$. Thus, in real life, the system cares very little about whether we break the symmetry of the molecule and put the electron on one of the sites or we keep the symmetry and put half an electron on each site—the corresponding energies are very much the same. In all approximations to DF theory which are constructed by using only the local density and its gradients, however, the situation is quite different. In addition to the energy terms considered above, the energy of the effective one-electron Hamiltonian of DF theory contains the terms

$$\tfrac{1}{2}\int \psi^2\ v\ \psi^2 + E_{xc}[\psi^2].$$

If we have found a functional which is very accurate for the atom and which achieves a high degree of cancellation of the so called self-interaction, this becomes

$$\tfrac{1}{2}\int \phi^2 \ v \ \phi^2 + E_{xc}[\phi^2] = U - J \approx 0$$

in the hydrogen-plus-proton case. Thus, for this case, the LDA provides a good answer. In the symmetric case, we instead obtain

$$2\tfrac{1}{2}\tfrac{1}{4}\int \phi^2 v \phi^2 + 2\tfrac{1}{2}\tfrac{1}{4}\int \phi_a^2 v \phi_b^2 + E_{xc}\big[\tfrac{1}{2}\big(\phi_a^2 + \phi_b^2 + 2\phi_a\phi_b\big)\big]$$
$$\approx \tfrac{1}{2}U + \frac{1}{4R} + 2E_{xc}\big[\tfrac{1}{2}\phi^2\big]$$

i.e., a quantity with a slow variation with R. This result is fundamentally incorrect, thus demonstrating that the correct functional must depend on the density in a more complicated way than just through its local value and its gradients. If we, *e.g.*, use the LDX approximation (Local-Density-eXchange-only) given in Section 4.5, we obtain an LD error of

$$\Delta E^{LD} = \tfrac{1}{2}U + \frac{1}{4R} - 2^{-1/3}J \approx 0.25/R - 0.3U,$$

and, since U is $\sim 8\,\mathrm{eV}$ for hydrogen, we obtain $\sim 2.4\,\mathrm{eV}$ overbinding in the LDA at large separation. Even though the other spurious term $(1/4R)$ partly reduces the error at smaller R, an effect might still remain at the equilibrium distance. (As a matter of fact, the LDA error in the binding energy of H_2^+ is only 0.15 eV at equilibrium [79].) Notice, however, that an essential ingredient of the mechanism discussed above is the existence of an unpaired electron. Consequently, the effect does not affect the binding energy of the dimers but it does affect their ionization potentials and also the binding energies of the dimer cations as demonstrated by Merkle *et al.* [80].

The accurate results achieved by the LDA for the 'billiard-ball' systems Be_2 (error $\sim 0.46\,\mathrm{eV}$) [64] and Mg_2 (error $\sim 0.12\,\mathrm{eV}$) [81], and the difficulty for the LDA to provide good answers in cases with near degeneracies is opposite to the situation encountered with Cl methods. The so called dynamical correlations responsible for, *e.g.*, binding in Be_2 require a very large number of configurations for their accurate description whereas the near-degeneracy cases are often taken care of by a few configurations. This immediately suggests a hybrid method [82] in which a local-density correction is added to a small Cl calculation.

The discussion above illuminates another peculiar feature of the LDA and approximations based on density gradients. Even though the LDA does not work well for a system with some particular symmetry, one can often find a "neighboring" system with nearly the same energy (*e.g.* a symmetry broken solution) for which the LDA provides an accurate answer. For example, the LDA gives a reasonable description of the H_2^+ molecule at large nuclear separation if we break the symmetry and put one electron on one of the protons. A more interesting example is furnished by the Cr_2 molecule whose bond length is accurately given by the LDA provided the molecule is assumed to be anti-ferromagnetic, *i.e.*, symmetry broken [83]. The binding energy is reasonable but too large as usual. In this case, one of the largest and most sophisticated Cl calculations is inferior to the LDA [84].

5. Gradient approximations

As we saw in the previous paragraph, the LDA is often quite adequate while, in other cases, a higher accuracy is desired. An error in a binding energy of the order of 20% or 1 eV is, *e.g.*, not acceptable in the study of chemical reactions. In this field we would like binding-energy errors to be of the order 0.05 eV or less. The simplicity of DF methods as compared to traditional many-body techniques has, however, spawned considerable efforts to improve on the LDA. It would, *e.g.*, mean a breakthrough in quantum chemistry if DF methods could be brought to the same level of accuracy as large-scale configuration-interaction (Cl) calculations for larger molecules. One could then confidently go on to treat very large molecules as well as extended systems which presently are beyond our computational capabilities.

Attempts to go beyond the LDA are based either on an improved description of the exchange-correlation hole (see Sections 2.1 and 2.3) in real space or on a description of exchange-correlation energies in reciprocal space usually leading to so-called generalized gradient approximations (GGA). Some times a mixture of the two approaches have been considered [77,78]. In recent years, the largest effort has gone into the reciprocal-space approach which so far has been the most successful *ab initio* DF method. As we shall see later, *ab initio* density functionals are still far from the 0.05-eV goal mentioned above but the results of new semi-phenomenological functionals are not too far away.

5.1. *Straightforward gradients*

As discussed in Section 3.5, the natural extension of the LDA is a straightforward gradient expansion. The starting point is then the homogeneous electron gas and the gradient terms and their coefficients are determined by the small wave-vector expansions of the density response functions of different orders pertinent to the gas. Due to the translational as well as the rotational invariance of the gas, one can show [47] that only rather few gradient terms exist in each order. For instance, all terms up to and including fourth-order gradients can be summarized as

$$E_{xc}[n] = E_{xc}^{LDA}[n] + \int B_{xc}^{(2)}(n(\boldsymbol{r}))\frac{|\nabla n(\boldsymbol{r})|^2}{|n(\boldsymbol{r})|^{4/3}}\,\mathrm{d}^3r$$
$$+ \int B_{xc}^{(4)}(n(\boldsymbol{r}))\frac{|\nabla^2 n(\boldsymbol{r})|^2}{|n(\boldsymbol{r})|^2}\,\mathrm{d}^3r$$
$$+ \int C_{xc}(n(\boldsymbol{r}))\frac{|\nabla n(\boldsymbol{r})|^2\nabla^2 n(\boldsymbol{r})}{|n(\boldsymbol{r})|^3}\,\mathrm{d}^3r$$
$$+ \int D_{xc}(n(\boldsymbol{r}))\frac{|\nabla n(\boldsymbol{r})|^4}{|n(\boldsymbol{r})|^4}\,\mathrm{d}^3r.$$

It is not difficult to realize that the terms containing the coefficients labeled B originate in linear response, the C-term comes from second-order response, and that involving D_{xc} from third-order response. The significant feature determining the origin is the total power of those densities subject to differentiation. The coefficient $B_{xc}^{(2)}$ has been calculated in an approximation involving the effects of both exchange and correlation whereas the coefficients $B_{xc}^{(4)}$ and C_{xc} are known in the exchange-only approximation. The coefficient D_{xc} is, however, yet unknown.

It is not hard to see that this kind of gradient expansion can never be applied to a finite system. Only the lowest-order gradient term will give a finite result for the exchange-correlation energy. The integrands of all the fourth-order terms will tend to some constant in the regions of exponentially decaying densities in the outskirts of any finite system, thus producing an infinite result in the subsequent integration. There will also be problems at the nucleus particularly from the term containing $\nabla^2 n$. Because of the Coulomb potential from the nucleus, the density will have a so called cusp at the nucleus. This means that the density has a term linear in the distance to the nucleus and that will cause a singularity of the form const/distance in the gradient part of the energy density. Consequently, the gradient correction will act as an additional contribution to the nuclear charge, a result which is definitely wrong from a physical point of view.

In a solid treated with pseudo potentials both the ‘surface’ problem and the ‘core’ problem go away and there is hope for obtaining accurate total energies from the gradients. Such tests have been carried out for different model solids treated with pseudo potentials and including only exchange energies. Of course, in the case of exchange only, comparisons with exact results can be made which allows for an precise assessment of the quality of a straightforward gradient expansion. And the gradient expansion of the correlation energy is not expected to behave in vary a different way in this regard. These test demonstrate the extraordinary quality of the total energies of ‘pseudo’ solids obtainable from such expansions. This can be seen in Table XII which displays only the smaller exchange part of the energies of a number of metals. The errors are of the order of 1 Milli-Rydberg or less and much smaller than those resulting from the use of common GGA:s to be discussed later. In Table XII, the quantities labeled ε_x^{LDX}, ε_x^{GE1}, and ε_x^{GE2} refer to the contribution from the LDA in the exchange only version and the contributions including linear and second order responses—respectively.

The significant features of all the systems presented in Table XII is that they are metals and there are no band gaps. Unfortunately, the very positive impression conveyed by Table XII is drastically altered when we turn to systems with band gaps. This is demonstrated in Table XIII displaying similar data as in Table XII but now for systems all having band gaps. In the first case labeled Si, we have to the homogeneous gas added a pseudo potential which results in a band structure very close to that of real bulk silicon [85]. The case labeled SiX is similar to Silicon but the pseudo potential has been scaled up resulting in a larger band gap. In these two examples the density has non-negligible Fourier components also on multiples of the fundamental harmonics of the lattice potential (V_{111}, V_{220}, V_{311}) and it could be argued that gradients also from third- and higher order response theory should have been included in the gradient expansion. This criticism can, however, not be raised against the two remaining cases labeled M1 and M2 in which care has been taken to avoid larger amplitudes on any of the more rapidly varying density components. Thus, these cases have slowly varying densities but still possess band gaps. And in all four cases the energy errors are an order of magnitude larger as compared to the metallic cases and actually worse than those of the GGA:s as exemplified by that of Becke [11]. The conclusion is now inescapable. The straightforward gradient expansion breaks down as soon as a band gap appears. This should, however, not come as a surprise. The entire gradient expansion originates in response theory, i.e., in perturbation theory which is known to break down in the presence of band gaps [86]. The appearance of a band gap is actually an extremely non-local property of the system and can never be predicted from a knowledge of the density or its gradients in one point in space. In my opinion, even a GGA can never be made very accurate and independent on the system to which it is applied, unless the information about the presence of a band gap is, somehow, fed into the construction of the GGA. It is thus more surprising that the GGA works so well for the systems with band gaps in Table XIII.

Table XII. Calculated exchange energies in Rydberg per electron for different values of the effective potential V. V$_{111}$ *(=* V$_{200}$*) in units of the Fermi energy,* ε_F.

V_{111}	Exact	ε_x^{LDX}	$\varepsilon_x^{LDX}+\varepsilon_x^{GE1}$	$\varepsilon_x^{LDX}+\varepsilon_x^{GE1}+\varepsilon_x^{GE2}$	GGA (Becke)
0.023313	−0.3303	−0.3302	−0.3304	−0.3304	−0.3306
0.069938	−0.3376	−0.3354	−0.3376	−0.3377	−0.3390
0.116563	−0.3515	−0.3460	−0.3518	−0.3517	−0.3550
0.163188	−0.3704	−0.3613	−0.3720	−0.3711	−0.3770
0.186500	−0.3809	−0.3702	−0.3836	−0.3820	−0.3892
0.209813	−0.3919	−0.3796	−0.3959	−0.3931	−0.4019
0.233126	−0.4030	−0.3893	−0.4086	−0.4043	−0.4147
0.256438	−0.4140	−0.3991	−0.4215	−0.4155	−0.4275
0.279751	−0.4248	−0.4089	−0.4346	−0.4264	−0.4402

5.2. *The kinetic energy*

In this subsection we will discuss a topic which is somewhat outside our stated task of finding successively better approximations to the exchange-correlation energies of inhomogeneous electronic systems. We will, for a moment, digress and discuss gradient approximations also to their non-interacting kinetic energies. We have chosen to do so because this topic is closely connected to the straightforward gradient expansion discussed in the previous subsection. Moreover, interest in this topic has increased over the past decade due to the desire to treat successively larger systems where it becomes difficult and expensive to solve even the simple one-electron Schrödinger problem of standard DFT. If we can construct accurate approximations to the kinetic energy from just the density alone we would have a theory without the need for wave functions and a large part of the computational effort of standard DFT would disappear.

Such a scheme has been a goal for many researcher since the time of Thomas and Fermi [2,3] and the gradient expansion of the kinetic energy including gradients up to order six has been known for years (see e.g. Ref. [36]). Tests mainly on atoms have, however, shown that the kinetic

Table XIII. Exchange energies in Rydberg per electron for Si and Si-related models

Model	Exact	ε_x^{LDX}	$\varepsilon_x^{LDA}+\varepsilon_x^{GE1}$	$\varepsilon_x^{LDX}+\varepsilon_x^{GE1}+\varepsilon_x^{GE2}$	GGA (Becke)
Si	−0.5321	−0.5041	−0.5249	−0.5173	−0.5263
SiX	−0.5539	−0.5186	−0.5545	−0.5320	−0.5468
M1	−0.5340	−0.5107	−0.5250	−0.5215	−0.5291
M2	−0.5550	−0.5267	−0.5471	−0.5393	−0.5495

energies obtainable from gradient expansions do not have the accuracy to which we have been accustomed previously in these notes (e.g. of the LDA). Learning about the accurate exchange energies I had obtained from straightforward gradient expansions in metallic system, Malcolm Stott [87] suggested that I should try the same technique also to the full kinetic energies. The main new idea which would allow us to disregard the negative results previously obtained for atoms was that we proposed to apply the gradient expansion only to pseudo atoms and not to real atoms. After all, most s-p bonded solids and even transition metal compounds are nowadays treated with pseudo potentials and plane-wave codes. Thus, we exposed the electron gas to a periodic array of local pseudo potentials and calculated the kinetic energy both from full one-electron band theory and from the straightforward known gradient expansion. In Table XIV we display the astonishingly accurate result for our first test case simulating an Aluminum metal as far as concerns the average electron density and the valence band structure. In these calculations we included up to fourth-order gradients although the expansion is known at least up to sixth order. Table XIV also shows the results for our pseudo Silicon system discussed in the previous subsection. The case of Silicon again demonstrates the breakdown of perturbation theory as soon as band gaps occur although the error is only 0.007 Ry/electron.

In Table XIV, the column marked t_{TF} is the Thomas–Fermi approximation which is equivalent to the LDA for the kinetic energy. The column marked $t_{TF}+t_{GE1}$ is the result obtained including gradients to fourth order from linear response theory and the column marked $t_{TF}+t_{GE1}+t_{GE2}$ contains results including all fourth-order gradients. Thus, the data demonstrate the relatively lesser importance of higher-order responses for these particular systems.

In order to obtain the results in Table XIV, we have calculated the gradient corrections from the accurate density obtained from the full band structure calculation. Thus, from a computational point of view, nothing is gained by this procedure which still requires the calculation of all the occupied wave functions. Therefore, we went one step further and minimized the total-energy expression containing only the density and up to fourth-order gradients. The minimization was carried out using a Car–Parrinello code [88] for minimization through simulated annealing. From older test on atoms it is well known (see, e.g., Ref. [36]) that a minimization of the gradient-expanded total energy is relatively unstable and gives singular or bad densities which, in turn, produce inaccurate total energies. And the problem is aggravated by including gradients of high order. This was actually one of the reasons why we decided not to go to sixth-order gradients. The rather small energy contributions from higher-order gradient terms (see Table XIV) in our test systems suggest that the minimization procedure could include only second-order gradients in order to make it more stable. Then, as a last step, one could apply also the fourth-order gradients to the minimizing density. Possibilities of this kind could be investigated as the method becomes more refined but, at the moment, we were quite pleased to find that the minimization procedure, at odds with the atomic experience, led to rather accurate energies of our pseudo systems. The error in Aluminum, some 20 milli Ry, is still a little too high but might be remedied by introducing GGA:s also for the kinetic energy. In Silicon, the original error from the gradient expansion was actually almost canceled by the minimization procedure which we, however, consider to be a lucky fluke. Nevertheless, we think that the results presented here offer a lot of hope for future total-energy calculations in solids without the use of wave functions—at least in systems which can be described by means of pseudo potentials.

Table XIV. Kinetic energies in Rydberg per electron for pseudo Si and pseudo Al.

Model	Exact	t_{TF}	$t_{TF}+t_{GE1}$	$t_{TF}+t_{GE1}+t_{GE2}$	t(minimized)
Al	0.466	0.436	0.465	0.464	0.486
Si	0.716	0.692	0.725	0.723	0.717

5.3. *The LPM scheme*

In Section 3.5 and in the previous section we have obviously made a strong case against the use of gradients as a tool for going beyond the LDA—except in the vary favorable case of metals. As we shall see here, the situation is not that bad. One of the first attempts to improve on the LDA by means of gradients but without much theoretical support was due to Herman *et al.* [89]. They added the second-order gradient correction with a variable coefficient $B_x^{(2)}$ chosen to reproduce atomic exchange energies. In this way, they were able to accurately reproduce Hartree–Fock energies of atoms but they had to use a coefficient which was rather far from the value predicted on the basis of gradient expansions for the electron gas. This represents the first so called Generalized Gradient Approximation (GGA) and this kind of approach is used extensively even today. The exchange-correlation (xc) energy is written as some function of the density and its first gradient including a large number of adjustable parameters. These are subsequently determined by requiring the expression to reproduce nearly exact results in a number of known cases. Embarrassingly enough, this rather *ad hoc* method is still probably the most accurate approach within DFT although results are never consistently good. One often encounters surprises when treating systems different in character as compared to those of the reference group.

But there are certainly more fundamental approaches toward improvements via GGA:s. As we discussed in Section 2.3, xc energies can be obtained from a modeling of the exchange-correlation (xc) hole in real space. An obvious complementary idea is to go to reciprocal space by means of Fourier transforms and to study xc energies coming from different regions of momentum space. From the discussion in Section 3.5 we also know that there is a close connection between momenta and gradients of the density—at least in the linear regime. We are thus led to believe that a decomposition of xc energies with regard to momentum transfers could, somehow, be converted into expressions involving gradients. This is actually a viable idea as we will now argue.

The first attempts in this direction is due to Langreth, Perdew, Mehl, and Hu [8,9,56] (LPM). I consider their work as the start of the activity to construct GGA:s from first principles. It would carry too far to present the rather complicated theory underlying the LPM correction to the LDA. We will here be content with presenting some of their key ideas and some results for different physical systems. Langreth and Perdew [8] start from the basic equation for the exchange-correlation energy of the inhomogeneous system (see Section 2.3). In order to find the interaction-averaged pair-correlation function $\tilde{g}$ of the inhomogeneous system they first use the RPA and compute the frequency dependent density-density response function to second order in a perturbing external potential w. Within the RPA, the required integration over the strength of the Coulomb interaction can be carried out analytically and from the zero-temperature version of the fluctuation-dissipation theorem [46] they then obtain E_{xc} for the slightly inhomogeneous system in the form of an integral in reciprocal ($\boldsymbol{k}-$) space. In this way, they can study the contributions from different wave vectors both to the LDA (E_{xc}^{LD}) and to the xc-energy obtained by adding the lowest order gradient correction. They find that the gradient correction works well for large wave vectors but strongly overestimates the contributions to the xc-energy at small $\boldsymbol{k}$. The crossover occurs around a $\boldsymbol{q}$-vector approximately equal to $q = |\nabla n(\boldsymbol{r})|/(6n(\boldsymbol{r}))$. In principle, their full result for E_{xc} could be used in actual calculations although the computational effort would be considerable. Therefore, in order to obtain an easily applicable scheme based on density gradients, Langreth, and Mehl [9] suggested an approximation to the full results by Langreth and Perdew. They proposed to keep the full gradient correction down to a cut-off $\sim q$ and to neglect the gradient correction altogether below this cut-off. In real space this results in the following approximation for E_{xc} :

$$E_{xc}[n] = E_{xc}^{LD}[n] + a\int [n(\boldsymbol{r})]^{-4/3}[\nabla n(\boldsymbol{r})]^2\left\{e^{-F} - \tfrac{7}{18}\right\}\mathrm{d}^3r$$

where

$$F = b\cdot|\nabla n(\boldsymbol{r})|\cdot[n(\boldsymbol{r})]^{-7/6};\quad a = \frac{\pi}{8}(3\pi^2)^{-4/3};\quad b = (9\pi)^{1/6}f$$

in atomic units, *i.e.*, Hartree.

The quantity f is a "fudge" factor of order 0.15–0.17 and results are relatively insensitive to the choice of f in this range. Notice that surface energies indeed are sensitive to the choice of f, indicating, not a breakdown of the above theory, but rather of the simple cut-off procedure leading to the practical formula above.

It is important to stress one of the basic assumptions underlying the parameterized version of the wave vector decomposition due to LPM. It is assumed that the physical system under study is dominated by a single length scale approximately given by $1/q$. This is probably correct in an atom where that length would be, *e.g.*, the radius of an atomic orbital. As a result, the LPM scheme yields quite accurate total energies for atoms, the error being of the order of a few parts in a thousand. As it turns out, the major source of error comes from the treatment of exchange. For this part of the energy the LPM approximation gives an error of $\sim 3\%$ which should be compared to an error of the order of $\sim 10\%$ in the LDA. In the case of atomic correlation energies only, the LPM errors are $\sim 10\%$ compared to the infamous $> 100\%$ overestimate resulting from the LDA.

In molecules, on the other hand, there is clearly at least one additional important length scale namely that of the molecular bond. As a result, the performance of the LPM scheme is not nearly as good in the case of binding energies of molecules as illustrated in Table VI. Also, in this case, there does not seem to be any advantage in treating exchange exactly and only leave the correlation part of the binding energies to the LPM scheme (see Table XVI).

As pointed out in Section 3.5, approximations based on a theory for systems with relatively small and slow density variations like, *e.g.*, the LDA and different gradient corrections to it, can perhaps not be expected to work well for strongly inhomogeneous systems like atoms and small molecules. Thus, tests on the ground-state properties of solids are highly desirable. Unfortunately, such tests are rare in the literature. The first such test was carried out by von Barth and Car [57] in bulk silicon and they obtained a cohesive energy of 4.89 eV compared to 5.19 eV in the LDA and 4.63 eV from experiment. The lattice parameter of the LPM calculation agrees with experiment and is 0.5% larger than that of the LDA. Within the numerical accuracy, the bulk moduli of the LDA and the LPM calculations both agree with experiment. Later von Barth and Pedroza [90] tried the LPM approximation in bulk beryllium and obtained a cohesive energy of 3.20 eV compared to 3.65 eV in the LDA and 3.32 eV from experiment. More recently Bagno *et al.* [91] have applied the LPM scheme to the structural properties of several elemental solids including transition metals in which the general tendency of the LDA to over-bind is particularly pronounced. Their results are similar to those reported above. The over-binding is reduced and the lattice parameters are slightly improved by the LPM scheme. In particular, this scheme gives the correct ground state of iron, *i.e.*, a ferromagnetic bcc structure whereas the LDA erroneously predicts a paramagnetic fcc structure.

We can summarize our experience of the performance of the LPM scheme in solids and molecules by saying that this scheme is superior to the LDA and that errors of the latter are typically reduced by a factor of two by the LPM approximation. Thus, the new scheme is an improvement but does certainly not represent a breakthrough within DF theory.

Before ending this section on the LPM scheme, we must mention a few technical details of relevance to the practical application of the scheme. In most cases, especially in finite systems, it is necessary to have a spin-polarized version of the theory. Such a generalization has been worked out by Hu and Langreth [56]. For self-consistent calculations the

Table XV. The errors (in eV) in the binding energies of the first-row dimers as obtained from different density functionals defined below. Δ is the average absolute error for each functional.

	Li_2	Be_2	B_2	C_2	N_2	O_2	F_2	Δ
LDA	−0.1	0.5	0.8	1.0	1.7	2.4	1.7	1.2
LPM	−0.5	0.3	0.2	−0.2	0.3	1.2	0.7	0.5
PW86	–	–	0.2	−0.1	–	0.7	0.5	0.4
PW91	−0.1	0.3	0.3	−0.1	0.7	1.0	0.7	0.5
B86	−0.1	0.1	−0.5	−0.8	0.2	0.2	0.0	0.3
B I	−0.1	0.1	−0.3	−0.7	0.3	0.3	0.2	0.3
B II	−0.3	–	–	–	0.4	0.8	0.5	0.5
B III	−0.3	–	–	–	−0.1	0.2	−0.1	0.2

LDA the local-density approximation. Data from Ref. [63].
LPM the functional by Langreth, Perdew, Mehl and Hu [8,9,56]. Data from Ref. [64].
PW86 the older functional by Perdew and Wang [77,78]. Data from Ref. [103].
PW91 the functional by Perdew *et al.* [97] preceding the PBE [61]. Data from Ref. [80].
B86 an older exchange approximation by Becke [64] plus correlation from Stoll *et al.* [107]. Data from Ref. [64].
B I the functional of “Becke: Thermo-chemistry I” [106]. Data from Ref. [80].
B II the functional of “Becke: Thermo-chemistry II” [109]. Data from Ref. [109].
B III the functional of “Becke: Thermo-chemistry III” [111]. Data from Ref. [111].

Table XVI. The errors (in eV) in the correlation-energy contributions to the binding energies of the first-row dimers as obtained from different correlation functionals defined below. Δ is the average absolute error for each functional.

	Li_2	Be_2	B_2	C_2	N_2	O_2	F_2	Δ
LDC	−0.1	−0.3	−1.9	−4.1	−2.4	−2.9	−2.5	2.0
LPMC	0.1	0.1	−0.8	−2.7	−1.0	−1.2	−1.4	1.0
PW91C	−0.3	−0.2	−1.5	−3.9	−2.4	−2.1	−2.4	1.8

LDC the local-density approximation for correlation energies. Data from Savin *et al.* in Ref. [33].
LPMC the correlation part of the functional by Langreth *et al.* [8,9,56]. Data from Savin *et al.* in Ref. [33].
PW91C the predecessor [97] to the PBE [61,62] by Perdew *et al.* Data from Ref. [119].

potential corresponding to the parameterized LPM scheme, *i.e.*, the functional derivative of E_{xc} with respect to the density, must be derived and an explicit formula involving second spatial derivatives of the density can be found in Ref. [9].

The LPM theory for the slightly inhomogeneous electron gas is based on the RPA. According to Langreth *et al.*, [9,92] there is a large cancellation between local and nonlocal contributions to the xc-energy also beyond the RPA and, in order not to lose this advantage by using a theory beyond the RPA for just the LDA, the RPA version of the latter should be used in conjunction with the LPM gradient correction.

As a byproduct of their wave vector analysis Langreth *et al.* found two very useful criteria determining the validity of the LDA and of the straightforward gradient correction. They found that the LDA is valid when

$$\frac{|\nabla n(\boldsymbol{r})|}{k_F(\boldsymbol{r})n(\boldsymbol{r})} \ll 6$$

whereas the gradient expansion is valid when

$$\frac{|\nabla n(\boldsymbol{r})|}{k_{TF}(\boldsymbol{r})n(\boldsymbol{r})} \ll 1.$$

Here, $k_F(\boldsymbol{r})$ is the local Fermi momentum ($k_F^3 = 3\pi^2 n$) and $k_{TF}(\boldsymbol{r})$ is the inverse of the local screening length, *i.e.*, the Thomas–Fermi wave vector given by $k_{TF} = 2\sqrt{k_F/\pi}$. In real systems the last criterion is usually an order of magnitude more severe than the first which explains why the LDA can provide reasonable answers while the inclusion of gradient terms will make things worse.

5.4. *The Perdew–Wang scheme*

In the discussion of the LPM scheme we noted that the largest errors in that scheme originates in their treatment of exchange. For this part of the energy Langreth *et al.* essentially use a straightforward gradient correction. Being concerned about the shortcomings of the previous treatment Perdew [93,94] introduced a new idea comprising a combination of real- and reciprocal space arguments. In Ref. [54] as well as in the work of many investigators, the success of the LDA is attributed to various exact properties of its exchange-correlation (xc) hole like the sum rule or the negativity of the exchange hole. In the wave vector analysis of Langreth *et al.* we can say that we study and approximate the xc-hole in reciprocal space and it becomes less evident how to implement, *e.g.*, the sum rule. A particular approximation for the xc-energy in reciprocal space can, however, often be transformed into an approximation for the xc-hole in real space based on the density and its gradients. Perdew showed, *e.g.*, that the second order gradient expansion for the xc-energy does not correspond to an xc-hole that obeys the sum rule. This is yet another way in which we can understand why the LDA in some cases can yield an accurate result when the addition of the lowest order gradient term can destroy that accuracy.

From the work by Gross and Dreizler [95] on gradient expansions it is not difficult to obtain one for the exchange hole. A second order expansion for the hole does not obey either the sum rule or the negativity condition (See Section 2.1). Therefore, Perdew [93] multiplied the result by Gross and Dreizler by an appropriate cut-off function to ensure the satisfaction of both the sum rule and the negativity condition. The resulting approximation for the exchange hole looked somewhat complicated and difficult to use but in a later paper by Perdew and Wang [77], the result was simplified by partial integration and parametrized for easy use in DF calculations. Their functional for the exchange energy became

$$E_x[n] = A_x \int [n(\boldsymbol{r})]^{4/3} F_x(s) \mathrm{d}^3 r,$$

where A_x is a constant determined by the electron-gas limit and given by

$$A_x = -\frac{3}{4\pi}(3\pi^2)^{1/3}.$$

The quantity $F_x(s)$ is a function of the dimensionless density gradient

$$s = \frac{|\nabla n(\boldsymbol{r})|}{2k_F(\boldsymbol{r})n(\boldsymbol{r})}.$$

The function F_x is originally obtained numerically in order for the x-hole to obey the sum rule and the negativity condition and later fitted to an analytical formula:

$$F_x(s) = (1 + as^2/m + bs^4 + cs^6)^m$$

with the constants

$$a = \tfrac{7}{81}; \qquad b = 14; \qquad c = \tfrac{1}{5}; \qquad m = \tfrac{1}{15}.$$

Energies are given in Hartrees. Notice that $F_x(s) = 1$ corresponds to the LDA for exchange only. Notice also that many systems are spin-polarized, particularly those of finite extent and that, therefore, the spin-polarized versions of the functionals must be used. In the case of exchange only this is easily obtained as

$$E_x[n_{\uparrow|}, n_\downarrow] = \tfrac{1}{2}\{E_x[2n_{\uparrow|}] + E_x[2n_\downarrow]\},$$

As a side issue, we finally notice that the generalized gradient approximation by LPM of the previous subsection, and that of Becke [11,64,96] (see next subsection) as well as, really, all GGA:s can be cast in the same form. And the different GGA:s (excluding, of course, the Meta-GGA:s involving also the Laplacian of the density) differ only in their particular form of the function $F_{xc}(s)$.

The particular parametrization of the Perdew–Wang scheme presented above has the property that an incorrect second-order gradient term is obtained in the slowly varying limit (small s). In this context, "incorrect" refers to a deviation from the gradient coefficient given exactly by Engel and Vosko [49]. Since the small-q limit of the exchange-only kernel $K_x(\boldsymbol{q})$ (see Section 3.5) is better described over a larger range of q-values by a correct second-order gradient coefficient for exchange, the function $F_x(s)$ in the Perdew–Wang scheme above should be modified to give this slowly varying limit. Setting $a = 10/81$ in the parametrization above will achieve the desired result. In later work [97], Perdew *et al.* made additional modifications of the function $F_x(s)$ and essentially made it much more similar to the Becke exchange approximation [11] (see also below). The slightly modified functional of Ref. [97] was tried in solid Li and Na and in atoms and small molecules [80,97]. This new GGA performed better but not much better.

Having improved on the LPM description of non-local contributions to exchange energies, Perdew [78] proceeded to introduce some beyond-RPA effects into the LPM scheme for the correlation energies. Apart from taking away a spurious contribution of exchange origin added by Langreth *et al.* in order to separate exchange and correlation in *finite* systems, Perdew's modification essentially amounts to a rescaling of the LPM result in order to retrieve the correct beyond RPA second-order gradient coefficient calculated by Geldart and Rasolt [51]. The explicit formula is very similar to the LPM formula of the previous subsection, but without the 7/18:th. We refer the reader to Ref. [78] for details. Having a beyond RPA result for the non-local correlation energy allowed Perdew to add the state-of-the-art version of the LDA for the correlation energy, *e.g.*, a parametrized version of the Monte–Carlo results for electron-gas correlation energies by Ceperley and Alder [41].

The Perdew–Wang (PW86) correction to the LDA described above has been employed in several electronic structure calculations of solids. Kong *et al.* [98] calculated the ground-state properties of solid Al, Si, and C and concluded that the PW86 functional gives a substantial reduction of the overbinding characteristic of the LDA. The bulk moduli are almost the same as those of the LDA but there was a small increase in the lattice parameters resulting in closer agreement with experiment. Notice, however, that these quantities are very close to experiment also in the LDA. Bagno *et al.* [91] tried both the PW86 functional and the LPM functional on solid K, Ca, V, Fe, and Cu. Their conclusions were similar to those of Kong *et al.* The LDA errors in the binding energies are reduced by typically a factor of two and the lattice parameters and the bulk moduli are also improved by the two gradient corrected schemes. In particular these schemes predict the correct ground state of iron, *i.e.*, ferromagnetic bcc. On the other hand one finds that one of the well known failures of the LDA, *i.e.*, the prediction of metallic ground states for the transition metal oxides FeO and CoO, are not corrected by the PW86 functional [99]. Comparing the PW86 and the LPM functionals, Bagno *et al.* found that the first had a slight edge on the latter although they noted a tendency of the PW86 functional to over-correct the LDA errors especially for the more weakly bound systems. The same tendency has been reported by Garcia *et al.* [100] in a study of the cohesive properties of solid Al, Si, Ge, GaAs, Nb, and Pd. In particular these researchers report that the PW86 functional in many cases gives lattices which are too soft. Similar conclusions have been reached by others [101,102].

As far as molecules are concerned, there is some experience with the PW86 functional [103]. It seems clear, however, that this functional performs slightly better than the LPM functional also in the molecular case and that it thus represents a substantial improvement on the LDA. As will be discussed in the next section, this is a somewhat spurious result most likely of unphysical origin.

As mentioned above, further improvements on the Perdew–Wang scheme was made in Ref. [94]. There, Perdew and collaborators incorporated the sum rule for the correlation hole, *i.e.*, the fact that this hole must contain zero charge. This follows from the sum rule for the total hole and the unit charge contained in the Fermi (or exchange) hole. They have also modified their approximation to the exchange energy through a different choice of the function $F_x(s)$ as discussed above. The latest $F_x(s)$ behaves as $1 + 10/81\ s^2$ at small s corresponding to the coefficient of Engel and Vosko [49] for the straightforward second-order gradient correction for the exchange energy of the homogeneous electron gas. Probably inspired by the success of one of Becke's [11] gradient corrections for exchange (see next section), the function $F_x(s)$ closely follows the Becke approximation at intermediate s. In contrast to the latter approximation, $F_x(s)$ is made to vanish as $const/s^2$ at large s in order to obey certain inequalities for exact exchange. The sum rule on the x-hole is not enforced but is obeyed rather closely by the resulting functional referred to as PW91 in what follows. The PW91 functional was tested by Perdew and his collaborators on solid Li and Na. In both cases the PW91 functional corrects the unusually large LDA errors in the lattice parameters of these systems (0.11 Å in Li and 0.17 Å in Na). The PW91 functional has, by now, been tried in many systems. It is rather similar to the PW86 functional and, in solids, it gives results of comparable accuracy or slightly better. Also in molecules the PW91 functional is comparable to the PW86 functional as reported by Merkle, Savin, and Preuss [80].

The experiences from the functionals of GGA type which we have referred to as PW86 and PW91 have been collected by Perdew and his collaborators in the GGA-functional called PBE [61,62]. This functional is today a standard ingredient in most readily available density-functionals codes for solids and molecules. It performs well in both finite and infinite systems without being specialized for either application. This is in contrast to many new functionals of Quantum Chemistry which most often are designed for molecular applications. It differs rather little from the PW91 but it is much easier to program and its derivation is much simpler. A particular advantage is that it gives a rather good representation over a range of momenta (q) of the static linear density response function of the electron gas as indicated by quantum Monte Carlo calculations [104,105].

We will now attempt to summarize our experience from the gradient corrected density functional methods. From a theoretical point of view, many of these schemes are a result of a careful analysis of exchange and correlation effects in a weakly perturbed electron gas (basically linear response theory). Through the use of sum rules and other consistency requirements, effects of higher order responses are brought into the approximations. The results are encouraging and usually much superior to the LDA. Errors in binding energies are reduced by a factor of two to three or even more. In the case of properties which are very well described already by the LDA the picture is less consistent and gradient corrections can lead to worse results although the difference is marginal. In going from the LPM gradient corrected functional to the PW86, the PW91, and later to the PBE and, especially, to the Meta-GGA functionals (to be discussed in Section 5.6) there is an increase in theoretical sophistication. Tests on different systems and for different physical properties indicate a minor and inconsistent improvement going from the LPM to the PW91 scheme. Some additional improvement, a simpler derivation and a smoother expression and thus a less singular potential is obtained by going on to the PBE. A further improvement results from the introduction of the Meta-GGA. The gradient corrected schemes are marginally more difficult to apply as compared to the LDA and in view of aforementioned properties we recommend that the later functionals, *e.g.* the PBE or the Meta-GGA, to be used in all DF calculations.

In Section 4.5, several of the failures of the LDA was blamed on the the inability of the LDA to account for the s-p or the s-d transfer energies. Tests on the gradient corrected functionals [32,76] show that they suffer from the same deficiency but still yield much better binding energies. This somewhat unexpected result is the basis for our comment in Section 4.5 that the s-d transfer problem may not even be the major source of error in the LDA.

In small molecules it is relatively easy to compute the exact exchange energy thus leaving only the much smaller correlation energy to be treated by DF methods. Such a separation leads to disastrous results as can be seen in Table XVI. The usual argument for treating exchange and correlation together is based on the large cancellation occurring between exchange and correlation energies due to the much smaller extent of the xc-hole as compared to the exchange and correlation holes taken separately. This argument is certainly valid in extended systems but its validity is doubtful for small systems where already the size of the system limits the extent of both holes. The bad results shown in Table XVI and resulting from a separate treatment of exchange and correlation, rather indicate a poor description of the correlation hole at intermediate distances—a region not amenable to gradient corrections based on electron-gas theory. At this stage, the relatively good results obtained for the combined DF treatment of exchange plus correlation must be considered as fortuitous. Indeed, as we shall see in the next section, Becke [106] has managed to produce accurate correlation energies for a large number of molecules using an approximation based on arguments relevant only to exchange energies.

5.5. *The Becke approach*

In the previous subsections we have described attempts to design approximations beyond the LDA which are based on a careful analysis of the effects of exchange and correlation in inhomogeneous systems. From early on, Becke has taken a different approach to DF theory. In principle we seek to map out the functional $E_{xc}[n]$ in systems of interest which, in Becke's case, have been atomic and molecular systems. Suppose that we have designed a functional based on, *e.g.*, the density, its different gradients, and different integrals thereof, containing a large number of parameters. Suppose further, that we were able to determine those parameters in such a way that the chosen functional, within a certain accuracy, reproduces the exact E_{xc} for a large number of physical systems with rather different properties with regard to exchange and correlation. Then, we would certainly believe that we had obtained an accurate description of the true $E_{xc}[n]$, at least for densities not deviating in a qualitative way from those already studied. We would feel comfortable in applying the constructed functional to unknown systems and we would have some confidence in the results thus obtained. There are certainly theoreticians who would be prone to scorn the described procedure but we find such attitudes unwise. First of all, the semi-phenomenological procedure will enable us to treat more complicated systems with much greater ease as compared to *ab initio* methods. And we could even attack systems which presently are beyond computational feasibility and we could concentrate on other physical aspects like, *e.g.*, molecular reactions. Secondly, we would have learned something about the functional $E_{xc}[n]$. For instance, if our chosen functional form contains only the density and its first spatial derivatives we would know that it must be possible to cast a sophisticated *ab initio* theory in a form containing the same basic ingredients.

When we, after this philosophical digression, return to Becke's work we must stress that a considerable amount of physical insight is needed in order to invent a reasonable functional form for $E_{xc}[n]$. A better functional form leads to fewer adjustable parameters and Becke's functionals do contain few parameters. He started with the observation that when density gradients become large, gradient corrections are irrelevant. Concentrating on the exchange energy, Becke [64] designed a functional which, at small gradients, amounts to a normal second-order gradient correction with an adjustable coefficient and which becomes an adjustable constant at large gradients. As a matter of fact most of Becke's gradient corrections for exchange can be expressed in terms of an enhancement factor $F_x(s)$ times the LDA or LDX as discussed in Section 5.4 in connection with the Perdew–Wang scheme. The first one [64] used

$$F_x(s) = 1 + \frac{\beta s^2}{1 + \gamma s^2}.$$

The parameters β and γ were adjusted so that this particular form gave an accurate fit to the exchange energies of a large number of atoms ($\beta = 0.2351, \gamma = 0.2431$). To this exchange energy must be added some approximation for the correlation energy and in his first semi-empirical approach toward molecular energies Becke chose one due to Stoll *et al.* [107] This approximation is much superior to the LDA for correlation energies of finite systems but is has the wrong slowly varying limit. The resulting atomization energies of the first row dimers can be studied in Table XV. The maximum error is 0.8 eV and the average error for the seven molecules is 0.3 eV. The corresponding figures for the LDA are 2.4 eV and 1.2 eV. Considering the simplicity of the approximation this represents an extraordinary achievement which only recently has been surpassed by a rather specialized GGA based on many adjustable parameters.

Based on ideas concerning the behavior of exchange energies in strongly inhomogeneous systems (large gradients) Becke [96] later proposed a minor modification of the above F_x amounting to raising the denominator to the 4/5 power and changing β and γ but, to our knowledge, this functional has never been tested in molecules. This is certainly not true about the next gradient correction for the exchange energy by Becke [11] which reads

$$F_x(s) = 1 + \frac{\beta s^2}{1 + (2.25/\pi)\, \beta s \, \mathrm{arc\,sin\,h}(\gamma s)}.$$

Here, γ is a well defined constant equal to $\gamma = 2(6\pi^2)^{1/3}$ (a conversion between Becke's $x = |\nabla \rho| \cdot \rho^{-4/3}$ and Perdew's $s = |\nabla n|/(2k_F n)$) and only β is an adjustable parameter which is determined by a fit to atomic exchange energies. A value of $\beta = 0.2743$ gives an average exchange-energy error of only 0.11% for six noble-gas atoms—a truly remarkable result considering the simplicity of the above ansatz. The chosen form for F might appear peculiar but it is constructed to give the correct inverse-distance dependence of the exchange-energy density far outside a finite system. In this region, the density decays as $e^{-\lambda \cdot r}$ and it is not difficult to see from the properties of the inverse sine hyperbolic function that this leads to a contribution of the form

$$-\frac{1}{2}\int_{r>R} \frac{n(\mathbf{r})}{r} \mathrm{d}^3 r$$

to $E_x[n]$ at large R. In fact, this particular exchange-energy functional is presently the only available functional with this correct behavior. But we are a little doubtful about the value of this property since the energy contributions from the large-R regions are small due to the very small density in these regions. Nevertheless, this "ansatz" by Becke has proven extremely successful. Becke [106] has tested the functional by calculating the atomization energies of the 55 molecules of the so-called Gaussian-1 data base of Pople *et al.* [12,108] For this test, Becke used the normal LDA for the correlation energy—an approximation known to overestimate the correlation energy by a factor two to three. One can only speculate why Becke refrained from adding one of the existing gradient-type correlation corrections to the LDA. In any case, his results were astonishingly good. The average absolute error was only 0.16 eV (3.7 kcal/mol) and the maximum error was 0.48 eV. The corresponding errors of the LDA are 1.6 eV and 3.7 eV, *i.e.*, an order of magnitude larger. Since there is a large overestimate of the correlation energies in the test, Becke's exchange functional must, somehow, make up for the difference. It is an interesting task for future research to understand how this cancellation of errors comes about. The Cl-oriented *ab initio* results for the same set of molecules have average errors of the order of 1 kcal/mol [12,108] and we clearly see that DF methods are approaching the same quality but with much less effort.

In other tests on atoms and molecules, Becke found that the tested functional had an unsatisfactory performance in the case of electron non-conserving processes like, *e.g.*, ionization. Becke [109] then decided to add to his previous functional the, at the time, most recent gradient correction by Perdew *et al.* [97] (PW91C) for the correlation energy. The approximation for the exchange energy remained the same—the arcsinh-ansatz. Becke also enlarged his data base by including many ionization potentials and several proton affinities and redid the test, again with a very impressive result [109]. The average absolute error and the maximum error for the atomization energies of the molecules of the Gaussian-1 data base was found to be 0.25 eV and 0.8 eV respectively. Thus, an "improved" description of correlation effects results in slightly larger binding-energy errors but there is a big gain as far as ionization potentials are concerned. These show an average error of only 0.15 eV and a maximum error of 0.44 eV. The corresponding numbers for the LDA are 0.23 eV and 0.62 eV for 42 ionization potentials.

These results represent, on the average, close to an order of magnitude improvement over the LDA and this is indeed encouraging. We are closer to achieving "chemical" accuracy but there is still some way to go. Occasional errors of the order of 0.8 eV cannot be tolerated. The Becke functionals are specially designed for finite systems (atoms and molecules) and they have the wrong slowly varying limit. What this would mean for solids or larger molecules is hard to say. Since only the gradient terms are in error, we would not expect the Becke functional to be any worse than the LDA in extended systems. From a theoretical point of view, however, we would prefer a functional with all the correct limits built in but no such functional can presently overshadow the one by Becke in smaller molecules.

As mentioned several times in these notes, exchange energies are much larger than correlation energies even in extended systems like metals. This immediately suggests that one should treat exchange exactly and, *e.g.*, use a GGA for only the correlation energy. So far, attempts in in this direction have failed as can be seen in Table XVI. In 1993, Becke [110] suggested a hybrid method in which half of the exchange energy is treated exactly leaving the other half plus the correlation contribution to approximations of the GGA-type. Theoretical support for this scheme can again be obtained from the basic formula of Section 2.3 for the xc-energy sometimes referred to as the "adiabatic connection" formula. Using the bare LDA for the correlation contributions Becke found that this new scheme had an accuracy similar to his previous functionals for the atomization energies of the Gaussian-1 data base. In one further step Becke [111] included some of the correlation correction of Perdew *et al.* (PW91C) and decided to treat the amount of exact exchange as a fitting parameter. His "hybrid" functional can be summarized as

$$\begin{aligned} E_{xc}[n] =& E_{xc}^{LD}[n] + a_o \cdot (E_x[n] - E_x^{LD}[n]) \\ & + a_x \cdot \Delta E_x^{B88} + a_c \cdot \Delta E_c^{PW91} \end{aligned}$$

where a_o, a_x, and a_c are semi empirical coefficients to be determined by an appropriate fit to experimental data. The symbol Δ signifies the gradient part of the corresponding approximation, B88 refers to the Becke [11] gradient correction for exchange (the arcsinh-ansatz), and PW91 refers to the mentioned gradient corrections for the correlation energy by Perdew *et al.* [97] The functional $E_x[n]$ is the exact exchange-energy functional, *i.e.*, the Hartree-Fock exchange-energy expression, here evaluated using the DF orbitals, and E_x^{LD} is the LDA version of only the exchange energy. The functional defined by the above equation is, in the literature, referred to as the B3PW91 functional where the "3" referes to the term "thermochemistry III" appearing in the title of Ref. [111]. The parameters a_o, a_x, and a_c of the B3PW91 were determined by fitting to the Gaussian-1 data base including also 42 ionization energies and 8 proton affinities and the resulting optimum parameters were found to be $a_o = 0.20$, $a_x = 0.72$, and $a_c = 0.81$. These values resulted in an average absolute error of only 0.10 eV (2.4 kcal/mol) for the atomization energies with a maximum error of 0.33 eV. These results are now about a factor of two away from the proclaimed goal of chemical accuracy meaning errors of the order of 1 kcal/mol.

Results like those just discussed had a strong impact on the community of quantum chemists and in the beginning of the nineties the use of DF methods started to spread rapidly within this community. [112,113] New gradient corrections for the correlation energy were constructed by Becke [114] and by Lee, Yang, and Parr (LYP) [115]. Both these functionals have been tested on small molecules by Miehlich *et al.* [112] Johnsson and coworkers [113] have also applied the LYP correlation functional in conjunction with the Becke gradient correction for exchange in an extensive test on a large number of smaller molecules. The conclusion from these tests is that the performance of the Becke correlation functional, the LYP functional and the previously discussed PW91 functional for correlation energies is rather similar.

The most widespread functional among the chemists is probably the one called B3LYP. It is a again a so called hybrid functional meaning that some portion of exact exchange is mixed into the expression for the exchange-correlation energy. It actually looks almost identical to the Becke functional B3PW91 displayed above except for the fact that the gradient part of the correlation energy is replaced by the LYP expression. [115] Using the paramaters (a_o, a_x, and a_c) optimized by Becke also in the B3LYP functional, Curtiss and coworkers [116–118] have peformed extensive tests of the latter as applied to many properties of a wide range of molecules. In molecular applications, the B3LYP rather consistently outperforms most other density functionals allthogh the difference to, *e.g.*, the B3PW91 is marginal.

The tests reveal, however, that results deteriorate with the size of the molecules. This should, however, come as no surprise since the the well known anomalies caused in extended systems by an unscreened exchange is expected to ruin the results of the hybrid functionals in solids and compromise their performance in larger molecules. Still the success of the hybrid functionals B3LYP and B3PW91 indicate that some sort of non-local exchange should be incorporated into future more accurate functionals—non-localities above those that can be accounted for by mere gardients.

5.6. *The Meta-GGA*

As we have seen above, generalized gradient approximations (GGA:s) constructed from only the density and its first spatial gradient are still some distance away from chemical accuracy (1 kcal/mol). And the efforts to construct GGA:s range from *ab initio* and parameter-free theory to advanced fitting schemes involving a large number of parameters. The situation is summarized by a very nice argument due to Philipsen and Baerends [120]. They first observe that the binding energies of molecules and solids are dominated by exchange effects and then go on to write the exchange energy like we did in our discussion of the Perdew–Wang scheme in Section 5.4. One can write all GGA:s involving only the density and its first gradient in the form

$$E_x[n] = \int n(\boldsymbol{r})\varepsilon_x((n(\boldsymbol{r}))F_x(s(\boldsymbol{r}))\ \mathrm{d}^3r$$

where $\varepsilon_x(n)$ is the exchange energy per electron of the homogeneous electron gas of density n and $s(\boldsymbol{r})$ is the reduced dimensionless density gradient as defined in Section 5.4. Different GGA:s only differ in their choice of enhancement factor $F_x(s)$ where $F_x = 1$ corresponds to the LDX, i.e., to the exchange-only version of the LDA. Finally, their investigations show that the binding energies originate in regions of space where the reduced density s lies in the interval $0.5 < s < 2.5$. Regions of very small gradients—the slowly varying regime—or of large gradients—the asymptotic region outside atoms—contribute little to the binding energies. In the important interval for s, most GGA:s have a rather smoothly increasing enhancement factor and if there were such a thing as an "exact F_x", it would certainly be very easy to model that with a few parameters. On the other hand, distorting F_x to obtain exact exchange energies for a few systems would certainly worsen the results for others. We can thus safely conclude that the whole idea of modeling exchange energies by the density and its first gradient is of limited validity. Moreover, given the failure of extensive fitting procedures in reaching chemical accuracy, we have probably already reached the limits of this approach.

In view of the discussion in the previous paragraph, it appears rather strange that there exist so few attempts at incorporating also more complicated but still local dependencies on the density like, e.g., the Laplacian of the density or the kinetic-energy density. Suggestions in this direction have not been lacking [47,93,121] and some quantum chemists have included the Laplacian in their attempts to fit chemical data to results from Meta-GGA:s [122–124]. From a fundamental point of view, gradient corrections to the LDA involving the Laplacian originate in second-order response theory about which little is known. And from a more pragmatical point of view the fitted corrections based on the Laplacian did not lead to a breakthrough in DFT although the results from these attempts clearly demonstrated the advantage of including also the Laplacian when constructing phenomenological gradient corrections [122]. The theoretical situation improved when the coefficient involving the Laplacian to lowest order was calculated by Svendsen and von Barth [47] in the Exchange-Only approximation (EXO). This inspired Perdew and coworkers [53] to apply their brand-marked philosophy of designing gradient corrections by incorporating as many known exact results as possible into the construction. In addition to the Laplacian, the resulting Meta-GGA had the kinetic-energy density as an extra ingredient. But the latter was expanded in gradients up to and including the Laplacian of the density thereby allowing Perdew *et al.* to express the Laplacian in the kinetic-energy density. In this way, they could substitute the latter for the former. The considerable advantage of this procedure stems from the fact that the kinetic-energy density, as opposed to the Laplacian, is well behaved both close to the nuclei and in the outskirts of finite or semi infinite systems. The result is a Meta-GGA with the correct expansion in the slowly varying regime but without the known singularities of the Laplacian. In the former respect, their Meta-GGA is actually superior to their previous and widely used GGA—the PBE [61].

Several tests of this Meta-GGA in variety of systems ranging from localized to extended have revealed [39,53] a substantial improvement over, e.g., the PBE [61]. The results for the atomization energies of a number of molecules can be seen in Table XVII.

We see from Table XVII the relatively better ability of the Meta-GGA, as compared to the PBE, to distinguish between molecules having double and triple bonds. The results of the Meta-GGA are now comparable to those obtained by Becke [110,111] in his procedure to include a certain suitably fitted percentage of the exact exchange energy into his functionals as discussed in Section 5.5. And this has been achieved through the use of fundamental theory. It should also be remembered that Becke's functionals are designed for finite systems in particular and do not perform well in extended systems. The same remark can be made about most of those functionals containing a large number of parameters and fitted to an

Table XVII. Atomization energies (in kcal/mole) calculated using the LDA, the PBE-GGA [61], and the Meta-GGA [53] of Perdew and coworkers compared to experimental data (Exp.) stripped of the energy of the zero-point motion.

Molecule	LDA	PBE-GGA	Meta-GGA	Exp.
H_2	113.3	104.6	114.5	109.5
LiH	61.1	53.5	58.4	57.8
CH_4	462.6	419.8	421.1	419.3
NH_3	337.3	301.7	298.8	297.4
OH	124.2	109.8	107.8	106.4
H_2O	266.6	234.2	230.1	232.2
HF	162.3	142.0	138.7	140.8
Li_2	23.8	19.9	22.5	24.4
LiF	156.1	138.6	128.0	138.9
Be_2	12.8	9.8	4.5	3.0
C_2H_2	460.3	414.9	401.2	405.4
C_2H_4	632.7	571.5	561.5	562.6
HCN	360.8	326.1	311.8	311.9
CO	298.9	268.8	256.0	259.3
N_2	266.9	243.2	229.2	228.5
NO	198.4	171.9	158.5	152.9
O_2	174.9	143.7	131.4	120.5
F_2	78.2	53.4	43.2	38.5
P_2	143.0	121.1	117.8	117.3
Cl_2	82.9	65.1	59.4	58.0
mean abs. error	31.7	7.9	3.1	–

extensive body of chemical data, even though some of these outperform the present Meta-GGA in molecules [122].

The performance of the Meta-GGA has also been tested in calculations of the surface energies of metallic systems [39,53], again with very encouraging results. Part of this success is certainly due to the present Meta-GGA being correct in the slowly varying limit, a property not shared by most other GGA:s.

Still, as we also see from Table XVII, errors in atomization energies of molecules can easily amount to half an eV or more. This is actually a factor of ten worse than our proclaimed goal of so called 'chemical accuracy'. Neither have any of the semi-empirical functionals been able to consistently achieve this goal. In my opinion, the hope of capturing the complexity of the full static many-body problem in terms of only the density and a few of its gradients, is futile. Nevertheless, I recommend the Meta-GGA described above as the standard approximation to be used in all DFT calculations of substances, finite as well as infinite, about which little is known before hand.

6. Real space methods

My comments throughout the sections on the GGA:s probably convey the message that we are at ways end as far as these methods are concerned. And this was actually my intention because this is my present working hypothesis. The fitting schemes involving the density and the first-order gradient and being constructed by means of neural networks give average errors around 2 kcal/mol but occasional errors ten times as large in localized systems. Including also the Laplacian in the fitting procedures might reduce these errors further. Judging from the discussed (see Section 5.6) experience from the Meta-GGA—maybe down to average errors of ~ 1.0–1.5 kcal/mol and occasional errors of the order of 10 kcal/mol. Adding even higher-order gradients is likely to result in successively smaller improvements.

Given the extreme computational advantage in using DF methods and the need for higher accuracy, how do we proceed? I do certainly not have the answer to this question but there is another route which has not been followed until the end. In Section 2.3, we discussed the LDA from the point of view of modeling the exchange-correlation (xc) hole in real space and we will now explore further possibilities along this track.

6.1. *Early attempts*

The construction of the LDA was described formally in Section 2.3 but we might gain some understanding by dressing the construction in words. We might recall that the xc energy is the total energy associated with each electron interacting with its xc hole. We obtained the LDA by cutting the xc hole around each electron out of a constant density given by the value n of the density at that electron. Moreover, the hole contained precisely one electron and had a spherical shape and an extent given by the r_s-value ($4\pi r_s^3 n = 3$) at the reference electron. When the density varies rather rapidly from point to point it would seem more appropriate to dig the hole out of some effective density $\bar{n}$ obtained by averaging the correct density around the reference electron. Such a procedure was suggested by Gunnarsson, Jonson and Lundqvist (GJL) [54] who also suggested that the extent of the hole be governed by the same average density $\bar{n}$. The resulting expression for the xc energy then becomes

$$E_{xc}^{AD}[n] = \int n(\mathbf{r})\varepsilon_{xc}(\bar{n}(\mathbf{r}))\ \mathrm{d}^3r$$

where the average density $\bar{n}$ is obtained from

$$\bar{n}(\mathbf{r}) = \int W(\mathbf{r} - \mathbf{r}')\ n\ (\mathbf{r}')\ \mathrm{d}^3r'$$

with some appropriate weight function $W(\mathbf{r})$ obeying the sum rule

$$\int W(\mathbf{r})\ \mathrm{d}^3r = 1$$

dictated by particle conservation. Here, as before, $\varepsilon_{xc}(n)$ is the xc energy per electron of the homogeneous electron gas. As in the case of the LDA, this means that the pair-correlation function is still modeled by that of the gas. The weighting function $W(\mathbf{r})$ can certainly be chosen in many different ways but, in analogy with the construction of gradient expansions, GJL suggested to choose W such that the new theory exactly reproduces the static linear density response function of the homogeneous electron gas. They showed that this requirement actually leads to a unique determination of the weight function W provided some extra minor conditions of physical origin are imposed. Consequently, the weight function W becomes a unique property of the homogeneous gas. It is a function of the length r of the radius vector and of the homogeneous density n which can be tabulated once and for all. In order to obtain the simple formula for the xc energy shown above, the density argument in the weight function W must again be chosen to be the resulting average density $\bar{n}(\mathbf{r})$ at the position $(\mathbf{r})$ of the reference electron.

Before discussing the numerical consequences of the new scheme, which by GJL was referred to as the Average Density (AD) approximation, we will also introduce the Weighted Density (WD) approximation by the same authors [54]. This approximation was independently suggested by Alonso and Girifalco [125] and GJL called it the Weighted Density (WD) approximation. This scheme is rather similar in spirit to the AD approximation. The xc hole is still modeled by means of the pair-correlation function of the homogeneous gas but the density pre-factor is the correct pre-factor $n(\mathbf{r}')$ of the exact expression for the xc energy. Thus the xc hole is dug out of the true density of the inhomogeneous system and the requirement that the hole must contain precisely one electron is, for each reference electron at $\mathbf{r}$, guaranteed by choosing an appropriate density argument in the pair-correlation function of the gas. The expression for the xc energy E_{xc}^{WD} in the WD approximation becomes

$$E_{xc}^{WD}[n] = \tfrac{1}{2}\int n(\mathbf{r})n(\mathbf{r}')\{\tilde{g}_h(\mathbf{r} - \mathbf{r}'; \bar{n}(\mathbf{r})) - 1\}v(\mathbf{r} - \mathbf{r}')\ \mathrm{d}^3r\ \mathrm{d}^3r'$$

where the density $\bar{n}$, for each point $\mathbf{r}$ in space, has to be determined such that

$$\int n(\mathbf{r}')\{\tilde{g}_h(\mathbf{r} - \mathbf{r}'; \bar{n}(\mathbf{r})) - 1\}\ \mathrm{d}^3r' = -1.$$

From these expressions we immediately notice their asymmetry with respect to their spatial arguments $\boldsymbol{r}$ and $\boldsymbol{r}'$—in contrast to the symmetry displayed by the exact expression for the xc energy. The asymmetry is also build into the AD approximation and we remind the reader (see Section 2.3) that also the LDA suffers from the same ailment. It might not mean so much as far as the resulting value for the energy is concerned since the value of the integral remains the same if we replace $\boldsymbol{r}$ by $\boldsymbol{r}'$ in the density argument of the pair-correlation functional. But it makes a big difference when we later want to construct the resulting xc potential by taking a functional derivative of the xc energy with respect to the density.

Both the AD and the WD schemes certainly have more physically correct features added to them as compared to the LDA and we would thus expect them to outperform the latter. The WD approximation was first tested by Alonso and Girifalco in a calculation of the exchange energies of a number of atoms. Much to their liking they found a large reduction of the $\sim 10\%$ exchange-energy errors one is used to from the LDA. Unfortunately, the WD approximation over corrected the LDA exchange energies. When GJL later included also the correlation effects they found that the WD scheme destroys the nice cancellation occurring within the LDA between the errors coming from exchange and correlation. As a result the atomic xc energies of the WD scheme are actually worse than those of the LDA. Also the AD scheme gives reasonable atomic exchange energies—better than those of the LDA but somewhat worse than those of the WD scheme. Unfortunately, the AD approximation gives a very bad description of the atomic correlation energies (too small in magnitude). Nevertheless, the total xc energies are, on the average, a factor of two better than those of the LDA. Considering the extra effort and superior physical reasoning involved in the construction of the AD and WD schemes, the atomic test results are a clear disappointment. Using a technique called shell partitioning, GJL managed to improve considerably on the atomic results of both the AD and the WD approximations. But this theory is very inelegant and very much dependent on the system under study. The concept of shells is rather well established in atoms but become ill-defined in pseudo solids and we do not wish to pursue this line of work any further.

GJL also tested both the AD and the WD schemes in the calculation of surface energies. Unfortunately, the results were very unphysical, a fact which they blamed on the difference between the exact xc holes and of the approximate ones extending too far into the bulk regions. Both the AD and the WD schemes give, however, a much better description of the surface region as compared to the LDA with its exponentially decaying xc potential. The xc potentials of both schemes have an image-like behavior although the coefficients are wrong. Likewise, both schemes give a one-over-distance dependence of the xc potentials outside any finite system—in contrast to the exponentially decaying potential of the LDA. Like in the case of the surface, the coefficients are, however, not correct. In my opinion, this failure is in both cases connected to the asymmetry of the expressions for the xc energies in the two schemes (see the discussion above).

In a later publication [126], Gunnarsson and Jones abandoned the idea of trying to model the xc hole by means of the pair-correlation function of the homogeneous gas. Instead, they choose a very localized function of the form

$$\tilde{g}(\boldsymbol{r},\boldsymbol{r}') = g_o(|\boldsymbol{r}-\boldsymbol{r}'|;\bar{n}(\boldsymbol{r}))$$

where

$$g_o(r;n) = 1 - A(n)\{1-\exp[-\lambda^5(n)/r^5]\}.$$

By requiring this ansatz to give the correct xc energy and normalization of the xc hole for the homogeneous electron gas, the two parameters A and λ become unique functions of the density n. As in the WD scheme, the density $\bar{n}(\boldsymbol{r})$ is then determined for each position $\boldsymbol{r}$ of the reference electron, by requiring the xc hole of the inhomogeneous system to contain precisely one electron. The power five on the distance dependence in the exponent is actually chosen such that the resulting approximation gives an image-like behavior with the correct coefficient outside a metallic surface. We notice again that this scheme lacks the symmetry discussed above and is thus likely to show the same inconsistency between energy density and potential that we noted in all the previous cases—be it the LDA or the AD or WD approximations.

The new scheme gives atomic xc energies which are almost an order of magnitude more accurate than those of the LDA—with errors typically less than 1%. And with the more rapid decay of the xc hole and the correct image behavior outside metallic surfaces, the new scheme stands a chance of producing reasonable surface energies. As far as we know, however, this has never been tested. But it has been tested on atomic ionization potentials where it actually does worse than the LDA. Notice that ionization potentials are differences between two ground-state energies and therefore obtainable from two DF calculations. The fact that the new scheme does not do well for a property mainly associated with the more slowly varying valence electrons suggest that the model is much too simplified and contains too little physics.

6.2. *The screened exchange model*

The relatively bad results, at least in comparison with the successes of the GGA:s, obtained from the early attempts to model exchange and correlation in real space probably caused these methods to fall into disrepute. An additional discouraging factor is that the real-space methods invariably seem to require more computational work in order to obtain the energy and, particularly, the potential. As we have argued many times previously in these notes, the latter objection is not a serious one since we can always rely on the variational property of the energy functional to calculate the total energy of any new functional by applying it to the self-consistent density of a GGA or even an LDA calculation.

We are, however, not prone to giving up easily and my postdoc Robert van Leeuwen and I decided to have another go at it in 1996 [127]. In the spirit of John Perdew, we decided to try and build in as many known correct properties as we could without making the theory inapplicable to real systems of interest. Here is a list of the properties which guided or most recent construction:

- Exchange energies are a factor of four to thirty larger than correlation energies where the former factor refers to metallic sodium and the latter to many atoms.
- In many atoms, the term structure, i.e. the energies of low lying excitations of definite symmetry, can be accurately predicted on the basis of simple term-dependent Hartree–Fock (HF) theory provided the Slater Coulomb (F_k) and exchange (G_k) integrals are reduced by 25% in value [75]. This means that the effects of different symmetries are reasonably well described by HF theory.
- In larger systems, the effect of correlations is, among other things, to reduce the range of the xc hole. In an extended system the pair-correlation function of HF theory decays as an oscillating function over distance to the third power. When correlation effects are included the decay changes to one over distance to the fifth power.
- In Section 4.5 we discussed the shortcomings of the LDA relative to HF theory in describing the so called s-p transfer energies. On this basis we argued that the nodal structure of the one-electron orbitals should be included in the construction of an accurate functional for the xc energy.
- In the prototype molecule H_2, HF theory gives an exchange hole the shape of which is independent of the position of the reference electron and which puts half an electron on each atom when they are pulled apart. In reality, the xc hole is relatively independent of the position of the reference electron if this is close to one of the protons. But at larger proton separation, the xc hole has almost all its weight on the atom where the reference electron sits. It seems impossible to model this behavior using any function of $\mathbf{r}-\mathbf{r}'$ reminiscent of the electron gas.

These considerations resulted in the following model for the xc functional which we coined the Screened-Exchange Model (SEM):

$$E_{xc}^{SEM}[n] = -\tfrac{1}{4}\int |n(\mathbf{r},\mathbf{r}')|^2 A(\mathbf{r})A(\mathbf{r}')B(\mathbf{r}-\mathbf{r}')\,\mathrm{d}^3r\,\mathrm{d}^3r'$$

where the quantity $n(\mathbf{r},\mathbf{r}')$ is the one-particle density matrix of Kohn–Sham orbitals,

$$n(\mathbf{r},\mathbf{r}') = 2\sum_{k}^{occ}\phi_k(\mathbf{r})\phi_k^*(\mathbf{r}').$$

The correlation factor $A\cdot A\cdot B$ has two ingredients:

(*i*) The electron-gas-like component B serves the purpose of contracting the exact exchange hole in a manner similar to this effect in the electron gas with an effective density given by the geometric mean of the densities at $\mathbf{r}$ and $\mathbf{r}'$ Thus, we choose

$$B(r;n) = \left[1 + a(n)r/r_s + b(n)(r/r_s)^2\right]\mathrm{e}^{-c(n)r/r_s}$$

with an effective density given by (as usual $4\pi r_s^3 n = 3$)

$$n = \bar{n} = \sqrt{n(\mathbf{r})n(\mathbf{r}')}.$$

Two of the density-dependent parameters a, b, and c are chosen such that the model becomes exact for the electron gas (xc energy and sum rule), and the third can, e.g., be determined such that the theory gives the exact total energy for the He atom.

(*ii*) The factor A is based on a new idea and is what allows the xc hole to move to one side of a dimer when the reference electron goes there. It is determined by the sum rule prescribing one unit of charge in the xc hole:

$$A(\mathbf{r})\int |n(\mathbf{r},\mathbf{r}')|^2 A(\mathbf{r}')B(|\mathbf{r}-\mathbf{r}'|;\bar{n}(\mathbf{r},\mathbf{r}'))\,\mathrm{d}^3r' = 2n(\mathbf{r}).$$

Consequently, the correlation factor $A(\mathbf{r})$ is determined by by an integral equation which has to be solved at each point in space. Fortunately, the factor $A(\mathbf{r})$ varies slowly between one and two and few iterations are usually needed for determining this factor.

In comparison to previous models this one has a number of improvements:

- It contains the exact exchange energy.
- It is fully symmetric in the arguments $\mathbf{r}$ and $\mathbf{r}'$.
- It gives the correct energy densities and potentials outside finite systems and outside metallic surfaces.
- It gives a correct description of the dissociation of the Hydrogen molecule.
- It is exact in the slowly varying limit.

Unfortunately, we have not had the time to pursue this research much further but the model definitely appears to hold a lot of promise. In Fig. 1 and Fig. 2 we compare the correlation hole and the xc hole of the H_2 molecule at its equilibrium distance (~1.4 Bohr) to the corresponding quantities from the Screened Exchange Model. In both cases we show a plane through the molecular axis and the reference electron is on that axis and close to the Hydrogen atom to the right in the figures (at ~0.7 Bohr). As seen, the

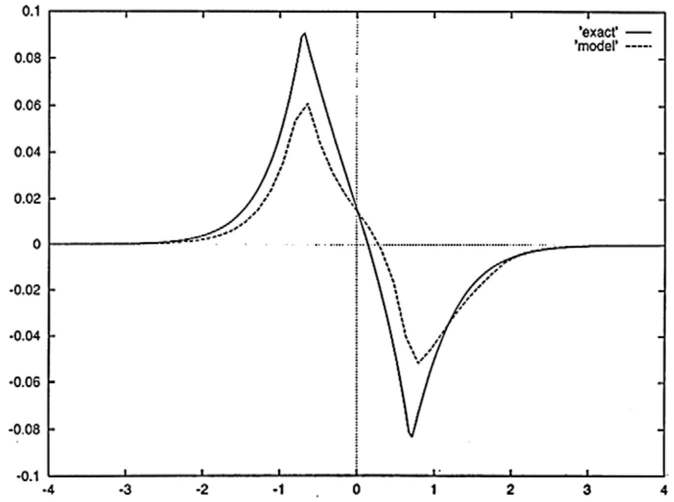

Fig. 1. A comparison of the true correlation hole of the H_2 molecule at the equilibrium distance to that (dotted) of the Screened Exchange model. The reference electron is close to the hydrogen atom to the right (at ~0.7 Bohr).

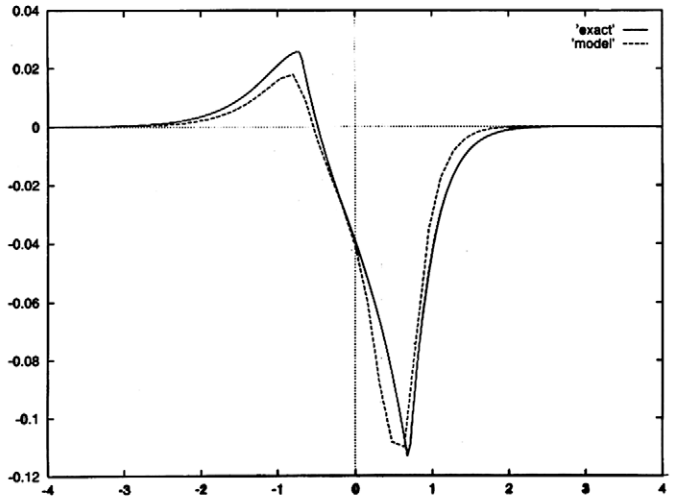

Fig. 2. A comparison of the true exchange-correlation hole of the H_2 molecule at equilibrium distance to that (dotted) of the Screened Exchange Model. The reference electron is on the molecular axis close to the hydrogen atom to the right (at ~0.7 Bohr).

screened-exchange model gives a very realistic description of the exact xc hole in the H_2 molecule.

One might also wonder if this model would stand a chance at describing also the very interesting van-der-Waal's (vdW) effects. After all, it seems to have several correct long-range features build into it. Our gut feeling at the moment is that the model is too simple minded for dealing with these extremely long-ranged effects. For such effects we refer to recent work by Langreth, Lundqvist and collaborators (see e.g. Ref. [128] and references therein). Their theory is, however, especially designed for treating the vdW effects and, with regard to correlations, it does not cover the whole range of distances from the vdW region into the interior of atoms.

7. Conclusions

We will end these lecture notes by voicing our prejudices concerning the present remaining problems within static DFT.

- *The band-gap problem.*
  The nature of correlations and also of the gradient part of the exchange energy is drastically altered by the introduction of a band gap. This represents an extremely non-local effect which we can never hope to capture using, e.g., gradients. It must be explicitly fed into any very accurate functional for the xc energy [129]. Most likely also into the correlation factor of the Screened-Exchange Model of Section 6.2.
- *The symmetry problem.*
  In finite systems, a difference in the symmetry of two states can have a strong effect on their energies while their densities can be very similar. None of the existing GGA:s or Meta-GGA:s have this effect built into them. This points toward orbital-dependent functionals.
- *The long-range problem.*
  There are two kinds of long-range problems—the van-der-Waal's problem and the extremely long-range problem associated with, e.g., the build up of a static polarization in an extended system. As in the case of band gaps, the latter problem might be solvable by feeding the polarization directly into the xc functional. There is, however, an interesting new approach in which this problem can be dealt with through the use of local approximations but within current density-functional (CDFT) theory rather than within standard DFT [130].
  There is so far no unified approach to the standard correlation problem in solids and molecules and the treatment of the van-der-Waal's interactions. Such an approach is more likely to be found among orbital-dependent functionals than among the GGA:s.
- *The near degeneracy problem.*
  This problem was discussed in Section 4.5 and is certainly outside any approach based on gradients. It could, perhaps be dealt with through models of the xc hole based on several determinants. But we do, at the moment, not know how to make this into a system independent approach. After all, a few determinants within one set of orbitals is an infinity of determinants in another.
- *The surface problem.*
  We are here referring to the problem of finding the xc energies coming from the surfaces of solids, from low-density regions in solids, or from the outskirts of atoms and molecules. This is the problem to which the GGA:s or the Meta-GGA:s are best suited and these are probably already accurate enough provided we do not require more of them than just that.
- *The core problem.*
  This is a very difficult problem within DFT which fortunately is rather irrelevant. We deal with this problem by either relying on a cancellation of errors or through the use of pseudo potentials. A word of caution, however. We are soon moving into the area of non-local xc functionals. These are non-local in the sense that, e.g., the xc energy density depends not only on the local value of the density and its different gradients but rather on some integrated property of the density. It is then not at all clear that such a theory can be pseudized to a very high degree of accuracy.

Thank you for your attention!

References

1. Gombás, "Die Statischtische Theorie des Atoms und Ihre Anwendungen," (Springer Verlag, Wien, 1949).
2. Thomas, L. H., Proc. Camb. Phil. Soc. **23**, 542 (1926).
3. Fermi, E., Z. Phys. **48**, 73 (1928).
4. Hohenberg, P. and Kohn, W., Phys. Rev. **136**, B864 (1964).
5. Kohn, W. and Sham, L. J., Phys. Rev. **140**, A1133 (1965).
6. Slater, J. C., Phys. Rev. **81**, 381 (1951).
7. Moruzzi, V. L., Janak, J. F. and Williams, A. R., "Calculated Electronic Properties of Metals," (Pergamon, New York 1978).
8. Langreth, D. C. and Perdew, J., Phys. Rev. B **21**, 5469 (1980).
9. Langreth, D. C. and Mehl, M. J., Phys. Rev. B **28**, 1809 (1983); ibid. Phys. Rev. B **29**, 2310E (1984).
10. Perdew, J. P., Burke, K. and Ernzerhof, M., Phys. Rev. Lett., **77**, 3865 (1996).
11. Becke, A. D., Phys. Rev. A **38**, 3098 (1988).
12. Pople, J. A., Head-Gordon, M., Fox, D. J., Raghavachari, K. and Curtiss, L. A., J. Chem. Phys. **90**, 5622 (1989).
13. von Barth, U. and Hedin, L., J. Phys. C **5**, 1629 (1972).
14. Pant, M. M. and Rajagopal, A. K., Solid State Commun. **10**, 1157 (1972).
15. Hamann, D. R., Schlüter, M. and Chiang, C., Phys. Rev. Lett. **43**, 1494 (1979).

16. von Barth, U. and Gelatt, C. D., Phys. Rev. B **21**, 2222 (1980).
17. von Barth, U. and Car, R., "Readily Constructed Norm-Conserving Pseudo Potentials and Exact Transferability," unpublished work, (IBM Thomas J. Watson Research Center, New York 1982). The theory is described in Ref. [90].
18. Peuckert, V., J. Phys. C **11**, 4945 (1978).
19. Runge, E. and Gross, E. K. U., Phys. Rev. Lett. **52**, 997 (1984).
20. Rajagopal, A. K. and Callaway, J., Phys. Rev. B **7**, 1912 (1973).
21. Rajagopal, A. K., J. Phys. C **11**, L943 (1978).
22. McDonald, A. H. and Vosko, S. H., J. Phys. C **12**, 2977 (1978).
23. Mermin, N. D., Phys. Rev. **137 A**, 1441 (1965).
24. Kreibich, T., and Gross, E. K. U., Phys. Rev. Lett. **86**, 2984 (2001).
25. Rajagopal, A. K., "Advances in Chemical Physics," (Edited by I. Prigogine and S. A. Rice) (Wiley, New York 1980) **41**, p. 59.
26. Williams, A. R. and von Barth, U., "Applications of Density-Functional Theory to Atoms, Molecules, and Solids," in Theory of the Inhomogeneous Electron Gas, (Edited by by S. Lundqvist and N. H. March), Physics of Solids and Liquids Series, (Plenum, New York 1983).
27. von Barth, U., "An Overview of Density-Functional Theory in Many-Body Phenomena at Surfaces," (Edited by D. Langreth and H. Suhi), (Academic Press, New York 1984) p. 3.
28. von Barth, U., "Density-Functional Theory for Solids in The Electronic Structure of Complex Systems" (Edited. by P. Phariseau and W. Temmerman) NATO ASI Series B: Physics (Plenum, New York 1984), **113**, p. 67.
29. Dahl, J. P. and Avery, J., ed. "Local-Density Approximations in Quantum Chemistry and Solid State Physics," (Plenum, New York 1984).
30. Callaway, J. and March, N. H., Solid State Phys. **38**, 135 (Edited by H. Ehrenreich, F. Seitz, and D. Turnball), (Academic Press, New York 1984).
31. Almbladh, C.-O. and von Barth, U., "Density-Functional Theory of Excitation Energies in Density-Functional Methods in Physics," (Edited by R. M. Dreizler and J. de Providencia), NATO ASI Series B: Physics, (Plenum, New York 1985), **123**, p. 209.
32. von Barth, U., Chem. Scripta **26**, 449 (1986).
33. Erdahl, R. and Smith, V. H., ed. "Density Matrices and Density Functionals," Proc. A. J. Coleman Symposium, (D. Reidel Publ. Co., Dordrecht 1987).
34. Jones, R. O. and Gunnarsson, O., Rev. Mod. Phys. **61**, 689 (1989).
35. Parr, R. G. and Yang, W., "Density-Functional Theory of Atoms and Molecules," (Oxford University Press, New York 1989).
36. Dreizler, R. M. and Gross, E. K. U., "Density-Functional Theory: An Approach to the Quantum Many-Body Problem," (Springer-Verlag, Berlin 1990).
37. Labanowski, J. K. and Andzelm, J. W., ed. of "Density-Functional Methods in Chemistry," (Springer-Verlag, New York 1991).
38. Perdew, J. P. and Kurth, S., "Density-Functionals for Non-Relativistic Coulomb Systems," in Density Functionals: Theory and Applications, (Edited by D. P. Joubert), Lecture Notes in Physics, (Springer, Berlin 1998), Vol. 500.
39. Kurth, S., Perdew, J. P. and Blaha, P., Int. J. Quantum Chem. **75**, 889 (1999).
40. Levy, M., Proc. Natl. Acad. Sci. USA, **76**, 6062 (1979).
41. Ceperley, D. M. and Alder, B. J., Phys. Rev. Lett. **45**, 566 (1980).
42. Vosko, S. H., Wilk, L. and Nusair, M., Can. J. Phys. **58**, 1200 (1980).
43. Perdew, J. P. and Zunger, A., Phys. Rev. B **23**, 5048 (1981).
44. Lindhard, J., Kgl. Danske Mat. Fys. Medd. **28**, 8 (1954).
45. Sham, L. J. and Schlüter, M., Phys. Rev. Lett. **51**, 1888 (1983).
46. Pines, D. and Nozières, P., "The Theory of Quantum Liquids," (Benjamin, New York 1966).
47. Svendsen, P. S. and von Barth, U., Phys. Rev. B **54**, 17402 (1996).
48. Geldart, D. J. W. and Taylor, R., Can. J. Phys. **48**, 167 (1970).
49. Engel, E. and Vosko, S. H., Phys. Rev. B **42**, 4940 (1990).
50. Ma, S.-K. and Brueckner, K. A., Phys. Rev. **165**, 18 (1968).
51. Geldart, D. J. W. and Rasolt, M., Phys. Rev. B **13**, 1477 (1976).
52. Rasolt, M., Phys. Rev. B **16**, 3234 (1977).
53. Perdew, J. P., Kurth, S., Zupan, A. and Blaha, P., Phys. Rev. Lett. **82**, 2544 (1999).
54. Gunnarsson, O., Jonson, M. and Lundqvist, B. I., Phys. Rev. B **20**, 3136 (1979).
55. Müller, J. E., Jones, R. O. and Harris, J., J. Chem. Phys. **79**, 1874 (1983).
56. Hu, C. D. and Langreth, D. C., Physica Scripta **32**, 391 (1985).
57. von Barth, U. and Car, R., "Ground State Properties of Bulk Si as a Test of the Langreth–Perdew–Mehl Scheme for Exchange and Correlation," unpublished work, IBM (Thomas J. Watson Research Center, New York 1982).
58. Janak, J. F., Moruzzi, V. L. and Williams, A. R., Phys. Rev. B **12**, 1257 (1975).
59. Janak, J. F. and Williams, A. R., Phys. Rev. B **14**, 4199 (1976).
60. Moruzzi, V. L., Williams, A. R. and Janak, J. F., Phys. Rev. B **15**, 2854 (1977).
61. Perdew, J. P., Burke, K. and Ernzerhof, M., Phys. Rev. Lett. **77**, 3865 (1996).
62. Perdew, J. P. Burke, K. and Ernzerhof, M., Phys. Rev. Lett. **78**, 1396E (1997).
63. Becke, A. D., Phys. Rev. A **33**, 2786 (1986).
64. Becke, A. D., J. Chem. Phys. **84**, 4524 (1986).
65. Williams, A. R., Kübler, J. and Gelatt, C. D., Phys. Rev. B **19**, 6094 (1979).
66. Gelatt, C. D., Williams, A. R. and Moruzzi, V. L., Phys. Rev. B **27**, 2005 (1983).
67. Janak, J. F. and Williams, A. R., Phys. Rev. B **14**, 4199 (1976).
68. Janak, J. F., Phys. Rev. B **16**, 255 (1977).
69. Janak, J. F., Solid State Commun. **25**, 53 (1978).
70. Janak, J. F., Phys. Rev. B **20**, 2206 (1979).
71. Terakura, K., Oguchi, T., Williams, A. R. and Kübler, J., Phys. Rev. B **30**, 4734 (1984).
72. Wang, C. S., Klein, B. M. and Krakauer, H., Phys. Rev. Lett. **54**, 1852 (1985).
73. Godby, R. W. and Needs, R. J., Phys. Rev. Lett. **62**, 1169 (1989).
74. Gunnarsson, O. and Jones, R. O., Phys. Rev. B **31**, 7588 (1985).
75. Slater, J. C., "Quantum Theory of Atomic Structure" (McGraw-Hill, New York 1960) Vol. I.
76. Kutzler, F. W. and Painter, G. S., Phys. Rev. B **43**, 6865 (1991).
77. Perdew, J. P. and Yue Wang, Phys. Rev. B **33**, 8800 (1986).
78. Perdew, J. P., Phys. Rev. B **33**, 8822 (1986).
79. Gunnarsson, O. and Johansson, P., Int. J. Quant. Chem. **X**, 307 (1976).
80. Merkle, R., Savin, A. and Preuss, H., J. Chem. Phys. **97**, 9216 (1992).
81. Mlynarski, P. and Salahub, D. R., Phys. Rev. B **43**, 1399 (1991).
82. Savin, A., Int. J. Quant. Chem.: Quant. Chem. Sympos. **22**, 59 (1988).
83. Bernholc, J. and Holzwarth, N. A. W., Phys. Rev. Lett. **50**, 1451 (1983).
84. Goodgame, M. M. and Goddard III, W. A., Phys. Rev. Lett. **48**, 135 (1982).
85. Cohen, M. and Bergstrasser, T. K., Phys. Rev. **141**, 789 (1966).
86. Harrison, W. A., "Pseudo-Potentials in the Theory of Metals," (Benjamin, New York 1966).
87. von Barth, U. and Stott, M., "Density-functional calculation without wave functions for pseudo systems," unpublished work, (Queen's University 1999).
88. Car, R. and Parrinello, M., Phys. Rev. Lett. **55**, 2471 (1985).
89. Herman, F., van Dyke, J. P. and Ortenburger, I. B., Phys. Rev. Lett. **22**, 807 (1969).
90. von Barth, U. and Pedroza, A. C., Physica Scripta **32**, 353 (1985).
91. Bagno, P., Jepsen, O. and Gunnarsson, O., Phys. Rev. B **40**, 1997 (1989).
92. Hu, C. D. and Langreth, D. C., Phys. Rev. B **33**, 943 (1986).
93. Perdew, J. P., Phys. Rev. Letters **55**, 1665 (1985).
94. Perdew, J. P., Physica B **172**, 1 (1991).
95. Gross, E. K. U. and Dreizler, R. M., Z. Phys. A **302**, 103 (1981).
96. Becke, A. D., J. Chem. Phys. **85**, 7184 (1986).
97. Perdew, J. P. *et al.*, Phys. Rev. B **46**, 6671 (1992).
98. Kong, X. J., Chan, C. T., Ho, K. M. and Ye, Y. Y., Phys. Rev. B **42**, 9357 (1990).
99. Leung, T. C., Chan, C. T. and Harmon, B. N., Phys. Rev. B **44**, 2923 (1991).
100. Garcia, A., Jing Zhu, Ch. Elsässer, Louie, S. G. and Cohen, M. L., Phys. Rev. B **46**, 9829 (1992).
101. Barbiellini, B., Moroni, E. G. and Jarlborg, T., J. Phys. Condens. Matter **2**, 7597 (1990).
102. Körling, M. and Hägglund, J., Phys. Rev. B **45**, 13293 (1992).
103. Kutzler, F. W. and Painter, G. S., Phys. Rev. Lett. **59**, 1285 (1987).
104. Bowen, C., Sugiyama, G. and Alder, B. J., Phys. Rev. B **50**, 14838 (1994).

105. Moroni, S., Ceperley, D. M. and Senatore, G., Phys. Rev. Lett. **75**, 689 (1995).
106. Becke, A. D., J. Chem. Phys. **96**, 2155 (1992).
107. Stoll, H., Pavlidou, C. M. E. and Preuss, H., Theoret. Chim. Acta. **49**, 143 (1978).
108. Curtiss, L. A., Jones, C., Trucks, G. W., Raghavachari, K. and Pople, J. A., J. Chem. Phys. **93**, 2537 (1990).
109. Becke, A. D., J. Chem. Phys. **97**, 9173 (1992).
110. Becke, A. D., J. Chem. Phys. **98**, 1372 (1993).
111. Becke, A. D., J. Chem. Phys. **98**, 5648 (1993).
112. Miehlich, B., Savin, A., Stoll, H. and Preuss, H., Chem. Phys. Lett. **157**, 200 (1989).
113. Johnsson, B. G., Gill, P. M. W. and Pople, J. A., J. Chem. Phys. **98**, 5612 (1993).
114. Becke, A. D., J. Chem. Phys. **88**, 1053 (1987).
115. Lee, C. L., Yang, W. and Parr, R. G., Phys. Rev. B **37**, 785 (1988).
116. Curtiss, L. A., Raghavachari, K., Redfern, P. C. and Pople, J. A., J. Chem. Phys. **106**, 1063 (1997).
117. Curtiss, L. A., Redfern, P. C., Raghavachari, K. and Pople, J. A., J. Chem. Phys. **109**, 42 (1998).
118. Curtiss, L. A., Raghavachari, K., Redfern, P. C. and Pople, J. A., J. Chem. Phys. **112**, 7374 (2000).
119. Merkle, R., Savin, A. and Preuss, H., Chem. Phys. Lett. **194**, 32 (1992).
120. Philipsen, P. H. T. and Baerends, E. J., Phys. Rev. B **54**, 5326 (1996).
121. Ghosh, S. K. and Parr, R. G., Phys. Rev. A **34**, 785 (1986).
122. van Voorhis, T. and Scuseria, G. E., J. Chem. Phys. **109**, 400 (1998).
123. Pryonov, E. I., Ruiz, E., Vela, A. and Salahub, D. R., Int. J. Quantum Chem. **29**, 61 (1995).
124. Becke, A. D., J. Chem. Phys. **109**, 2092 (1998).
125. Alonso, J. A. and Girifalco, L. A., Phys. Rev. B **17**, 3735 (1978).
126. Gunnarsson, O. and Jones, R. O., Physica Scripta **21**, 394 (1980).
127. van Leeuwen, R. and von Barth, U., "The Screened-Exchange Model within Density-Functional Theory," unpublished work (Lund University 1996).
128. Andersson, Y. *et al*., in "Electronic Density Functional Theory: Recent Progress and New Directions," (Edited by by J. F. Dobson, G. Vignale, and M. P. Das), (Plenum, New York 1997), p. 243.
129. I tried to sell this idea to Andreas Savin during the Quantum Chemistry Meeting in Montreal in 1985 but we never got the opportunity to collaborate on this exciting topic. Savin, Joe Krieger, and collaborators have done some work along these lines on atoms in: J. B. Krieger, J. Chen, G. J. Iafrate, and A. Savin, "Construction of an Accurate Self-Interaction Corrected Correlation Energy Functional Based on an Electron Gas with a Band Gap," p. 463, in the Proceedings of Electron Correlation and Material Properties (Edited By A. Gonis and N. Kioussis), (Plenum, New York 1999).
130. de Boeij, P. L., Kootstra, F., Berger, J. A., van Leeuwen, R. and Snijders, J. G., J. Chem. Phys. **115**, 1995 (2001).