

6 The Quest for Approximate Exchange-Correlation Functionals

In the previous chapter we introduced the Kohn-Sham formalism which allows an exact treatment of most of the contributions to the electronic energy of an atomic or molecular system, including the major fraction of the kinetic energy. All remaining – unknown – parts are collectively folded into the exchange-correlation functional $E_{XC}[\rho]$. These include the non-classical portion of the electron-electron interaction along with the correction for the self-interaction and the component of the kinetic energy not covered by the non-interacting reference system. Obviously, the whole endeavor of applying the Kohn-Sham scheme as a tool to get a grip on the Schrödinger equation makes sense only if explicit approximations to this functional are available. The quality of the density functional approach hinges solely on the accuracy of the chosen approximation to E_{XC} . Hence, the quest of finding better and better functionals is at the very heart of density functional theory. In the following we will review the current state of the art regarding approximate functionals for E_{XC} . We start out by showing that unlike in conventional wave function based methods, in density functional theory there is no systematic way towards improved approximate functionals, which in fact represents one of the major drawbacks associated with this approach. Then, we introduce the adiabatic connection, which provides the link between the non-classical potential energy of exchange and correlation and the E_{XC} functional of the Kohn-Sham scheme, with special emphasis on the corresponding hole function. The simple concept of the local density approximation based on the uniform electron gas, which represents the bedrock of almost all current functionals, is discussed. Even though this physical model performs better than anticipated, it is not accurate enough for chemical applications. Hence, ideas about how one can go beyond that approximation have been put forward by many researchers. We will develop the connection from the local density approximation to the more sophisticated generalized gradient approximation up to the nowadays so popular hybrid functionals. These general strategies are realized in many different individual functionals and the most widespread representatives as well as new developments for both classes are presented. We continue with a discussion of the problems due to the self-interaction of the charge density and to the behavior of the corresponding exchange-correlation potentials in the long range asymptotic region. Both aspects are inherent to all approximate exchange-correlation functionals and give rise to unwanted effects. The strengths and weaknesses of the various approaches will be discussed and we conclude this chapter with an assessment as to where future developments might lead.

6.1 Is There a Systematic Strategy?

Before we start looking at possible approximations to E_{XC} we need to address whether there will be some kind of guidance along the way. If we consider conventional, wave function based methods for solving the electronic Schrödinger equation, the quality of the

results solely depends on our choice of the approximate wave function. From basic concepts of linear algebra we in fact know the prescription for how the true wave function should be constructed in principle, such as in the full configuration interaction scheme, characterized by an expansion of both the one and the many particle problem in a complete, i. e., essentially infinite basis. Even though this can never be realized because the resulting equations would be much too complicated to be ever solved, this prescription shows us the way how the approximate wave functions can be improved step by step in a systematic manner. Unfortunately a similar beacon guiding us along the way towards our final, albeit unreachable destination does not exist in density functional theory. The origin of this sobering statement is simply that the explicit form of the exact functional is a total mystery to us. Not only is the physics underlying the success of current functionals far from being fully understood, we simply do not have the faintest idea how to arrive at approximations which are closer to the exact functional. All searching for better functionals relies largely on physical or mathematical intuition and has a strong ‘trial and error’ component. There are, however, a few physical constraints which a reasonable functional has to fulfill. Among those a prominent place is held by the sum rules valid for the exact exchange-correlation holes as outlined in Chapter 2. Of course, the more closely the model hole that emerges from an approximate exchange-correlation functional resembles the true hole, the better this functional will be able to account for the non-classical effects. Other properties include the cusp condition of the correlation hole at zero separation of the two electrons, certain scaling conditions of the exchange and correlation energies and asymptotic properties of the corresponding exchange-correlation potentials, etc.¹⁸ However, one should not expect too much help from such formal boundary conditions, since one of the baffling peculiarities of approximate density functional theory is that functionals which strictly meet these requirements are not necessarily better than others that do not. In fact, some of the most successful approximate functionals violate several of these conditions. It is therefore of immense importance to carefully study the performance of a particular functional with respect to a suitable set of reference data (we will elaborate on these decisive aspects in much detail in the second part of this book). Indeed, the most stringent tests currently available for new functionals are completely empirical and involve the comparison with accurate reference data, such as atomization, ionization and reaction energies, structural data and the like. The most frequently used set of energetic reference data is probably the so-called G2 thermochemical data base which contains more than 50 experimentally well established atomization energies of small molecules containing main group elements originally collected by Curtiss et al., 1991. The ability to reproduce the energetics of this data base or extensions to it has become the de facto standard for measuring the accuracy of a new computational method. The target accuracy is the so-called chemical accuracy which corresponds to an average absolute error of about 0.1 eV or 2 kcal/mol. However, this is a very ambitious goal and up to now only very few and very expensive traditional quantum chemical strategies are able to achieve this kind of accuracy. The performance of approximate

¹⁸ We will not digress on that matter here but rather refer the interested reader to the relevant literature, see Perdew and Burke, 1996, or Adamo, di Matteo, and Barone, 1999.

functionals with regard to the G2 and related reference sets will occupy us in significant detail in Section 9.1. We should also mention in this context that the energy delivered by a particular functional is not the ultimate probe for its quality. The exchange-correlation energy results from the integral over the exchange-correlation potential and in principle the correct energy can be obtained even from an erroneous potential because of a fortuitous error cancellation. Therefore, a more physically motivated test is provided by a point-by-point comparison between the model exchange-correlation potential and accurate potentials derived from high-quality correlated wave function based calculations. However, such accurate potentials are rarely available and this kind of validation is seldom used.

6.2 The Adiabatic Connection

The purpose of this chapter is the illustration of the ways how a good approximation to the exact exchange-correlation functional of Kohn-Sham theory can be found. But before we proceed we need to take a second look at this very quantity and relate E_{XC} with the concept of exchange-correlation holes introduced earlier. The hole functions we discussed in Chapter 2 contained all information about the non-classical contributions to the *potential energy* due to the electron-electron interaction, E_{ncl} . However, we saw in the preceding chapter that E_{XC} as defined in the framework of Kohn-Sham theory also accounts for T_C , the difference between the *kinetic energy* of the real, fully interacting system, T , and the kinetic energy T_S related to the non-interacting reference system,

$$E_{XC}[\rho] = \{T[\rho] - T_S[\rho]\} + E_{ncl}[\rho] = T_C[\rho] + E_{ncl}[\rho]. \quad (6-1)$$

Thus, the information about $T[\rho] - T_S[\rho]$ must be somehow folded into the corresponding hole functions. To do this, imagine that we connect the two systems central for the KS scheme (i. e. the non-interacting reference with no $1/r_{ij}$ electron-electron interaction and the real one where this interaction is operative with full strength) by gradually increasing the *coupling strength parameter* λ from 0 to 1:

$$\hat{H}_\lambda = \hat{T} + V_{ext}^\lambda + \lambda \sum_i \sum_{j>i}^N \frac{1}{r_{ij}}. \quad (6-2)$$

For each λ the effective external potential V_{ext}^λ is adapted such that the density always equals the density of the fully interacting system. Hence $\rho(\vec{r})$ is independent of the actual value of λ . Clearly, for $\lambda = 0$, we recover the Hamilton operator for the non-interacting reference system, and $V_{ext}^{\lambda=0} = V_S$, while $\lambda = 1$ is the regular expression for the real system

with $V_{ext}^{\lambda=1} + \sum_i \sum_{j>i}^N \frac{1}{r_{ij}} = V_{eff} = \int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2 + V_{XC}(\vec{r}_1) - \sum_A \frac{Z_A}{r_{1A}}$. Equation (6-2) describes

how these two endpoints are smoothly connected through a continuum of artificial, partially interacting systems. Borrowing from thermodynamics this path is called the *adi-*

adiabatic connection. In terms of the adiabatic connection the energy of the interacting system $E_{\lambda=1}$ can be expressed as the following integral

$$E_{\lambda=1} - E_{\lambda=0} = \int_0^1 dE_{\lambda}, \text{ and thus } E_{\lambda=1} = \int_0^1 dE_{\lambda} + E_{\lambda=0} \quad (6-3)$$

To utilize this relation we now need an explicit expression for dE_{λ} . To this end we investigate how the total energy E_{λ} changes upon an infinitesimal change in the coupling strength λ . This energy is the expectation value of the corresponding Hamiltonian

$$\hat{H}_{\lambda} = dV_{\text{ext}}^{\lambda} + d\lambda \sum_i^N \sum_{j>i}^N \frac{1}{r_{ij}}, \quad (6-4)$$

and, using the hole formalism, can be expressed as

$$\begin{aligned} dE_{\lambda} = & \int \rho(\vec{r}) dV_{\text{ext}}^{\lambda} d\vec{r} + \frac{1}{2} d\lambda \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \\ & + \frac{1}{2} d\lambda \iint \frac{\rho(\vec{r}_1)h_{\text{XC}}^{\lambda}(\vec{r}_1; \vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \end{aligned} \quad (6-5)$$

Inserting equation (6-5) in the integral of equation (6-3) leads to

$$\begin{aligned} E_{\lambda=1} - E_{\lambda=0} = & \int \rho(\vec{r}) \left[V_{\text{ext}}^{\lambda=1} - V_{\text{ext}}^{\lambda=0} \right] d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \\ & + \frac{1}{2} \iint \int_0^1 \frac{\rho(\vec{r}_1)h_{\text{XC}}^{\lambda}(\vec{r}_1; \vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 d\lambda \end{aligned} \quad (6-6)$$

where we have made use of the λ -independence of the density $\rho(\vec{r})$. Replacing $V_{\text{ext}}^{\lambda=0}$ and $V_{\text{ext}}^{\lambda=1}$ by V_S and V_{eff} , respectively and using the energy expression for the non-interacting Kohn-Sham system,

$$E_{\lambda=0} = T_S + \int \rho(\vec{r}) V_S d\vec{r} \quad (6-7)$$

and defining the *coupling-strength integrated exchange-correlation hole* \bar{h}_{XC} as

$$\bar{h}_{\text{XC}}(\vec{r}_1; \vec{r}_2) \equiv \int_0^1 h_{\text{XC}}^{\lambda}(\vec{r}_1; \vec{r}_2) d\lambda \quad (6-8)$$

we finally arrive at the following equation for the energy of the real, interacting system

$$\begin{aligned}
E_{\lambda=1} = T_S + \int \rho(\vec{r}) V_{\text{eff}} d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1) \rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \\
+ \frac{1}{2} \iint \frac{\rho(\vec{r}_1) \bar{h}_{\text{XC}}(\vec{r}_1; \vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2
\end{aligned} \tag{6-9}$$

Thus, the important take-home message is that the exchange-correlation energy of the Kohn-Sham scheme can be expressed through the coupling-strength integrated exchange-correlation hole \bar{h}_{XC} . If we know this hole, we know the exchange-correlation energy as demonstrated in the following expression

$$E_{\text{XC}} = \frac{1}{2} \iint \frac{\rho(\vec{r}_1) \bar{h}_{\text{XC}}(\vec{r}_1; \vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2. \tag{6-10}$$

What does this mean? If we compare equation (6-9) which we just derived with its counterpart which can be deduced from equation (2-19) of Chapter 2 (by adding to the E_{ee} term of the electron-electron interaction described in that expression the kinetic energy and the contribution from the external potential)

$$\begin{aligned}
E = T + \int \rho(\vec{r}) V_{\text{eff}} d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1) \rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 \\
+ \frac{1}{2} \iint \frac{\rho(\vec{r}_1) h_{\text{XC}}(\vec{x}_1; \vec{x}_2)}{r_{12}} d\vec{x}_1 d\vec{x}_2
\end{aligned} \tag{6-11}$$

where T is the *true* kinetic energy of the fully interacting system and the last term the non-classical contribution to the electron-electron repulsion, we see that the integration over the coupling-strength elegantly transfers the difference between T and T_S , i. e. the part of the kinetic energy not covered by the non-interacting reference system, into the exchange-correlation hole. In other words, when going from equation (6-11) to equation (6-9) we drastically simplify the expression for the kinetic energy (T) by reducing it to the kinetic energy of the non-interacting reference system (T_S). The price we pay for this is a further complication in the exchange-correlation hole brought about by the additional integration over the coupling strength parameter λ , i. e., we replace h_{XC} by \bar{h}_{XC} . Importantly, this integration has no effect on the formal properties of the exchange-correlation hole discussed in Chapter 2, the sum rules, cusp conditions, etc. that apply to h_{XC} apply as well to \bar{h}_{XC} . For details on this very fundamental concept see the contributions of Gunnarsson and Lundqvist, 1976, Harris, 1984, Becke, 1988a, 1995, and Jones and Gunnarsson, 1989.

6.3 From Holes to Functionals

As already alluded to above, the analysis of the properties of model hole functions that emerge from approximate exchange-correlation functionals is a major tool for assessing

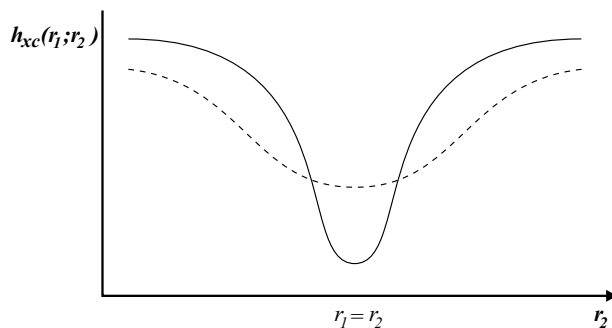


Figure 6-1. Fermi holes of different depths for the on-top density.

the suitability of such functionals. Recall that in Chapter 2 we separated the total exchange-correlation hole into two components, the Fermi and Coulomb holes, with the former being by far the most important contribution to the total hole. The Fermi hole was identified as a non-positive quantity which contains exactly one elementary charge. This is a fairly stringent restriction. For example, for ‘normal’ positions, where the hole is concentrated around the reference electron it follows that the ‘deeper’ the hole gets for $\vec{r}_2 \rightarrow \vec{r}_1$ (where it assumes $-\rho(\vec{r}_1)$) the less will it extend into space, i. e., the shorter its range, as schematically shown in Figure 6-1. We should note that the exchange-correlation hole for the special case that $\vec{r}_2 = \vec{r}_1$ is called the *on-top hole*, which has attracted considerable attention lately, see, e. g., Perdew et al., 1997, and Burke, Perdew, and Ernzerhof, 1998.

On the other hand, the Coulomb hole integrates to zero and can be negative as well as positive. Hence this sum rule is of only minor help. There is one additional, very important aspect in this context. The exact hole functions are highly asymmetric entities and it will be very difficult for any approximate hole to recover all the subtle details of its six-dimensional shape (remember that the hole depends on the coordinates of two electrons). However, the expression for the exchange-correlation energy given in equation (6-10) does not rely on the angular details of \bar{h}_{xc} because of the clearly isotropic character of the Coulomb interaction represented by the $1/r_{ij}$ operator and only depends on the distance between any two electrons. Hence, our approximate hole only has to model the *spherically averaged* exact hole about each reference point, which is significantly less complicated (but still complicated enough).

6.4 The Local Density and Local Spin-Density Approximations

In this section we introduce the model system on which virtually all approximate exchange-correlation functionals are based. At the center of this model is the idea of a hypothetical *uniform electron gas*. This is a system in which electrons move on a positive background charge distribution such that the total ensemble is electrically neutral. The number of elec-

trons N as well as the volume V of the gas are considered to approach infinity, while the electron density, i. e., N/V remains finite, $N \rightarrow \infty$, $V \rightarrow \infty$, $N/V = \rho$ and attains a constant value everywhere. Physically, such a situation resembles the model of an idealized metal consisting of a perfect crystal of valence electrons and positive cores where the cores are smeared out to arrive at a uniform positive background charge. Indeed, the uniform electron gas is a fairly good physical model for simple metals such as sodium. On the other hand, we should note from the start that this model system, which is also known under the label of the homogeneous electron gas, is pretty far from any realistic situation in atoms or molecules, which are usually characterized by rapidly varying densities. The reason why the uniform electron gas has such a prominent place in density functional theory is that it is the only system for which we know the form of the exchange and correlation energy functionals exactly or at least to very high accuracy. We actually already met the exchange functional of this model system in Chapter 3 when we briefly discussed the Dirac exchange functional that appears in the Thomas-Fermi-Dirac method. The idea to use this model for approximating E_{XC} in the Kohn-Sham scheme was already included in the original paper by Kohn and Sham, 1965. Let us tackle the problem now from a slightly different point of view. Central to this model is the assumption that we can write E_{XC} in the following, very simple form

$$E_{XC}^{LDA}[\rho] = \int \rho(\vec{r}) \epsilon_{XC}(\rho(\vec{r})) d\vec{r}. \quad (6-12)$$

Here, $\epsilon_{XC}(\rho(\vec{r}))$ is the exchange-correlation energy per particle of a uniform electron gas of density $\rho(\vec{r})$. This energy per particle is weighted with the probability $\rho(\vec{r})$ that there is in fact an electron at this position in space. Writing E_{XC} in this way defines the *local density approximation*, LDA for short. The quantity $\epsilon_{XC}(\rho(\vec{r}))$ can be further split into exchange and correlation contributions,

$$\epsilon_{XC}(\rho(\vec{r})) = \epsilon_X(\rho(\vec{r})) + \epsilon_C(\rho(\vec{r})). \quad (6-13)$$

The exchange part, ϵ_X , which represents the exchange energy of an electron in a uniform electron gas of a particular density is, apart from the pre-factor, equal to the form found by Slater in his approximation of the Hartree-Fock exchange (Section 3.3) and was originally derived by Bloch and Dirac in the late 1920's:

$$\epsilon_X = -\frac{3}{4} \sqrt{\frac{3 \rho(\vec{r})}{\pi}}. \quad (6-14)$$

Inserting equation (6-14) into equation (6-12) retrieves the $\rho^{4/3}$ dependence of the exchange energy indicated in equation (3-5). This exchange functional is frequently called *Slater exchange* and is abbreviated by S . No such explicit expression is known for the correlation part, ϵ_C . However, highly accurate numerical quantum Monte-Carlo simulations of the homogeneous electron gas are available from the work of Ceperly and Alder, 1980.

On the basis of these results various authors have presented analytical expressions of ϵ_C based on sophisticated interpolation schemes. The most widely used representations of ϵ_C are the ones developed by Vosko, Wilk, and Nusair, 1980, while the most recent and probably also most accurate one has been given by Perdew and Wang, 1992. The common short hand notation for the former implementations of the correlation functional is VWN. Hence, instead of the abbreviation LDA, which defines the model of the local density approximation, one frequently finds the acronym SVWN to identify the particular functional. Note that in their paper Vosko, Wilk, and Nusair report several expressions for ϵ_C . VWN usually implies that the correlation energy density of the homogeneous electron gas has been obtained in the random phase approximation (RPA), while the somewhat less frequently used VWN5 variant (note that this is the one recommended by the authors) denotes the use of the parameterization scheme based upon the results of Ceperly and Alder. Even though these two VWN functionals in most cases perform similarly (Hertwig and Koch, 1997. But see Section 9.4 for examples where VWN and VWN5 perform differently) one should be cautious about which flavor of the VWN functional is actually implemented in the corresponding computer program in order to avoid confusion. Before we go on, we pause for a minute to make a general remark on the nomenclature found in the literature to name a particular functional. While there is no strict rule, most authors now term the functionals as ‘XC’ where X stands for the exchange part and C for the correlation part as described by the initial letter of the names of the corresponding authors. The letters are augmented by a year, if the same authors developed more than one functional. If the exchange and correlation parts are due to the same authors, the letters are usually given only once.

In the preceding chapter we mentioned that approximate functionals are usually also expressed in an unrestricted version, where not the electron density $\rho(\vec{r})$, but the two spin densities, $\rho_\alpha(\vec{r})$ and $\rho_\beta(\vec{r})$, with $\rho_\alpha(\vec{r}) + \rho_\beta(\vec{r}) = \rho(\vec{r})$ are employed as the central input. Even though from a puristic theoretical point of view the exact functional will not depend on the spin densities (as long as the external potential is spin-independent), approximations to it will benefit from the additional flexibility of having two instead of one variable. In particular, for open-shell situations with an unequal number of α and β electrons, functionals of the two spin densities consistently lead to more accurate results. But also for certain situations with an even number of electrons, such as the H_2 molecule at larger separation, the unrestricted functionals perform significantly better because they allow symmetry breaking. Up to this point the local density approximation was introduced as a functional depending solely on $\rho(\vec{r})$. If we extend the LDA to the unrestricted case, we arrive at the *local spin-density approximation*, or LSD. Formally, the two approximations differ only that instead of equation (6-12) we now write

$$E_{XC}^{LSD}[\rho_\alpha, \rho_\beta] = \int \rho(\vec{r}) \epsilon_{XC}(\rho_\alpha(\vec{r}), \rho_\beta(\vec{r})) d\vec{r}. \quad (6-15)$$

Just as for the simple, spin compensated situation where $\rho_\alpha(\vec{r}) = \rho_\beta(\vec{r}) = \frac{1}{2}\rho(\vec{r})$, there are related expressions for the exchange and correlation energies per particle of the uniform electron gas characterized by $\rho_\alpha(\vec{r}) \neq \rho_\beta(\vec{r})$, the so-called spin polarized case. The degree of spin polarization is often measured through the spin-polarization parameter

$$\xi = \frac{\rho_{\alpha}(\vec{r}) - \rho_{\beta}(\vec{r})}{\rho(\vec{r})}. \quad (6-16)$$

ξ attains values from 0 (spin compensated) to 1 (fully spin polarized, i. e., all electrons have only one kind of spin). For details see in particular Appendix E of Parr and Yang, 1989. In the following we do not differentiate between the local and the local spin-density approximation and use the abbreviation LDA for both, unless otherwise noted.

How do we interpret the LDA for the exchange-correlation functional? Let us consider the general case of an open-shell atom or molecule. At a certain position \vec{r} in this system we have the corresponding spin densities $\rho_{\alpha}(\vec{r})$ and $\rho_{\beta}(\vec{r})$. In the local spin-density approximation we now take these densities and insert them into equation (6-15) obtaining $E_{XC}(\vec{r})$. Thus, we associate with the densities $\rho_{\alpha}(\vec{r})$ and $\rho_{\beta}(\vec{r})$ the exchange and correlation energies and potentials that a homogeneous electron gas of equal, but constant density and the same spin polarization ξ would have. This is now repeated for each point in space and the individual contributions are summed up (integrated) as schematically indicated in Figure 6-2. Obviously, this approximation hinges on the assumption that the exchange-correlation potentials depend only on the *local* values of $\rho_{\alpha}(\vec{r})$ and $\rho_{\beta}(\vec{r})$.

This is a very drastic approximation since, after all, the density in our actual system is certainly anything but constant and does not even come close to the situation characteristic of the uniform electron gas. As a consequence, one might wonder whether results obtained with such a crude model will be of any value at all. Somewhat surprisingly then, experience tells us that the local (spin) density approximation is actually not that bad, but rather deliv-

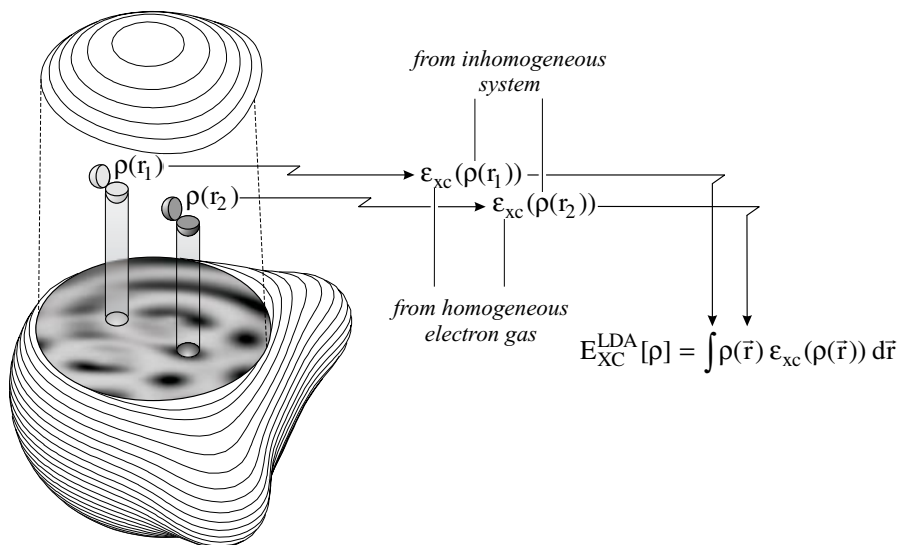


Figure 6-2. The local density approximation.

ers results that are comparable to or even better than the Hartree-Fock approximation. It has proven particularly successful for the determination of molecular properties such as equilibrium structures, harmonic frequencies or charge moments as we will discuss in more detail in later chapters. However, we should not get overexcited about such observations, because a look at energetical details, such as bond energies, immediately shows that for such properties the performance of the LDA is rather poor. If we take the average unsigned deviation from the experimental atomization energies from the G2 data set as an indicator, the LDA deviates by 36 kcal/mol! On the other hand, we need to put this into perspective because the deviation of the HF method is even substantially larger: its error is a hefty 78 kcal/mol, more than twice as large as the error of the local density approximation! While the HF approximation typically underestimates atomization energies, the LDA errs in the opposite direction, giving rise to the notorious overbinding tendency of this approximation, which we will discuss in Chapter 9.

In anticipation of the future discussion on how to improve on the local density approximation, we need to ask ourselves, what are the reasons that the LDA works better than expected from the underlying physical model of the homogeneous electron gas? The clue for an understanding seems to be that the exchange-correlation hole of the uniform electron gas, which is being used as model for the exact hole in the local density approximation, satisfies most of the important relations established for the true hole. Among those are the sum-rules, the behavior for vanishing inter-electronic distance of the exchange part ($h_X(\vec{r}_2 \rightarrow \vec{r}_1; \vec{r}_1) = -\rho(\vec{r}_1)$) and the correlation part (cusp condition), as well as the property of the exchange hole to be negative everywhere. Of course, the LDA model hole and the exact hole differ in many details. The overbinding tendency of the local density approximation can be rationalized in terms of the exchange hole properties. The LDA hole is spherically symmetric and always attached to the reference electron while the exact hole has a pronounced angular structure. In the bonding region between two atoms, the LDA model hole resembles the exact exchange hole, which becomes more isotropic (and symmetric with respect to the reference electron) than in the separated atoms. In an atom, the exact exchange hole is displaced toward the nucleus, whereas the LDA hole remains as it was in the molecular bond: centered on its reference electron. The neglect of this displacement in the LDA causes significant deviations from the exact differential exchange energy upon bond formation, with substantial errors on the atomic asymptote of an atomization process. In other words: the centered LDA exchange hole is a better approximation for the more homogeneous molecular density than for the more inhomogeneous density of atoms. This causes in particular the exchange energy of the molecular system to be too negative, that is, causes a dramatic overbinding (Ernzerhof, Perdew, and Burke, 1997).

Fortunately, only the spherically averaged exchange-correlation hole is of relevance for the exchange-correlation energy, as shown in the preceding section. The agreement between the spherically averaged LDA and exact holes is indeed much better – which is among the reasons why the LDA works at all – and the homogeneous electron gas in fact provides a reasonable first approximation to spherically averaged exchange-correlation holes of real systems. However, we should keep in mind that the LDA hole $h_{XC}^{LDA}(\vec{r}_1; \vec{r}_2)$ will have its highest degree of accuracy for small distances between the reference and the other

electron because in the local density approximation we treat the exchange-correlation hole around \vec{r}_1 as if the neighborhood were part of a homogenous electron gas of constant density. Clearly, in a real system with considerably varying charge density, this assumption will deteriorate the larger the distance between the reference electron at \vec{r}_1 and the other one at \vec{r}_2 is.

6.5 The Generalized Gradient Approximation

The only moderate accuracy that the local (spin) density approximation delivers is certainly insufficient for most applications in chemistry. Hence, for the many years in which the LDA was the only approximation available for E_{XC} , density functional theory was mostly employed by solid-state physicists and hardly had any impact in computational chemistry. The situation changed significantly in the early eighties when the first successful extensions to the purely local approximation were developed. The logical first step in that direction was the suggestion of using not only the information about the density $\rho(\vec{r})$ at a particular point \vec{r} , but to supplement the density with information about the *gradient* of the charge density, $\nabla\rho(\vec{r})$ in order to account for the non-homogeneity of the true electron density. In other words, we interpret the local density approximation as the first term of a Taylor expansion of the uniform density and expect to obtain better approximations of the exchange-correlation functional by extending the series with the next lowest term. Thus, we arrive at (with σ and σ' indicating α or β spin)

$$E_{XC}^{GEA}[\rho_\alpha, \rho_\beta] = \int \rho \, \varepsilon_{XC}(\rho_\alpha, \rho_\beta) \, d\vec{r} + \sum_{\sigma, \sigma'} \int C_{XC}^{\sigma, \sigma'}(\rho_\alpha, \rho_\beta) \frac{\nabla \rho_\sigma}{\rho_\sigma^{2/3}} \frac{\nabla \rho_{\sigma'}}{\rho_{\sigma'}^{2/3}} \, d\vec{r} + \dots \quad (6-17)$$

This form of functional is termed the *gradient expansion approximation* (GEA) and it can be shown that it applies to a model system where the density is not uniform but very slowly varying. Unfortunately, and at first glance counterintuitively, if utilized to solve real molecular problems the GEA does not lead to the desired improved accuracy but frequently performs even worse than the simple local density approximation. The reason for this failure is that the exchange-correlation hole associated with a functional such as in equation (6-17) has lost many of the properties which made the LDA hole physically meaningful. For example, the sum rules do not apply any more and the exchange hole is not restricted to be negative for any pair $\vec{r}_1; \vec{r}_2$. Thus, the dependence between the depth of the on-top hole and its extension is lost and the holes as well as the corresponding exchange-correlation energies will be much more erratic. This shows again that it is not so much the model system of the uniform electron gas but much more so the fact that the corresponding exchange-correlation hole system obeys most of the rules of the real system which is responsible for the success of the local density and local spin-density approximations.

In a very elegant (or shall we say brute force) way, this problem was solved by straightforwardly enforcing the restrictions valid for the true holes also for the hole of the beyond-LDA functionals. If there are parts in the GEA exchange holes which violate the requirement of being negative everywhere, just set them to zero. And, in order to have the correct sum rule behavior, well, let us simply truncate the exchange and correlation holes such that $h_X(\vec{r}_1; \vec{r}_2)$ and $h_C(\vec{r}_1; \vec{r}_2)$ contain one and zero electron charges, respectively. Functionals that include the gradients of the charge density and where the hole constraints have been restored in the above manner are collectively known as *generalized gradient approximations* (GGA). These functionals are the workhorses of current density functional theory and can be generically written as

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla \rho_\alpha, \nabla \rho_\beta) d\vec{r}. \quad (6-18)$$

As we will see presently, several suggestions for the explicit dependence of the integrand f on the densities and their gradients exist, including semiempirical functionals which contain parameters that are calibrated against reference values rather than being derived from first principles. In practice, E_{XC}^{GGA} is usually split into its exchange and correlation contributions

$$E_{XC}^{GGA} = E_X^{GGA} + E_C^{GGA} \quad (6-19)$$

and approximations for the two terms are sought individually.

Let us take a closer look at gradient-corrected exchange functionals in order to illustrate the general ideas. In particular, the reader should convince him- or herself that we are dealing with mathematically complex constructs which have been chosen such that the desired boundary conditions which the functionals and corresponding hole functions should satisfy are fulfilled and a satisfactory performance results. One should be aware that it is not the physics but the results obtained from them which dictate the choice of the mathematical constructs. In fact, some of these functionals are not even based on any physical model. In other words, the actual form of E_X^{GGA} and E_C^{GGA} usually does not assist the understanding of the physics these functionals try to describe. This underlines the pragmatic character so typical for approximate density functional theory in general.

We rewrite the exchange part of E_{XC}^{GGA} as

$$E_X^{GGA} = E_X^{LDA} - \sum_{\sigma} \int F(s_{\sigma}) \rho_{\sigma}^{4/3}(\vec{r}) d\vec{r}. \quad (6-20)$$

The argument of the function F is the *reduced density gradient* for spin σ

$$s_{\sigma}(\vec{r}) = \frac{|\nabla \rho_{\sigma}(\vec{r})|}{\rho_{\sigma}^{4/3}(\vec{r})}. \quad (6-21)$$

s_{σ} is to be understood as a local inhomogeneity parameter. It assumes large values not only for large gradients, but also in regions of small densities, such as the exponential tails

far from the nuclei. Likewise, small values of s_σ occur for small gradients, typical for bonding regions, but also for regions of large density. For example, the combination of large density gradients and large densities close to the nuclei typically leads to values of s_σ in this region which are in between the reduced density gradients in the bonding and tail regions, respectively. Of course, the homogeneous electron gas is characterized by $s_\sigma = 0$ everywhere. Finally, a word on why we divide by the $4/3$ power of ρ and not just by ρ itself. This is needed to make s_σ a dimensionless quantity: the dimension of the density is the inverse dimension of volume and hence $[r]^{-3}$. Its gradient has therefore dimensions of $[r]^{-4}$. But this is just the same dimension that $\rho^{4/3}$ has, because of $([r]^{-3})^{4/3} = [r]^{-4}$ and we arrive at the desired dimensionless reduced gradient.

For the function F two main classes of realizations have been put forward (see in particular Adamo, di Matteo, and Barone, 1999). The first one is based on a GGA exchange functional developed by Becke, 1988b. As outlined above, this functional is abbreviated simply as B (sometimes one also finds $B88$)

$$F^B = \frac{\beta s_\sigma^2}{1 + 6\beta s_\sigma \sinh^{-1} s_\sigma}. \quad (6-22)$$

β is an empirical parameter that was determined to 0.0042 by a least-squares fit to the exactly known exchange energies of the rare gas atoms He through Rn. In addition to the sum rules, this functional was designed to recover the exchange energy density asymptotically far from a finite system.

Functionals which are related to this approach include among others the recent FT97 functional of Filatov and Thiel, 1997, the PW91 exchange functional (Perdew, 1991, and Burke, Perdew, and Wang, 1998) and the CAM(A) and CAM(B) functionals developed by Handy and coworkers (Laming, Termath, and Handy, 1993).

The second class of GGA exchange functionals use for F a rational function of the reduced density gradient. Prominent representatives are the early functionals by Becke, 1986 (B86) and Perdew, 1986 (P), the functional by Lacks and Gordon, 1993 (LG) or the recent implementation of Perdew, Burke, and Ernzerhof, 1996 (PBE). As an example, we explicitly write down F of Perdew's 1986 exchange functional, which, just as for the more recent PBE functional, is free of semiempirical parameters:

$$F^{P86} = \left(1 + 1.296 \left(\frac{s_\sigma}{(24\pi^2)^{1/3}} \right)^2 + 14 \left(\frac{s_\sigma}{(24\pi^2)^{1/3}} \right)^4 + 0.2 \left(\frac{s_\sigma}{(24\pi^2)^{1/3}} \right)^6 \right)^{1/15}. \quad (6-23)$$

The corresponding gradient-corrected correlation functionals have even more complicated analytical forms and cannot be understood by simple physically motivated reasonings. We therefore refrain from giving their explicit expressions and limit ourselves to a more qualitative discussion of the most popular functionals. Among the most widely used choices is the correlation counterpart of the 1986 Perdew exchange functional, usually termed P or $P86$. This functional employs an empirical parameter, which was fitted to the

correlation energy of the neon atom. A few years later Perdew and Wang, 1991, refined their correlation functional, leading to the parameter free PW91. Another, nowadays even more popular correlation functional is due to Lee, Yang, and Parr, 1988 (LYP). Unlike all the other functionals mentioned so far, LYP is not based on the uniform electron gas but is derived from an expression for the correlation energy of the helium atom based on an accurate, correlated wave function presented in the context of wave function based theory by Colle and Salvetti, 1975. The LYP functional contains one empirical parameter. It differs from the other GGA functionals in that it contains some local components. We should note that all these correlation functionals are based on systems that only include *dynamical*, i. e., short range correlation effects (the uniform electron gas or the helium atom). Non-dynamical effects are not covered by these functionals, a property that we will come back to in the next section.

In principle, each exchange functional could be combined with any of the correlation functionals, but only a few combinations are currently in use. The exchange part is almost exclusively chosen to be Becke's functional which is either combined with Perdew's 1986 correlation functional or the Lee, Yang, Parr one – levels usually abbreviated as BP86 and BLYP, respectively. Sometimes also the PW91 correlation functional is employed, corresponding to BPW91. To be fair, all these flavors of gradient-corrected KS-density functional theory deliver results of similar quality as demonstrated by several studies which assess the performance of these functional. However, in this chapter we will predominantly concentrate on the more formal theoretical aspects of functionals and postpone a detailed view on the actual performance of modern functionals to our discussion in Part B.

We finally note a semantic detail. GGA functionals are frequently termed *non-local* functionals in the literature. This is a somewhat misleading and actually sloppy terminology that should be avoided. In our discussion of the Hartree-Fock scheme in Section 1.3 we introduced the difference between local and non-local operators and showed that the classical Coulomb potential is a local one while the HF exchange contribution represents a typical non-local potential. According to this discussion, all GGA functionals are perfectly local in the mathematical sense: the value of the functional at a point \vec{r} depends only on information about the density $\rho(\vec{r})$, its gradient $\nabla\rho(\vec{r})$, and possibly other information at this very point and is absolutely independent of properties of $\rho(\vec{r}')$ at points $\vec{r}' \neq \vec{r}$. Calling these functionals 'non-local' is only motivated by the fact that these functionals go beyond the 'local' density approximation and of course the observation that knowledge of the gradients is the first step towards accounting for the inhomogeneity of the real density; nevertheless it is sloppy physicists' jargon.

6.6 Hybrid Functionals

We have repeatedly indicated that usually the exchange contributions are significantly larger in absolute numbers than the corresponding correlation effects. Therefore, an accurate expression for the exchange functional in particular is a prerequisite for obtaining meaningful results from density functional theory. However, we have seen in Chapters 1 and 5 that the

exchange energy of a Slater determinant can be computed exactly, recall equations (5-17) or (5-18). Thus, why do we bother with complicated, but nevertheless only approximate exchange functionals at all? The straightforward and seemingly most appropriate strategy for arriving at a most accurate exchange-correlation energy seems to be to use the exact exchange energy of equation (5-18) and rely on approximate functionals only for the part missing in the HF picture, i. e., the electron correlation,

$$E_{XC} = E_X^{\text{exact}} + E_C^{\text{KS}}. \quad (6-24)$$

If applied to atoms this concept indeed delivers promising results. Unfortunately, and at first glance very surprisingly, it does not live up to the expectation at all if applied to molecules and chemical bonding. Against the G2 reference set we noted a mean absolute error of 78 kcal/mol for the Hartree-Fock level (i. e., exact exchange only, where we assume that the HF and KS orbitals are similar). While the inclusion of correlation through an appropriate functional in the spirit of equation (6-24) indeed cuts the error down to 32 kcal/mol, this is nevertheless a disappointing result if we consider that the errors associated with the currently used E_{XC} functionals of the GGA type, where both exchange and correlation are approximated, are in the order of only 5-7 kcal/mol.

What are the reasons for this significant failure of the exact exchange/density functional correlation combination in molecular calculations? For an analysis let us recall our discussion from Chapter 2 about the properties and shapes of the exact hole functions of the H_2 molecule, and in particular have another look at Figure 2-2. We saw that in this simple case the exchange hole corresponds to one half of the density of the σ_g occupied molecular orbital and corresponds to the removal of half an electron from the vicinity of each nucleus. It is completely delocalized and independent from the position of the reference electron. However, the exact full hole is relatively localized, in particular for extended internuclear distances where left-right correlation prevails. To salvage this overall characteristic of the total hole, the exact exchange hole has to be complemented by the correlation hole, which by itself is also delocalized. Thus, both components taken individually are bad representations of the whole, in particular the non-local exchange hole can in no way account for the effects that occur upon bond stretching or in similar situations. As an aside, we note that this is a manifestation of the fact that the separation of E_{XC} in individual exchange and correlation contributions is actually artificial and is only a consequence of the use of a particular reference system, i. e., a single Slater determinant. We re-emphasize that a clear physical meaning can only be attributed to the undivided exchange-correlation energy and hole. If we turn to the approximate holes given by the local density approximation as well as its gradient-corrected extensions we note that they are by construction based on a local model. Thus, these functionals implicitly assume that both, the exchange and the correlation hole are localized holes because all properties are determined by the density and its gradient at one particular point in space. Pictorially speaking, the approximate functionals E_{XC}^{appr} only ‘see’ their direct neighborhood and are completely ‘unaware’ of what is going on farther away. But this explains why the simple ansatz of equation (6-24) is bound to fail if applied to molecules. We combine the exact, delocalized exchange hole with a localized

model hole for correlation. Because the cancellation between the two individual holes cannot take place (as discussed above and in Section 2.3), the resulting total hole has the wrong characteristics. On the other hand, approximate exchange-correlation holes based on the uniform electron gas are again local and are therefore a better model for the exact hole than equation (6-24). Actually, Becke, 1995, Gritsenko, Schipper, and Baerends, 1997, and Schipper, Gritsenko, and Baerends, 1998b and 1999, as well as others pointed out and verified numerically that current density functionals for exchange with their localized holes effectively reproduce the sum of exact exchange *and* non-dynamical correlation while the corresponding correlation functionals represent only the effects of dynamical electron correlation. By the way, the approximate correlation hole provided by conventional, wave function based techniques indeed has the required long-range characteristics for dealing with non-dynamical correlation, usually accomplished through including energetically low-lying Slater determinants into the wave function. In fact, there are attempts to already include these long-range correlation effects into the ‘exact’ E_X contribution of equation (6-24) through a multi-configurational SCF (MCSCF) ansatz. Combining this ‘exchange plus non-dynamical correlation’ portion with a local density functional for dynamical correlation should be more appropriate (see, e. g., the recent reports by Leininger et al., 1997, Borowski et al., 1998, or Gräfenstein and Cremer, 2000). The two major problems with this approach are its significantly increased computational costs and that double-counting of correlation effects cannot be completely excluded.

Rather than pursuing this approach further, we follow a different avenue to exploit exact exchange outlined by Becke, 1993a and 1993b. The theoretical justification of this approach can be extracted from the adiabatic connection sketched in Section 6.2 above. We recall from equation (6-10) that the exchange-correlation energy of the Kohn-Sham scheme is obtained through the coupling-strength integrated exchange-correlation hole. This equation is of course equivalent to the following expression (6-25), where we integrate over the λ -dependent exchange-correlation potential energy, which is nothing else than the non-classical contribution to the electron-electron interaction for different values of λ (note that E_{ncl} corresponds to the pure potential energy contributions, dependent on λ . Only the integration over λ introduces the kinetic energy part into E_{XC})

$$E_{\text{XC}} = \int_0^1 E_{\text{ncl}}^{\lambda} d\lambda. \quad (6-25)$$

Let us explore first the nature of the integrand E_{ncl}^{λ} for the limiting cases. At $\lambda = 0$ we are dealing with an interaction free system, and the only component which is not included in the classical term is due to the antisymmetry of the fermion wave function. Thus, $E_{\text{ncl}}^{\lambda=0}$ is composed of exchange only, there is no correlation whatsoever.¹⁹ Hence, the $\lambda = 0$ limit of the integral in equation (6-25) simply corresponds to the exchange contribution of a Slater determinant, as for example, expressed through equation (5-18). Remember, that $E_{\text{ncl}}^{\lambda=0}$ can

¹⁹ Keep in mind that dynamic electron correlation is always connected to the fact that electrons interact as charged species.

be computed exactly, once the KS orbitals are available. On the other hand, for $\lambda = 1$, the non-classical contributions are those of the fully interacting system, containing exchange as well as electron correlation parts. This interacting exchange-correlation energy is not known, but can be approximated – more or less satisfactorily – by any E_{XC} functional. The true exchange-correlation energy is given by the integral of equation (6-25) and we know its value for $\lambda = 0$ exactly and have pretty good approximations for $\lambda = 1$. To exactly evaluate this integral, however, we would need E_{ncl}^λ for intermediate values of λ . But this information is not available and we must try to find approximations to this integral. Alternatively, we can analyze this integral from a slightly different point of view. We have seen above that the model holes of the LDA or GGA schemes are reference point centered and relatively localized. Hence, they provide a crude simulation of left-right correlation. As we have discussed at length, this is a desirable feature for describing the hole of the interacting system, which is also localized. On the other hand, at $\lambda = 0$ all there is, is exchange with its delocalized hole and our localized model holes are completely inadequate. Therefore, in terms of holes, to describe the $\lambda = 0$ end of the integration it appears plausible to mix in a certain amount of the pure, exact exchange hole into the overall hole.

Let us be specific. The simplest approximation to solve equation (6-25) is to assume that E_{ncl}^λ is a linear function in λ . This leads to

$$E_{XC}^{HH} = \frac{1}{2} E_{XC}^{\lambda=0} + \frac{1}{2} E_{XC}^{\lambda=1}, \quad (6-26)$$

and corresponds to the situation shown schematically in Figure 6-3a. Using the LDA exchange-correlation functional for $E_{ncl}^{\lambda=1}$, equation (6-26) represents the so-called *half-and-half* (HH) combination of ‘exact’ exchange and density functional exchange-correlation as introduced by Becke, 1993a. In fact this approach showed a promising performance. The absolute average error with respect to the G2 atomization energies amounts to 6.5 kcal/mol, and rivals the value of 5.7 kcal/mol for the gradient-corrected BPW91, if basis-set free, fully numerical results are utilized. The next step taken by Becke, 1993b was to introduce semiempirical coefficients to determine the weights of the various components in this scheme leading to the following extension of equation (6-26):

$$E_{XC}^{B3} = E_{XC}^{LSD} + a(E_{XC}^{\lambda=0} - E_X^{LSD}) + bE_X^B + cE_C^{PW91}. \quad (6-27)$$

In this equation we have three parameters. The amount of exact exchange in the functional is determined through a , while b and c control the contributions of exchange and correlation gradient corrections to the local density approximation. As indicated in equation (6-27), Becke utilized his 1988 exchange functional and Perdew and Wang’s 1991 correlation functional in his original approach. The three empirical parameters were chosen such that the atomization and ionization energies as well as the proton affinities included in the G2 data base and some total energies were optimally reproduced. This led to $a = 0.20$, $b = 0.72$, and $c = 0.81$. Hence, the amount of exact exchange was reduced relative to the earlier half-and-half scheme, indicative of a large slope of E_{XC}^λ at $\lambda = 0$, see Figure

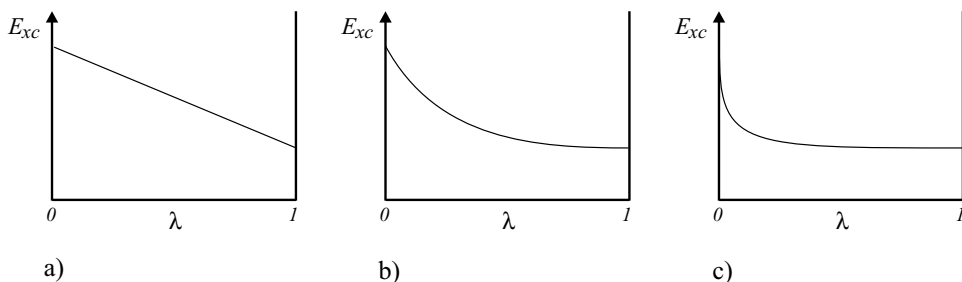


Figure 6-3. λ -dependence of E_{XC} .

6-3b. Most importantly, this three-parameter fit reduced the average absolute error in the G2 atomization energies significantly to only about 2–3 kcal/mol, already very close to the target accuracy of 2 kcal/mol. Of course one should keep in mind that the parameters a , b , and c were fitted to exactly these data and it is a priori completely unclear whether a similarly good performance can also be expected in general. Functionals of this sort, where a certain amount of exact exchange is incorporated are frequently called DFT/HF *hybrid* functionals, because they represent a hybrid between pure density functionals for exchange and exact Hartree-Fock exchange. They are also sometimes referred to as ACM functionals, where the acronym stands for *adiabatic connection method*.

Currently, the most popular hybrid functional is known as B3LYP and was suggested by Stephens et al., 1994. While it is of very similar spirit to the original form proposed, as given in equation (6-27), in B3LYP, the PW91 correlation functional is replaced by the LYP functional. The values of the three parameters were directly taken from Becke's original paper. Thus, the B3LYP exchange-correlation energy expression is (with a , b , and c just as above)

$$E_{XC}^{B3LYP} = (1 - a) E_X^{LSD} + a E_X^{\lambda=0} + b E_X^{B88} + c E_C^{LYP} + (1 - c) E_C^{LSD}. \quad (6-28)$$

For the B3LYP functional an unsigned error with respect to the G2 data base of only slightly above 2 kcal/mol was determined.

Since their incarnation in the early nineties these hybrid functionals experienced an unprecedented success (Raghavachari, 2000). In particular the B3LYP functional was an absolute shooting star and soon developed into by far the most popular and most widely used functional. This amazing success was fueled by the surprisingly good performance B3LYP and related functionals demonstrated in many chemical applications, including such difficult areas as open-shell transition-metal chemistry as we will discuss in much detail in later parts of this book.

More recent developments by Becke, 1996a reduced the number of parameters to one

$$E_{XC}^{B1} = E_{XC}^{DFT} + a(E_X^{\lambda=0} - E_X^{DFT}) \quad (6-29)$$

where the amount of exact exchange was empirically determined as $a = 0.28$ if for E_{XC}^{DFT} a combination of the standard Becke exchange (B) and a new correlation functional due to Becke (B95) was inserted. This functional is commonly referred to as B1B95. Its average absolute error with respect to the G2 atomization energies is only around 2 kcal/mol. However, there are certain technical disadvantages with this functional. Most notably, it depends explicitly on the kinetic energy density in addition to the density and its gradient, which complicates the implementation into standard molecular structure computer programs. This might be the reason, why the new B1B95 has attracted less interest than its three-parameter predecessor, even though inclusion of the kinetic energy density into exchange-correlation functionals seems to have gained a lot of attention lately, as outlined below. In the last part of his series of papers on density-functional thermochemistry, Becke, 1997, introduced a new type of exchange-correlation functional which was based on an elaborate fitting procedure. The exchange-correlation functional was separated into several parts, i. e., exchange, like-spin correlation and unlike-spin correlation and an additional amount of exact, Hartree-Fock exchange, i. e.,

$$E_{XC}^{B97} = E_X^{\alpha\alpha} + E_X^{\beta\beta} + E_C^{\alpha\alpha} + E_C^{\beta\beta} + E_C^{\alpha\beta} + c_X^{HF} E_X^{HF} \quad (6-30)$$

Each component, with the exception of the HF exchange, is expressed in a power series involving the density and the reduced density gradient. These expansions were terminated at second order, since otherwise unphysical, overfitted functionals were obtained. The resulting ten linear coefficients were optimized by a least-square fit to energetical data from the G2 set. Note, that the optimal parameters were determined in a fully numerical (i. e., basis set free), non-self-consistent procedure using LDA densities. If measured against the G2 training set, average absolute and maximum errors of atomization energies of only 1.8 and 5.5 kcal/mol, respectively, were obtained. The amount of exact exchange was determined at 20 % in this B97 functional. As Becke noted, this kind of accuracy is probably as far as one can get with conventional gradient-corrected GGA functionals for exchange and correlation and a certain (but fixed, see below) amount of exact, i. e. $\lambda = 0$, exchange. A year later, Schmider and Becke 1998a, reparameterized the B97 functional with respect to the extended G2 set. The resulting B98 functional retains the good absolute average (1.9 kcal/mol) and maximum errors (9.1 kcal/mol) also for this larger and more demanding training set.²⁰ Hamprecht et al., 1998, reparameterized the original B97 functional in a self-consistent procedure, i. e., with densities optimized within the same functional and using a TZ2P basis set and termed the resulting B97 flavor B97-1. In the same paper these authors also suggest an extension to the B97 idea. They additionally require in the parameterization scheme that the functional also reproduces nuclear gradients for molecules (i. e., zero if calculations are performed at equilibrium geometries) and, probably even more importantly, that it yields accurate exchange-correlation potentials, a property that will become

²⁰ Schmider and Becke, 1998a, presented various parameterizations in their paper, which differed in the choice of data included in the fitting. The quoted performance applies to their parameter set 2c, where 148 heats of formation, 42 ionization energies, 25 electron affinities, 8 proton affinities and 10 total energies were included in the training set.

of interest in Section 6-8. Because the least-square procedure now includes much more information, the power series was extended to forth order, increasing the number of parameters to 15. However, unlike in the B97 or B98 schemes, the resulting HCTH functional is a pure GGA functional and contains no exact exchange. In a subsequent paper (Boese et al., 2000), the training set for the parametrization was extended in particular by anions and weakly bound dimers leading to the HCTH/120 and HCTH/147 functionals (the numbers indicate the number of systems used in the training sets. In the original HCTH parametrization a total of 93 systems was used).

Before closing this section let us mention that a fraction of about 20–25 % exact exchange as realized in the above functionals seems to be reasonable also on purely theoretical grounds as shown by Perdew, Ernzerhof, and Burke, 1996, and Burke, Ernzerhof, and Perdew, 1997. These authors proposed parameter-free hybrid functionals of the general form, where the amount of exact exchange has been derived as 25 % from theoretical reasonings through a perturbation theory argument,

$$E_{XC}^{\text{hybrid}} = E_{XC}^{\text{GGA}} + 0.25(E_X^{\text{HF}} - E_X^{\text{GGA}}). \quad (6-31)$$

If the PBE exchange-correlation functional is chosen as the GGA component, the PBE1PBE model emerges (some authors prefer to call this functional PBE0). As shown by Adamo and Barone, 1999, PBE1PBE shows promising performance for all important properties, being competitive with the most reliable, empirically parameterized current functionals. However, while about a quarter of exact exchange is reasonable for most regular systems, it should be clear that in general this parameter is certainly not universal but depends on the actual situation. This can be impressively demonstrated using our standard guinea-pig, the H_2 molecule and its molecular ion, H_2^+ . Let us consider first the neutral hydrogen molecule. As we have seen in Section 2.3, as the distance of the two nuclei increases, the total exchange-correlation hole gets more and more localized. For infinite distance it is strictly localized and removes exactly one electron from the proton where the reference electron is located while it is zero at the other nucleus. In this situation, the amount of exact exchange in a restricted calculation must go to zero as $r \rightarrow \infty$: the correlation holes of approximate functionals are localized. Mixing in any fraction of the delocalized exact exchange hole would therefore lead to an unphysically delocalized total hole (see also, Gritsenko, van Leeuwen, and Baerends, 1996). The situation is completely different for the corresponding radical cation, the one-electron system H_2^+ . Here, the exchange-correlation hole obviously contains only the exchange part, which is completely delocalized over the molecule, independent of the internuclear distance. Due to their inherent local character none of the current approximate exchange functionals is capable of correctly representing this situation. In order to describe this delocalized hole a hybrid with 100 % exact exchange would be needed, as discussed by Sodupe et al., 1999. Generally speaking, in situations, where the $\lambda = 0$ limit is represented by degenerate or near-degenerate ground states, the local exchange density functional is a good approximation throughout the whole integration, including $\lambda = 0$. In other words, the λ -dependence of E_{XC}^λ is characterized under these circumstances by an extreme slope (approaching $-\infty$) at $\lambda = 0$ and the solution

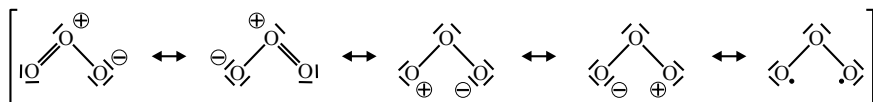


Figure 6-4. Resonance structures of O_3 .

Table 6-1. Harmonic frequencies and experimental fundamentals for ozone [cm^{-1}]. Deviations from the experimental result [%] are given in parentheses.

Method	bending (a_1)	antisym. stretch (b_2)	sym. stretch (a_1)
Hartree-Fock	870 (+21.5)	1419 (+30.3)	1541 (+35.8)
MP2	747 (+4.3)	2211 (+203.0)	1170 (+3.1)
B3LYP	750 (+4.7)	1205 (+10.7)	1259 (+10.9)
<i>m</i> PW1PW	778 (+8.7)	1296 (+19.0)	1323 (+16.6)
BLYP	688 (−3.9)	991 (−9.0)	1135 (± 0)
BP	708 (−1.1)	1054 (−3.2)	1179 (+3.9)
Experiment	716	1089	1135

for $\lambda = 0$ does not contribute to the integral as indicated in Figure 6-3c. A typical example is provided by the ozone molecule which is known to be pathological because of near-degeneracy effects. The ionic and biradical resonance structures indicated in Figure 6-4 both contribute significantly to the overall wave function.

The most sensitive properties in that respect are the vibrational frequencies, in particular the antisymmetric O-O stretching vibration (of b_2 symmetry). Along this vibrational mode the relative weights of the two main contributors of Figure 6-4 to the wave function change. Conventional methods, such as HF or the MP2 approach in particular, where dynamical electron correlation is estimated through second order perturbation theory, fail completely. But also hybrid functionals such as B3LYP or the very recent, one-parameter *m*PW1PW scheme predict harmonic frequencies of O_3 which are in much less harmony with the experimental data than the results obtained from plain GGA protocols, such as BLYP or BP86. Table 6-1 summarizes theoretically predicted harmonic frequencies for ozone from representative computational models employing a flexible cc-pVQZ basis set to expand the KS orbitals.

6.7 Self-Interaction

There is one more problem which is typical for approximate exchange-correlation functionals. Consider the simple case of a one electron system, such as the hydrogen atom. Clearly, the energy will only depend on the kinetic energy and the external potential due to the nucleus. With only one single electron there is absolutely no electron-electron interaction in such a system. This sounds so trivial that the reader might ask what the point is. But

consider the energy expression for a one electron system in the Kohn-Sham scheme (which is no different from the general equation (5-14)),

$$E[\rho(\vec{r})] = T_s[\rho] + J[\rho] + E_{XC}[\rho] + E_{Ne}[\rho]. \quad (6-32)$$

The classical electrostatic repulsion term is

$$J[\rho] = \frac{1}{2} \iint \frac{\rho(\vec{r}_1) \rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2. \quad (6-33)$$

This term does not exactly vanish for a one electron system since it contains the spurious interaction of the density with itself. Hence, for equation (6-32) to be correct, we must demand that $J[\rho]$ exactly equals minus $E_{XC}[\rho]$ such that the wrong *self-interaction* is cancelled

$$\frac{1}{2} \iint \frac{\rho(\vec{r}_1) \rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 d\vec{r}_2 = -E_{XC}[\rho]. \quad (6-34)$$

And that is where the trouble begins.

We saw in Section 1.3 that by construction the exchange term of the Hartree-Fock model indeed exactly neutralizes the unwanted portion of $J[\rho]$. In particular, for a one electron system equation (6-34) is satisfied and the HF scheme is therefore free of self-interaction errors. On the other hand, in any realization of the Kohn-Sham density functional scheme we have to employ approximations to the exchange-correlation energy which are independent of $J[\rho]$ and we should not expect equation (6-34) to hold. In fact, none of the currently used exchange-correlation functionals is self-interaction free. In Table 6-2 we have summarized the results for the hydrogen atom as obtained with typical exchange-correlation functionals employing a large cc-pV5Z basis set.

We see that the self-interaction error, $J[\rho] + E_{XC}[\rho]$, is in all cases in the order of $10^{-3} E_h$ or a few hundredths of an eV. In addition, the data in Table 6-2 reiterate some of the facts that we noted before. B3LYP, BP86 and BPW91 yield total energies below the exact result of $-0.5 E_h$, in an apparent contradiction to the variational principle (see discussion in Sec-

Table 6-2. Energy components [E_h] of various functionals for the hydrogen atom.

Functional	E_{tot}	$J[\rho]$	$E_X[\rho]$	$E_C[\rho]$	$J[\rho] + E_{XC}[\rho]$
SVWN	-0.49639	0.29975	-0.25753	-0.03945	0.00277
BLYP	-0.49789	0.30747	-0.30607	0.0	0.00140
B3LYP	-0.50243	0.30845	-0.30370 ^a	-0.00756	-0.00281
BP86	-0.50030	0.30653	-0.30479	-0.00248	-0.00074
BPW91	-0.50422	0.30890	-0.30719	-0.00631	-0.00460
HF	-0.49999	0.31250	-0.31250	0.0	0.0

^a Includes 0.06169 E_h from exact exchange.

tion 4.5). It is also noteworthy that of the three correlation functionals, only LYP yields the correct result of zero correlation energy for a single electron (i. e., it is self-interaction free), all others deviate non-negligibly from zero.

Of course, this self-correction error is not limited to one electron systems, where it can be identified most easily, but applies to all systems. Perdew and Zunger, 1981, suggested a *self-interaction corrected* (SIC) form of approximate functionals in which they explicitly enforced equation (6-34) by subtracting out the unphysical self-interaction terms. Without going into any detail, we just note that the resulting one-electron equations for the SIC orbitals are problematic. Unlike the regular Kohn-Sham scheme, the SIC-KS equations do not share the same potential for all orbitals. Rather, the potential is orbital dependent which introduces a lot of practical complications. As a consequence, there are hardly any implementations of the Perdew-Zunger scheme for self-interaction correction.

Surprisingly, while application of the Perdew-Zunger self-interaction correction improves the results for atoms as expected, this does not necessarily carry over to ground state energies and geometries of molecules, where the self-interaction corrected scheme may even lead to a deterioration of the results as compared to regular approximate Kohn-Sham calculations, as reported by Goedecker and Umrigar, 1997. The reasons for this behavior are, however, not fully understood. Similarly, to what extent the unphysical self-interaction affects the results of density functional calculations in general is not completely clear yet, but it certainly can sometimes have severe consequences. Among the most spectacular examples is the difficulty that approximate exchange-correlation functionals experience when the dissociation of radicals consisting of two identical moieties are studied as pointed out by several authors; e. g., by Merkle, Savin, and Preuss, 1992, and more recently by Bally and Sastry, 1997, and Zhang and Yang, 1998. Sodupe et al., 1999, for example show that Kohn-Sham calculations predict the wrong order of stability for the two low-lying structural isomers of the $(\text{H}_2\text{O})_2^+$ dimer, overestimating the stability of the symmetric $\text{H}_2\text{O} \cdots \text{OH}_2$ by some 17 kcal/mol. Even for systems as simple as the one-electron hydrogen molecular ion, H_2^+ , the dissociation curve is significantly in error leading to much too small binding energies. The origin of the huge self-interaction error as the H-H bond stretches is the incapability of the intrinsically localized model holes of approximate functionals to describe the delocalized exchange hole of $(\text{H} \cdots \text{H})^+$ as $r_{\text{H-H}} \rightarrow \infty$, see also the clear discussion in Perdew and Ernzerhof, 1998. According to Zhang and Yang, 1998, such problems are always to be expected for situations where non-integer number of electrons are involved. The H_2^+ dissociation offers a prototype for this scenario since it leads to a delocalized state according to $\text{H}^{+0.5} \cdots \text{H}^{+0.5}$ as R_{HH} increases. In general, such cases are to be expected if the ionization energy of one dissociation partner differs by only a small amount from the electron affinity of the other partner. For $\text{H}_2^+ \rightarrow \text{H} + \text{H}^+$ this criterion is perfectly satisfied since the ionization energy of H and the electron affinity of H^+ are identical. These authors go on to speculate, that similar problems should also surface if transition states of chemical reactions with stretched bonds or certain charge-transfer complexes are studied with approximate exchange-correlation functionals. That self-interaction indeed plays a decisive role in the vicinity of transition structures and may heavily affect reaction barriers has been demonstrated, e. g., by Csonka and Johnson, 1998: the barrier for the seemingly simple

hydrogen abstraction reaction, $\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H}_2$, is raised by approximately 8–9 kcal/mol by inclusion of the self-interaction correction and brings the computed results into much better agreement with the experimental activation barrier, as we will discuss in more detail in Chapter 13. Their paper also includes an instructive and yet concise outline of the Perdew-Zunger procedure.

8 Asymptotic Behavior of Exchange-Correlation Potentials

While the behavior of the exchange-correlation potential V_{XC} (recall from equation (2-16) that the exchange-correlation potential V_{XC} is defined as the functional derivative of the exchange-correlation energy E_{XC} with respect to the charge density ρ : $V_{\text{XC}} = \frac{\delta E_{\text{XC}}}{\delta \rho}$) at distances far from the atom or molecule seems of little importance at first glance, it turned out to be critical for properties which depend not only on the quality of the occupied Kohn-Sham orbitals but also on how well the virtual orbitals are described. Typical examples include atomic electron affinities and properties related to the response of the system to an electromagnetic field, such as polarizabilities, or excitation energies to energetically high-lying states, in particular Rydberg states, computed from the poles of the frequency dependent polarizability (i. e., in the TDDFT scheme). How should the asymptotic V_{XC} look like? We know that for an N -electron system the Coulomb potential $V_{\text{C}}(\vec{r}_1) = \int \frac{\rho(\vec{r}_2)}{r_{12}} d\vec{r}_2$ behaves like N/r when $r \rightarrow \infty$. In order to cancel the unphysical self-interaction in the Coulomb term, the exchange-correlation potential (since correlation effects are much more short-ranged, it usually suffices to analyze only the exchange potential in the asymptotic region) must therefore have a $-1/r$ dependence at large r . The electron far away from the molecule now sees the correct net $(N+1-Z)$ charge (Z being the positive nuclear charge). There are two problems with current popular exchange functionals in this context. First, none of the corresponding potentials has the correct $-1/r$ behavior, they rather all decrease exponentially, i. e., much too fast. As a consequence, these approximate potentials are less attractive than the exact one at large r . The second problem is more subtle and much more difficult to grasp without a detailed theoretical analysis. In the following we will limit ourselves to a pictorial description of the problem without dwelling too much into the physical background. In a celebrated (see Zhang and Yang, 2000) and often quoted paper, Perdew et al., 1982, extended Kohn-Sham density functional theory to fractional electron numbers. In that context they showed that the exchange-correlation potential actually jumps by a constant as the number of electrons passes through an integer. This phenomenon is known as the *derivative discontinuity* in DFT. As a corollary to this it can be shown that none of the currently available approximate functionals, which are all characterized by a continuous potential with respect to variations in the number of electrons is able to model this behavior. This in turn has the rather unexpected consequence that an accurate continuous potential should not vanish asymptotically. Rather, as shown for example by Tozer and Handy, 1998, the asymptotic potential should obey