Density Functional Theory Approach

The path to the exact non-relativistic description of electrons in molecules involving first Hartree-Fock calculations and then post-Hartree-Fockk correction using configuration interaction (or coupled cluster expansion) is in principle clear and well defined. But, the problem is that the computational effort increases so rapidly with the number of electrons that it becomes impossible to include more than a few tens of electrons, i.e. only for small molecules can this approach be taken. Various approximations can be applied to larger molecules, but it is then less clear how accurate the results are. The golden standard of quantum chemistry today is taken to be the coupled cluster approximation with single and double excitation Slater determinants included included as well as a perturbative correction due to triple excitations, so called CCSD(T). But, the computational effort of this approximation (which is sometimes not accurate enough) scales as the number of electrons to the seventh power, N^7.

The basic problem here is the enormous complexity of the many electron wave function. For a molecule containing N electrons, the wave function depends on $3N$ variables, $\Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, ...\mathbf{r}_N)$, where $\mathbf{r}_i$ specifies the location of electron i using for example three Cartesian coordinates or spherical polar coordinates. In order to record such a function, one would need at least a few parameters for each coordinate, a minimum of say 3 parameters. This means that a wave function for an N-electron system would require the specification of 3^{3N} parameters. A medium size molecule can easily contain N=1000 electrons, in which case this number of parameters becomes $3^{3N} \approx 10^{1500}$. This is an enormous number and there is no way any super computer will be able to save such a large number of parameters in any foreseeable future. Note that the number of barions in the Universe, the most elementary particles that make up all matter, is estimated to be 10^{80}. The many electron wave function is, therefore, impossible to even record, let alone calculate, except for small molecules. This led early on to speculations whether there was some other formulation of quantum mechanics that could bypass the evaluation of the wave function. Walter Kohn has expressed this sentiment by the following statement: "In general the many-electron wave function $\Psi(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, ...\mathbf{r}_N)$, for a system of N electrons is not a legitimate scientific concept, when $N \geq N_0$ where $N_0 \approx 10^3$."(see: W.Kohn, Nobel Lecture: Electronic structure of matter—wave functions and density functionals, Reviews of Modern Physics, Vol. 71, 1999).

Early on, Thomas and Fermi attempted to replace the wave function with a much simpler quantity, namely the electron density, $\rho(\mathbf{r})$ which is a function of only three variables. It specifies the density of electrons in the three dimensional space irrespective of the identity of the electrons, i.e. irrespective of which electron (or electrons) is found in the particular location specified. The electron density can be obtained from the magnitude squared of the wave function by integrating over all coordinates except for those corresponding to one of the electrons. Due to the antisymmetry, it does not matter which electron coordinates are retained, so one can pick electron 1 for example and carry out $N - 1$ integrals over variables $\mathbf{r}_2, \mathbf{r}_3, ...\mathbf{r}_N$ to get

$$\rho(\mathbf{r}) = N \int \cdots \int |\Psi(\mathbf{r}, \mathbf{r}_2, \mathbf{r}_3, \ldots \mathbf{r}_N)|^2 \, d\mathbf{r}_2 d\mathbf{r}_3 \ldots d\mathbf{r}_N.$$

Apparently, a great deal of information is lost by including only the magnitude of the wave function and by carrying out the integrations. The electron density does not seem to contain enough information to provide, for example, a value of the energy of the system. This can be illustrated with an example of a one-dimensional plane wave, i.e. a wave function for a free particle, where $\Psi(x) = Ae^{-ikx}$. Here, k is the wave number which is inversely proportional to the wave length, $k \propto 1/\lambda$. The kinetic energy is proportional to k^2 but the magnitude squared of the wave function is $|\Psi(x)|^2 = |A|^2$ which is independent of k, so the essential information needed to determine the kinetic energy is lost when the magnitude of the complex wave function is evaluated. Thomas-Fermi theory did not get used much. While it could describe atoms to some extent, chemical bonding was not accounted for, i.e. separated atoms where predicted to have lower energy than molecules.

Kohn and his co-workers Hohenberg and Sham developed an approach which is referred to as density functional theory (DFT) and it can be considered to be a formal exactification of Thomas-Fermi theory. There,

the energy of an electronic system is obtained as a functional of the electron density. The practical implementations of DFT are approximate and it is not clear how to reach the exact result in a well defined limit, but most electronic structure calculations today are nevertheless carried out using DFT. The computational effort in DFT calculations scales as the number of electrons to the third power, N^3, so this method can be applied to much larger systems than QCISD and CCSD(T), for example, and cab be applied in studies of systems with up to a few hundred atoms. But, DFT not only makes calculations of larger systems possible, it also helps understand properties of many electron systems in terms of quantities that can be measured, such as the electron density. In practical implementations of DFT, the Kohn-Sham approach, a mapping of the many electron system onto a system of independent electrons is used, and the concept of orbitals is retained. It turns out that Kohn-Sham orbitals are quite similar to Hartree-Fock orbitals, but the accuracy of the calculated total energy is much improved over the Hartree-Fock approximation. In this way, DFT helps push the orbital based approach to higher accuracy in terms of, for example, bond energy in molecules. Interestingly, even though the derivation of HF indicates that it gives the best orbital based approximation, it turns out that Kohn-Sham DFT gives an even better orbital based approximation. Unfortunately, the orbital energies are, however, typically less accurate in DFT than in HF as estimates of ionization energy and electron affinity.

Hohenberg-Kohn theorem

The corner stone of DFT is the Hohenberg-Kohn (HK) theorem. It provides motivation for pursuing the development of approximate energy functionals. It proves that for a given ground state electron density, $\rho(\mathbf{r})$, a unique corresponding external potential, V, can in principle be found. In calculations of electrons in molecules, this external potential corresponds to the Coulomb potential between an electron and the positive nuclei of the atoms in the molecule. The other terms in the Hamiltonian can then be added, i.e. the kinetic energy, T, and the electron-electron interaction, U. With the full Hamiltonian, the Schrödinger equation can in principle be solved to obtain the wave function and the energy. Therefore, a path exists for using only information about the ground state electron density in order to obtain the value of the ground state energy. The ground state energy is, thus, in principle a functional of the ground state electron density, $E[\rho(\mathbf{r})]$. The figure below illustrates the connection.
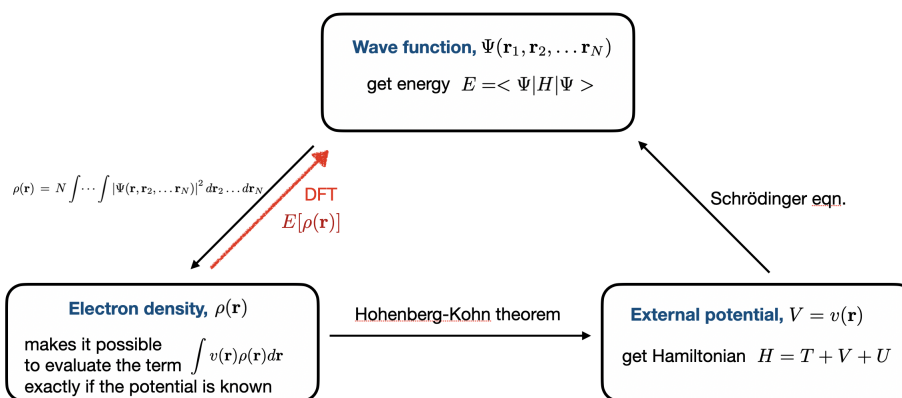
An illustration of the connection between wave function, electron density and external potential through Schrödinger equation and Hohenberg-Kohn theorem. T denotes the kinetic energy operator, V the external potential (here, the Coulomb interaction between an electron and the atomic nuclei) and U the electron-electron interaction.

While the HK theorem is elegant and applies to the ground state of any electronic system, the existence of such an energy functional for molecules is, in hindsight, not surprising. At the atomic nuclei, the Coulomb interaction between an electron and the nucleus goes to negative infinity as $-1/r$. As a result, the electron wavefunction and also the corresponding electron density has at each atom nucleus. Given an electron density in a molecule, one can identify from the cusps the location of the nuclei of the atoms in the molcule. Furthermore, the shape of the cusps can reveal the charge on the nuclei. The electron-nuclei Coulomb

potential is thereby obtained from the electron density, $V = v(\mathbf{r})$. By adding T and U, which are the same for all electronic systems, the full Hamiltonian can be constructed and the Schrödinger equation solved, in principle, to give the energy. Thus, the exact functional is in fact known for molecules, but it involves solving the Schrödinger equation, so it provides no benefit compared to the traditional approach. This is, however, taken as a motivation for constructing energy functionals that are analytical or semi-analytical approximations to the exact functional. The development of DFT functionals is an active field of research and many new functional approximations continue to be produced (perhaps too many?).

The HK theorem is simply based on the variational principle, $< \Psi|H|\Psi > \geq E_0$, where E_0 is the energy of the ground state and Ψ is a normalized trial function. It is easiest to formulate it assuming the ground state is non-degenerate, and this will be done here. An extension to degenerate ground states can be formulated as well. Also, spin is not included in the formulation presented here. Assuming this is a closed shell system with all spins paired, the spin has been integrated out. But, this can be extended to include spin explicitly.

HK theorem: The ground-state density $\rho(\mathbf{r})$ *of a bound system of interacting electrons in some external potential* $v(\mathbf{r})$ *determines this potential uniquely.* Note: Here "uniquely" means "up to an additive constant".

Proof by contradiction: Let $\rho(\mathbf{r})$ be the non-degenerate ground-state density of N electrons in the potential $v_1(\mathbf{r})$, corresponding to ground eigenstate $|\Psi_1 >$, and energy E_1. This means that

$$E_1 \ = \ < \Psi_1|H_1|\Psi_1 > = \int v_1(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \ < \Psi_1|T + U|\Psi_1 >$$

where H_1 is the Hamiltonian corresponding to v_1. Now assume there exists another potential $v_2(\mathbf{r})$ (that is not simply equal to $v_1(\mathbf{r})$ plus a constant), corresponding to ground state eigenvalue E_2 and eigenstate $|\Psi_2 >$ that has the same electron density $\rho(\mathbf{r})$. Then,

$$E_2 \ = \ < \Psi_2|H_2|\Psi_2 > = \int v_2(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \ < \Psi_2|T + U|\Psi_2 > .$$

The variational principle for $|\Psi_1 >$ gives

$$E_1 < \ < \Psi_2|H_1|\Psi_2 > = \int v_1(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + \ < \Psi_2|T+U|\Psi_2 > = \int (v_2(\mathbf{r})+v_1(\mathbf{r})-v_2(\mathbf{r}))\,\rho(\mathbf{r})d\mathbf{r} + \ < \Psi_2|T+U|\Psi_2 >$$

so

$$E_1 < E_2 + \int (v_1(\mathbf{r}) - v_2(\mathbf{r}))\,\rho(\mathbf{r})d\mathbf{r}.$$

Similarly, by reversing the role of 1 and 2, the variational principle gives

$$E_2 < E_1 + \int (v_2(\mathbf{r}) - v_1(\mathbf{r}))\,\rho(\mathbf{r})d\mathbf{r}.$$

By adding up the two inequalities one gets

$$E_1 + E_2 \ < \ E_2 + E_1$$

which is a contradiction. Therefore, it is not possible that the density is the same for the ground states corresponding to the two potentials $v_1(\mathbf{r})$ and $v_2(\mathbf{r})$.

The theorem shows that two different potentials cannot correspond to the same electron density. Thus, it is in principle possible to find the potential that corresponds to a given electron density. However, the theorem does not ensure that such a potential exists, only that if it exists, it is unique (to within an arbitrary constant). The question whether there exists such a potential is referred to as the 'v-representability' issue. Some cases have been identified where a given electron density does not correspond to any potential function and the density is then said not to be v-representable.

Variational principle based on electron density

The variational principle can be extended to be based on the electron density rather than the wave function. To every trial function ψ corresponds an electron density ρ, obtained by integrating over all coordinates

except those of one of the electrons and multiplying by N. The variational calculation becomes a two step procedure. First, a trial density, $\tilde{\rho}$, is fixed and the set of all trial functions that correspond to this trial density are denoted as Ψ^{α}_{ρ}. Here, α is some parameter identifying the different functions in the set. An energy minimization over this set of functions is carried out to give a constrained minimum of the energy as

$$E_v[\tilde{\rho}] \,=\, \min_{\alpha} <\Psi^{\alpha}_{\tilde{\rho}}|H|\Psi^{\alpha}_{\tilde{\rho}}> \,=\, \int v(\mathbf{r})\tilde{\rho}(\mathbf{r})d\mathbf{r} \,+\, F[\tilde{\rho}]$$

where the functional F is defined as

$$F[\tilde{\rho}] \,\equiv\, \min_{\alpha} <\Psi^{\alpha}_{\tilde{\rho}}|T+U|\Psi^{\alpha}_{\tilde{\rho}}> \,.$$

F is a universal functional of the electron density and requires no knowledge of the potential, v.

In the second step, minimize $E_v[\tilde{\rho}]$ over all trial densities $\tilde{\rho}$

$$E \,=\, \min_{\tilde{\rho}} E_v[\tilde{\rho}] \,=\, \min_{\tilde{\rho}} \left\{ \int v(\mathbf{r})\tilde{\rho}(\mathbf{r})d\mathbf{r} \,+\, F[\tilde{\rho}] \right\}.$$

Thereby, the problem of finding the minimum of $<\Psi|H|\Psi>$ with respect to the 3N-dimensional trial function Ψ has been transformed into a *seemingly* simpler problem of finding the minimum of $E_v[\tilde{\rho}]$ with respect to a three-dimensional function $\tilde{\rho}$. But, the definition of F leads us back to a minimization with respect to 3N-dimensional trial wave functions. Nevertheless, a significant formal progress has been made, as the problem now involves evaluation of the contribution due to the potential, the part that is different for different systems, in terms of just the three-dimensional electron density, while the harder problem, the minimization of F, is still in terms of wave functions, but this is the same functional for all electronic systems. The goal is then to find a suitable approximation for F and this is done by analyzing a simple model systems, namely the uniform and nearly uniform electron gas.

Kohn-Sham equations

While the HK theorem states that it is in principle possible to obtain the ground state energy of an electronic system from a given electron density, assuming that the corresponding external potential exists (v-representability), it does not provide any clue as to how to write this functional except for the long and roundabout way that involves solving the Schrödinger equation (as indicated in the figure above). Kohn and Sham presented a practical implementation which is generally referred to as density functional theory (DFT) or more specifically as Kohn-Sham density functional theory (KS-DFT).

The challenge is to find a practical way to evaluate the functional $F[\rho]$, that is the kinetic energy of the electrons and the electron-electron interaction. Orbitals are introduced in the following way. For a given electron density, $\rho(\mathbf{r})$, of a system with N electrons, a set of N orbitals, $\phi_j(\mathbf{r})$ with $j = 1, \ldots, N$, is found in such a way that

$$\rho(\mathbf{r}) \,=\, \sum_{j=1}^{N} |\phi_j(\mathbf{r})|^2.$$

While the electron density corresponds to a system of interacting electrons and a many-electron wave function that cannot in general be represented by a superposition of orbitals, this set of orbitals, referred to as Kohn-Sham orbitals (different from Hartree-Fock orbitals), represents a mapping of the true system onto a fictitious system of non-interacting electrons with the same electron density. The kinetic energy of the system of interacting electrons is then approximated by the kinetic energy of the non-interacting electron system

$$T[\rho] \,\approx\, T_s[\rho] \,\equiv\, -\frac{1}{2}\sum_{j} \nabla^2 \phi_j(\mathbf{r}).$$

The electron-electron interaction, U, is divided into a classical Coulomb interaction, U_C, and an exchange and correlation interaction, U_{xc}, where the terms exchange and correlation energy have the meaning that was introduced in Hartree-Fock theory. The goal set out by Kohn and Sham is to write everything as much as

possible in terms of the total electron density, even though the concept of orbitals is introduced to evaluate the kinetic energy, so the estimate of the classical Coulomb interaction is taken to be

$$U_C = \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \, d\mathbf{r} d\mathbf{r}'.$$

Each of the two integrals is over all space. This estimate of the classical Coulomb interaction using the total electron density clearly involves an error in that the electrons interact with themselves. This can be seen easily by applying it to a system with only one electron. The electron density is non-zero over an extended region because of the quantum delocalization and so the double integral will give a non-zero value, representing a Coulomb interaction of the electron with itself. This is referred to as the Coulomb self-interaction error. But, the second term, U_{xc}, should then in principle remove this error. A similar formulation of the classical Coulomb interaction can be used in Hartree-Fock theory, but then the exchange energy contains the same self-interaction error with an opposite sign and the two errors cancel out. One of the main limitations of the KS-DFT formulation is that the self-interaction error does not cancel out because the Coulomb self-interaction is highly non-local, involves a double integral over all space, while the exchange energy is estimated with a local or semi-local approximation (see below). The correction to the kinetic energy estimate and the Coulomb energy estimate as well as the contributions of exchange and correlation is formally written as E_{xc} and the in principle exact expression of the ground state energy of the system written as

$$E[\rho] = T_s[\rho] + V[\rho] + U_C[\rho] + E_{xc}[\rho].$$

The problem of finding an expression for the energy functional has been shifted to finding the functional E_{xc}.

The variational principle based on electron density can be applied and $E[\rho]$ minimized subject to the constraint that the orbitals are orthonormal

$$\int \phi_j^*(\mathbf{r})\phi_k(\mathbf{r}) \, d\mathbf{r} - \delta_{jk} = 0$$

Using the Lagrange method, the constraints are multiplied by constants ϵ_{jk} and added to $E[\rho]$ to give an objective function

$$E[\rho] - \epsilon_{jk} \left(\int \phi_j^*(\mathbf{r})\phi_k(\mathbf{r}) \, d\mathbf{r} - \delta_{jk} \right)$$

which is then minimized by finding the ρ that makes the objective function stationary with respect to any arbitrary infinitesimal change in the orbitals $\delta\phi_k$. Since ρ is represented by a sum over contributions from orbitals, the optimization procedure results in equations for the optimal orbitals, analogous to the way variational optimization of the Hartree-Fock energy gives the Hartree-Fock equations for the orbitals. Also, as in HF theory, the energy does not change when some linear combination of the occupied orbitals is used instead, i.e. the energy is unitary invariant. The convenient choice of orbitals is then made for which the Lagrance multiplier matrix, ϵ_{jk}, is diagonal.

The equations obtained for the Kohn-Sham orbitals are

$$\left(-\frac{1}{2}\nabla^2 + v_{eff}(\mathbf{r}) - \epsilon_j \right) \phi_j(\mathbf{r}) = 0$$

and are referred to as the Kohn-Sham equations. The effective potential is the same for all the orbitals

$$v_{eff}(\mathbf{r}) = v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \, d\mathbf{r}' + v_{xc}(\mathbf{r})$$

where v_{xc} is the functional derivative of E_{xc}

$$v_{xc}(\mathbf{r}) \equiv \frac{\partial}{\partial\tilde{\rho}(\mathbf{r})} E_{xc}[\tilde{\rho}(\mathbf{r})]|_{\tilde{\rho}=\rho}.$$

The resulting KS orbitals are different from HF orbitals in that they effectively take into account the correlation energy, the missing term in the total energy given by HF theory. Kohn, however, emphasized

that the KS orbitals are fictitious, only a mathematical tool for representing the electron density, the KS orbital being 'density optimal' while HF orbitals are 'energy optimal'. Nevertheless, KS orbitals turn out in most cases to be quite similar to HF orbitals and are thereby roughly equally useful or not useful.

It can be shown that the eigenvalue, ϵ_N, corresponding to the highest occupied orbital obtained from the exact functional corresponds to the lowest ionization energy, but the Koopman rule does not in other respects apply in DFT.

Functional approximations

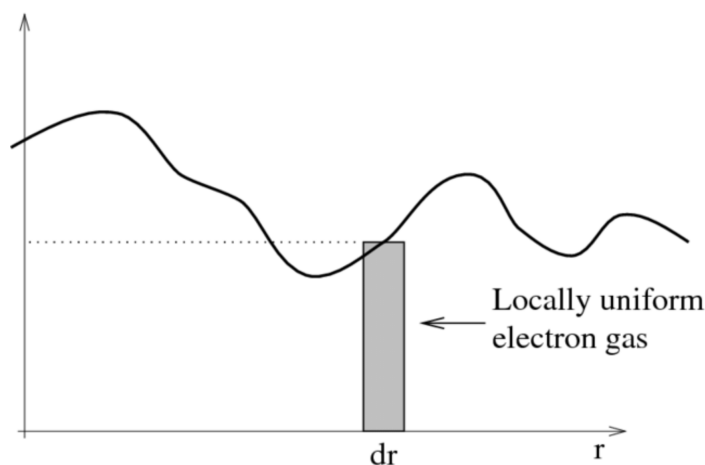### Local density approximation

The remaining challenge, and a big challenge indeed, is to find a viable approximation for E_{xc}. Kohn and Sham presented what is referred to as the local density approximation (LDA). The simplest many electron system that can be solved exactly is the uniform electron gas, i.e. a system of electrons with a constant external potential so as the make the system charge neutral. The total energy can be evaluated as a function of the density of the electrons, $E(\rho)$, as well as the two approximate terms $T_s(\rho)$ and $U_C(\rho)$. The unknown term, $E_{xc}[\rho]$ can, therefore, be evaluated for this system and by dividing by the number of electrons, N, an e_{xc} per electron can be defined as

$$e_{xc}(\rho) = \frac{(E(\rho) - T_s(\rho) - U_C(\rho))}{N}.$$

For a non-uniform electron density, $\rho(\mathbf{r})$, the local approximation then involves approximating a given electron density locally as uniform electron gas within a small volume element, evaluating its contribution to E_{xc} as $e_{xc}[\rho(\mathbf{r})]\rho(\mathbf{r})$ and then integrating over all space

$$E_{xc}[\rho(\mathbf{r})] \; = \; \int e_{xc}[\rho(\mathbf{r})]\rho(\mathbf{r}) \, d\mathbf{r}.$$

This approximation is illustrated in the figure below.

An illustration of the local density approximation to the E_{xc} term in the Kohn-Sham energy functional. For each volume element located at $\mathbf{r}$, its contribution to E_{xc} is estimated as the contribution per electron, $e_{xc}[\rho]$ times the number of electrons in the volume element, $\rho(\mathbf{r}) \, d\mathbf{r}$.

The energy of the uniform electron gas is divided up into exchange and correlation contributions. The exchange contribution can be evaluated analytically, but the correlation contribution has been evaluated numerically for a range in electron density using Monte Carlo simulations (Ceperley and Alder 1980). Note the difference between HF where the exchange of electrons is highly non-local, involving all electrons in the system irrespective of the distance between them, while the effect of exchange in LDA is only local, dependent only on the electron density at that point in space. The electrons are 'near-sighted' in LDA. The fact that

functional groups on molecules, such as an -OH group or -COOH groups, have similar properties irrespective of what atoms are present in the molecule a few bond lengths away, illustrates the fact that electrons are indeed near-sighted and that HF which includes infinite range exchange does in the end not provide the optimal orbital based description of many electron system, despite the apparently rigorous derivation from the variational principle using a single Slater determinant trial function.

The LDA approximation is most applicable to solids where the electron density is slowly varying, in particular for simple metals. It was applied extensively in solid state physics for a couple of decades but did not get applied much in chemistry since it overestimates the energy of chemical bonds. The problem is that the electron density in molecules varies greatly, especially in the tails of the electron density. The energy of atoms, where such tails involve a relatively larger volume of space is relatively too high compared with the energy of molecules, thus the overestimation of the binding energy.

Generalized gradient approximation

The accuracy of the functional approximation can be improved significantly by also taking into account the derivative of the electron density. Functionals that include information about the derivative(s) of the electron density are referred to as semi-local. The correction due to the derivative can be estimated from an electron gas with sinusoidal variation in the electron density. An important aspect of this is to make sure the exchange-correlation hole is properly accounted for. The estimate of E_{xc} is within this generalized gradient approximation (GGA) is

$$E_{xc}[\rho(\mathbf{r})] \ = \ \int f_{xc}^{1}[\rho(\mathbf{r}), |\nabla\rho(\mathbf{r})|]\rho(\mathbf{r})\, d\mathbf{r}.$$

A successful gradient based functional was presented by Becke and is referred to as BLYP. The correlation part of the functional is there taken from the work of Lee, Yang and Parr (thus the LYP part of the acronym). This functional is often used in calculations of molecules. In condensed matter studies, a functional due to Perdew and co-workers is more often used, the PBE functional.

Meta generalized gradient approximation

One can go further in the expansion of the electron density in terms of derivatives and include also the Laplacian. This is referred to as meta-GGA approximation and there the estimate of E_{xc} is formally

$$E_{xc}[\rho(\mathbf{r})] \ = \ \int f_{xc}^{2}[\rho(\mathbf{r}), |\nabla\rho(\mathbf{r})|, \nabla^{2}\rho(\mathbf{r})]\rho(\mathbf{r})\, d\mathbf{r}.$$

The most commonly used meta-GGA functional is the TPSS functional, and now more recently the more complex SCAN functional.

Hybrid functional approximations

It turns out that errors in the energy obtained from GGA and meta-GGA functionals are typically of opposite sign to errors in the energy obtained with HF. For example, the binding energy of molecules is overestimated in GGA while it is underestimated in HF. Also, GGA has a tendency to overly favor delocalized electronic states over localized states, while HF favors localized states. Not surprisingly, improved approximations can therefore be obtained by mixing GGA with HF. Since the orbitals that are obtained are strictly speaking not HF orbitals it is more common to refer to such an approach as a mix of GGA and exact exchange. The most commonly used functional in chemistry calculations is the B3LYP functional which includes 20% weight of exact exchange. Similarly, a mix of meta-GGA and exact exchange can be formulated as well. Such functionals are referred to as hybrid functionals. The problem is that there are many different ways of creating hybrid functionals and there is an ever increasing number of such approximations presented, some highly optimized for a particular set of systems, while the accuracy for other systems at the same time gets worse. DFT calculations are increasingly becoming semi-empirical (while they are often referred to as *ab initio* or 'first principles') in the sense that the functionals are developed to produce results that are in agreement with experimental data or high level wave function based calculations for small systems.