2 Theory and Practice of Density-Functional Theory

Peter E. Blöchl

Institute for Theoretical Physics

Clausthal University of Technology

Contents

1   Introduction

On the nanoscale, materials around us have surprisingly simple structures: The standard model of solid state physics and chemistry only knows of two types of particles, namely the nuclei making up the periodic table and the electrons. Only one kind of interaction between them needs to be considered, namely the electrostatic interaction. Even magnetic forces are important only in rare occasions. All other fundamental particles and interactions are irrelevant for chemistry.

The behavior of these particles can be described by the Schrödinger equation (or better the relativistic Dirac equation), which is easily written down. However, the attempt to solve this equation for any system of interest fails miserably due to what Walter Kohn termed the exponential wall [1].

To obtain an impression of the powers of the exponential wall, imagine the wave function of a N_2 molecule, having two nuclei and fourteen electrons. For N particles, the Schrödinger equation is a partial differential equation in $3N$ dimensions. Let us express the wave function on a grid with about 100 points along each spatial direction and let us consider two spin states for each electron. Such a wave function is represented by $2^{14} \, 100^{3 \times 16} \approx 10^{100}$ complex numbers. A data server for this amount of data, made of current terabyte hard disks, would occupy a volume with a diameter of 10^{10} light years!

Treating the nuclei as classical particles turned out to be a good approximation, but the quantum nature of the electrons cannot be ignored. A great simplification is to describe electrons as non-interacting quasi particles. Instead of one wave function in $3N$ dimensions, one only needs to describe N wave functions in three dimensions each, a dramatic simplification from 10^{100} to 10^7 numbers.

While the independent-particle model is very intuitive, and while it forms the basis of most text books on solid-state physics, materials physics, and chemistry, the Coulomb interaction between electrons is clearly not negligible.

Here, density-functional theory [2, 3] comes to our rescue: it provides a rigorous mapping from interacting electrons onto a system of non-interacting electrons. Unfortunately, the exact mapping is utterly complicated, and this is where all the complexity goes. Luckily, there are simple approximations that are both intuitive and surprisingly accurate. Furthermore, with the help of clever algorithms, density-functional calculations can be performed on current computers for large systems with several hundred atoms in a unit cell or a molecule. The microscopic insight gained from density functional calculations is a major source of progress in solid state physics, chemistry, material science, and biology.

In the first part of this article, I will try to familiarize the novice reader with the basics of density-functional theory, provide some guidance into common approximations and give an idea of the type of problems that can be studied with density functional theory.

Beyond this article, I recommend the insightful review articles on density functional theory by Jones and Gunnarsson [4], Baerends [5], von Barth [6], Perdew [7], Yang [8], and their collaborators.

Solving the one-particle Schrödinger equation, which results from density-functional theory,

for real materials is a considerable challenge. Several avenues have been developed to their solution. This is the field of electronic structure methods, which will be discussed in the second part of this article. This part is taken from earlier versions by Clemens Först, Johannes Kästner and myself [9, 10].

2 Basics of density-functional theory

The dynamics of the electron wave function is governed by the Schrödinger equation $i\hbar\partial_t|\Psi\rangle = \hat{H}|\Psi\rangle$ with the N-particle Hamiltonian $\hat{H}$.

$$\hat{H} = \sum_{j=1}^{N}\left(\frac{-\hbar^2}{2m_e}\vec{\nabla}_j^2 + v_{ext}(\vec{r}_j)\right) + \frac{1}{2}\sum_{i\neq j}^{N}\frac{e^2}{4\pi\epsilon_0|\vec{r}_i - \vec{r}_j|} \ . \tag{1}$$

With m_e we denote the electron mass, with ϵ_0 the vacuum permittivity, e is the elementary charge and $\hbar$ is the Planck quantum divided by 2π. The Coulomb potentials of the nuclei have been combined into an external potential $v_{ext}(\vec{r})$.

All N-electron wave functions $\Psi(\vec{x}_1, \ldots, \vec{x}_N)$ obey the Pauli principle, that is they change their sign, when two of its particle coordinates are exchanged.

We use a notation that combines the position vector $\vec{r} \in \mathbb{R}^3$ of an electron with its discrete spin coordinate $\sigma \in \{\uparrow, \downarrow\}$ into a single vector $\vec{x} := (\vec{r}, \sigma)$. Similarly, we use the notation of a four-dimensional integral $\int d^4x := \sum_\sigma \int d^3r$ for the sum over spin indices and the integral over the position. With the generalized symbol $\delta(\vec{x}-\vec{x'}) := \delta_{\sigma,\sigma'}\delta(\vec{r}-\vec{r'})$ we denote the product of Kronecker delta of the spin coordinates and Dirac's delta function for the positions. While, at first sight, it seems awkward to combine continuous and discrete numbers, this notation is less error prone than the notation that treats the spin coordinates as indices, where they can be confused with quantum numbers. During the first reading, the novice can ignore the complexity of the spin coordinates, treating $\vec{x}$ like a coordinate. During careful study, he will nevertheless have the complete and concise expressions.

One-particle reduced density matrix and two-particle density

In order to obtain the ground state energy $E = \langle\Psi|\hat{H}|\Psi\rangle$ we need to perform 2^N integrations in $3N$ dimensions each, i.e.

$$E \ = \ \int d^4x_1 \cdots \int d^4x_N \ \Psi^*(\vec{x}_1, \ldots, \vec{x}_N)\hat{H}\Psi(\vec{x}_1, \ldots, \vec{x}_N) \ . \tag{2}$$

However, only two different types of integrals occur in the expression for the energy, so that most of these integrations can be performed beforehand leading to two quantities of physical significance.

- One of these quantities is the one-particle reduced density matrix $\rho^{(1)}(\vec{x}, \vec{x'})$, which allows one to evaluate all expectation values of one-particle operators such as the kinetic energy

and the external potential,

$$\rho^{(1)}(\vec{x}, \vec{x'}) \; := \; N \int d^4 x_2 \ldots \int d^4 x_N \; \Psi(\vec{x}, \vec{x}_2, \ldots, \vec{x}_N) \Psi^*(\vec{x'}, \vec{x}_2, \ldots, \vec{x}_N) \; . \quad (3)$$

- The other one is the two-particle density $n^{(2)}(\vec{r}, \vec{r'})$, which allows to determine the inter-action between the electrons,

$$n^{(2)}(\vec{r}, \vec{r'}) \; := \; N(N-1) \sum_{\sigma, \sigma'} \int d^4 x_3 \ldots \int d^4 x_N \; |\Psi(\vec{x}, \vec{x'}, \vec{x}_3, \ldots, \vec{x}_N)|^2 \; . \quad (4)$$

If it is confusing that there are two different quantities depending on two particle coordinates, note that the one-particle reduced density matrix $\rho^{(1)}$ depends on two $\vec{x}$-arguments of the same particle, while the two-particle density $n^{(2)}$ depends on the positions of two different particles. With these quantities the total energy is

$$
\begin{aligned}
E \;=\;& \int d^4 x' \int d^4 x \; \delta(\vec{x'} - \vec{x}) \left(\frac{-\hbar^2}{2m_e} \vec{\nabla}^2 + v_{ext}(\vec{r}) \right) \rho^{(1)}(\vec{x}, \vec{x'}) \\
& + \frac{1}{2} \int d^3 r \int d^3 r' \; \frac{e^2 n^{(2)}(\vec{r}, \vec{r'})}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} \; ,
\end{aligned}
\quad (5)
$$

where the gradient of the kinetic energy operates on the first argument $\vec{r}$ of the density matrix.

One-particle reduced density matrix and natural orbitals

In order to make oneself familiar with the one-particle reduced density matrix, it is convenient to diagonalize it. The eigenstates $\varphi_n(\vec{r})$ are called natural orbitals [11] and the eigenvalues $\bar{f}_n$ are their occupations. The index n labeling the natural orbitals may stand for a set of quantum numbers.

The density matrix can be written in the form

$$\rho^{(1)}(\vec{x}, \vec{x'}) = \sum_n \bar{f}_n \varphi_n(\vec{x}) \varphi_n^*(\vec{x'}) \; . \quad (6)$$

The natural orbitals are orthonormal one-particle orbitals, i.e.

$$\int d^4 x \; \varphi_m^*(\vec{x}) \varphi_n(\vec{x}) = \delta_{m,n} \; . \quad (7)$$

Due to the Pauli principle, occupations are non-negative and never larger than one [12]. The natural orbitals already point the way to the world of effectively non-interacting electrons.

The one-particle density matrix provides us with the electron density

$$n^{(1)}(\vec{r}) = \sum_\sigma \rho^{(1)}(\vec{x}, \vec{x}) = \sum_\sigma \sum_n \bar{f}_n \varphi_n^*(\vec{x}) \varphi_n(\vec{x}) \; . \quad (8)$$

With the natural orbitals, the total energy Eq. 5 obtains the form

$$E = \sum_n \bar{f}_n \int d^4 x \; \varphi_n^*(\vec{x}) \frac{-\hbar^2}{2m} \vec{\nabla}^2 \varphi_n(\vec{x}) \;\; + \;\; \int d^3 r \; v_{ext}(\vec{r}) n^{(1)}(\vec{r})$$

$$+ \; \frac{1}{2} \int d^3 r \int d^3 r' \; \frac{e^2 n^{(2)}(\vec{r}, \vec{r'})}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} \; . \quad (9)$$

Two-particle density and exchange-correlation hole

The physical meaning of the two-particle density $n^{(2)}(\vec{r}, \vec{r'})$ is the following: For particles that are completely uncorrelated, meaning that they do not even experience the Pauli principle, the two particle density would be[1] the product of one-particle densities, i.e. $n^{(2)}(\vec{r}, \vec{r'}) = n^{(1)}(\vec{r})n^{(1)}(\vec{r'})$. If one particle is at position $\vec{r}_0$, the density of the remaining $N - 1$ particles is the conditional density

$$\frac{n^{(2)}(\vec{r}_0, \vec{r})}{n^{(1)}(\vec{r}_0)} \; .$$

The conditional density is the electron density seen by one of the electrons at $\vec{r}_0$. This observer electron obviously only sees the remaining $N - 1$ electrons.

It is convenient to express the two-particle density by the hole function $h(\vec{r}, \vec{r'})$, i.e.

$$n^{(2)}(\vec{r}, \vec{r'}) = n^{(1)}(\vec{r})\left[n^{(1)}(\vec{r'}) + h(\vec{r}, \vec{r'})\right] \; . \tag{10}$$

One electron at position $\vec{r}$ does not "see" the total electron density $n^{(1)}$ with N electrons, but only the density of the $N - 1$ other electrons, because it does not see itself. The hole function $h(\vec{r}_0, \vec{r})$ is simply the difference of the total electron density and the electron density seen by the observer electron at $\vec{r}_0$.

The division of the two-particle density in Eq. 10 suggests that we split the electron-electron interaction into the so-called Hartree energy

$$E_H \stackrel{\text{def}}{=} \frac{1}{2} \int d^3r \int d^3r' \; \frac{e^2 n^{(1)}(\vec{r})n^{(1)}(\vec{r'})}{4\pi\epsilon_0|\vec{r} - \vec{r'}|} \tag{11}$$

and the *potential energy of exchange and correlation*

$$U_{xc} \stackrel{\text{def}}{=} \int d^3r \; n^{(1)}(\vec{r})\frac{1}{2} \int d^3r' \; \frac{e^2 \; h(\vec{r}, \vec{r'})}{4\pi\epsilon_0|\vec{r} - \vec{r'}|} \; . \tag{12}$$

Keep in mind that U_{xc} is *not* the exchange correlation energy. The difference is a kinetic energy correction that will be discussed later in Eq. 19.

The hole function has a physical meaning: An electron sees the total density minus the electrons accounted for by the hole. Thus each electron not only experiences the electrostatic potential of the total electron density $n^{(1)}(\vec{r})$, but also the attractive potential of its own exchange correlation hole $h(\vec{r}_0, \vec{r})$.

A few facts for this hole density are apparent:

1. Because each electron of a N-electron system sees $N-1$ other electrons, the hole function integrates to exactly minus one electron

$$\int d^3r \; h(\vec{r}_0, \vec{r}) = -1 \tag{13}$$

   irrespective of the position $\vec{r}_0$ of the observing electron.

---

[1]This is correct only up to a term that vanishes in the limit of infinite particle number.
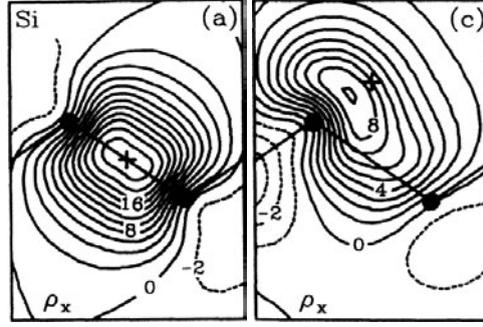
Fig. 1: *Exchange hole in silicon. The cross indicates the position of the observer electron. The black spheres and the lines indicate the atomic positions and bonds in the (110) plane. Reprinted figure with permission from Mark S. Hybertsen and Steven G. Louie, Physical Review B 34, 5390 (1986). Copyright 1986 by the American Physical Society.*

2. The density of the remaining $N - 1$ electrons can not be larger than the total electron density. This implies

$$h(\vec{r}_0, \vec{r}) \geq -n^{(1)}(\vec{r}) . \tag{14}$$

3. Due to the Pauli principle, no other electron with the same spin as the observer electron can be at the position $\vec{r}_0$. Thus the on-top hole $h(\vec{r}_0, \vec{r}_0)$ obeys the limits [13]

$$-\frac{1}{2}n^{(1)}(\vec{r}_0) \geq h(\vec{r}_0, \vec{r}_0) \geq -n^{(1)}(\vec{r}_0) . \tag{15}$$

4. Assuming locality, the hole function vanishes at large distances from the observer electron at $\vec{r}_0$, i.e.

$$h(\vec{r}_0, \vec{r}) \to 0 \qquad \text{for} \qquad |\vec{r} - \vec{r}_0| \to \infty . \tag{16}$$

With locality I mean that the density does not depend on the position or the presence of an observer electron, if the latter is very far away.

A selfmade functional

It is fairly simple to make our own density functional[2]: For a given density, we choose a simple shape for the hole function, such as a spherical box. Then we scale the value and the radius such that the hole function integrates to -1, and that its value is opposite equal to the spin density at its center. The electrostatic potential of this hole density at its center is the exchange-correlation energy for the observer electron. Our model has an exchange correlation energy[3] of

$$U_{xc}[n^{(1)}] \approx -\frac{1}{2} \int d^3r \, n^{(1)}(\vec{r}) \left(\frac{3}{4} \frac{e^2}{4\pi\epsilon_0} \sqrt[3]{\frac{2\pi}{3}} \left(n^{(1)}(\vec{r}) \right)^{\frac{1}{3}} \right) \sim \int d^3r \left(n^{(1)}(\vec{r}) \right)^{\frac{4}{3}} . \tag{17}$$

---

[2] A functional $F[y]$ maps a function $y(x)$ to a number F. It is a generalization of the function $F(\vec{y})$ of a vector $\vec{y}$, where the vector index of $\vec{y}$ is turned into a continuous argument x.

[3] For this model we do not distinguish between the energy of exchange and correlation and its potential energy contribution
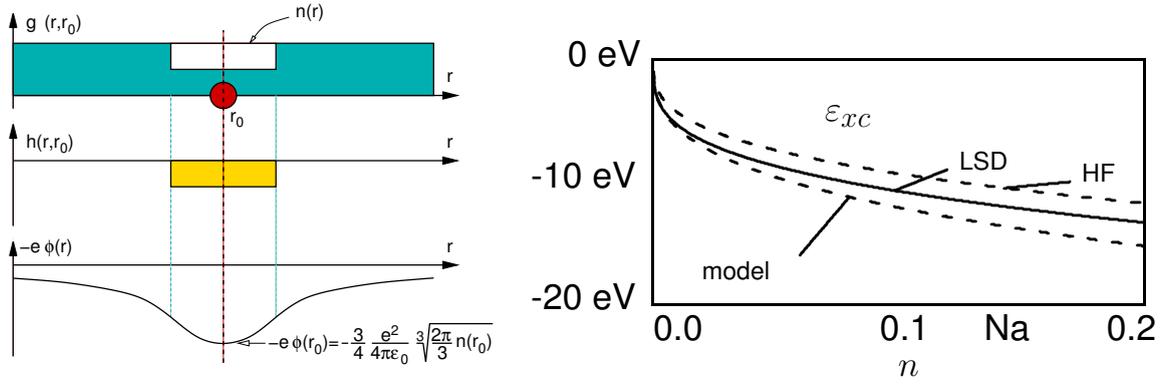
Fig. 2: *Left: Scheme to demonstrate the construction of the exchange correlation energy from a simple model. Right: exchange correlation energy per electron ϵ_{xc} as function of electron density from our model, Hartree-Fock approximation and the exact result. The symbol "Na" indicates the density of Sodium.*

The derivation is an elementary exercise and is given in the appendix. The resulting energy per electron ϵ_{xc} is given on the right-hand side of Fig. 2 indicated as "model" and compared with the exact result indicated as "LSD" and the Hartree-Fock result indicated as "HF" for a homogeneous electron gas.

The agreement with the correct result, which is surprisingly good for such a crude model, provides an idea of how robust the density-functional theory is with respect to approximations. While this model has been stripped to the bones, it demonstrates the way physical insight enters the construction of density functionals. Modern density functionals are far more sophisticated and exploit much more information [14], but the basic method of construction is similar.

Kinetic energy

While the expression for the kinetic energy in Eq. 9 seems familiar, there is a catch to it. In order to know the natural orbitals and the occupations we need access to the many-particle wave function or at least to its reduced density matrix.

A good approximation for the kinetic energy of the interacting electrons is the kinetic energy functional $T_s[n^{(1)}]$ of the ground state of non-interacting electrons with the same density as the true system. It is defined by

$$
\begin{aligned}
T_s[n^{(1)}] \;=\; \min_{\{f_n \in [0,1], |\psi_n\rangle\}} \Bigg\{ & \sum_n f_n \int d^4x \; \psi_n^*(\vec{x}) \frac{-\hbar^2 \vec{\nabla}^2}{2m} \psi_n(\vec{x}) \\
& + \int d^3r \; v_{eff}(\vec{r}) \left(\left[\sum_n f_n \sum_\sigma \psi_n^*(\vec{x}) \psi_n(\vec{x}) \right] - n^{(1)}(\vec{r}) \right) \\
& - \sum_{n,m} \Lambda_{m,n} \Big(\langle \psi_n | \psi_m \rangle - \delta_{n,m} \Big) \Bigg\} .
\end{aligned}
\tag{18}
$$

Note that $f_n \neq \bar{f}_n$ and that the so-called Kohn-Sham orbitals $\psi_n(\vec{x})$ differ[4] from the natural

---

[4]To be precise, Kohn-Sham orbitals are the natural orbitals for non-interacting electrons of a given density.

orbitals $\varphi_n(\vec{x})$. Natural orbitals and Kohn-Sham wave functions are fairly similar, while the occupations f_n of Kohn-Sham orbitals differ considerably from those $\bar{f}_m$ of the natural orbitals. The effective potential $v_{eff}(\vec{r})$ is the Lagrange multiplier for the density constraint. $\Lambda_{m,n}$ is the Lagrange multiplier for the orthonormality. Diagonalization of $\boldsymbol{\Lambda}$ yields a diagonal matrix with the one-particle energies on the diagonal.

This kinetic energy $T_s[n^{(1)}]$ is a unique functional of the density, which is the first sign that we are approaching a density-functional theory. Also it is the introduction of this kinetic energy, where we made for the first time a reference to a ground state. Density functional theory as described here is inherently a ground-state theory.

Why does the true kinetic energy of the interacting system differ from that of the non-interacting energy? Consider the hole function of a non-interacting electron gas. When inserted into Eq. 12 for U_{xc} the potential energy of exchange and correlation, we obtain a contribution to the total energy that is called exchange energy. The interaction leads to a second energy contribution that is called correlation energy. Namely, when the interaction is switched on, the wave function is deformed in such a way that the Coulomb repulsion between the electrons is reduced. This makes the hole function more compact. However, there is a price to pay when the wave functions adjust to reduce the Coulomb repulsion between the electrons, namely an increase of the kinetic energy: Pushing electrons away from the neighborhood of the reference electrons requires that work be performed against the kinetic pressure of the electron gas, which raises the kinetic energy. Thus, the system has to find a compromise between minimizing the electrostatic repulsion of the electrons and increasing its kinetic energy. As a result, the correlation energy has a potential-energy contribution and a kinetic-energy contribution.

This tradeoff can be observed in Fig. 2. The correct exchange correlation energy is close to our model at low densities, while it becomes closer to the Hartree-Fock result at high densities. This is consistent with the fact that the electron gas can easily be deformed at low densities, while the deformation becomes increasingly costly at high densities due to the larger pressure of the electron gas.

The difference between T_s and the true kinetic energy is combined with the potential energy of exchange and correlation U_{xc} from Eq. 12 into the exchange correlation energy E_{xc}, i.e.

$$E_{xc} = U_{xc} + \sum_n \bar{f}_n \int d^4x \, \varphi_n^*(\vec{x}) \frac{-\hbar^2}{2m} \vec{\nabla}^2 \varphi_n(\vec{x}) - T_s[n^{(1)}] \ . \tag{19}$$

Note, that the $\phi_n(\vec{x})$ and the $\bar{f}_n$ are natural orbitals and occupations of the interacting electron gas, and that they differ from the Kohn-Sham orbitals $\psi_n(\vec{x})$ and occupations f_n.

They are however different from the natural orbitals of interacting electrons at the same density.

Total energy

The total energy obtains the form

$$
\begin{aligned}
E \;=\; & \min_{|\Phi\rangle,\{|\psi_n\rangle, f_n \in [0,1]\}} \Bigg\{ \sum_n f_n \int d^4x \; \psi_n^*(\vec{x}) \frac{-\hbar^2}{2m} \vec{\nabla}^2 \psi_n(\vec{x}) \\
& + \int d^3r \; v_{eff}(\vec{r}) \left(\left[\sum_n f_n \sum_\sigma \psi_n^*(\vec{x})\psi_n(\vec{x}) \right] - n(\vec{r}) \right) + \int d^3r \; v_{ext}(\vec{r}) n^{(1)}(\vec{r}) \\
& + \frac{1}{2} \int d^3r \int d^3r' \; \frac{e^2 n^{(1)}(\vec{r}) n^{(1)}(\vec{r'})}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} + E_{xc} - \sum_{n,m} \Lambda_{m,n} \Big(\langle \psi_n | \psi_m \rangle - \delta_{n,m} \Big) \Bigg\} \; . \quad (20)
\end{aligned}
$$

In order to evaluate the total energy with Eq. 20, we still have to start from the many-particle wave function $|\Phi\rangle$. Only the many-particle wave function allows us to evaluate the one-particle density $n^{(1)}(\vec{r})$ and the exchange correlation energy E_{xc}. Kohn-Sham orbitals $|\psi_n\rangle$ and occupations f_n are obtained by an independent minimization for each density.

If, however, we were able to express the exchange-correlation energy E_{xc} as a functional of the density alone, there would be no need for the many-particle wave function at all and the terrors of the exponential wall would be banned. We could minimize Eq. 20 with respect to the density, Kohn-Sham orbitals and their occupations.

Let us, for the time being, simply assume that $E_{xc}[n^{(1)}]$ is a functional of the electron density and explore the consequences of this assumption. Later, I will show that this assumption is actually valid.

The minimization in Eq. 20 with respect to the one-particle wave functions yields the Kohn-Sham equations

$$
\left[\frac{-\hbar^2}{2m_e} \vec{\nabla}^2 + v_{eff}(\vec{r}) - \epsilon_n \right] \psi_n(\vec{x}) = 0 \qquad \text{with} \qquad \int d^4x \; \psi_m(\vec{x})\psi_n(\vec{x}) = \delta_{m,n} \; . \quad (21)
$$

The Kohn-Sham energies ϵ_n are the diagonal elements of the Lagrange multiplier $\mathbf{\Lambda}$, when the latter is forced to be diagonal.

The requirement that the derivative of the total energy Eq. 20 with respect to the density vanishes, yields an expression for the effective potential

$$
v_{eff}(\vec{r}) = v_{ext}(\vec{r}) + \int d^3r' \; \frac{e^2 n^{(1)}(\vec{r'})}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} + \frac{\delta E_{xc}[n^{(1)}]}{\delta n^{(1)}(\vec{r})} \; . \quad (22)
$$

Both equations, together with the density constraint

$$
n^{(1)}(\vec{r}) = \sum_n f_n \sum_\sigma \psi_n^*(\vec{x})\psi_n(\vec{x}) \; , \quad (23)
$$

form a set of coupled equations, that determine the electron density and the total energy. This set of coupled equations, Eqs. 21, 22, and 23, is what is solved in the so-called self-consistency loop. Once the set of self-consistent equations has been solved, we obtain the electron density and we can evaluate the total energy.
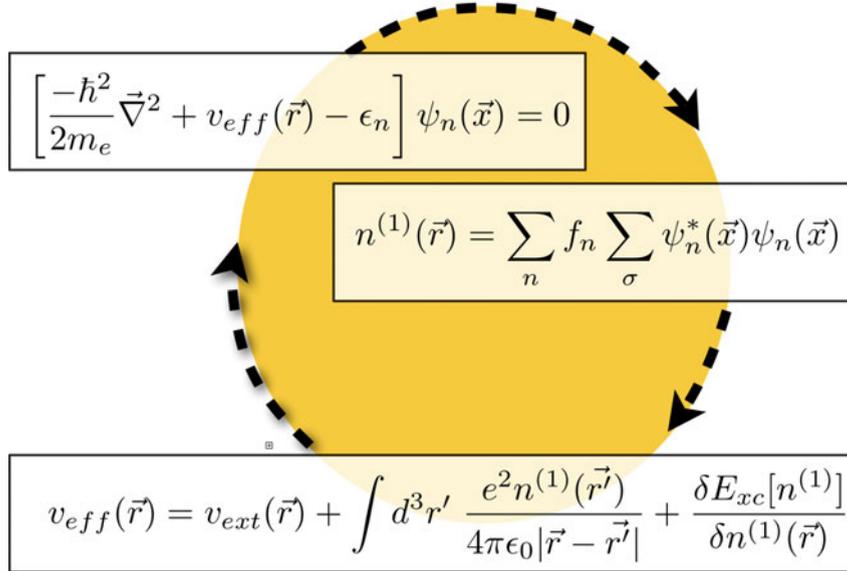
$$\left[\frac{-\hbar^2}{2m_e} \vec{\nabla}^2 + v_{eff}(\vec{r}) - \epsilon_n \right] \psi_n(\vec{x}) = 0$$

$$n^{(1)}(\vec{r}) = \sum_n f_n \sum_\sigma \psi_n^*(\vec{x})\psi_n(\vec{x})$$

$$v_{eff}(\vec{r}) = v_{ext}(\vec{r}) + \int d^3r' \frac{e^2 n^{(1)}(\vec{r'})}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} + \frac{\delta E_{xc}[n^{(1)}]}{\delta n^{(1)}(\vec{r})}$$

Fig. 3: *Self-consistency cycle.*

In practice, one often makes the assumption that the non-interacting electrons in the effective potential closely resemble the true interacting electrons, and extracts a wealth of other physical properties from the Kohn-Sham wave functions $|\psi_n\rangle$ and the Kohn-Sham energies ϵ_n. However, there is little theoretical backing for this approach and, if it fails, one should not blame density functional theory!

Is there a density functional?

The argument leading to the self-consistent equations, Eqs. 21, 22, and 23, relied entirely on the hope that exchange correlation functional can be expressed as a functional of the electron density. In fact, this can easily be shown, if we restrict us to ground state densities. The proof goes back to the seminal paper by Levy [15, 16].

Imagine that one could construct all fermionic many-particle wave functions. For each of these wave functions, we can determine in a unique way the electron density

$$n^{(1)}(\vec{r}) = N \sum_\sigma \int d^3x_2 \ldots \int d^3x_N \ |\Psi(\vec{x}, \vec{x}_2, \ldots \vec{x}_N)|^2 \ . \tag{24}$$

Having the electron densities, we sort the wave functions according to their density. For each density, I get a mug $M[n^{(1)}]$ that holds all wave functions with that density, which is written on the label of the mug.

Now we turn to each mug $M[n^{(1)}]$ in sequence and determine for each the wave function with the lowest energy. Because the external potential energy is the same for all wave functions with the same density, we need to consider only the kinetic energy operator $\hat{T}$ and the operator $\hat{W}$ of
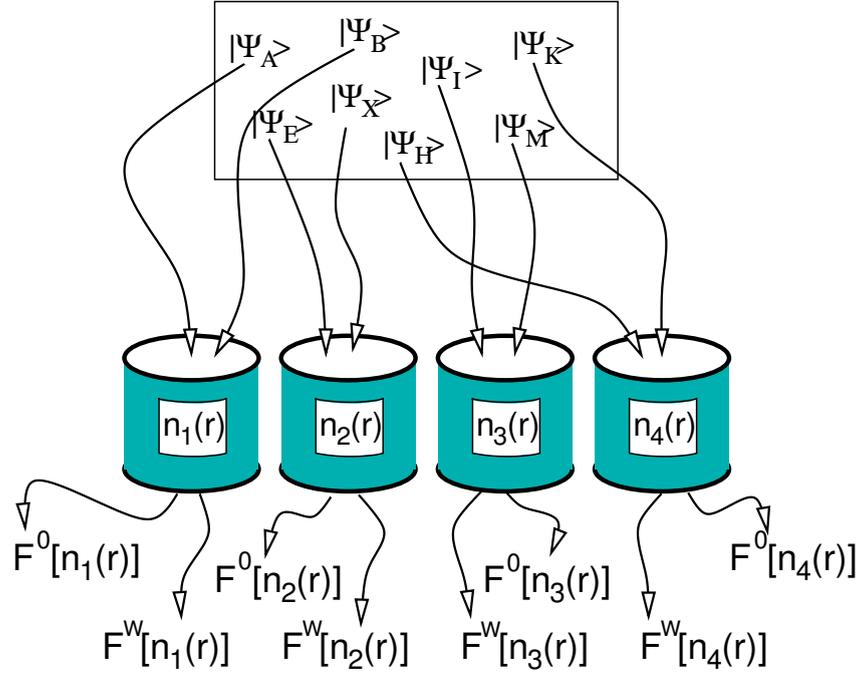
Fig. 4: *Illustration for Levy's proof that there exists a density functional.*

the electron-electron interaction, and we do not need to consider the external potential.

$$F^{\hat{W}}[n^{(1)}] = \min_{|\Psi\rangle \in M[n^{(1)}]} \langle \Psi | \hat{T} + \hat{W} | \Psi \rangle \tag{25}$$

$F^{\hat{W}}[n^{(1)}]$ is the universal density functional. It is universal in the sense that it is an intrinsic property of the electron gas and absolutely independent of the external potential.

Next, we repeat the same construction as that for a universal density functional, but now we leave out the interaction $\hat{W}$ and consider only the kinetic energy $\hat{T}$.

$$F^0[n^{(1)}] = \min_{|\Psi\rangle \in M[n^{(1)}]} \langle \Psi | \hat{T} | \Psi \rangle \tag{26}$$

The resulting functional $F^0[n^{(1)}]$ is nothing but the kinetic energy of non-interacting electrons $T_s[n^{(1)}]$.

Now we can write down the total energy as functional of the density

$$E[n^{(1)}] = F^{\hat{W}}[n^{(1)}] + \int d^3r \, v_{ext}(\vec{r}) n^{(1)}(\vec{r}) \tag{27}$$

When we compare Eq. 27 with Eq. 20, we obtain an expression for the exchange correlation energy.

$$E_{xc}[n^{(1)}] = F^{\hat{W}}[n^{(1)}(\vec{r})] - F^0[n^{(1)}(\vec{r})] - \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 n^{(1)}(\vec{r}) n^{(1)}(\vec{r'})}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} \tag{28}$$

This completes the proof that the exchange correlation energy is a functional of the electron density. The latter was the assumption for the derivation of the set of self-consistent equations, Eqs. 21, 22, and 23 for the Kohn-Sham wave functions $\psi_n(\vec{x})$.

With this, I finish the description of the theoretical basis of density-functional theory. We have seen that the total energy can rigorously be expressed as a functional of the density or, in practice, as a functional of a set of one-particle wave functions, the Kohn-Sham wave functions and their occupations. Density functional theory per se is not an approximation and, in contrast to common belief, it is not a mean-field approximation. Nevertheless, we need to introduce approximations to make density functional theory work. This is because the exchange correlation energy $E_{xc}[n^{(1)}]$ is not completely known. These approximations will be discussed in the next section.

3   Jacob's ladder of density functionals

The development of density functionals is driven by mathematical analysis of the exact exchange correlation hole [14, 7], physical insight and numerical benchmark calculations on real systems. The functionals evolved in steps from one functional form to another, with several parameterizations at each level. Perdew pictured this development by Jacob's ladder leading up to heaven [17, 7]. In his analogy the different rungs of the ladder represent the different levels of density functionals leading to the unreachable, ultimately correct functional.

LDA, the big surprise

The first density functionals used in practice were based on the local-density approximation (LDA). The hole function for an electron at position $\vec{r}$ has been approximated by the one of a homogeneous electron gas with the same density as $n^{(1)}(\vec{r})$. The exchange correlation energy for the homogeneous electron gas has been obtained by quantum Monte Carlo calculations [18] and analytic calculations [19]. The local density approximation has been generalized early to local spin-density approximation (LSD) [20].

Truly surprising was how well the theory worked for real systems. Atomic distances could be determined within a few percent of the bond length and energy differences in solids were surprisingly good.

This was unexpected, because the density in real materials is far from homogeneous. Gunnarsson and Lundquist [21] explained this finding with sumrules that are obeyed by the local density approximation: Firstly, the exchange correlation energy depends only on the spherical average of the exchange correlation hole. Of the radial hole density only the first moment contributes, while the second moment is fixed by the sum-rule that the electron density of the hole integrates to -1. Thus we can use

$$\int d^3r \; \frac{e^2 h(\vec{r}_0, \vec{r})}{4\pi\epsilon_0 |\vec{r} - \vec{r}_0|} = -\frac{e^2}{4\pi\epsilon_0} \frac{\int_0^\infty dr \; r \langle h(\vec{r}_0, \vec{r'}) \rangle_{|\vec{r'} - \vec{r}_0| = r}}{\int_0^\infty dr \; r^2 \langle h(\vec{r}_0, \vec{r'}) \rangle_{|\vec{r'} - \vec{r}_0| = r}} \tag{29}$$

where the angular brackets imply the angular average of $\vec{r'} - \vec{r}_0$. This dependence on the hole density is rather insensitive to small changes of the hole density. Even for an atom, the *spherically averaged* exchange hole closely resembles that of the homogeneous electron gas [4].

The main deficiency of the LDA was the strong overbinding with bond energies in error by about one electron volt. On the one hand, this rendered LDA useless for most applications in chemistry. On the other hand, the problem was hardly visible in solid state physics where bonds are rarely broken, but rearranged so that errors cancelled.

GGA, entering chemistry

Being concerned about the large density variations in real materials, one tried to include the first terms of a Taylor expansion in the density gradients. These attempts failed miserably. The culprit has been a violation of the basic sum rules as pointed out by Perdew [22]. The cure was a cutoff for the gradient contributions at high gradients, which lead to the class of generalized gradient approximations (GGA) [23].

Becke [24] provides an intuitive description for the workings of GGA's, which I will sketch here in a simplified manner: Becke uses an ansatz $E_{xc} = \int d^3r \, A(n(\vec{r}))F(x(\vec{r}))$ for the exchange-correlation energy where $n(\vec{r})$ is the local density and $x = |\vec{\nabla}n|/n^{\frac{4}{3}}$ is a dimensionless reduced gradient. Do not confuse this symbol with the combined position-and-spin coordinate $\vec{x}$. The function A is simply the LDA expression and $F(x)$ is the so-called enhancement factor. The large-gradient limit of $F(x)$ is obtained from a simple physical argument:
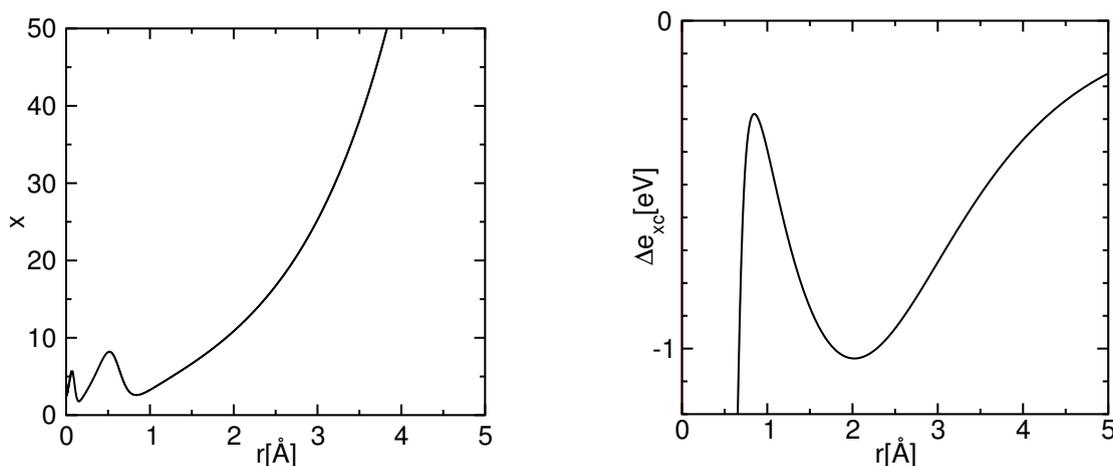

Fig. 5: *Left figure: reduced density gradient $x = |\vec{\nabla}n|/n^{\frac{4}{3}}$ of a silicon atom as function of distance from the nucleus demonstrating that the largest reduced gradients occur in the exponential tails. Right figure: additional contribution from the gradient correction (PBE versus PW91 LDA) of the exchange correlation energy per electron. The figure demonstrates that the gradient correction stabilizes the tails of the wave function. The covalent radius of silicon is at 1.11 Å.*

Somewhat surprisingly, the reduced gradient is largest not near the nucleus but in the exponentially decaying charge-density tails as shown in Fig. 5. For an electron that is far from an atom, the hole is on the atom, because a hole can only be dug where electrons are. Thus the Coulomb interaction energy of the electron with its hole is $-\frac{e^2}{4\pi\epsilon_0 r}$, where r is the distance of the reference

electron from the atom. As shown in appendix B, the enhancement factor can now be obtained by enforcing this behavior for exponentially decaying densities.

As a result, the exchange and correlation energy per electron in the tail region of the electron density falls of with the inverse distance in GGA, while it has a much faster, exponential decay in the LDA. Thus, the the tail region is stabilized by GGA. This contribution acts like a negative "surface energy".

When a bond between two atoms is broken, the surface is increased. In GGA this bond-breaking process is more favorable than in LDA, and, hence, the bond is weakened. Thus the GGA cures the overbinding error of the LDA.

These gradient corrections greatly improved the bond energies and made density functional theory useful also for chemists. The most widely distributed GGA functional is the Perdew-Burke-Ernzerhof (PBE) functional [25].

Meta GGA's

The next level of density functionals are the so-called meta GGA's [26–28] that include not only the gradient of the density, but also the second derivatives of the density. These functionals can be reformulated so that the additional parameter is the kinetic energy density instead of the second density derivatives. Perdew recommends his TPSS functional [29].

Hybrid functionals

Another generation of functionals are hybrid functionals [30, 31], which replace some of the exchange energy by the exact exchange

$$E_X^{HF} = -\frac{1}{2}\sum_{m,n}\bar{f}_m\bar{f}_n\int d^4x\int d^4x'\frac{e^2\psi_m^*(\vec{x})\psi_n(\vec{x})\psi_n^*(\vec{x'})\psi_m(\vec{x'})}{4\pi\epsilon_0|\vec{r}-\vec{r'}|} \tag{30}$$

where $\bar{f}_n$ and the $\psi_n(\vec{x})$ are the Kohn-Sham occupations and wave functions, respectively. The motivation for this approach goes back to the adiabatic connection formula [32, 33, 21]

$$E_{xc}[n(\vec{r})] = \int_0^1 d\lambda\, U_{xc}^{\lambda\hat{W}}[n(\vec{r})] = \int d^3r\, n(\vec{r})\int_0^1 d\lambda\frac{1}{2}\int d^3r'\frac{h_\lambda(\vec{r},\vec{r'})}{4\pi\epsilon|\vec{r}-\vec{r'}|} \tag{31}$$

which expresses the exchange correlation energy as an integral of the potential energy of exchange and correlation over the interaction strength. Here the interaction in the Hamiltonian is scaled by a factor λ, leading to a λ-dependent universal functional $F^{\lambda\hat{W}}[n^{(1)}]$. The interaction energy can be expressed by

$$\begin{aligned}F^{\hat{W}}[n] &= F^0[n] + \int_0^1 d\lambda\,\frac{d}{d\lambda}F^{\lambda\hat{W}}[n]\\ &= T_s[n] + \frac{1}{2}\int d^3r\int d^3r'\,\frac{e^2n(\vec{r})n(\vec{r'})}{4\pi\epsilon_0|\vec{r}-\vec{r'}|} + \int_0^1 d\lambda\, U_{xc}^{\lambda\hat{W}}[n]\end{aligned} \tag{32}$$

which leads via Eq. 28 to Eq. 31. Using perturbation theory, the derivative of $F^{\lambda\hat{W}}[n]$ simplifies to the expectation value of the interaction $\langle\Psi(\lambda)|\hat{W}|\Psi(\lambda)\rangle$, which is the potential energy of exchange and correlation evaluated for a many-particle wave function obtained for the specified given interaction strength.

The underlying idea of the hybrid functionals is to interpolate the integrand between the end points. In the non-interacting limit, i.e. for $\lambda = 0$ the integrand $U_{xc}^{\lambda\hat{W}}$ is exactly given by the exact exchange energy of Eq. 30. For the full interaction, on the other hand, the LDA or GGA functionals are considered correctly. Thus a linear interpolation would yield

$$E_{xc} = \frac{1}{2}\left(U_{xc}^0 + U_{xc}^{\hat{W}}\right) = \frac{1}{2}\left(E_X^{HF} + U_{xc}^{\hat{W}}\right) = E_{xc}^{GGA} + \frac{1}{2}\left(E_X^{HF} - E_X^{GGA}\right). \qquad (33)$$

Depending on whether the λ-dependence is a straight line or whether it is convex, the weight factor may be equal or smaller than $\frac{1}{2}$. Perdew [34] has given arguments that a factor $\frac{1}{4}$ would actually be better than a factor $\frac{1}{2}$.

Hybrid functionals perform substantially better than GGA functionals regarding binding energies, band gaps and reaction energies. However, they are flawed for the description of solids. The reason is that the exact exchange hole in a solid is very extended. These long-range tails are screened away quickly when the interaction is turned on, because they are cancelled by the correlation. Effectively, we should use a smaller mixing factor for the long range part of the exchange hole. This can be taken into account, by cutting off the long-range part of the interaction for the calculation of the Hartree-Fock exchange [35]. This approach improves the results for band gaps while reducing the computational effort [36].

The effective cancellation of the long-ranged contribution of exchange with a similar contribution from correlation, which is also considered properly already in the LDA, is one of the explanation for the superiority of the LDA over the Hartree-Fock approximation.

The most widely used hybrid functional is the B3LYP functional [37], which is, however, obtained from a parameter fit to a database of simple molecules. The functional PBE0 [38, 39] is born out of the famous PBE GGA functional and is a widely distributed parameter-free functional.

LDA+U and local hybrid functionals

Starting from a completely different context, Anisimov et. al. [40] introduced the so-called LDA+U method, which, as described below, has some similarities to the hybrid functionals above.

The main goal was to arrive at a proper description of transition metal oxides, which tend to be Mott insulators, while GGA calculations predict them often to be metals. The remedy was to add a correlation term[5] [41] borrowed from the Hubbard model and to correct the resulting

---

[5]The expression given here looks unusually simple. This is due to the notation of spin orbitals, which takes care of the spin indices.

double counting of the interactions by E_{dc}.

$$E \;=\; E^{GGA} + \frac{1}{2} \sum_R \sum_{\alpha,\beta,\gamma,\delta \in \mathcal{C}_R} U_{\alpha,\beta,\gamma,\delta} \left(\rho_{\gamma,\alpha}\rho_{\delta,\beta} - \rho_{\delta,\alpha}\rho_{\gamma,\beta} \right) - E_{dc} \tag{34}$$

$$U_{\alpha,\beta,\gamma,\delta} \;=\; \int d^4x \int d^4x' \frac{e^2 \chi_\alpha^*(\vec{x}) \chi_\beta^*(\vec{x'}) \chi_\gamma(\vec{x}) \chi_\delta(\vec{x'})}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} \tag{35}$$

$$\rho_{\alpha,\beta} \;=\; \langle \pi_\alpha | \psi_n \rangle f_n \langle \psi_n | \pi_\beta \rangle \;, \tag{36}$$

where $|\chi_\alpha\rangle$ are atomic tight-binding orbitals and $|\pi_\alpha\rangle$ are their projector functions.[6] The additional energy is a Hartree-Fock exchange energy, that only considers the exchange for specified sets of local orbitals. The exchange term does only consider a subset of orbitals $\mathcal{C}_R$ for each atom R and it ignores the contribution involving orbitals centered on different atoms.

Novak et al. [42] made the connection to the hybrid functionals explicit and restricted the exact exchange contribution of a hybrid functional to only a shell of orbitals. While in the LDA+U method the bare Coulomb matrix elements are reduced by a screening factor, in the hybrid functionals it is the mixing factor that effectively plays the same role. Both LDA+U and the local hybrid method have in common that they radically remove the contribution of off-site matrix elements of the interaction. Tran et al. [43] applied this method to transition metal oxides and found results that are similar to those of the full implementation of hybrid functionals.

Van der Waals interactions

One of the major difficulties for density functionals is the description of van der Waals forces, because it is due to the quantum mechanical synchronization of charge fluctuations on distinct molecules. I refer the reader to the work made in the group of Lundqvist [44–46].

4   Benchmarks, successes and failures

The development of density functionals has profited enormously from careful benchmark studies. The precondition is a data set of test cases for which reliable and accurate experimental data exist. The most famous data sets are the G1 and G2 databases [47–50] that have been set up to benchmark quantum-chemistry codes. Becke [51, 52, 31, 53, 54] set a trend by using these large sets of test cases for systematic studies of density functionals. In order to separate out the accuracy of the density functionals, it is vital to perform these calculations on extremely accurate numerical methods. Becke used basis set free calculations that were limited to small molecules, while being extremely accurate. Paier et. al. [55–57, 36] have later performed careful comparisons of two methods, Gaussian and the projector augmented-wave method, to single out the error of the electronic structure method.

---

[6]Projector functions obey the biorthogonality condition $\langle \chi_\alpha | \pi_\beta \rangle = \delta_{\alpha,\beta}$. Within the sub-Hilbert space of the tight-binding orbitals, i.e. for wave functions of the form $|\psi\rangle = \sum_\alpha |\chi_\alpha\rangle c_\alpha$, the projector functions decompose the wave function into tight binding orbitals, i.e. $|\psi\rangle = \sum_\alpha |\chi_\alpha\rangle \langle \pi_\alpha | \psi \rangle$. A similar projection is used extensively in the projector augmented-wave method described later.

Overall, the available density functionals predict molecular structures very well. Bond distances agree with the experiment often within one percent. Bond angles come out within a few degrees. The quality of total energies depends strongly on the level of functionals used. On the LDA level bonds are overestimated in the 1 eV range, on the GGA level these errors are reduced to a about 0.3 eV, and hybrid functionals reduce the error by another factor of 2. The ultimate goal is to reach chemical accuracy, which is about 0.05 eV. Such an accuracy allows to predict reaction rates at room temperature within a factor of 10.

Band gaps are predicted to be too small with LDA and GGA. The so-called band gap problem has been one of the major issues during the development of density functionals. Hybrid functionals clearly improve the situation. A problem is the description of materials with strong electron correlations. For LDA and GGA many insulating transition metal oxides are described as metals. This changes again for the hybrid functionals, which turns them into antiferromagnetic insulators, which is a dramatic improvement.

5 Electronic structure methods

In this second part of my lecture notes, I will address the problem of how to solve the Kohn-Sham equations and how to obtain the total energy and other observables. It is convenient to use a slightly different notation: Instead of treating the nuclei via an external potential, we combine all electrostatic interactions into a single double integral.

This brings the total energy into the form

$$
\begin{aligned}
E\Big[\{\psi_n(\vec{r})\}, \{\vec{R}_R\}\Big] &= \sum_n f_n \langle \psi_n | \frac{\hat{\vec{p}}^2}{2m_e} | \psi_n \rangle \\
&+ \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 \Big(n(\vec{r}) + Z(\vec{r}) \Big) \Big(n(\vec{r'}) + Z(\vec{r'}) \Big)}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} + E_{xc}[n] ,
\end{aligned}
\tag{37}
$$

where $Z(\vec{r}) = -\sum_R \mathcal{Z}_R \delta(\vec{r} - \vec{R}_R)$ is the nuclear charge density expressed in electron charges. Z_R is the atomic number of a nucleus at position $\vec{R}_R$.

The electronic ground state is determined by minimizing the total energy functional $E[\Psi_n]$ of Eq. 37 at a fixed ionic geometry. The one-particle wave functions have to be orthogonal. This constraint is implemented with the method of Lagrange multipliers. We obtain the ground-state wave functions from the extremum condition for

$$
Y\Big[\{|\psi_n\rangle\}, \boldsymbol{\Lambda}\Big] = E\Big[\{|\psi_n\rangle\}\Big] - \sum_{n,m} \Big[\langle \psi_n | \psi_m \rangle - \delta_{n,m}\Big] \Lambda_{m,n}
\tag{38}
$$

with respect to the wavefunctions and the Lagrange multipliers $\Lambda_{m,n}$. The extremum condition for the wavefunctions has the form

$$
\hat{H}|\psi_n\rangle f_n = \sum_m |\psi_m\rangle \Lambda_{m,n} ,
\tag{39}
$$

where $\hat{H} = \frac{1}{2m_e}\hat{\vec{p}}^2 + \hat{v}_{\text{eff}}$ is the effective one-particle Hamilton operator.

The corresponding effective potential depends itself on the electron density via

$$v_{eff}(\vec{r}) = \int d^3r' \frac{e^2\left(n(\vec{r'}) + Z(\vec{r'})\right)}{4\pi\epsilon_0|\vec{r} - \vec{r'}|} + \mu_{xc}(\vec{r}) , \tag{40}$$

where $\mu_{xc}(\vec{r}) = \frac{\delta E_{xc}[n(\vec{r})]}{\delta n(\vec{r})}$ is the functional derivative of the exchange and correlation functional. After a unitary transformation that diagonalizes the matrix of Lagrange multipliers $\boldsymbol{\Lambda}$, we obtain the Kohn-Sham equations

$$\hat{H}|\psi_n\rangle = |\psi_n\rangle\epsilon_n . \tag{41}$$

The one-particle energies ϵ_n are the eigenvalues of the matrix with the elements $\Lambda_{n,m}(f_n + f_m)/(2f_nf_m)$ [58].

The one-electron Schrödinger equations, namely the Kohn-Sham equations given in Eq. 21, still pose substantial numerical difficulties: (1) in the atomic region near the nucleus, the kinetic energy of the electrons is large, resulting in rapid oscillations of the wavefunction that require fine grids for an accurate numerical representation. On the other hand, the large kinetic energy makes the Schrödinger equation stiff, so that a change of the chemical environment has little effect on the shape of the wavefunction. Therefore, the wavefunction in the atomic region can be represented well already by a small basis set. (2) In the bonding region between the atoms the situation is opposite. The kinetic energy is small and the wavefunction is smooth. However, the wavefunction is flexible and responds strongly to the environment. This requires large and nearly complete basis sets.

Combining these different requirements is non-trivial and various strategies have been developed.

- The atomic point of view has been most appealing to quantum chemists. Basis functions are chosen that resemble atomic orbitals. This choice exploits that the wavefunction in the atomic region can be described by a few basis functions, while the chemical bond is described by the overlapping tails of these atomic orbitals. Most techniques in this class are a compromise of, on the one hand, a well adapted basis set, where the basis functions are difficult to handle, and, on the other hand, numerically convenient basis functions such as Gaussians, where the inadequacies are compensated by larger basis sets.

- Pseudopotentials regard an atom as a perturbation of the free electron gas. The most natural basis functions for the free electron gas are plane waves. Plane-wave basis sets are in principle complete and suitable for sufficiently smooth wavefunctions. The disadvantage of the comparably large basis sets required is offset by their extreme numerical simplicity. Finite plane-wave expansions are, however, absolutely inadequate to describe the strong oscillations of the wavefunctions near the nucleus. In the pseudopotential approach the Pauli repulsion by the core electrons is therefore described by an effective potential that expels the valence electrons from the core region. The resulting wavefunctions are smooth

and can be represented well by plane waves. The price to pay is that all information on the charge density and wavefunctions near the nucleus is lost.

- Augmented-wave methods compose their basis functions from atom-like wavefunctions in the atomic regions and a set of functions, called envelope functions, appropriate for the bonding in between. Space is divided accordingly into atom-centered spheres, defining the atomic regions, and an interstitial region in between. The partial solutions of the different regions are matched with value and derivative at the interface between atomic and interstitial regions.

The projector augmented-wave method is an extension of augmented wave methods and the pseudopotential approach, which combines their traditions into a unified electronic structure method.

After describing the underlying ideas of the various approaches, let us briefly review the history of augmented wave methods and the pseudopotential approach. We do not discuss the atomic-orbital based methods, because our focus is the PAW method and its ancestors.

6 Augmented wave methods

The augmented wave methods have been introduced in 1937 by Slater [59]. His method was called augmented plane-wave (APW) method. Later Korringa [60], Kohn and Rostoker [61] modified the idea, which lead to the so-called KKR method. The basic idea behind the augmented wave methods has been to consider the electronic structure as a scattered-electron problem: Consider an electron beam, represented by a plane wave, traveling through a solid. It undergoes multiple scattering at the atoms. If, for some energy, the outgoing scattered waves interfere destructively, so that the electrons can not escape, a bound state has been determined. This approach can be translated into a basis-set method with energy- and potential-dependent basis functions. In order to make the scattered wave problem tractable, a model potential had to be chosen: The so-called muffin-tin potential approximates the true potential by a potential that is spherically symmetric in the atomic regions and constant in between.

Augmented-wave methods reached adulthood in the 1970s: O. K. Andersen [62] showed that the energy dependent basis set of Slater's APW method can be mapped onto one with energy independent basis functions by linearizing the partial waves for the atomic regions with respect to their energy. In the original APW approach, one had to determine the zeros of the determinant of an energy dependent matrix, a nearly intractable numerical problem for complex systems. With the new energy independent basis functions, however, the problem is reduced to the much simpler generalized eigenvalue problem, which can be solved using efficient numerical techniques. Furthermore, the introduction of well defined basis sets paved the way for full-potential calculations [63]. In that case, the muffin-tin approximation is used solely to define the basis set $|\chi_i\rangle$, while the matrix elements $\langle\chi_i|H|\chi_j\rangle$ of the Hamiltonian are evaluated with the full potential.

In the augmented wave methods one constructs the basis set for the atomic region by solving the radial Schrödinger equation for the spherically averaged effective potential

$$\left[\frac{-\hbar^2}{2m_e}\vec{\nabla}^2 + v_{eff}(\vec{r}) - \epsilon\right]\phi_{\ell,m}(\epsilon,\vec{r}) = 0$$

as function of the energy. Note that a partial wave $\phi_{\ell,m}(\epsilon,\vec{r})$ is an angular-momentum eigenstate and can be expressed as a product of a radial function and a spherical harmonic. The energy-dependent partial wave is expanded in a Taylor expansion about some reference energy $\epsilon_{\nu,\ell}$

$$\phi_{\ell,m}(\epsilon,\vec{r}) = \phi_{\nu,\ell,m}(\vec{r}) + (\epsilon - \epsilon_{\nu,\ell})\dot{\phi}_{\nu,\ell,m}(\vec{r}) + O((\epsilon - \epsilon_{\nu,\ell})^2) ,$$

where $\phi_{\nu,\ell,m}(\vec{r}) = \phi_{\ell,m}(\epsilon_{\nu,\ell},\vec{r})$. The energy derivative of the partial wave $\dot{\phi}_\nu(\vec{r}) = \left.\frac{\partial\phi(\epsilon,\vec{r})}{\partial\epsilon}\right|_{\epsilon_{\nu,\ell}}$ is obtained from the energy derivative of the Schrödinger equation

$$\left[\frac{-\hbar^2}{2m_e}\vec{\nabla}^2 + v_{eff}(\vec{r}) - \epsilon_{\nu,\ell}\right]\dot{\phi}_{\nu,\ell,m}(\vec{r}) = \phi_{\nu,\ell,m}(\vec{r}) .$$

Next, one starts from a regular basis set, such as plane waves, Gaussians or Hankel functions. These basis functions are called envelope functions $|\tilde{\chi}_i\rangle$. Within the atomic region they are replaced by the partial waves and their energy derivatives, such that the resulting wavefunction $\chi_i(\vec{r})$ is continuous and differentiable. The augmented envelope function has the form

$$\chi_i(\vec{r}) = \tilde{\chi}_i(\vec{r}) - \sum_R \theta_R(\vec{r})\tilde{\chi}_i(\vec{r}) + \sum_{R,\ell,m} \theta_R(\vec{r})\left[\phi_{\nu,R,\ell,m}(\vec{r})a_{R,\ell,m,i} + \dot{\phi}_{\nu,R,\ell,m}(\vec{r})b_{R,\ell,m,i}\right] . \quad (42)$$

$\theta_R(\vec{r})$ is a step function that is unity within the augmentation sphere centered at $\vec{R}_R$ and zero elsewhere. The augmentation sphere is atom-centered and has a radius about equal to the covalent radius. This radius is called the muffin-tin radius, if the spheres of neighboring atoms touch. These basis functions describe only the valence states; the core states are localized within the augmentation sphere and are obtained directly by a radial integration of the Schrödinger equation within the augmentation sphere.

The coefficients $a_{R,\ell,m,i}$ and $b_{R,\ell,m,i}$ are obtained for each $|\tilde{\chi}_i\rangle$ as follows: The envelope function is decomposed around each atomic site into spherical harmonics multiplied by radial functions

$$\tilde{\chi}_i(\vec{r}) = \sum_{\ell,m} u_{R,\ell,m,i}(|\vec{r} - \vec{R}_R|)Y_{\ell,m}(\vec{r} - \vec{R}_R) . \quad (43)$$

Analytical expansions for plane waves, Hankel functions or Gaussians exist. The radial parts of the partial waves $\phi_{\nu,R,\ell,m}$ and $\dot{\phi}_{\nu,R,\ell,m}$ are matched with value and derivative to $u_{R,\ell,m,i}(|\vec{r}|)$, which yields the expansion coefficients $a_{R,\ell,m,i}$ and $b_{R,\ell,m,i}$.

If the envelope functions are plane waves, the resulting method is called the linear augmented plane-wave (LAPW) method. If the envelope functions are Hankel functions, the method is called linear muffin-tin orbital (LMTO) method.

A good review of the LAPW method [62] has been given by Singh [64]. Let us now briefly mention the major developments of the LAPW method: Soler [65] introduced the idea of additive

augmentation: While augmented plane waves are discontinuous at the surface of the augmentation sphere if the expansion in spherical harmonics in Eq. 42 is truncated, Soler replaced the second term in Eq. 42 by an expansion of the plane wave with the same angular momentum truncation as in the third term. This dramatically improved the convergence of the angular momentum expansion. Singh [66] introduced so-called local orbitals, which are non-zero only within a muffin-tin sphere, where they are superpositions of ϕ and $\dot{\phi}$ functions from different expansion energies. Local orbitals substantially increase the energy transferability. Sjöstedt [67] relaxed the condition that the basis functions are differentiable at the sphere radius. In addition she introduced local orbitals, which are confined inside the sphere, and that also have a kink at the sphere boundary. Due to the large energy cost of kinks, they will cancel, once the total energy is minimized. The increased variational degree of freedom in the basis leads to a dramatically improved plane-wave convergence [68].

The second variant of the linear methods is the LMTO method [62]. A good introduction into the LMTO method is the book by Skriver [69]. The LMTO method uses Hankel functions as envelope functions. The atomic spheres approximation (ASA) provides a particularly simple and efficient approach to the electronic structure of very large systems. In the ASA the augmentation spheres are blown up so that the sum of their volumes is equal to the total volume. Then, the first two terms in Eq. 42 are ignored. The main deficiency of the LMTO-ASA method is the limitation to structures that can be converted into a closed packed arrangement of atomic and empty spheres. Furthermore, energy differences due to structural distortions are often qualitatively incorrect. Full potential versions of the LMTO method, that avoid these deficiencies of the ASA have been developed. The construction of tight binding orbitals as superposition of muffin-tin orbitals [70] showed the underlying principles of the empirical tight-binding method and prepared the ground for electronic structure methods that scale linearly instead of with the third power of the number of atoms. The third generation LMTO [71] allows to construct true minimal basis sets, which require only one orbital per electron pair for insulators. In addition, they can be made arbitrarily accurate in the valence band region, so that a matrix diagonalization becomes unnecessary. The first steps towards a full-potential implementation, that promises good accuracy, while maintaining the simplicity of the LMTO-ASA method, are currently under way. Through the minimal basis-set construction the LMTO method offers unrivaled tools for the analysis of the electronic structure and has been extensively used in hybrid methods combining density functional theory with model Hamiltonians for materials with strong electron correlations [72].

7   Pseudopotentials

Pseudopotentials have been introduced to (1) avoid describing the core electrons explicitly and (2) to avoid the rapid oscillations of the wavefunction near the nucleus, which normally require either complicated or large basis sets.

The pseudopotential approach can be traced back to 1940 when C. Herring invented the orthogonalized plane-wave method [73]. Later, Phillips [74] and Antončík [75] replaced the or-

thogonality condition by an effective potential, which mimics the Pauli repulsion by the core electrons and thus compensates the electrostatic attraction by the nucleus. In practice, the potential was modified, for example, by cutting off the singular potential of the nucleus at a certain value. This was done with a few parameters that have been adjusted to reproduce the measured electronic band structure of the corresponding solid.

Hamann, Schlüter and Chiang [76] showed in 1979 how pseudopotentials can be constructed in such a way that their scattering properties are identical to that of an atom to first order in energy. These first-principles pseudopotentials relieved the calculations from the restrictions of empirical parameters. Highly accurate calculations have become possible especially for semiconductors and simple metals. An alternative approach towards first-principles pseudopotentials by Zunger and Cohen [77] even preceded the one mentioned above.

The idea behind the pseudopotential construction

In order to construct a first-principles pseudopotential, one starts out with an all-electron density-functional calculation for a spherical atom. Such calculations can be performed efficiently on radial grids. They yield the atomic potential and wavefunctions $\phi_{\ell,m}(\vec{r})$. Due to the spherical symmetry, the radial parts of the wavefunctions for different magnetic quantum numbers m are identical.

For the valence wavefunctions one constructs pseudo wavefunctions $|\tilde{\phi}_{\ell,m}\rangle$. There are numerous ways [78–81] to construct those pseudo wavefunctions: Pseudo wave functions are identical to the true wave functions outside the augmentation region, which is called core region in the context of the pseudopotential approach. Inside the augmentation region the pseudo wavefunction should be nodeless and have the same norm as the true wavefunctions, that is $\langle\tilde{\phi}_{\ell,m}|\tilde{\phi}_{\ell,m}\rangle = \langle\phi_{\ell,m}|\phi_{\ell,m}\rangle$ (compare Figure 6).

From the pseudo wavefunction, a potential $u_\ell(\vec{r})$ can be reconstructed by inverting the respective Schrödinger equation, i.e.

$$\left[-\frac{\hbar^2}{2m_e}\vec{\nabla}^2 + u_\ell(\vec{r}) - \epsilon_{\ell,m} \right] \tilde{\phi}_{\ell,m}(\vec{r}) = 0 \;\Rightarrow\; u_\ell(\vec{r}) = \epsilon + \frac{1}{\tilde{\phi}_{\ell,m}(\vec{r})} \cdot \frac{\hbar^2}{2m_e}\vec{\nabla}^2\tilde{\phi}_{\ell,m}(\vec{r}) \;.$$

This potential $u_\ell(\vec{r})$ (compare Figure 6), which is also spherically symmetric, differs from one main angular momentum ℓ to the other. Note, that this *inversion of the Schrödinger equation* works only if the wave functions are nodeless.

Next we define an effective pseudo Hamiltonian

$$\hat{\tilde{H}}_\ell = -\frac{\hbar^2}{2m_e}\vec{\nabla}^2 + v_\ell^{ps}(\vec{r}) + \int d^3r' \frac{e^2\left(\tilde{n}(\vec{r'}) + \tilde{Z}(\vec{r'})\right)}{4\pi\epsilon_0|\vec{r} - \vec{r'}|} + \mu_{xc}([n(\vec{r})],\vec{r}) \;, \tag{44}$$

where $\mu_{xc}(\vec{r}) = \delta E_{xc}[n]/\delta n(\vec{r})$ is the functional derivative of the exchange and correlation energy with respect to the electron density. Then, we determine the pseudopotentials v_ℓ^{ps} such that the pseudo Hamiltonian produces the pseudo wavefunctions, that is

$$v_\ell^{ps}(\vec{r}) = u_\ell(\vec{r}) - \int d^3r' \frac{e^2\left(\tilde{n}(\vec{r'}) + \tilde{Z}(\vec{r'})\right)}{4\pi\epsilon_0|\vec{r} - \vec{r'}|} - \mu_{xc}([\tilde{n}(\vec{r})],\vec{r}) \;. \tag{45}$$
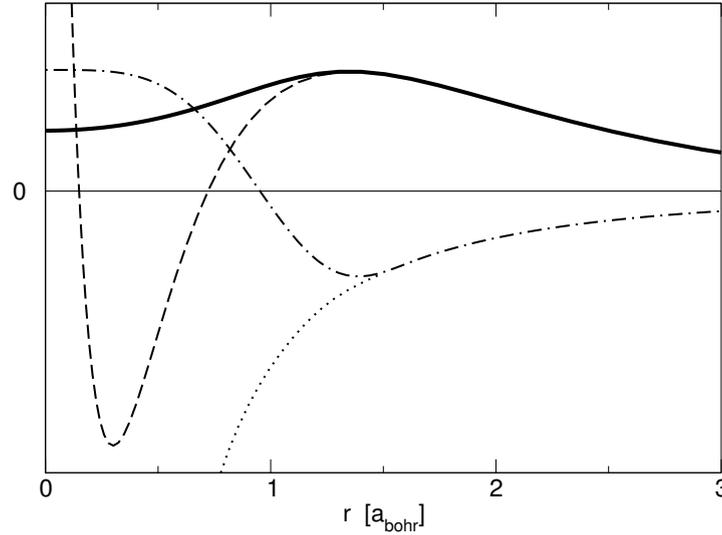
Fig. 6: *Illustration of the pseudopotential concept for the example of the 3s wavefunction of Si. The solid line shows the radial part of the pseudo wavefunction $\tilde{\phi}_{\ell,m}$. The dashed line corresponds to the all-electron wavefunction $\phi_{\ell,m}$, which exhibits strong oscillations at small radii. The angular momentum dependent pseudopotential u_ℓ (dash-dotted line) deviates from the all-electron potential v_{eff} (dotted line) inside the augmentation region. The data are generated by the fhi98PP code [82].*

This process is called "unscreening".

$\tilde{Z}(\vec{r})$ mimics the charge density of the nucleus and the core electrons. It is usually an atom-centered, spherical Gaussian that is normalized to the charge of nucleus and core of that atom. In the pseudopotential approach, $\tilde{Z}_R(\vec{r})$ does not change with the potential. The pseudo density $\tilde{n}(\vec{r}) = \sum_n f_n \tilde{\psi}_n^*(\vec{r}) \tilde{\psi}_n(\vec{r})$ is constructed from the pseudo wavefunctions.

In this way, we obtain a different potential for each angular momentum channel. In order to apply these potentials to a given wavefunction, the wavefunction must first be decomposed into angular momenta. Then each component is applied to the pseudopotential v_ℓ^{ps} for the corresponding angular momentum.

The pseudopotential defined in this way can be expressed in a semi-local form

$$v^{ps}(\vec{r}, \vec{r}') = \bar{v}(\vec{r})\delta(\vec{r} - \vec{r}') + \sum_{\ell,m}\left[Y_{\ell,m}(\vec{r})\,[v_\ell^{ps}(\vec{r}) - \bar{v}(\vec{r})]\,\frac{\delta(|\vec{r}| - |\vec{r}'|)}{|\vec{r}|^2}Y_{\ell,m}^*(\vec{r}') \right] \ . \quad (46)$$

The local potential $\bar{v}(\vec{r})$ only acts on those angular momentum components that are not already considered explicitly in the non-local, angular-momentum dependend pseudopotentials v_ℓ^{ps}. Typically it is chosen to cancel the most expensive nonlocal terms, the one corresponding to the highest physically relevant angular momentum.

The pseudopotential $v^{ps}(\vec{r}, \vec{r'})$ is non-local as its depends on two position arguments, $\vec{r}$ and $\vec{r'}$. The expectation values are evaluated as a double integral

$$\langle \tilde{\psi}|\hat{v}_{ps}|\tilde{\psi}\rangle = \int d^3r \int d^3r' \ \tilde{\psi}^*(\vec{r})v^{ps}(\vec{r}, \vec{r}')\tilde{\psi}(\vec{r}') \quad (47)$$

The semi-local form of the pseudopotential given in Eq. 46 is computationally expensive. Therefore, in practice one uses a separable form of the pseudopotential [83–85]

$$\hat{v}^{ps} \approx \sum_{i,j} \hat{v}^{ps}|\tilde{\phi}_i\rangle \left[\langle \tilde{\phi}_j|\hat{v}^{ps}|\tilde{\phi}_i\rangle \right]_{i,j}^{-1} \langle \tilde{\phi}_j|\hat{v}^{ps} . \tag{48}$$

Thus, the projection onto spherical harmonics used in the semi-local form of Eq. 46 is replaced by a projection onto angular momentum dependent functions $\hat{v}^{ps}|\tilde{\phi}_i\rangle$.

The indices i and j are composite indices containing the atomic-site index R, the angular momentum quantum numbers ℓ, m and an additional index α. The index α distinguishes partial waves with otherwise identical indices R, ℓ, m when more than one partial wave per site and angular momentum is allowed. The partial waves may be constructed as eigenstates of the pseudopotential $\hat{v}^{ps}_\ell$ for a set of energies.

One can show that the identity of Eq. 48 holds by applying a wavefunction $|\tilde{\psi}\rangle = \sum_i |\tilde{\phi}_i\rangle c_i$ to both sides. If the set of pseudo partial waves $|\tilde{\phi}_i\rangle$ in Eq. 48 is complete, the identity is exact. The advantage of the separable form is that $\langle \tilde{\phi}|\hat{v}^{ps}$ is treated as one function, so that expectation values are reduced to combinations of simple scalar products $\langle \tilde{\phi}_i|\hat{v}^{ps}|\tilde{\psi}\rangle$.

The total energy of the pseudopotential method can be written in the form

$$E = \sum_n f_n \langle \tilde{\psi}_n| \frac{\hat{\vec{p}}^2}{2m_e}|\tilde{\psi}_n\rangle + E_{self} + \sum_n f_n \langle \tilde{\psi}_n|\hat{v}_{ps}|\tilde{\psi}_n\rangle$$

$$+ \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 \left(\tilde{n}(\vec{r}) + \tilde{Z}(\vec{r}) \right)\left(\tilde{n}(\vec{r}') + \tilde{Z}(\vec{r}') \right)}{4\pi\epsilon_0|\vec{r} - \vec{r}'|} + E_{xc}[\tilde{n}(\vec{r})] . \tag{49}$$

The constant E_{self} is adjusted such that the total energy of the atom is the same for an all-electron calculation and the pseudopotential calculation.

For the atom, from which it has been constructed, this construction guarantees that the pseudopotential method produces the correct one-particle energies for the valence states and that the wave functions have the desired shape.

While pseudopotentials have proven to be accurate for a large variety of systems, there is no strict guarantee that they produce the same results as an all-electron calculation, if they are used in a molecule or solid. The error sources can be divided into two classes:

- Energy transferability problems: Even for the potential of the reference atom, the scattering properties are accurate only in given energy window.

- Charge transferability problems: In a molecule or crystal, the potential differs from that of the isolated atom. The pseudopotential, however, is strictly valid only for the isolated atom.

The plane-wave basis set for the pseudo wavefunctions is defined by the shortest wave length $\lambda = 2\pi/|\vec{G}|$, where $\vec{G}$ is the wave vector, via the so-called plane-wave cutoff $E_{PW} = \frac{\hbar^2 G_{max}^2}{2m_e}$ with $G_{max} = \max\{|\vec{G}|\}$. It is often specified in Rydberg (1 Ry=$\frac{1}{2}$ H≈13.6 eV). The plane-wave cutoff is the highest kinetic energy of all basis functions. The basis-set convergence can systematically be controlled by increasing the plane-wave cutoff.

The charge transferability is substantially improved by including a nonlinear core correction [86] into the exchange-correlation term of Eq. 49. Hamann [87] showed how to construct pseudopotentials also from unbound wavefunctions. Vanderbilt [88, 89] generalized the pseudopotential method to non-normconserving pseudopotentials, so-called ultra-soft pseudopotentials, which dramatically improve the basis-set convergence. The formulation of ultra-soft pseudopotentials has already many similarities with the projector augmented-wave method. Truncated separable pseudopotentials suffer sometimes from so-called ghost states. These are unphysical core-like states, which render the pseudopotential useless. These problems have been discussed by Gonze [90]. Quantities such as hyperfine parameters that depend on the full wavefunctions near the nucleus, can be extracted approximately [91]. Good reviews of pseudopotential methodology have been written by Payne *et al.* [92] and Singh [64].

In 1985 R. Car and M. Parrinello published the ab-initio molecular dynamics method [93]. Simulations of the atomic motion have become possible on the basis of state-of-the-art electronic structure methods. Besides making dynamical phenomena and finite temperature effects accessible to electronic structure calculations, the ab-initio molecular dynamics method also introduced a radically new way of thinking into electronic structure methods. Diagonalization of a Hamilton matrix has been replaced by classical equations of motion for the wavefunction coefficients. If one applies friction, the system is quenched to the ground state. Without friction truly dynamical simulations of the atomic structure are performed. By using thermostats [94–97], simulations at constant temperature can be performed. The Car-Parrinello method treats electronic wavefunctions and atomic positions on an equal footing.

8 Projector augmented-wave method

The Car-Parrinello method had been implemented first for the pseudopotential approach. There seemed to be insurmountable barriers against combining the new technique with augmented wave methods. The main problem was related to the potential dependent basis set used in augmented wave methods: the Car-Parrinello method requires a well defined and unique total energy functional of atomic positions and basis set coefficients. Furthermore the analytic evaluation of the first partial derivatives of the total energy with respect to wave functions, $\frac{\partial E}{\partial \langle \psi_n |} = \hat{H} | \psi_n \rangle f_n$, and atomic positions, the forces $\vec{F}_j = -\vec{\nabla}_j E$, must be possible. Therefore, it was one of the main goals of the PAW method to introduce energy and potential independent basis sets, which were as accurate as the previously used augmented basis sets. Other requirements have been: (1) The method should at least match the efficiency of the pseudopotential approach for Car-Parrinello simulations. (2) It should become an exact theory when converged and (3) its convergence should be easily controlled. We believe that these criteria have been met, which explains why the use of the PAW method has become widespread today.

Transformation theory

At the root of the PAW method lies a transformation that maps the true wavefunctions with their complete nodal structure onto auxiliary wavefunctions that are numerically convenient. We aim for smooth auxiliary wavefunctions, which have a rapidly convergent plane-wave expansion. With such a transformation we can expand the auxiliary wave functions into a convenient basis set such as plane waves, and evaluate all physical properties after reconstructing the related physical (true) wavefunctions.

Let us denote the physical one-particle wavefunctions as $|\psi_n\rangle$ and the auxiliary wavefunctions as $|\tilde{\psi}_n\rangle$. Note that the tilde refers to the representation of smooth auxiliary wavefunctions and n is the label for a one-particle state and contains a band index, a k-point and a spin index. The transformation from the auxiliary to the physical wave functions is denoted by $\hat{\mathcal{T}}$, i.e.

$$|\psi_n\rangle = \hat{\mathcal{T}}|\tilde{\psi}_n\rangle \ . \tag{50}$$

Now we express the constrained density functional F of Eq. 38 in terms of our auxiliary wavefunctions

$$F\left[\{\hat{\mathcal{T}}|\tilde{\psi}_n\rangle\}, \{\Lambda_{m,n}\}\right] = E\left[\{\hat{\mathcal{T}}|\tilde{\psi}_n\rangle\}\right] - \sum_{n,m}\left[\langle\tilde{\psi}_n|\hat{\mathcal{T}}^\dagger\hat{\mathcal{T}}|\tilde{\psi}_m\rangle - \delta_{n,m}\right]\Lambda_{m,n} \ . \tag{51}$$

The variational principle with respect to the auxiliary wavefunctions yields

$$\hat{\mathcal{T}}^\dagger\hat{H}\hat{\mathcal{T}}|\tilde{\psi}_n\rangle = \hat{\mathcal{T}}^\dagger\hat{\mathcal{T}}|\tilde{\psi}_n\rangle\epsilon_n \ . \tag{52}$$

Again, we obtain a Schrödinger-like equation (see derivation of Eq. 41), but now the Hamilton operator has a different form, $\hat{\tilde{H}} = \hat{\mathcal{T}}^\dagger\hat{H}\hat{\mathcal{T}}$, an overlap operator $\hat{\tilde{O}} = \hat{\mathcal{T}}^\dagger\hat{\mathcal{T}}$ occurs, and the resulting auxiliary wavefunctions are smooth.

When we evaluate physical quantities, we need to evaluate expectation values of an operator $\hat{A}$, which can be expressed in terms of either the true or the auxiliary wavefunctions, i.e.

$$\langle\hat{A}\rangle \ = \ \sum_n f_n\langle\psi_n|\hat{A}|\psi_n\rangle = \sum_n f_n\langle\tilde{\psi}_n|\hat{\mathcal{T}}^\dagger\hat{A}\hat{\mathcal{T}}|\tilde{\psi}_n\rangle \ . \tag{53}$$

In the representation of auxiliary wavefunctions we need to use transformed operators $\hat{\tilde{A}} = \hat{\mathcal{T}}^\dagger\hat{A}\hat{\mathcal{T}}$. As it is, this equation only holds for the valence electrons. The core electrons are treated differently, as will be shown below.

The transformation takes us conceptionally from the world of pseudopotentials to that of augmented wave methods, which deal with the full wavefunctions. We will see that our auxiliary wavefunctions, which are simply the plane-wave parts of the full wavefunctions, translate into the wavefunctions of the pseudopotential approach. In the PAW method the auxiliary wavefunctions are used to construct the true wavefunctions and the total energy functional is evaluated from the latter. Thus it provides the missing link between augmented wave methods and the pseudopotential method, which can be derived as a well-defined approximation of the PAW method.

In the original paper [58], the auxiliary wavefunctions were termed pseudo wavefunctions and the true wavefunctions were termed all-electron wavefunctions, in order to make the connection more evident. We avoid this notation here, because it resulted in confusion in cases where the correspondence is not clear-cut.

Transformation operator

So far we have described how we can determine the auxiliary wave functions of the ground state and how to obtain physical information from them. What is missing is a definition of the transformation operator $\hat{\mathcal{T}}$.

The operator $\hat{\mathcal{T}}$ has to modify the smooth auxiliary wave function in each atomic region, so that the resulting wavefunction has the correct nodal structure. Therefore, it makes sense to write the transformation as identity plus a sum of atomic contributions $\hat{\mathcal{S}}_R$

$$\hat{\mathcal{T}} = \hat{1} + \sum_R \hat{\mathcal{S}}_R. \tag{54}$$

For every atom, $\hat{\mathcal{S}}_R$ adds the difference between the true and the auxiliary wavefunction.

The local terms $\hat{\mathcal{S}}_R$ are defined in terms of solutions $|\phi_i\rangle$ of the Schrödinger equation for the isolated atoms. This set of partial waves $|\phi_i\rangle$ will serve as a basis set so that, near the nucleus, all relevant valence wavefunctions can be expressed as superposition of the partial waves with yet unknown coefficients as

$$\psi(\vec{r}) = \sum_{i \in R} \phi_i(\vec{r}) c_i \quad \text{for} \quad |\vec{r} - \vec{R}_R| < r_{c,R} . \tag{55}$$

With $i \in R$ we indicate those partial waves that belong to site R.

Since the core wavefunctions do not spread out into the neighboring atoms, we will treat them differently. Currently we use the frozen-core approximation, which imports the density and the energy of the core electrons from the corresponding isolated atoms. The transformation $\hat{\mathcal{T}}$ shall produce only wavefunctions orthogonal to the core electrons, while the core electrons are treated separately. Therefore, the set of atomic partial waves $|\phi_i\rangle$ includes only valence states that are orthogonal to the core wavefunctions of the atom.

For each of the partial waves we choose an auxiliary partial wave $|\tilde{\phi}_i\rangle$. The identity

$$\begin{aligned}
|\phi_i\rangle &= (\hat{1} + \hat{\mathcal{S}}_R)|\tilde{\phi}_i\rangle \quad \text{for} \quad i \in R \\
\hat{\mathcal{S}}_R|\tilde{\phi}_i\rangle &= |\phi_i\rangle - |\tilde{\phi}_i\rangle
\end{aligned} \tag{56}$$

defines the local contribution $\hat{\mathcal{S}}_R$ to the transformation operator. Since $\hat{1} + \hat{\mathcal{S}}_R$ should change the wavefunction only locally, we require that the partial waves $|\phi_i\rangle$ and their auxiliary counterparts $|\tilde{\phi}_i\rangle$ are pairwise identical beyond a certain radius $r_{c,R}$.

$$\phi_i(\vec{r}) = \tilde{\phi}_i(\vec{r}) \quad \text{for} \quad i \in R \quad \text{and} \quad |\vec{r} - \vec{R}_R| > r_{c,R} \tag{57}$$

Note that the partial waves are not necessarily bound states and are therefore not normalizable unless we truncate them beyond a certain radius $r_{c,R}$. The PAW method is formulated such that

the final results do not depend on the location where the partial waves are truncated, as long as this is not done too close to the nucleus and identical for auxiliary and all-electron partial waves.

In order to be able to apply the transformation operator to an arbitrary auxiliary wavefunction, we need to be able to expand the auxiliary wavefunction locally into the auxiliary partial waves

$$\tilde{\psi}(\vec{r}) = \sum_{i \in R} \tilde{\phi}_i(\vec{r}) c_i = \sum_{i \in R} \tilde{\phi}_i(\vec{r}) \langle \tilde{p}_i | \tilde{\psi} \rangle \quad \text{for} \quad |\vec{r} - \vec{R}_R| < r_{c,R} , \tag{58}$$

which defines the projector functions $|\tilde{p}_i\rangle$. The projector functions probe the local character of the auxiliary wave function in the atomic region, and examples are shown in Figure 7. From Eq. 58 we can derive $\sum_{i \in R} |\tilde{\phi}_i\rangle \langle \tilde{p}_i| = 1$, which is valid within $r_{c,R}$. It can be shown by insertion, that the identity Eq. 58 holds for any auxiliary wavefunction $|\tilde{\psi}\rangle$ that can be expanded locally into auxiliary partial waves $|\tilde{\phi}_i\rangle$, if

$$\langle \tilde{p}_i | \tilde{\phi}_j \rangle = \delta_{i,j} \quad \text{for} \quad i, j \in R . \tag{59}$$

Note that neither the projector functions nor the partial waves need to be mutually orthogonal. The projector functions are fully determined with the above conditions and a closure relation that is related to the unscreening of the pseudopotentials (see Eq. 90 in [58]).
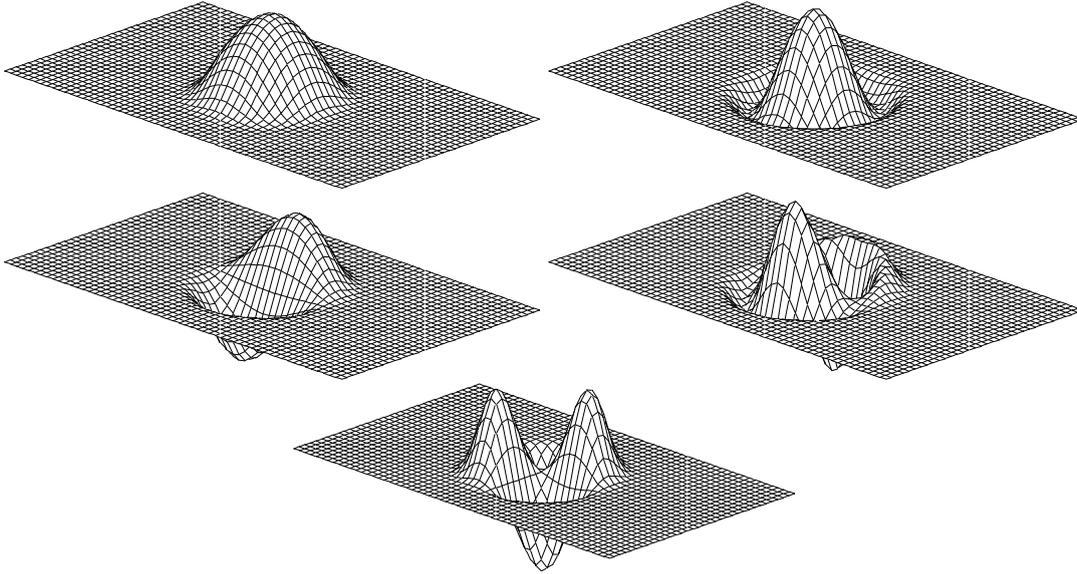

Fig. 7: *Projector functions of the chlorine atom. Top: two s-type projector functions, middle: p-type, bottom: d-type.*

By combining Eq. 56 and Eq. 58, we can apply $\hat{\mathcal{S}}_R$ to any auxiliary wavefunction.

$$\hat{\mathcal{S}}_R |\tilde{\psi}\rangle = \sum_{i \in R} \hat{\mathcal{S}}_R |\tilde{\phi}_i\rangle \langle \tilde{p}_i | \tilde{\psi} \rangle = \sum_{i \in R} \Big(|\phi_i\rangle - |\tilde{\phi}_i\rangle \Big) \langle \tilde{p}_i | \tilde{\psi} \rangle . \tag{60}$$

Hence, the transformation operator is

$$\hat{\mathcal{T}} = \hat{1} + \sum_i \Big(|\phi_i\rangle - |\tilde{\phi}_i\rangle \Big) \langle \tilde{p}_i| , \tag{61}$$

where the sum runs over all partial waves of all atoms. The true wave function can be expressed as

$$|\psi\rangle = |\tilde{\psi}\rangle + \sum_i \left(|\phi_i\rangle - |\tilde{\phi}_i\rangle \right) \langle \tilde{p}_i|\tilde{\psi}\rangle = |\tilde{\psi}\rangle + \sum_R \left(|\psi_R^1\rangle - |\tilde{\psi}_R^1\rangle \right) \tag{62}$$

with

$$|\psi_R^1\rangle = \sum_{i \in R} |\phi_i\rangle \langle \tilde{p}_i|\tilde{\psi}\rangle \tag{63}$$

$$|\tilde{\psi}_R^1\rangle = \sum_{i \in R} |\tilde{\phi}_i\rangle \langle \tilde{p}_i|\tilde{\psi}\rangle . \tag{64}$$

In Fig. 8 the decomposition of Eq. 62 is shown for the example of the bonding p-σ state of the Cl_2 molecule.
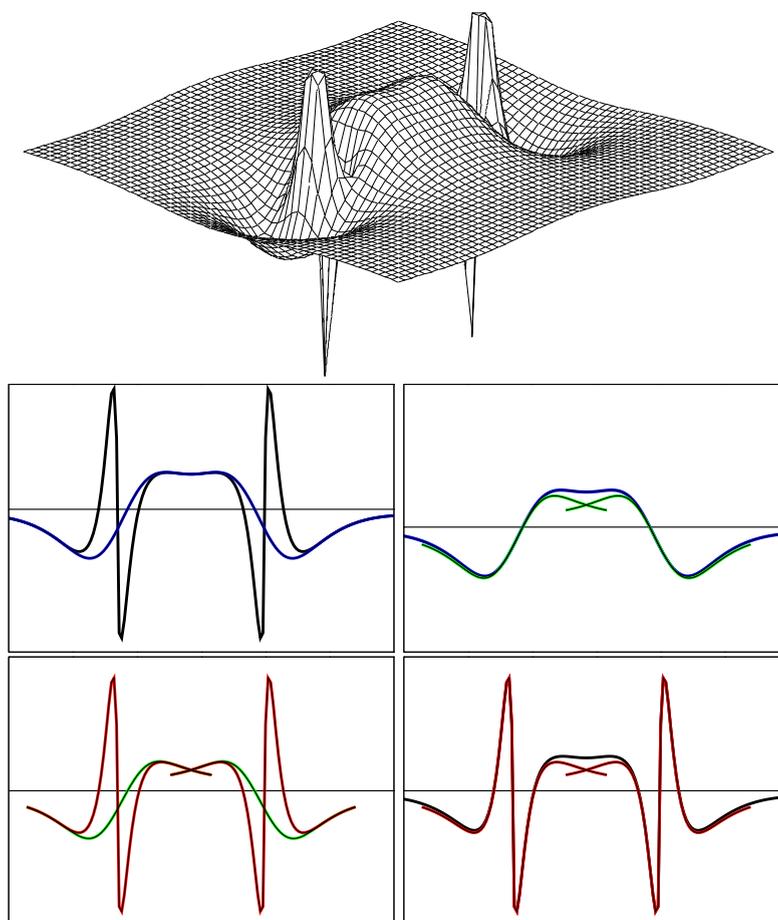

Fig. 8: *Bonding p-σ orbital of the Cl_2 molecule and its decomposition into auxiliary wavefunction and the two one-center expansions. Top-left: True and auxiliary wave function; top-right: auxiliary wavefunction and its partial wave expansion; bottom-left: the two partial wave expansions; bottom-right: true wavefunction and its partial wave expansion.*

To understand the expression Eq. 62 for the true wave function, let us concentrate on different regions in space. (1) Far from the atoms, the partial waves are, according to Eq. 57, pairwise

identical so that the auxiliary wavefunction is identical to the true wavefunction, that is $\psi(\vec{r}) = \tilde{\psi}(\vec{r})$. (2) Close to an atom R, however, the auxiliary wavefunction is, according to Eq. 58, identical to its one-center expansion, that is $\tilde{\psi}(\vec{r}) = \tilde{\psi}_R^1(\vec{r})$. Hence the true wavefunction $\psi(\vec{r})$ is identical to $\psi_R^1(\vec{r})$, which is built up from partial waves that contain the proper nodal structure. In practice, the partial wave expansions are truncated. Therefore, the identity of Eq. 58 does not hold strictly. As a result, the plane waves also contribute to the true wavefunction inside the atomic region. This has the advantage that the missing terms in a truncated partial wave expansion are partly accounted for by plane waves. This explains the rapid convergence of the partial wave expansions. This idea is related to the additive augmentation of the LAPW method of Soler [65].

Frequently, the question comes up, whether the transformation Eq. 61 of the auxiliary wavefunctions indeed provides the true wavefunction. The transformation should be considered merely as a change of representation analogous to a coordinate transform. If the total energy functional is transformed consistently, its minimum will yield auxiliary wavefunctions that produce the correct wave functions $|\psi\rangle$.

Expectation values

Expectation values can be obtained either from the reconstructed true wavefunctions or directly from the auxiliary wave functions

$$
\begin{aligned}
\langle \hat{A} \rangle &= \sum_n f_n \langle \psi_n | \hat{A} | \psi_n \rangle + \sum_{n=1}^{N_c} \langle \phi_n^c | \hat{A} | \phi_n^c \rangle \\
&= \sum_n f_n \langle \tilde{\psi}_n | \hat{\mathcal{T}}^\dagger \hat{A} \hat{\mathcal{T}} | \tilde{\psi}_n \rangle + \sum_{n=1}^{N_c} \langle \phi_n^c | \hat{A} | \phi_n^c \rangle ,
\end{aligned} \tag{65}
$$

where f_n are the occupations of the valence states and N_c is the number of core states. The first sum runs over the valence states, and second over the core states $|\phi_n^c\rangle$.

Now we can decompose the matrix element for a wavefunction ψ into its individual contributions according to Eq. 62.

$$
\begin{aligned}
\langle \psi | \hat{A} | \psi \rangle &= \langle \tilde{\psi} + \sum_R (\psi_R^1 - \tilde{\psi}_R^1) | \hat{A} | \tilde{\psi} + \sum_{R'} (\psi_{R'}^1 - \tilde{\psi}_{R'}^1) \rangle \\
&= \underbrace{\langle \tilde{\psi} | \hat{A} | \tilde{\psi} \rangle + \sum_R \left(\langle \psi_R^1 | \hat{A} | \psi_R^1 \rangle - \langle \tilde{\psi}_R^1 | \hat{A} | \tilde{\psi}_R^1 \rangle \right)}_{\text{part 1}} \\
&+ \underbrace{\sum_R \left(\langle \psi_R^1 - \tilde{\psi}_R^1 | \hat{A} | \tilde{\psi} - \tilde{\psi}_R^1 \rangle + \langle \tilde{\psi} - \tilde{\psi}_R^1 | \hat{A} | \psi_R^1 - \tilde{\psi}_R^1 \rangle \right)}_{\text{part 2}} \\
&+ \underbrace{\sum_{R \neq R'} \langle \psi_R^1 - \tilde{\psi}_R^1 | \hat{A} | \psi_{R'}^1 - \tilde{\psi}_{R'}^1 \rangle}_{\text{part 3}}
\end{aligned} \tag{66}
$$

Only the first part of Eq. 66 is evaluated explicitly, while the second and third parts of Eq. 66 are neglected, because they vanish for sufficiently local operators as long as the partial wave expansion is converged: The function $\psi_R^1 - \tilde{\psi}_R^1$ vanishes per construction beyond its augmentation region, because the partial waves are pairwise identical beyond that region. The function $\tilde{\psi} - \tilde{\psi}_R^1$ vanishes inside its augmentation region if the partial wave expansion is sufficiently converged. In no region of space are both functions $\psi_R^1 - \tilde{\psi}_R^1$ and $\tilde{\psi} - \tilde{\psi}_R^1$ simultaneously nonzero. Similarly the functions $\psi_R^1 - \tilde{\psi}_R^1$ from different sites are never non-zero in the same region in space. Hence, the second and third parts of Eq. 66 vanish for operators such as the kinetic energy $\frac{-\hbar^2}{2m_e}\vec{\nabla}^2$ and the real space projection operator $|r\rangle\langle r|$, which produces the electron density. For truly nonlocal operators the parts 2 and 3 of Eq. 66 would have to be considered explicitly. The expression, Eq. 65, for the expectation value can therefore be written with the help of Eq. 66 as

$$
\begin{aligned}
\langle\hat{A}\rangle &= \sum_n f_n\left(\langle\tilde{\psi}_n|\hat{A}|\tilde{\psi}_n\rangle + \langle\psi_n^1|\hat{A}|\psi_n^1\rangle - \langle\tilde{\psi}_n^1|\hat{A}|\tilde{\psi}_n^1\rangle\right) + \sum_{n=1}^{N_c}\langle\phi_n^c|\hat{A}|\phi_n^c\rangle \\
&= \sum_n f_n\langle\tilde{\psi}_n|\hat{A}|\tilde{\psi}_n\rangle + \sum_{n=1}^{N_c}\langle\tilde{\phi}_n^c|\hat{A}|\tilde{\phi}_n^c\rangle \\
&+ \sum_R\left(\sum_{i,j\in R} D_{i,j}\langle\phi_j|\hat{A}|\phi_i\rangle + \sum_{n\in R}^{N_{c,R}}\langle\phi_n^c|\hat{A}|\phi_n^c\rangle\right) \\
&- \sum_R\left(\sum_{i,j\in R} D_{i,j}\langle\tilde{\phi}_j|\hat{A}|\tilde{\phi}_i\rangle + \sum_{n\in R}^{N_{c,R}}\langle\tilde{\phi}_n^c|\hat{A}|\tilde{\phi}_n^c\rangle\right),
\end{aligned}
\tag{67}
$$

where $\mathbf{D}$ is the one-center density matrix defined as

$$
D_{i,j} = \sum_n f_n\langle\tilde{\psi}_n|\tilde{p}_j\rangle\langle\tilde{p}_i|\tilde{\psi}_n\rangle = \sum_n\langle\tilde{p}_i|\tilde{\psi}_n\rangle f_n\langle\tilde{\psi}_n|\tilde{p}_j\rangle.
\tag{68}
$$

The auxiliary core states, $|\tilde{\phi}_n^c\rangle$ allow us to incorporate the tails of the core wavefunction into the plane-wave part, and therefore assure that the integrations of partial wave contributions cancel exactly beyond r_c. They are identical to the true core states in the tails, but are a smooth continuation inside the atomic sphere. It is not required that the auxiliary wave functions are normalized.

Following this scheme, the electron density is given by

$$
n(\vec{r}) = \tilde{n}(\vec{r}) + \sum_R\left(n_R^1(\vec{r}) - \tilde{n}_R^1(\vec{r})\right)
\tag{69}
$$

$$
\tilde{n}(\vec{r}) = \sum_n f_n\tilde{\psi}_n^*(\vec{r})\tilde{\psi}_n(\vec{r}) + \tilde{n}_c(\vec{r})
$$

$$
n_R^1(\vec{r}) = \sum_{i,j\in R} D_{i,j}\phi_j^*(\vec{r})\phi_i(\vec{r}) + n_{c,R}(\vec{r})
$$

$$
\tilde{n}_R^1(\vec{r}) = \sum_{i,j\in R} D_{i,j}\tilde{\phi}_j^*(\vec{r})\tilde{\phi}_i(\vec{r}) + \tilde{n}_{c,R}(\vec{r}),
\tag{70}
$$

where $n_{c,R}$ is the core density of the corresponding atom and $\tilde{n}_{c,R}$ is the auxiliary core density, which is identical to $n_{c,R}$ outside the atomic region, but smooth inside.

Before we continue, let us discuss a special point: The matrix elements of a general operator with the auxiliary wavefunctions may be slowly converging with the plane-wave expansion, because the operator $\hat{A}$ may not be well behaved. An example of such an operator is the singular electrostatic potential of a nucleus. This problem can be alleviated by adding an "intelligent zero": If an operator $\hat{B}$ is purely localized within an atomic region, we can use the identity between the auxiliary wavefunction and its own partial wave expansion

$$0 \;=\; \langle \tilde{\psi}_n | \hat{B} | \tilde{\psi}_n \rangle - \langle \tilde{\psi}_n^1 | \hat{B} | \tilde{\psi}_n^1 \rangle \,. \tag{71}$$

Now we choose an operator $\hat{B}$ so that it cancels the problematic behavior of the operator $\hat{A}$, but is localized in a single atomic region. By adding $\hat{B}$ to the plane-wave part and the matrix elements with its one-center expansions, the plane-wave convergence can be improved without affecting the converged result. A term of this type, namely $\hat{\bar{v}}$ will be introduced in the next section to cancel the Coulomb singularity of the potential at the nucleus.

Total energy

As with wavefunctions and expectation values, the total energy can be divided into three parts.

$$E\Big[\{|\tilde{\psi}_n\rangle\}, \{R_R\}\Big] \;=\; \tilde{E} + \sum_R \Big(E_R^1 - \tilde{E}_R^1 \Big) \tag{72}$$

The plane-wave part $\tilde{E}$ involves only smooth functions and is evaluated on equi-spaced grids in real and reciprocal space. This part is computationally most demanding, and is similar to the expressions in the pseudopotential approach.

$$
\begin{aligned}
\tilde{E} \;=\;& \sum_n \langle \tilde{\psi}_n | \frac{\hat{\vec{p}}^2}{2m_e} | \tilde{\psi}_n \rangle + \frac{1}{2} \int d^3r \int d^3r' \, \frac{e^2 \Big(\tilde{n}(\vec{r}) + \tilde{Z}(\vec{r}) \Big) \Big(\tilde{n}(\vec{r}') + \tilde{Z}(\vec{r}') \Big)}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} \\
&+ \int d^3r \, \bar{v}(\vec{r}) \tilde{n}(\vec{r}) + E_{xc}[\tilde{n}]
\end{aligned}
\tag{73}
$$

$\tilde{Z}(\mathbf{r})$ is an angular-momentum dependent core-like density that will be described in detail below. The remaining parts can be evaluated on radial grids in a spherical-harmonics expansion. The nodal structure of the wavefunctions can be properly described on a logarithmic radial grid that

becomes very fine near the nucleus,

$$
\begin{aligned}
E_R^1 &= \sum_{i,j \in R} D_{i,j} \langle \phi_j | \frac{\hat{\vec{p}}^2}{2m_e} | \phi_i \rangle + \sum_{n \in R}^{N_{c,R}} \langle \phi_n^c | \frac{\hat{\vec{p}}^2}{2m_e} | \phi_n^c \rangle \\
&+ \frac{1}{2} \int d^3 r \int d^3 r' \frac{e^2 \left(n^1(\vec{r}) + Z(\vec{r}) \right) \left(n^1(\vec{r'}) + Z(\vec{r'}) \right)}{|\vec{r} - \vec{r'}|} + E_{xc}[n^1]
\end{aligned}
\tag{74}
$$

$$
\begin{aligned}
\tilde{E}_R^1 &= \sum_{i,j \in R} D_{i,j} \langle \tilde{\phi}_j | \frac{\hat{\vec{p}}^2}{2m_e} | \tilde{\phi}_i \rangle + \frac{1}{2} \int d^3 r \int d^3 r' \frac{e^2 \left(\tilde{n}^1(\vec{r}) + \tilde{Z}(\vec{r}) \right) \left(\tilde{n}^1(\vec{r'}) + \tilde{Z}(\vec{r'}) \right)}{4\pi\epsilon_0 |\vec{r} - \vec{r'}|} \\
&+ \int d^3 r \, \bar{v}(\vec{r}) \tilde{n}^1(\vec{r}) + E_{xc}[\tilde{n}^1] \,.
\end{aligned}
\tag{75}
$$

The compensation charge density $\tilde{Z}(\vec{r}) = \sum_R \tilde{Z}_R(\vec{r})$ is given as a sum of angular momentum dependent Gauss functions, which have an analytical plane-wave expansion. A similar term occurs also in the pseudopotential approach. In contrast to the norm-conserving pseudopotential approach, however, the compensation charge of an atom $\tilde{Z}_R$ is non-spherical and constantly adapts instantaneously to the environment. It is constructed such that

$$
n_R^1(\vec{r}) + Z_R(\vec{r}) - \tilde{n}_R^1(\vec{r}) - \tilde{Z}_R(\vec{r})
\tag{76}
$$

has vanishing electrostatic multipole moments for each atomic site. With this choice, the electrostatic potentials of the augmentation densities vanish outside their spheres. This is the reason why there is no electrostatic interaction of the one-center parts between different sites.

The compensation charge density as given here is still localized within the atomic regions. A technique similar to an Ewald summation, however, allows it to be replaced by a very extended charge density. Thus we can achieve that the plane-wave convergence of the total energy is not affected by the auxiliary density.

The potential $\bar{v} = \sum_R \bar{v}_R$, which occurs in Eqs. 73 and 75 enters the total energy in the form of "intelligent zeros" described in Eq. 71

$$
0 = \sum_n f_n \left(\langle \tilde{\psi}_n | \bar{v}_R | \tilde{\psi}_n \rangle - \langle \tilde{\psi}_n^1 | \bar{v}_R | \tilde{\psi}_n^1 \rangle \right) = \sum_n f_n \langle \tilde{\psi}_n | \bar{v}_R | \tilde{\psi}_n \rangle - \sum_{i,j \in R} D_{i,j} \langle \tilde{\phi}_i | \bar{v}_R | \tilde{\phi}_j \rangle \,.
\tag{77}
$$

The main reason for introducing this potential is to cancel the Coulomb singularity of the potential in the plane-wave part. The potential $\bar{v}$ allows us to influence the plane-wave convergence beneficially, without changing the converged result. $\bar{v}$ must be localized within the augmentation region, where Eq. 58 holds.

Approximations

Once the total energy functional provided in the previous section has been defined, everything else follows: Forces are partial derivatives with respect to atomic positions. The potential is the derivative of the non-kinetic energy contributions to the total energy with respect to the density,

and the auxiliary Hamiltonian follows from derivatives $\tilde{H}|\tilde{\psi}_n\rangle$ with respect to auxiliary wave functions. The fictitious Lagrangian approach of Car and Parrinello [98] does not allow any freedom in the way these derivatives are obtained. Anything else than analytic derivatives will violate energy conservation in a dynamical simulation. Since the expressions are straightforward, even though rather involved, we will not discuss them here.

All approximations are incorporated already in the total energy functional of the PAW method. What are those approximations?

- Firstly we use the frozen-core approximation. In principle this approximation can be overcome.

- The plane-wave expansion for the auxiliary wavefunctions must be complete. The plane-wave expansion is controlled easily by increasing the plane-wave cutoff defined as $E_{PW} = \frac{1}{2}\hbar^2 G_{max}^2$. Typically we use a plane-wave cutoff of 30 Ry.

- The partial wave expansions must be converged. Typically we use one or two partial waves per angular momentum (ℓ, m) and site. It should be noted that the partial wave expansion is not variational, because it changes the total energy functional and not the basis set for the auxiliary wavefunctions.

We do not discuss here numerical approximations such as the choice of the radial grid, since those are easily controlled.

Relation to pseudopotentials

We mentioned earlier that the pseudopotential approach can be derived as a well defined approximation from the PAW method: The augmentation part of the total energy $\Delta E = E^1 - \tilde{E}^1$ for one atom is a functional of the one-center density matrix $\mathbf{D}$ defined in Eq. 68. The pseudopotential approach can be recovered if we truncate a Taylor expansion of ΔE about the atomic density matrix after the linear term. The term linear in $\mathbf{D}$ is the energy related to the nonlocal pseudopotential.

$$
\begin{aligned}
\Delta E(\mathbf{D}) &= \Delta E(\mathbf{D}^{at}) + \sum_{i,j} \frac{\partial \Delta E}{\partial D_{i,j}}\bigg|_{\mathbf{D}^{at}} (D_{i,j} - D_{i,j}^{at}) + O(\mathbf{D} - \mathbf{D}^{at})^2 \\
&= E_{self} + \sum_n f_n \langle \tilde{\psi}_n | \hat{v}^{ps} | \tilde{\psi}_n \rangle - \int d^3r\ \bar{v}(\vec{r})\tilde{n}(\vec{r}) + O(\mathbf{D} - \mathbf{D})^2\,, \quad (78)
\end{aligned}
$$

which can directly be compared with the total energy expression Eq. 49 of the pseudopotential method. The local potential $\bar{v}(\vec{r})$ of the pseudopotential approach is identical to the corresponding potential of the projector augmented-wave method. The remaining contributions in the PAW total energy, namely $\tilde{E}$, differ from the corresponding terms in Eq. 49 only in two features: our auxiliary density also contains an auxiliary core density, reflecting the nonlinear core correction of the pseudopotential approach, and the compensation density $\tilde{Z}(\vec{r})$ is non-spherical and depends on the wave function. Thus we can look at the PAW method also as a pseudopotential

method with a pseudopotential that adapts instantaneously to the electronic environment. In the PAW method, the explicit nonlinear dependence of the total energy on the one-center density matrix is properly taken into account.

What are the main advantages of the PAW method compared with the pseudopotential approach? Firstly all errors can be systematically controlled, so that there are no transferability errors. As shown by Watson [99] and Kresse [100], most pseudopotentials fail for high spin atoms such as Cr. While it is probably true that pseudopotentials can be constructed that cope even with this situation, a failure can not be known beforehand, so that some empiricism remains in practice: A pseudopotential constructed from an isolated atom is not guaranteed to be accurate for a molecule. In contrast, the converged results of the PAW method do not depend on a reference system such as an isolated atom, because PAW uses the full density and potential.

Like other all-electron methods, the PAW method provides access to the full charge and spin density, which is relevant, for example, for hyperfine parameters. Hyperfine parameters are sensitive probes of the electron density near the nucleus. In many situations they are the only information available that allows us to deduce atomic structure and chemical environment of an atom from experiment.

The plane-wave convergence is more rapid than in norm-conserving pseudopotentials and should in principle be equivalent to that of ultra-soft pseudopotentials [88]. Compared to the ultra-soft pseudopotentials, however, the PAW method has the advantage that the total energy expression is less complex and can therefore be expected to be more efficient.

The construction of pseudopotentials requires us to determine a number of parameters. As they influence the results, their choice is critical. Also the PAW methods provides some flexibility in the choice of auxiliary partial waves. However, this choice does not influence the converged results.

Recent developments

Since the first implementation of the PAW method in the CP-PAW code [58], a number of groups have adopted the PAW method. The second implementation was done by the group of Holzwarth [101], and the resulting PWPAW code is freely available [102]. This code is also used as a basis for the PAW implementation in the ABINIT project [104]. An independent PAW code has been developed by Valiev and Weare [103]. This implementation has entereed the NWChem code [107]. An independent implementation of the PAW method is that of the VASP code [100]. The PAW method has also been implemented by W. Kromen [106] into the EStCoMPP code of Blügel and Schroeder. Another implementation is in the Quantum Espresso code [105]. A real-space-grid based version of the PAW method is the code GPAW developed by Mortensen et al. [108].

Another branch of methods uses the reconstruction of the PAW method, without taking into account the full wavefunctions in the energy minimization. Following chemists' notation, this approach could be termed "post-pseudopotential PAW". This development began with the evaluation for hyperfine parameters from a pseudopotential calculation using the PAW reconstruc-

tion operator [109] and is now used in the pseudopotential approach to calculate properties that require the correct wavefunctions such as hyperfine parameters.

The implementation of the PAW method by Kresse and Joubert [100] has been particularly useful as they had an implementation of PAW in the same code as the ultra-soft pseudopotentials, so that they could critically compare the two approaches. Their conclusion is that both methods compare well in most cases, but they found that magnetic energies are seriously – by a factor two – in error in the pseudopotential approach, while the results of the PAW method were in line with other all-electron calculations using the linear augmented plane-wave method. As an aside, Kresse and Joubert claim incorrectly that their implementation is superior as it includes a term that is analogous to the non-linear core correction of pseudopotentials [110]: this term, however, is already included in the original version in the form of the pseudized core density.

Several extensions of the PAW have been done in the recent years: For applications in chemistry truly isolated systems are often of great interest. As any plane-wave based method introduces periodic images, the electrostatic interaction between these images can cause serious errors. The problem has been solved by mapping the charge density onto a point charge model, so that the electrostatic interaction could be subtracted out in a self-consistent manner [111]. In order to include the influence of the environment, the latter was simulated by simpler force fields using the quantum-mechanics–molecular-mechanics (QM-MM) approach [112].

In order to overcome the limitations of the density functional theory several extensions have been performed. Bengone [113] implemented the LDA+U approach into our CP-PAW code. Soon after this, Arnaud [114] accomplished the implementation of the GW approximation into our CP-PAW code. The VASP-version of PAW [115] and our CP-PAW code have now been extended to include a non-collinear description of the magnetic moments. In a non-collinear description, the Schrödinger equation is replaced by the Pauli equation with two-component spinor wavefunctions.

The PAW method has proven useful to evaluate electric field gradients [116] and magnetic hyperfine parameters with high accuracy [117]. Invaluable will be the prediction of NMR chemical shifts using the GIPAW method of Pickard and Mauri [118], which is based on their earlier work [119]. While the GIPAW is implemented in a post-pseudopotential manner, the extension to a self-consistent PAW calculation should be straightforward. An post-pseudopotential approach has also been used to evaluate core level spectra [120] and momentum matrix elements [121].

Acknowledgment

Appendices

A   Model exchange-correlation energy

We consider a model with a constant density and a hole function that describes a situation where all electrons of the same spin are repelled completely from a sphere centered at the reference electron

The hole function has the form

$$h(\vec{r}, \vec{r}_0) = \begin{cases} -\frac{1}{2}n(\vec{r}_0) & \text{for } |\vec{r} - \vec{r}_0| < r_h \\ 0 & \text{otherwise} \end{cases}$$

where $n(\vec{r})$ is the electron density and the hole radius $r_h = \sqrt[3]{\frac{2}{4\pi n}}$ is the radius of the sphere, which is determined such that the exchange correlation hole integrates to -1, i.e. $\frac{4\pi}{3}r_h^3\left(\frac{1}{2}n\right) = 1$.

The potential of a homogeneously charged sphere with radius r_h and one positive charge is

$$v(r) = \frac{e^2}{4\pi\epsilon_0}\begin{cases} -\frac{3}{2r_h} + \frac{1}{2r_h}\left(\frac{r}{r_h}\right)^2 & \text{for } r \leq r_h \\ -\frac{1}{r} & \text{for } r > r_h \end{cases}$$

where $r = |\vec{r} - \vec{r}_0|$.

With Eq. 12 we obtain for the potential contribution of the exchange correlation energy

$$U_{xc} = -\int d^3r\ n(\vec{r})v(r=0) = -\int d^3r\ \frac{e^2}{4\pi\epsilon_0}\frac{3}{4}\sqrt[3]{\frac{2\pi}{3}}\cdot n^{\frac{4}{3}}$$

B   Large-gradient limit of the enhancement factor

An exponentially decaying density

$$n(r) = \exp(-\lambda r) \tag{79}$$

has a reduced gradient

$$x := \frac{|\vec{\nabla}n|}{n^{\frac{4}{3}}} = \lambda\exp(+\frac{1}{3}\lambda r) \tag{80}$$

We make the following ansatz for the exchange correlation energy per electron

$$\epsilon_{xc}(n, x) = -Cn^{\frac{1}{3}}F(x) \tag{81}$$

where only the local exchange has been used and C is a constant.

Enforcing the long-distance limit of the exchange correlation energy per electron for exponentially decaying densities

$$\epsilon_{xc}((n(r), x(r)) = -\frac{1}{2}\frac{e^2}{4\pi\epsilon_0 r} \tag{82}$$

yields

$$F(x) = \frac{e^2}{4\pi\epsilon_0 r(x) 2C n^{\frac{1}{3}}(r(x))} \tag{83}$$

Using Eqs. 79 and 80, we express the radius and the density by the reduced gradient, i.e.

$$r(x) = -\frac{3}{\lambda}\left(\ln[\lambda] - \ln[x]\right) \tag{84}$$

$$n(x) = n(r(x)) = \lambda^3 x^{-3} , \tag{85}$$

and obtain

$$F(x) = \frac{e^2}{4\pi\epsilon_0 \left[-\frac{3}{\lambda}\left(\ln[\lambda] - \ln[x]\right)\right]\left[2C\lambda x^{-1}\right]} = \left(\frac{e^2}{4\pi\epsilon_0 \cdot 6C}\right)\frac{x^2}{x\ln(\lambda) - x\ln(x)}$$

$$\stackrel{x\to\infty}{\rightarrow} -\left(\frac{e^2}{4\pi\epsilon_0 \cdot 6C}\right)\frac{x^2}{x\ln(x)} \tag{86}$$

Now we need to ensure that $F(0) = 1$, so that the gradient correction vanishes for the homogeneous electron gas, and that $F(x) = F(-x)$ to enforce spin reversal symmetry. There are several possible interpolations for these requirements, but the simplest is

$$F(x) = 1 - \frac{\beta x^2}{1 + \frac{4\pi\epsilon_0}{e^2} \cdot 6C\beta x \cdot \mathrm{asinh}(x)} \tag{87}$$

This is the enhancement factor for exchange used by Becke in his B88 functional [24].

References

[1] W. Kohn, *Nobel Lectures, Chemistry, 1996-2000* (World Scientific, Singapore, 2003), Ch.: Electronic Structure of Matter - Wave functions and Density Functionals, p. 213

[2] P. Hohenberg, W. Kohn, Phys. Rev. **136**, B864 (1964)

[3] W. Kohn, L. Sham, Phys. Rev. **140**, A1133 (1965)

[4] R.O. Jones, O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989)

[5] E.J. Baerends, O.V. Gritsenko, J. Phys. Chem. A **101**, 5383 (1997)

[6] U. von Barth, Physica Scripta **109**, 9 (2004)

[7] J. Perdew, *et al.*, J. Chem. Phys. **123**, 62201 (2005)

[8] A.J. Cohen, P. Mori-Sánchez, W. Yang, Science **321**, 792 (2008)

[9] P.E. Blöchl, C.J. Först, J. Kästner, *Handbook for Materials Modeling* (Springer, 2005), chap. Electronic Structure Methods: Augmented Waves, Pseudopotentials and the Projector augmented wave method, p. 93

[10] P.E. Blöchl, C.J. Först, J. Schimpl, Bull. Mater. Sci. **26**, 33 (2003)

[11] P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955)

[12] A. Coleman, Rev. Mod. Phys. **35**, 668 (1963)

[13] K. Burke, J.P. Perdew, M. Ernzerhof, J. Chem. Phys. **109**, 3760 (1998)

[14] J.P. Perdew, K. Burke, Int. J. Quant. Chem. **57**, 309 (1996)

[15] M. Levy, Proc. Nat'l Acad. Sci. USA **76**, 6062 (1979)

[16] E. Lieb, Int. J. Quant. Chem. **24**, 243 (1983)

[17] J.P. Perdew, K. Schmidt, AIP Conf. Proc. **577**, 1 (2001)

[18] D.M. Ceperley, B. Alder, Phys. Rev. Lett. **45**, 566 (1980)

[19] S.K. Ma, K. Brueckner, Phys. Rev. **165**, 18 (1968)

[20] U. von Barth, L. Hedin, J. Phys. C: Solid State Phys. **5**, 1629 (1972)

[21] O. Gunnarsson, B.I. Lundquist, Phys. Rev. B **13**, 4274 (1976)

[22] J.P. Perdew, Phys. Rev. Lett. **55**, 1665 (1985)

[23] D.C. Langreth, M. Mehl, Phys. Rev. B **28**, 1809 (1983)

[24] A.D. Becke, Phys. Rev. A **38**, 3098 (1988)

[25] J.P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996)

[26] E. Proynov, E. Ruiz, A. Vela, D.R. Salahub, Int. J. Quant. Chem. **56, S29**, 61 (1995)

[27] T.V. Voorhis, G. Scuseria, J. Chem. Phys. **109**, 400 (1998)

[28] A.D. Becke, J. Chem. Phys. **109**, 2092 (1998)

[29] J. Tao, J.P. Perdew, V.N. Staroverov, G.E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003)

[30] A.D. Becke, J. Chem. Phys. **98**, 1372 (1993)

[31] A.D. Becke, J. Chem. Phys. **98**, 5648 (1993)

[32] J. Harris, R.O. Jones, J. Phys. F: Met. Phys. **4**, 1170 (1974)

[33] D.C. Langreth, J.P. Perdew, Sol. St. Commun. **17**, 1425 (1975)

[34] J.P. Perdew, M. Ernzerhof, K. Burke, J. Chem. Phys. **105**, 9982 (1996)

[35] J. Heyd, G. Scuseria, M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003)

[36] M. Marsman, J. Paier, A. Stroppa, G. Kresse, J. Phys. Chem. **20**, 64201 (2008)

[37] P. Stephens, F. Devlin, C. Chabalowski, M. Frisch, J. Phys. Chem. **98**, 11623 (1994)

[38] M. Ernzerhof, G. Scuseria, J. Chem. Phys. **110**, 5029 (99)

[39] C. Adamo, V. Barone, J. Chem. Phys. **110**, 6158 (1999)

[40] V.I. Anisimov, J. Zaanen, O.K. Andersen, Phys. Rev. B **44**, 943 (1991)

[41] A.I. Lichtenstein, V.I. Anisimov, J. Zaanen, Phys. Rev. B **52**, 5467 (1995)

[42] P. Novak, J. Kunes, L. Chaput, W.E. Pickett, Phys. Stat. Sol. B **243**, 563 (2006)

[43] F. Tran, P. Blaha, K. Schwarz, P. Novák, Phys. Rev. B **74**, 155108 (2006)

[44] M. Dion, H. Rydberg, E. Schröder, D.C. Langreth, B. I. Lundqvist,
Phys. Rev. Lett. **92**, 246401 (2004)

[45] T. Thonhauser, V.R. C.S. Li, A. Puzder, P. Hyldgaard, D. Langreth,
Phys. Rev. B **76**, 125112 (2007)

[46] K. Lee, E.D. Murray, L. Kong, B.I. Lundqvist, D.C. Langreth,
Phys. Rev. B **82**, 81101 (2010)

[47] J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari, L.A. Curtiss,
J. Chem. Phys. **90**, 5622 (1989)

[48] L.A. Curtiss, C. Jones, G.W. Trucks, K. Raghavachari, J.A. Pople,
     J. Chem. Phys. **93**, 2537 (1990)

[49] L.A. Curtiss, K. Raghavachari, P.C. Redfern, J.A. Pople,
     J. Chem. Phys. **106**, 1063 (1997)

[50] L.A. Curtiss, P.C. Redfern, K. Raghavachari, J.A. Pople, J. Chem. Phys. **109**, 42 (1998)

[51] A.D. Becke, J. Chem. Phys. **96**, 2155 (1992)

[52] A.D. Becke, J. Chem. Phys. **97**, 9173 (1992)

[53] A.D. Becke, J. Chem. Phys. **104**, 1040 (1996)

[54] A.D. Becke, J. Chem. Phys. **107**, 8554 (1997)

[55] J. Paier, R. Hirschl, M. Marsman, G. Kresse, J. Chem. Phys. **122**, 234102 (2005)

[56] J. Paier, *et al.*, J. Chem. Phys. **124**, 154709 (2006)

[57] J. Paier, *et al.*, J. Chem. Phys. **125**, 249901 (2006)

[58] P.E. Blöchl, Phys. Rev. B **50**, 17953 (1994)

[59] J.C. Slater, Phys. Rev. **51**, 846 (1937)

[60] J. Korringa, Physica **13**, 392 (1947)

[61] W. Kohn, N. Rostoker, Phys. Rev. **94**, 1111 (1954)

[62] O.K. Andersen, Phys. Rev. B **12**, 3060 (1975)

[63] H. Krakauer, M. Posternak, A.J. Freeman, Phys. Rev. B **19**, 1706 (1979)

[64] D.J. Singh, *Planewaves, pseudopotentials and the LAPW method* (Kluwer, 1994)

[65] J. Soler, A.R. Williams, Phys. Rev. B **40**, 1560 (1989)

[66] D.J. Singh, Phys. Rev. B **43**, 6388 (1991)

[67] E. Sjöstedt, L. Nordström, D.J. Singh, Solid State Commun. **113**, 15 (2000)

[68] G.K. H. Madsen, P. Blaha, K. Schwarz, E. Sjoestedt, L. Nordstroem, Phys. Rev. B **64**, 195134 (2001)

[69] H.L. Skriver, *The LMTO method: muffin-tin orbitals and electronic structure* (Springer, 1984)

[70] O.K. Andersen, O. Jepsen, Phys. Rev. Lett. **53**, 2571 (1984)

[71] O.K. Andersen, T. Saha-Dasgupta, S. Ezhof, Bull. Mater. Sci. **26**, 19 (2003)

[72] K. Held, *et al.*, NIC Series (John von Neumann Institute for Computing) **10**, 175 (2002)

[73] C. Herring, Phys. Rev. **57**, 1169 (1940)

[74] J. Phillips, L. Kleinman, Phys. Rev. **116**, 287 (1959)

[75] E. Antončík, J. Phys. Chem. Solids **10**, 314 (1959)

[76] D.R. Hamann, M. Schlüter, C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979)

[77] A. Zunger, M.L. Cohen, Phys. Rev. B **18**, 5449 (1978)

[78] G. Kerker, J. Phys. C: Solid State Phys. **13**, L189 (1980)

[79] G.B. Bachelet, D.R. Hamann, M. Schlüter, Phys. Rev. B **26**, 4199 (1982)

[80] N. Troullier, J. Martins, Phys. Rev. B **43**, 1993 (1991)

[81] J. Lin, A. Qteish, M. Payne, V. Heine, Phys. Rev. B **47**, 4174 (1993)

[82] M. Fuchs, M. Scheffler, Comp. Phys. Comm. **119**, 67 (1998)

[83] L. Kleinman, D.M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982)

[84] P.E. Blöchl, Phys. Rev. B **41**, 5414 (1990)

[85] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990)

[86] S.G. Louie, S. Froyen, M.L. Cohen, Phys. Rev. B **26**, 1738 (1982)

[87] D.R. Hamann, Phys. Rev. B **40**, 2980 (1989)

[88] D. Vanderbilt, Phys. Rev. B **41**, 17892 (1990)

[89] K. Laasonen, A. Pasquarello, R. Car, C. Lee, D. Vanderbilt,
     Phys. Rev. B **47**, 110142 (1993)

[90] X. Gonze, R. Stumpf, M. Scheffler, Phys. Rev. B **44**, 8503 (1991)

[91] C.V. de Walle, P.E. Blöchl, Phys. Rev. B **47**, 4244 (1993)

[92] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, J.D. Joannopoulos,
     Rev. Mod. Phys **64**, 11045 (1992)

[93] R. Car, M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985)

[94] S. Nose, J. Chem. Phys. **81**, 511 (1984)

[95] W.G. Hoover, Phys. Rev. A **31**, 1695 (1985)

[96] P.E. Blöchl, M. Parrinello, Phys. Rev. B **45**, 9413 (1992)

[97] P.E. Blöchl, Phys. Rev. B **65**, 104303 (2002)

[98] R. Car, M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985)

[99] S.C. Watson, E.A. Carter, Phys. Rev. B **58**, R13309 (1998)

[100] G. Kresse, J. Joubert, Phys. Rev. B **59**, 1758 (1999)

[101] N.A.W. Holzwarth, G.E. Matthews, R.B. Dunning, A.R. Tackett, Y. Zheng,
Phys. Rev. B **55**, 2005 (1997)

[102] A.R. Tackett, N.A.W. Holzwarth, G.E. Matthews,
Computer Physics Communications **135**, 329 (2001)

[103] M. Valiev, J.H. Weare, J. Phys. Chem. A **103**, 10588 (1999)

[104] M. Torrent, F. Jollet, F. Bottin, G. Zerah and X. Gonze, Comp. Mat. Sci. **42**, 337 (2008)

[105] P. Gianozzi *et al.*, J. Phys C: Condens. Matt. 21,395502 (2009)

[106] W. Kromen, *Die Projector Augmented Wave-Methode: Ein schnelles Allelektronenver-
fahren für die ab-initio-Molekulardynamik*, (PhD thesis, RWTH Aachen, 2001)

[107] T.L. Windus, E.J. Bylaska, M. Dupuis, S. Hirata and L. Pollack, et al. Lecture Notes in
Computer Science, 2003, Volume 2660, Computational Science ICCS 2003, 168 (2003)

[108] J. J. Mortensen, L. B. Hansen and K.W. Jacobsen, Phys. Rev. B **71**, 35109 (2005)

[109] C.G. Van de Walle, P.E. Blöchl, Phys. Rev. B **47**, 4244 (1993)

[110] S.G. Louie, S. Froyen, M.L. Cohen, Phys. Rev. B **26**, 1738 (1982)

[111] P.E. Blöchl, J. Chem. Phys. **103**, 7422 (1995)

[112] T.K. Woo, P.M. Margl, P.E. Blöchl, T. Ziegler, J. Phys. Chem. B **101**, 7877 (1997)

[113] O. Bengone, M. Alouani, P.E. Blöchl, J. Hugel, Phys. Rev. B **62**, 16392 (2000)

[114] B. Arnaud, M. Alouani, Phys. Rev. B. **62**, 4464 (2000)

[115] D. Hobbs, G. Kresse, J. Hafner, Phys. Rev. B **62**, 11556 (2000)

[116] H.M. Petrilli, P.E. Blöchl, P. Blaha, K. Schwarz, Phys. Rev. B **57**, 14690 (1998)

[117] P.E. Blöchl, Phys. Rev. B **62**, 6158 (2000)

[118] C.J. Pickard, F. Mauri, Phys. Rev. B. **63**, 245101 (2001)

[119] F. Mauri, B.G. Pfrommer, S.G. Louie, Phys. Rev. Lett. **77**, 5300 (1996)

[120] D.N. Jayawardane, C.J. Pickard, L.M. Brown, M.C. Payne,
Phys. Rev. B **64**, 115107 (2001)

[121] H. Kageshima, K. Shiraishi, Phys. Rev. B **56**, 14985 (1997)