Projector augmented wave method: *ab initio* molecular dynamics with full wave functions

PETER E BLÖCHL*, CLEMENS J FÖRST[†] and JOHANNES SCHIMPL

Clausthal University of Technology, Institute for Theoretical Physics, Leibnizstr.10, D-38678 Clausthal-Zellerfeld, Germany
[†]Also at Vienna University of Technology, Institute for Materials Chemistry, Getreidemarkt 9/165-TC, A-1060 Vienna, Austria

Abstract. A brief introduction to the projector augmented wave method is given and recent developments are reviewed. The projector augmented wave method is an all-electron method for efficient *ab initio* molecular dynamics simulations with full wave functions. It extends and combines the traditions of existing augmented wave methods and the pseudopotential approach. Without sacrificing efficiency, the PAW method avoids transferability problems of the pseudopotential approach and it has been valuable to predict properties that depend on the full wave functions.

Keywords. Electronic structure calculations; *ab initio* molecular dynamics; pseudopotentials; augmented wave methods.

1. Introduction

The main goal of electronic structure methods is to solve the Schrödinger equation for the electrons in a molecule or solid, to evaluate the resulting total energies, forces, response functions and other quantities of interest. In this paper we review the projector augmented wave (PAW) method (Blöchl 1994), an electronic structure method for *ab initio* molecular dynamics with full wave functions. The main goal of this paper is not to provide a particularly complete or detailed account of the methodology, but rather to lay out the underlying ideas. A more rigorous description can be found in the original paper (Blöchl 1994).

Density functional theory (Hohenberg and Kohn 1964; Kohn and Sham 1965) maps a description for interacting electrons onto one of non-interacting electrons in an effective potential. The remaining one-electron Schrödinger equation still poses substantial numerical difficulties: (i) in the atomic region near the nucleus, the kinetic energy of the electrons is large, resulting in rapid oscillations of the wave function 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 wave function. Therefore, the wave function in the atomic region can be represented well already by a small basis set and (ii) in the bonding region between the atoms the situation is opposite. The kinetic energy is small and the wave function is smooth.

However, the wave function 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.

- Most appealing to quantum chemists has been the atomic point of view. Basis functions that resemble atomic orbitals are chosen. They exploit that the wave function in the atomic region can be described by a few basis functions, while the bonding is described by the overlapping tails of these atomic orbitals. Most techniques in this class are a compromise of a well adapted basis set with complex matrix elements on the one hand 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 are plane waves. Plane waves are complete and well adapted to sufficiently smooth wave functions. The disadvantage of the large basis sets required is offset by the extreme simplicity to evaluate matrix elements. Finite plane wave expansions are, however, absolutely inadequate to describe the strong oscillations of the wave functions near the nucleus. In the pseudopotential approach the Pauli repulsion of the core electrons is therefore described by an effective potential that expels the valence electrons from the core region. The resulting wave functions are smooth and can be represented well by plane waves. The price to pay is that all information on the charge density and wave functions near the nucleus is lost.

---

*Author for correspondence

- Augmented wave methods compose their basis functions out of atom-like partial waves in the atomic regions and a set of functions, called envelope functions, appropriate for the bonding in between. Space is divided accordingly into atom-centred spheres, defining the atomic region, and an interstitial region for the bonds. The partial solutions of the different regions are matched 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 methods 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.

The augmented wave methods have been introduced in 1937 by Slater (Slater 1937) and later modified by Korringa (1947), Kohn and Rostokker (1954). They approached 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, a bound state has been determined. This approach can be translated into a basis set method with energy dependent 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 potential by a constant in the interstitial region and by a spherically symmetric potential in the atomic region.

The pseudopotential approach traces back to 1940 when C Herring invented the orthogonalized plane wave method (Herring 1940). Later, Phillips and Kleinman (1959) and Antoncik (1959) replaced the orthogonality condition by an effective potential, that 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.

Augmented wave and pseudopotential methods reached adulthood in the 1970s: At first Andersen (1975) 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 in energy. In the original APW approach the zeros of an energy dependent matrix had to be determined, which is problematic, if many states lie in a small energy region as 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. In that case the muffin–tin approximation is used solely to define the basis set. The matrix elements of the Hamiltonian are evaluated with the full potential.

Hamann *et al* (1979) showed 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. A main disadvantage of these pseudopotentials has been the large basis set size required especially for first-row and transition metal atoms.

In 1985, Car and Parrinello published the *ab initio* molecular dynamics method. 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 wave function coefficients. If one applies friction, the system is quenched to the ground state. Without friction truly dynamical simulations of the atomic structure are performed. Electronic wave functions and atomic positions are treated on equal footing.

The Car–Parrinello method had been implemented first for the pseudopotential approach. There seemed to be unsurmountable barriers against combining the new technique with augmented wave methods. The main problem was related to the potential dependent basis set used so far: the Car–Parrinello method requires a well defined and unique total energy functional of atomic positions and basis set coefficients. Therefore, it was one of the main goals of the PAW method to introduce energy and potential independent basis sets that were as accurate and numerically efficient as the previously used augmented basis sets. Other requirements have been: (i) The method should match the efficiency of the pseudopotential approach for Car–Parrinello simulations, (ii) it should become an exact theory when converged and (iii) its convergence should be easily controlled. We believe that these criteria have been met, which explains why the PAW method has become increasingly wide spread today.

We would like to point out that most of these seemingly singular developments did not come out of the blue, but the ideas seemed to have evolved in the community. In the case of the PAW method, similar ideas have been developed by Vanderbilt (1990) in the context of ultrasoft pseudopotentials. The first dynamical simulations using a semiempirical electronic structure method have

been performed by Wang and Karplus (1973). The first *ab initio* pseudopotentials have been published by Zunger and Cohen (1978) one year before Hamann *et al* (1979).

2. Transformation theory

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

Let us denote the physical one-particle wave functions as $|\psi_n\rangle$ and the auxiliary wave functions as $|\tilde{\psi}_n\rangle$. Note that the tilde refers to the representation of smooth auxiliary wave functions. 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 $\mathcal{T}$,

$$|\psi_n\rangle = \mathcal{T} |\tilde{\psi}_n\rangle. \tag{1}$$

We use here Dirac's Bra and Ket notation. A wave function, $\psi_n(\mathbf{r})$, corresponds to a ket $|\psi_n\rangle$, the complex conjugate wave function, $\psi_n^*(\mathbf{r})$ corresponds to a bra $\langle\psi_n|$, and a scalar product $\int d^3 r \psi_n^*(\mathbf{r})\psi_m(\mathbf{r})$ is written as $\langle\psi_n|\psi_m\rangle$. Vectors in the 3-d coordinate space are indicated by bold-faced symbols.

The electronic ground state is determined by minimizing a total energy functional $E[\psi_n]$ of the density functional theory. 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

$$F([\psi_n], \Lambda_{m,n}) = E[\psi_n] - \sum_{n,m}[\langle\psi_n|\psi_m\rangle - d_{n,m}]\Lambda_{n,m}, \tag{2}$$

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

$$H|\psi_n\rangle f_n = \sum_m |\psi_m\rangle\Lambda_{m,n}, \tag{3}$$

where the f_n are the occupation numbers and $H = -\frac{\hbar^2}{2m_e}\nabla^2 + v_{\text{eff}}(\mathbf{r})$ is the effective one-particle Hamilton operator.

After a unitary transformation that diagonalizes the matrix of Lagrange multipliers, $\Lambda_{m,n}$, we obtain the Kohn–Sham equations

$$H|\psi_n\rangle = |\psi_n\rangle e_n. \tag{4}$$

The one-particle energies, e_n, are the eigenvalues of $\Lambda_{n,m} \frac{f_n+f_m}{2f_n f_m}$.

Now we express the functional, F, in terms of our auxiliary wave functions

$$F([\mathcal{T}\tilde{\psi}_n], \Lambda_{m,n}) = E[\mathcal{T}\tilde{\psi}_n]$$
$$- \sum_{n,m}[\langle\tilde{\psi}_n|\mathcal{T}^\dagger\mathcal{T}|\tilde{\psi}_m\rangle - d_{n,m}]\Lambda_{n,m}. \tag{5}$$

The variational principle with respect to the auxiliary wave functions yields

$$\mathcal{T}^\dagger H\mathcal{T}|\tilde{\psi}_n\rangle = \mathcal{T}^\dagger\mathcal{T}|\tilde{\psi}_n\rangle e_n. \tag{6}$$

Again we obtain a Schrödinger-like equation, but now the Hamilton operator has a different form, $\mathcal{T}^\dagger H\mathcal{T}$, an overlap operator $\mathcal{T}^\dagger\mathcal{T}$ occurs and the resulting auxiliary wave functions are smooth.

When we evaluate physical quantities we need to evaluate expectation values of an operator A, which can be expressed in terms of either the true or the auxiliary wave functions.

$$\langle A\rangle = \sum_n f_n\langle\psi_n|A|\psi_n\rangle = \sum_n f_n\langle\tilde{\psi}_n|\mathcal{T}^\dagger A\mathcal{T}|\tilde{\psi}_n\rangle. \tag{7}$$

In the representation of auxiliary wave functions we need to use transformed operators, $\tilde{A} = \mathcal{T}^\dagger A\mathcal{T}$. 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 wave functions. We will see that our auxiliary wave functions, which are simply the plane wave parts of the full wave functions, translate into the wave functions of the pseudopotential approach. In the PAW method the auxiliary wave functions are used to construct the true wave functions 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 (Blöchl 1994), the auxiliary wave functions have been termed pseudo wave functions and the true wave functions have been termed all-electron wave functions, in order to make the connection more evident. We shall avoid this notation here, because it results in confusion in cases, where the correspondence is not clear cut.

3. 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, $\mathcal{T}$.

The operator, $\mathcal{T}$, has to modify the smooth auxiliary wave function in each atomic region, so that the resulting

wave function has the correct nodal structure. Therefore, it makes sense to write the transformation as identity plus a sum of atomic contributions, $\mathcal{S}_R$

$$\mathcal{T} = 1 + \sum_R \mathcal{S}_R. \tag{8}$$

For every atom, $\mathcal{S}_R$ adds the difference between the true and the auxiliary wave function. The index R is a label for an atomic site.

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

$$\mathcal{y}(\mathbf{r}) = \sum_{i \in R} f_i(\mathbf{r}) c_i \quad \text{for} \quad |\mathbf{r} - \mathbf{R}_R| < r_{c,R}. \tag{9}$$

The index i refers to a site index R, the angular momentum indices (l, m) and an additional index that differentiates partial waves with same angular momentum quantum numbers on the same site. With $i \in R$ we indicate those partial waves that belong to site R. $\mathbf{R}_R$ is the position of the nucleus of site R.

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.

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

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

$$|f_i\rangle = (1 + \mathcal{S}_R)|\tilde{f}_i\rangle \quad \text{for} \quad i \in R,$$
$$\mathcal{S}_R|\tilde{f}_i\rangle = |f_i\rangle - |\tilde{f}_i\rangle, \tag{10}$$

defines the local contribution of $\mathcal{S}_R$ to the transformation operator. Since $1 + \mathcal{S}_R$ shall change the wave function only locally, we require that the partial waves $|f_i\rangle$ and their auxiliary counterparts $|\tilde{f}_i\rangle$ are pairwise identical beyond a certain radius, r_c.

$$f_i(r) = \tilde{f}_i(r) \quad \text{for} \quad i \in R \quad \text{and} \quad |\mathbf{r} - \mathbf{R}_R| > r_{c,R}. \tag{11}$$

In order to be able to apply the transformation operator to an arbitrary auxiliary wave function, we need to be able

to expand the auxiliary wave function locally into the auxiliary partial waves

$$\tilde{\mathcal{y}}(\mathbf{r}) = \sum_{i \in R} \tilde{f}_i(\mathbf{r}) \langle \tilde{p}_i | \tilde{\mathcal{y}} \rangle \quad \text{for} \quad |\mathbf{r} - \mathbf{R}_R| < r_{c,R}, \tag{12}$$

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. Examples of projector functions are shown in figure 1. From (12) we can derive

$$\sum_i |\tilde{f}_i\rangle \langle \tilde{p}_i| = 1,$$

which is valid within r_c. It can be shown by insertion, that the identity (12) holds for any auxiliary wave function $|\tilde{\mathcal{y}}\rangle$ that can be expanded locally into auxiliary partial waves $|\tilde{f}_i\rangle$, if

$$\langle \tilde{p}_i | \tilde{f}_j \rangle = d_{i,j} \quad \text{for} \quad i, j \in R. \tag{13}$$

Note that neither the projector functions nor the partial waves need to be orthogonal among themselves.

By combining (10) and (12), we can apply $\mathcal{S}_R$ to any auxiliary wave function

$$\mathcal{S}_R|\tilde{\mathcal{y}}\rangle = \sum_{i \in R} \mathcal{S}_R|\tilde{f}_i\rangle\langle \tilde{p}_i|\tilde{\mathcal{y}}\rangle = \sum_{i \in R} (|f_i\rangle - |\tilde{f}_i\rangle)\langle \tilde{p}_i|\tilde{\mathcal{y}}\rangle. \tag{14}$$

Hence the transformation operator is

$$\mathcal{T} = 1 + \sum_i (|f_i\rangle - |\tilde{f}_i\rangle)\langle \tilde{p}_i|, \tag{15}$$

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

$$|\mathcal{y}\rangle = |\tilde{\mathcal{y}}\rangle + \sum_i (|f_i\rangle - |\tilde{f}_i\rangle)\langle \tilde{p}_i|\tilde{\mathcal{y}}\rangle$$
$$= |\tilde{\mathcal{y}}_n\rangle + \sum_R (|\mathcal{y}_R^1\rangle - |\tilde{\mathcal{y}}_R^1\rangle), \tag{16}$$
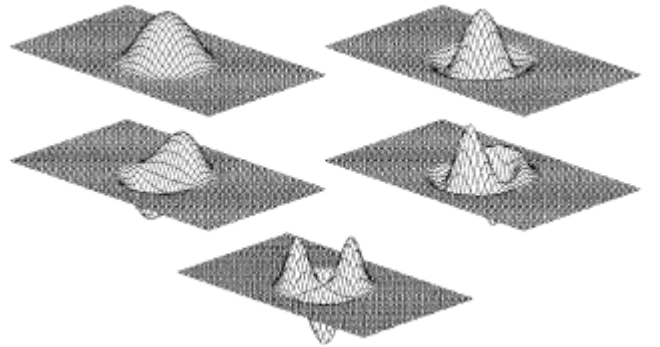

Figure 1. Top: projector functions of the Cl atom for two *s*-type partial waves, middle: *p*-type, bottom: *d*-type.

with

$$|\mathbf{y}_R^1\rangle = \sum_{i \in R} |\mathbf{f}_i\rangle\langle \tilde{p}_i |\tilde{\mathbf{y}}\rangle, \qquad (17)$$

$$|\tilde{\mathbf{y}}_R^1\rangle = \sum_{i \in R} |\tilde{\mathbf{f}}_i\rangle\langle \tilde{p}_i |\tilde{\mathbf{y}}\rangle. \qquad (18)$$

In figure 2 the decomposition of (16) is shown for the example of the bonding p–s state of the Cl_2 molecule.

To understand the expression for the true wave function, (16), let us concentrate on different regions in space. (i) Far from the atoms, the partial waves are, according to (11), pairwise identical so that the auxiliary wave function is identical to the true wave function $\mathbf{y}(\mathbf{r}) = \tilde{\mathbf{y}}(\mathbf{r})$ and (ii) close to an atom, however, the true wave function, $\mathbf{y}(\mathbf{r}) = \mathbf{y}_R^1(\mathbf{r})$ is built up from partial waves that contain the proper nodal structure, because the auxiliary wave function and its partial wave expansion are equal according to (12).

In practice the partial wave expansions are truncated. Therefore, the identity of (12) does not hold strictly. As a result the plane waves also contribute to the true wave function 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, which explains the rapid convergence of the partial wave expansions.

Frequently, the question comes up, whether the transformation (15) of the auxiliary wave functions indeed
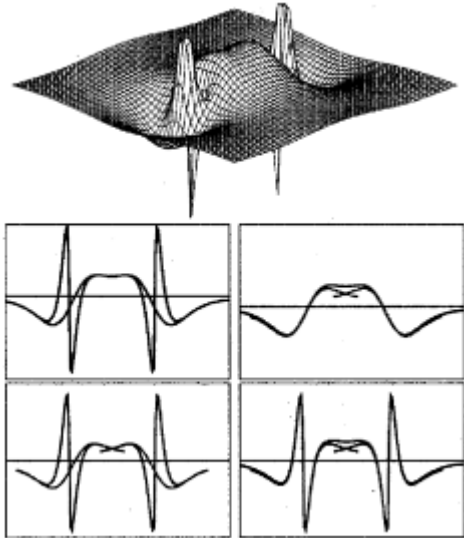

Figure 2. Bonding p–s orbital of the Cl_2 molecule and its decomposition of the wave function into auxiliary wave function and the two one-centre expansions. Top-left: True and auxiliary wave function; top-right: auxiliary wave function and its partial wave expansion; bottom-left: the two partial wave expansions; bottom-right: true wave function and its partial wave expansion.

provides the true wave function. 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 an auxiliary wave function that produces a correct wave function $|\mathbf{y}\rangle$.

4. Expectation values

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

$$\langle A \rangle = \sum_n f_n \langle \mathbf{y}_n | A | \mathbf{y}_n \rangle + \sum_{n=1}^{N_c} \langle \mathbf{f}_n^c | A | \mathbf{f}_n^c \rangle$$

$$= \sum_n f_n \langle \tilde{\mathbf{y}}_n | \mathcal{T}^\dagger A \mathcal{T} | \tilde{\mathbf{y}}_n \rangle + \sum_{n=1}^{N_c} \langle \mathbf{f}_n^c | A | \mathbf{f}_n^c \rangle, \qquad (19)$$

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 $|\mathbf{f}_n^c\rangle$.

Now we can decompose the matrix elements into their individual contributions according to (16)

$$\langle \mathbf{y}|A|\mathbf{y}\rangle = \left\langle \tilde{\mathbf{y}} + \sum_R (\mathbf{y}_R^1 - \tilde{\mathbf{y}}_R^1) \middle| A \middle| \tilde{\mathbf{y}} + \sum_{R'} (\mathbf{y}_{R'}^1 - \tilde{\mathbf{y}}_{R'}^1) \right\rangle$$

$$= \underbrace{\langle \tilde{\mathbf{y}}|A|\tilde{\mathbf{y}}\rangle + \sum_R (\langle \mathbf{y}_R^1|A|\mathbf{y}_R^1\rangle - \langle \tilde{\mathbf{y}}_R^1|A|\tilde{\mathbf{y}}_R^1\rangle)}_{\text{part 1}}$$

$$+ \underbrace{\sum_R (\langle \mathbf{y}_R^1 - \tilde{\mathbf{y}}_R^1|A|\tilde{\mathbf{y}} - \tilde{\mathbf{y}}_R^1\rangle + \langle \tilde{\mathbf{y}} - \tilde{\mathbf{y}}_R^1|A|\mathbf{y}_R^1 - \tilde{\mathbf{y}}_R^1\rangle)}_{\text{part 2}}$$

$$+ \underbrace{\sum_{R \neq R'} \langle \mathbf{y}_R^1 - \tilde{\mathbf{y}}_R^1|A|\mathbf{y}_{R'}^1 - \tilde{\mathbf{y}}_{R'}^1\rangle}_{\text{part 3}}. \qquad (20)$$

Only the first part of (20) is evaluated explicitly, while the second and third parts of (20) are neglected, because they vanish for sufficiently local operators as long as the partial wave expansion is converged: The function $\mathbf{y}_R^1 - \tilde{\mathbf{y}}_R^1$ vanishes per construction beyond some augmentation region, because the partial waves are pairwise identical beyond that region. The function, $\tilde{\mathbf{y}} - \tilde{\mathbf{y}}_R^1$, vanishes inside the augmentation region, if the partial wave expansion is sufficiently converged. In no region of space both functions $\mathbf{y}_R^1 - \tilde{\mathbf{y}}_R^1$ and $\tilde{\mathbf{y}} - \tilde{\mathbf{y}}_R^1$ are simultaneously nonzero. Similarly the functions $\mathbf{y}_R^1 - \tilde{\mathbf{y}}_R^1$ from different sites are never non-zero in the same region in space. Hence, the second and third parts of (20) vanish for operators such as the kinetic energy $\frac{-\hbar^2}{2m_e}\nabla^2$ and the

real space projection operator $|r\rangle\langle r|$, which produces the electron density. For truly nonlocal operators the second and third parts of (20) would have to be considered explicitly.

The expression, (20), for the expectation value can therefore be written as

$$\langle A \rangle = \sum_n f_n (\langle \tilde{\psi}_n \mid A \mid \tilde{\psi}_n \rangle + \langle \psi_n^1 \mid A \mid \psi_n^1 \rangle - \langle \tilde{\psi}_n^1 \mid A \mid \tilde{\psi}_n^1 \rangle)$$

$$+ \sum_{n=1}^{N_c} \langle \boldsymbol{f}_n^c \mid A \mid \boldsymbol{f}_n^c \rangle$$

$$= \sum_n f_n \langle \tilde{\psi}_n \mid A \mid \tilde{\psi}_n \rangle + \sum_{n=1}^{N_c} \langle \tilde{\boldsymbol{f}}_n^c \mid A \mid \tilde{\boldsymbol{f}}_n^c \rangle$$

$$+ \sum_R \left(\sum_{i,j \in R} D_{i,j} \langle \boldsymbol{f}_j \mid A \mid \boldsymbol{f}_i \rangle + \sum_{n \in R}^{N_{c,R}} \langle \boldsymbol{f}_n^c \mid A \mid \boldsymbol{f}_n^c \rangle \right)$$

$$- \sum_R \left(\sum_{i,j \in R} D_{i,j} \langle \tilde{\boldsymbol{f}}_j \mid A \mid \tilde{\boldsymbol{f}}_i \rangle + \sum_{n \in R}^{N_{c,R}} \langle \tilde{\boldsymbol{f}}_n^c \mid A \mid \tilde{\boldsymbol{f}}_n^c \rangle \right), \quad (21)$$

where $D_{i,j}$ is the one-centre density matrix defined as

$$D_{i,j} = \sum_n f_n \langle \tilde{\psi}_n \mid \tilde{p}_j \rangle \langle \tilde{p}_i \mid \tilde{\psi}_n \rangle = \sum_n \langle \tilde{p}_i \mid \tilde{\psi}_n \rangle f_n \langle \tilde{\psi}_n \mid \tilde{p}_j \rangle. \quad (22)$$

The auxiliary core states, $|\tilde{\boldsymbol{f}}_n^c\rangle$ allow to incorporate the tails of the core wave function into the plane wave part, and therefore assure, that the integrations of partial wave contributions cancel strictly 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.

For example, the electron density is given by

$$n(\mathbf{r}) = \tilde{n}(\mathbf{r}) + \sum_R (n_R^1(\mathbf{r}) - \tilde{n}_R^1(\mathbf{r})), \quad (23)$$

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

$$n_R^1(\mathbf{r}) = \sum_{i,j \in R} D_{i,j} \boldsymbol{f}_j^*(\mathbf{r}) \boldsymbol{f}_i(\mathbf{r}) + n_{c,R},$$

$$\tilde{n}_R^1(\mathbf{r}) = \sum_{i,j \in R} D_{i,j} \tilde{\boldsymbol{f}}_j^*(\mathbf{r}) \tilde{\boldsymbol{f}}_i(\mathbf{r}) + \tilde{n}_{c,R}, \quad (24)$$

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

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

$$0 = \langle \tilde{\psi}_n \mid B \mid \tilde{\psi}_n \rangle - \langle \tilde{\psi}_n^1 \mid B \mid \tilde{\psi}_n^1 \rangle. \quad (25)$$

Now we choose an operator B so that it cancels the problematic behaviour of the operator A, but is localized in a single atomic region. By adding B to the plane wave part and the matrix elements with its one-centre expansions, the plane wave convergence can be improved without affecting the converged result.

5. Total energy

Like wave functions and expectation values also the total energy can be divided into three parts

$$E([\tilde{\psi}_n], R_i) = \tilde{E} + \sum_R (E_R^1 - \tilde{E}_R^1). \quad (26)$$

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

$$\tilde{E} = \sum_n \left\langle \tilde{\psi}_n \left| \frac{-\hbar^2}{2m_e} \nabla^2 \right| \tilde{\psi}_n \right\rangle$$

$$+ \frac{e^2}{8\pi e_0} \int d^3 r \int d^3 r' \frac{[\tilde{n}(\mathbf{r}) + \tilde{Z}(\mathbf{r})][\tilde{n}(\mathbf{r}') + \tilde{Z}(\mathbf{r}')]}{|\mathbf{r} - \mathbf{r}'|}$$

$$+ \int d^3 r \, \tilde{n}(\mathbf{r}) \boldsymbol{e}_{xc}(\mathbf{r}, [\tilde{n}]) + \int d^3 r \, \bar{v}(\mathbf{r}) \tilde{n}(\mathbf{r}), \quad (27)$$

where $\tilde{Z}(\mathbf{r})$ is an angular 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 wave functions can be properly described on a logarithmic radial grid that becomes very fine near nucleus,

$$E_R^1 = \sum_{i,j \in R} D_{i,j} \left\langle \boldsymbol{f}_j \left| \frac{-\hbar^2}{2m_e} \nabla^2 \right| \boldsymbol{f}_i \right\rangle + \sum_{n \in R}^{N_{c,R}} \left\langle \boldsymbol{f}_n^c \left| \frac{-\hbar^2}{2m_e} \nabla^2 \right| \boldsymbol{f}_n^c \right\rangle$$

$$+ \frac{e^2}{8\pi e_0} \int d^3 r \int d^3 r' \frac{[n^1(\mathbf{r}) + Z(\mathbf{r})][n^1(\mathbf{r}') + Z(\mathbf{r}')]}{|\mathbf{r} - \mathbf{r}'|}$$

$$+ \int d^3r\, n^1(\mathbf{r})\boldsymbol{e}_{\mathrm{xc}}(\mathbf{r},[n^1]). \tag{28}$$

$$\tilde{E}_R^1 = \sum_{i,j \in R} D_{i,j} \left\langle \tilde{\boldsymbol{f}}_j \left| \frac{-\hbar^2}{2m_e} \nabla^2 \right| \tilde{\boldsymbol{f}}_i \right\rangle$$

$$+ \frac{e^2}{8\boldsymbol{p}\boldsymbol{e}_0} \int d^3r \int d^3r' \frac{[\tilde{n}^1(\mathbf{r}) + \tilde{Z}(\mathbf{r})][\tilde{n}^1(\mathbf{r}') + \tilde{Z}(\mathbf{r}')]}{|\mathbf{r} - \mathbf{r}'|}$$

$$+ \int d^3r\, \tilde{n}^1(\mathbf{r})\boldsymbol{e}_{\mathrm{xc}}(\mathbf{r},[\tilde{n}^1]) + \int d^3r\, \bar{v}(\mathbf{r})\tilde{n}^1(\mathbf{r}). \tag{29}$$

The nuclear charge density $-eZ(\mathbf{r})$ is defined as a sum of *d*functions on the nuclear sites, $Z(\mathbf{r}) = -\sum_R z_R \boldsymbol{d}\mathbf{r}{-}\mathbf{R})$, with the atomic numbers z_R. Note that the self energy of a point charge is infinite and must be subtracted out.

The compensation density $\tilde{Z}(\mathbf{r}) = \sum_R \tilde{Z}_R(\mathbf{r})$ is given as a sum of angular momentum dependent Gauss functions, which have an analytical Fourier transform. A similar term occurs also in the pseudopotential approach. In contrast to the norm-conserving pseudopotential approach however, the compensation charge is non-spherical and it is constantly adapting to the instantaneous environment. It is constructed such that the augmentation charge densities

$$n_R^1(\mathbf{r}) + Z_R(\mathbf{r}) - \tilde{n}_R^1(\mathbf{r}) - \tilde{Z}_R(\mathbf{r}), \tag{30}$$

have vanishing electrostatic multi-pole moments for each atomic site. As a result the sum of all one-centre contributions from one atom does not produce an electrostatic potential outside their own atomic region. This is the reason that the electrostatic interaction of the one-centre parts between different sites vanish.

The compensation charge density as given here is still localized within the atomic regions, but a technique similar to an Ewald summation allows to replace it by a very extended charge density. Thus we can achieve, that all functions in $\tilde{E}$ converge as fast as the auxiliary density itself.

The potential $\bar{v}$, which occurs in (27) and (29) enters the total energy in the form of a zero described in (25)

$$\sum_n f_n \langle \tilde{\boldsymbol{y}}_n | \left(\bar{v} - \sum_{i,j} |\tilde{p}_i\rangle\langle \tilde{\boldsymbol{f}}_i | \bar{v} | \tilde{\boldsymbol{f}}_j \rangle\langle \tilde{p}_j | \right) | \tilde{\boldsymbol{y}}_n \rangle. \tag{31}$$

The main reason for introducing this potential is that the self-consistent potential resulting from the plane wave part is not necessarily optimally smooth. The potential, $\bar{v}$, allows to influence the plane wave convergence beneficially, without changing the converged result. $\bar{v}$ must be localized within the augmentation region, where (12) holds.

6. 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 potential energy with respect to the density, and the Hamiltonian follows from derivatives with respect to wave functions. The fictitious Lagrangian approach of Car and Parrinello (1985) 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 wave functions must be complete. The plane wave expansion is controlled easily by increasing the plane wave cutoff defined as $E_{\mathrm{PW}} = (\hbar G_{\mathrm{max}})/2m_e$. 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 the partial wave expansion changes the total energy functional and not only the basis set.

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

We mentioned earlier that the pseudopotential approach can be derived as a well defined approximation from the PAW method: The augmentation part $\Delta E = E^1 - \tilde{E}^1$ is a functional of the one-centre density matrix, $D_{i,j}$, defined in (22). 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 to $D_{i,j}$ is the energy related to the nonlocal pseudopotential

$$\Delta E(D_{i,j}) = \Delta E(D_{i,j}^{\mathrm{at}}) + \sum_{i,j} (D_{i,j} - D_{i,j}^{\mathrm{at}}) \frac{\partial \Delta E}{\partial D_{i,j}}$$

$$+ O(D_{i,j} - D_{i,j}^{\mathrm{at}})^2$$

$$= E_{\mathrm{self}} + \sum_n f_n \langle \tilde{\boldsymbol{y}}_n | v_{nl} | \tilde{\boldsymbol{y}}_n \rangle + O(D_{i,j} - D_{i,j}^{\mathrm{at}})^2. \tag{32}$$

Thus we can look at the PAW method also as a pseudopotential method with a pseudopotential that adapts to the instantaneous electronic environment, because the expli-

cit nonlinear dependence of the total energy on the one-centre density matrix is properly taken into account.

What are the main advantages of the PAW method compared to the pseudopotential approach?

Firstly all errors can be systematically controlled so that there are no transferability errors. As shown by Watson and Carter (1998) and Kresse and Joubert (1999) 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 cannot 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 it uses the full density and potential.

The PAW method provides access to the full charge and spin density, which is relevant 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 to deduce atomic structure and chemical environment of an atom. There are reconstruction techniques for the pseudopotential approach, which however, are poor man's versions (Van de Walle and Blöchl 1993) of the PAW method.

The plane wave convergence is more rapid than in norm-conserving pseudopotentials and should in principle be equivalent to that of ultra-soft pseudopotentials (Vanderbilt 1990). Compared to the ultra-soft pseudopotentials, however, the PAW method has the advantage that the total energy expression is less complex and therefore is expected to be more efficient.

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

7. Recent developments

Since the first implementation of the PAW method in the CP–PAW code, a number of groups have adopted the PAW method. The second implementation was done by the group of Holzwarth (Holzwarth *et al* 1997). The resulting PWPAW code is freely available (Tackett *et al* 2001). This code is also used as a basis for the PAW implementation in the AbInit project (AbInit). An independent PAW code has been developed by Valiev and Weare (1999). Recently the PAW method has been implemented into the VASP code (Kresse and Joubert 1999). The PAW method has also been implemented by W. Kromen into the ESTCoMPP code of Blügel *et al* (2001).

Another branch of method uses the reconstruction of the PAW method, without taking into account the full wave functions in the self-consistency. Following chemist 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 reconstruction operator (Van de Walle and Blöchl 1993) and is now used in the pseudopotential approach to calculate properties that require the correct wave functions.

The implementation by Kresse and Joubert (1999) 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 with each other and LAPW calculations. Their conclusion is that both methods compare well in most cases, but they found that magnetic energies are seriously—by a factor of 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 a short note, Kresse and Joubert incorrectly claim that their implementation is superior as it includes a term that is analogous to the non-linear core correction of pseudopotentials (Louie *et al* 1982): 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 (Blöchl 1995). In order to include the influence of the environment, the latter was simulated by simpler force fields using the molecular-mechanics–quantum-mechanics (QM–MM) approach (Woo *et al* 1997, 2000).

In order to overcome the limitations of the density functional theory several extensions have been performed. Bengone *et al* (2000) implemented the LDA + U approach (Anisimov *et al* 1991) into the CP–PAW code. Soon after this, Arnaud and Alouani (2000) accomplished the implementation of the GW approximation into the CP–PAW code. The VASP-version of PAW (Hobbs *et al* 2000) and the CP–PAW code have now been extended to include a noncollinear description of the magnetic moments. In a noncollinear description the Schrödinger equation is replaced by the Pauli equation with two-component spinor wave functions.

The PAW method has proven useful to evaluate electric field gradients (Petrilli *et al* 1998) and magnetic hyperfine parameters with high accuracy (Blöchl 2000). Invaluable will be the prediction of NMR chemical shifts using the GIPAW method of Pickard and Mauri (2001), which is based on their earlier work (Mauri *et al* 1996). While the GIPAW is implemented in a post-pseudopote-

ntial manner, the extension to a self-consistent PAW calculation should be straightforward. An post-pseudo-potential approach has also been used to evaluate core level spectra (Jayawardane *et al* 2001) and momentum matrix elements (Kageshima and Shiraishi 1997).

Acknowledgement

References

AbInit http://www.mapr.ucl.ac.be/ABINIT/
Andersen O K 1975 *Phys. Rev.* **B12** 3060
Anisimov V I, Zaanen J and Andersen O K 1991 *Phys. Rev.* **B44** 943
Antoncik E 1959 *J. Phys. Chem. Solids* **10** 314
Arnaud B and Alouani M 2000 *Phys. Rev.* **B62** 4464
Bengone O, Alouani M, Blöchl P E and Hugel J 2000 *Phys. Rev.* **B62** 16392
Blöchl P E 1994 *Phys. Rev.* **B50** 17953
Blöchl P E 1995 *J. Chem. Phys.* **103** 7422
Blöchl P E 2000 *Phys. Rev.* **B62** 6158
Blügel S, Kromen W and Schroeder K 2001 *Spring meeting of the German Physical Society 2001*, *Hamburg*, *Germany*
Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
Hamann D R, Schlüter M and Chiang C 1979 *Phys. Rev. Lett.* **43** 1494
Herring C 1940 *Phys. Rev.* **57** 1169
Hobbs D, Kresse G and Hafner J 2000 *Phys. Rev.* **B62** 11556
Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864
Holzwarth N A W, Mathews G E, Dunning R B, Tackett A R and Zheng Y 1997 *Phys. Rev.* **B55** 2005
Jayawardane D N, Pickard C J, Brown L M and Payne M C 2001 *Phys. Rev.* **B64** 115107
Kageshima H and Shiraishi K 1997 *Phys. Rev.* **B56** 143985
Kohn W and Rostocker J 1954 *Phys. Rev.* **94** 1111
Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
Korringa J 1947 *Physica* (*Utrecht*) **13** 392
Kresse G and Joubert J 1999 *Phys. Rev.* **B59** 1758
Louie S G, Froyen S and Cohen M L 1982 *Phys. Rev.* **B26** 1738
Mauri F, Pfrommer B G and Louie S G 1996 *Phys. Rev. Lett.* **77** 5300
Petrilli H M, Blöchl P E, Blaha P and Schwarz K 1998 *Phys. Rev.* **B57** 14690
Phillips J C and Kleinman L 1959 *Phys. Rev.* **116** 287
Pickard C J and Mauri F 2001 *Phys. Rev.* **B63** 245101
Slater J C 1937 *Phys. Rev.* **51** 846
Tackett A R, Holzwarth N A W and Matthews G E 2001 *Comput. Phys. Commun.* **135** 329, 348
Valiev M and Weare J H 1999 *J. Phys. Chem.* **A103** 10588
Vanderbilt D 1990 *Phys. Rev.* **B41** 7892
Van de Walle C G and Blöchl P E 1993 *Phys. Rev.* **B47** 4244
Wang I S Y and Karplus M 1973 *J. Am. Chem. Soc.* **94** 8160
Watson S C and Carter E A 1998 *Phys. Rev.* **B58** R13309
Woo T K, Margl P M, Blöchl P E and Ziegler T 1997 *J. Phys. Chem.* **B101** 7879
Woo T K, Blöchl P E and Ziegler T 2000 *J. Mol. Struct.* (*Theochem.*) **506** 313
Zunger A and Cohen M 1978 *Phys. Rev.* **B18** 5449