Formulation of the Augmented Plane-Wave and Muffin-Tin Orbital Method

Takao Kotani[1], Hiori Kino[2], and Hisazumu Akai[3]

[1]*Department of Applied Mathematics and Physics, Tottori University, Tottori 680-8552, Japan*
[2]*National Institute for Materials Science, Tsukuba, Ibaraki 305-0047, Japan*
[3]*Institute of Solid State Physics, University of Tokyo, Kashiwa, Chiba 277-8581, Japan*

The augmented plane waves and the muffin-tin orbitals method (the PMT method) was proposed by Kotani and van Schilfgaarde in Phys. Rev. B **81**, 125117 (2010). It is a mixed basis all-electron full-potential method, which uses two types of augmented waves simultaneously, in addition to the local orbitals. In this paper, this mixed basis method is reformulated on the basis of a new formalism named as the 3-component formalism, which is a mathematically transparent version of the additive augmentation originally proposed by Soler and Williams in Phys. Rev. B **47**, 6784 (1993). Atomic forces are easily derived systematically. We discuss some problems in the mixed basis method and ways to manage them. In addition, we compare the method with the PAW method on the same footing. This PMT method is the basis for our new development of the quasiparticle self-consistent *GW* method in J. Phys. Soc. Jpn. **83**, 094711 (2014), available as the `ecalj` package at github.

1. Introduction

In the first-principles electronic-structure calculations based on the density functional theory in the LDA/GGA (local density approximation/generalized gradient approximation), a key element is the one-body problem solver, which should have high efficiency, accuracy and robustness. As such solvers, the linearized augmented plane wave (LAPW) method and the linearized muffin-tin orbital (LMTO) method were proposed by Andersen in 1975,[1] followed by many improvements and extensions.[2–7] Today, LAPW and LMTO have developed to be full-potential methods, which we treat in this paper. In these methods, wavefunctions are represented by superpositions of augmented waves. The LAPW uses the augmented plane waves (APWs) made of plane waves (PWs) as envelope functions; the LMTO uses the muffin-tin orbitals (MTOs) made of the atom-centered Hankel functions. Corresponding to these envelope functions, the APWs fit to the extended nature of wavefunctions; in contrast, the MTOs fit to the localized nature of wavefunctions. However, wavefunctions in real materials should have both natures.

This fact is reflected as shortcomings in these methods. In the case of the LAPW, it requires so many bases in order to represent sharp structures of wavefunctions just outside of muffin-tins. For example, $3d$ orbitals of transition metals are typical cases. Most of all the PWs used in the LAPW method are consumed only to reproduce the sharp structures. On the other hand, the LMTO is problematic in representing the extended nature of wavefunctions. For example, we sometimes need to put empty spheres between atoms. In addition, it is not simple to enlarge basis set systematically in order to check numerical convergence.

To overcome the above shortcomings, Kotani and van Schilfgaarde introduced a new linearized method named the APW and MTO method (the PMT method),[8] which is an all-electron (AE) full-potential mixed basis method using APWs and MTOs simultaneously. Because these two types of basis are complementary, we can overcome the shortcomings. To the best of our knowledge, no other mixed basis methods have used different types of augmented waves simultaneously in the full-potential methods.

A minimum description on the formalism of the PMT method is given in Ref. 8, which is based on Ref. 5 for a version of LMTO method by Methfessel et al. However, the formalism was not very transparent, mainly because it was not derived from the explicit total energy minimization. This makes the theoretical treatment of the PMT method somehow complicated. For example, it resulted in a complicated logic to derive atomic forces in Refs. 5 and 9. It was not easy to compare the PMT method with the projector augmented wave (PAW) methods[6,10] on the same footing. Thus, we should give a simple and clear formalism to the PMT method for its further developments rather than that in Refs. 5 and 8.

In this paper, we introduce a formalism, named the 3-component formalism, which is a mathematically transparent generalization of the additive augmentation given by Soler and Williams[11–13] (see discussion in Sect. VII in Ref. 6). We give a formalism of the PMT method based on the 3-component formalism. In the PAW method,[6] the total energy is minimized as a functional of pseudo wavefunctions. In the 3-component formalism, the minimization is formulated as for the wavefunctions represented in the *3-component space* defined in Sect. 2 under some constraints. This is somehow general in the sense that it allows us to use any types of basis (need not to be given by projectors); thus, it is suitable formulating mixed basis methods such as the PMT. Results of the PMT method applied to diatomic molecules from H_2 through Kr_2 are already given in Ref. 14. Considering the fact that the PMT method (even the LMTO itself) is already very good for describing solids,[5,8,15,16] the PMT method can be a candidate to perform full-potential calculations for molecules and solids in a unified manner, more efficiently than LAPW.

Note that we have already implemented the quasiparticle self-consistent *GW* (QSGW) method[15,17,18] in the PMT method;[19] this allows us to apply the QSGW method to a wide range of materials without empty spheres, while not being bothered with the difficulty in setting parameters of MTOs. We have used the QSGW method in Refs. 20 and 21. This kind of method was referred to by Kimes, Katak and Kresse,[22] who claimed that accurate and efficient *GW* calculations should be based on a method that uses both

types of bases (localized and extended bases in space) simultaneously.

In Sect. 2, we will give the 3-component formalism. Functional relations of physical quantities become transparent in such formalism. In Sect. 3, we give the formulation of the PMT method based on the 3-component formalism. Then we discuss problems in the PMT method and ways to overcome them, giving a comparison with the PAW method. The derivation of atomic forces given in Appendix B becomes very transparent in comparison with those given in Refs. 9 and 13.

2. 3-component Formalism

We assume a periodic boundary condition where real space (or unit cell) is specified by Ω. Ω is divided into the muffin-tin (MT) and the interstitial regions. The MTs are located at $\mathbf{R}_a$ with a radius R_a, where a is the index specifying a MT within Ω. $L \equiv (l, m)$ is the angular momentum index. We use units, $\hbar = 1$, the electron mass m_e, and the electron charge e. The spin index is not shown for simplicity.

Here we give the 3-component formalism as a general scheme for the augmented-wave methods, which include any types of augmented waves including the local orbitals.[7]

2.1 3-component space

Any augmented basis $F_i(\mathbf{r})$ consists of three components, where i is the index specifying basis function. $F_i(\mathbf{r})$ consists of the following three components:

(0) Smooth part (= envelope function) $F_{0i}(\mathbf{r})$
(1) True parts $F_{1i,a}(\mathbf{r})$ defined in MTs $|\mathbf{r}| \leq R_a$
(2) Counterparts $F_{2i,a}(\mathbf{r})$ defined in MTs $|\mathbf{r}| \leq R_a$ (canceling the smooth parts within MTs)

We call $F_{0i}(\mathbf{r})$, $F_{01,a}(\mathbf{r})$, and $F_{02,a}(\mathbf{r})$ as the 0th, 1st, and 2nd components of $F_i(\mathbf{r})$, respectively. The $F_{0i}(\mathbf{r})$ should be an analytic function in space or a linear combination of analytic functions. In the PMT method, $F_{0i}(\mathbf{r})$ are the PWs or the Bloch sums of the Hankel functions; specifically, we use atom-centered smooth Hankel functions (smHankels) instead of conventional Hankel functions, to avoid divergence at its center[5,23] [see Eq. (37)]. $F_{1i,a}(\mathbf{r})$ and $F_{2i,a}(\mathbf{r})$ are defined only at $|\mathbf{r}| \leq R_a$ (in the cases below, we sometimes take that these are zero at $|\mathbf{r}| > R_a$). In the sense that $F_{0i}(\mathbf{r})$ is analytic and the 1st and 2nd components are given on a dense radial mesh, the basis $F_i(\mathbf{r})$ is specified without any numerical inaccuracy.

$F_i(\mathbf{r})$ is a member in the *3-component space*, which is defined as a direct sum of linear spaces corresponding to the components (0), (1), and (2). Thus, F_i can be expressed as $F_i = \{F_{0i}(\mathbf{r}), \{F_{1i,a}(\mathbf{r})\}, \{F_{2i,a}(\mathbf{r})\}\}$ (curly brackets mean a set), where $F_{1i} \equiv \{F_{1i,a}(\mathbf{r})\}$ means a set as for the MT index a, and F_{2i} as well. However, in the following, we use slightly different expression instead:

$$F_i(\mathbf{r}) = F_{0i}(\mathbf{r}) \oplus \{F_{1i,a}(\mathbf{r})\} \ominus \{F_{2i,a}(\mathbf{r})\}$$
$$= F_{0i}(\mathbf{r}) \oplus F_{1i}(\mathbf{r}) \ominus F_{2i}(\mathbf{r}). \quad (1)$$

This makes the following expressions easy to read without any difference in their meanings. The symbols $\oplus$ and $\ominus$ mean nothing more than separators. We call a member in the 3-component space as a 3-component function in the following. Wavefunctions are also given as 3-component

functions. With the coefficients $\{\alpha_p^i\}$, wavefunctions can be written as

$$\psi_p(\mathbf{r}) = \sum_i \alpha_p^i F_i(\mathbf{r}), \quad (2)$$

where linear combinations are taken for each component. We represent electron density and so on as a 3-component function as well.

Note that the 3-component space is a mathematical construction, a model space: we have to specify how to map a 3-component function to a function in real space. For this purpose, we define $\mathcal{A}$-mapping (augmentation mapping) from a 3-component function to a function in real space,

$$\mathcal{A}[\psi_p(\mathbf{r})] \equiv \psi_{0p}(\mathbf{r}) + \sum_a \psi_{1p,a}(\mathbf{r} - \mathbf{R}_a)$$
$$- \sum_a \psi_{2p,a}(\mathbf{r} - \mathbf{R}_a). \quad (3)$$

This is nothing but a conventional augmentation where physically meaningful wavefunctions $\psi_p(\mathbf{r})$ should satisfy the following conditions:

(A) Within MTs ($|\mathbf{r}| < R_a$), $\psi_{2p,a}(\mathbf{r}) = \psi_{0p}(\mathbf{r} + \mathbf{R}_a)$.
(B) At MT boundaries ($|\mathbf{r}| = R_a$), $\psi_{1p,a}(\mathbf{r})$ and $\psi_{2p,a}(\mathbf{r})$ should have the same value and slope.

If (A) is satisfied, the contribution from ψ_{0p} within MTs perfectly cancels those of $\psi_{2p,a}$ in Eq. (3). The total energy in the DFT is given as a functional of eigenfunctions as $E[\{\psi_p(\mathbf{r})\}]$, where $\{\psi_p(\mathbf{r})\}$ are for occupied states. Our problem is to minimize this under the constraint of orthogonality of $\psi_p(\mathbf{r})$ with conditions (A) and (B) on $\{\psi_p(\mathbf{r})\}$. Local orbitals[7] are also treated as 3-component functions whose 0th and 2nd components are zero overall.

In conventional LAPW (e.g., see Refs. 3 and 4), (A) and (B) are very accurately satisfied. The 2nd component almost completely satisfies (A) with the use of spherical Bessel functions. The 1st component are expanded with very high l ($\gtrsim 8$) cutoff. Thus, the LAPW can be quite accurate. However, it can be expensive. (We also have null-vector problem. See Sect. 3.1.)

Thus, Soler and Williams[11] introduced additive augmentation. To make calculations efficient, we use condition (A′) as a relaxed version of condition (A):

(A′) Within MTs ($|\mathbf{r}| \leq R_a$), $\psi_{2p,a}(\mathbf{r}) \approx \psi_{0p}(\mathbf{r} + \mathbf{R}_a)$.

Then, we expect high-energy (high-frequency) contributions of eigenfunctions not included in the 1st or 2nd components to be accounted for by the 0th component. In practice, we can use a low l cutoff $\lesssim 4$ for both of the 1st and 2nd components. The LAPW package HiLAPW, developed by Oguchi et al.,[24] implemented a procedure of evaluating physical quantities from the basis given by Eq. (3) with condition (A′).

However, it is complicated to evaluate all quasilocal products such as the density and kinetic-energy density from $\mathcal{A}[F_i^*(\mathbf{r})]\mathcal{A}[F_j(\mathbf{r}')]$, since it contains cross terms that connect different components. Thus, Soler and Williams[11] gave a prescription to avoid the evaluation of the cross terms. With $\mathcal{A}$-mapping applied not to wavefunctions but to their products as in Sect. 2.2, we have a separable form of the total energy and all other physical quantities (no cross terms between components). This is based on the fact that the total energy in the separable form should be in agreement with the

true total energy only when conditions (A) and (B) are satisfied. As we see in the following, it is a good approximation to use (A′) instead of (A).

The above two important concepts, i.e., the additive augmentation and the separable form, were used in both LMTO and PAW.[5,6,10] They were originally introduced in Ref. 11.

Let us consider how to determine $F_{1i,a}$ and $F_{2i,a}$ for a given F_{0i}. As for $F_{2i,a}$, (A′) means that F_{0i} should be reproduced well within MTs. Generally, $F_{2i,a}(\mathbf{r})$ can be represented as

$$F_{2i,a}(\mathbf{r}) \equiv \sum_{k,L} C_{akL}^{i} P_{akL}(\mathbf{r}), \qquad (4)$$

where k is the index for the radial degree of freedom. We introduce the truncation parameters $k_{\max,a}$ and $l_{\max,a}$; we assume that the sum in Eq. (4) is taken for $k \leq k_{\max,a}$ and $l \leq l_{\max,a}$; when $k_{\max,a}$ and $l_{\max,a}$ become infinite, we assume that condition (A) is satisfied. Even when these truncation parameters are finite, $F_{2i,a}$ should reproduce the low-energy (low-frequency) parts of F_{0i} well. The set $\{P_{akL}(\mathbf{r})\}$ can contain any functions in principle; as explained in Sect. 3, the parts of smHankel functions within their central MT spheres are treated as members of $\{P_{akL}(\mathbf{r})\}$.[25] $F_{1i,a}(\mathbf{r})$ is given from Eq. (4) with the replacement of $P_{akL}(\mathbf{r})$ with $\widetilde{P}_{akL}(\mathbf{r})$. Here, $\widetilde{P}_{akL}(\mathbf{r})$ is a linear combination of partial waves so as to have the same value and slope with $P_{akL}(\mathbf{r})$ at $|\mathbf{r}| = R_a$. With this replacement, we have

$$F_{1i,a}(\mathbf{r}) = \sum_{k,L} C_{akL}^{i} \widetilde{P}_{akL}(\mathbf{r}). \qquad (5)$$

2.2 Augmentation for product of 3-component functions

Let us give a prescription for evaluating physical quantities for wavefunctions satisfying conditions (A′) and (B). First, we define the diagonal product of 3-component functions as

$$F_i^*(\mathbf{r})F_j(\mathbf{r}') \equiv F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r}') \oplus \{F_{1i,a}^*(\mathbf{r})F_{1j,a}(\mathbf{r}')\}$$

$$\ominus \{F_{2i,a}^*(\mathbf{r})F_{2j,a}(\mathbf{r}')\}, \qquad (6)$$

where we have no cross terms between different components. We apply $\mathcal{A}$-mapping in Eq. (3) to this product as

$$\mathcal{A}[F_i^*(\mathbf{r})F_j(\mathbf{r}')]$$

$$= F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r}') + \sum_a F_{1i,a}^*(\mathbf{r} - \mathbf{R}_a)F_{1j,a}(\mathbf{r}' - \mathbf{R}_a)$$

$$- \sum_a F_{2i,a}^*(\mathbf{r} - \mathbf{R}_a)F_{2j,a}(\mathbf{r}' - \mathbf{R}_a). \qquad (7)$$

We will use $\mathcal{A}[F_i^*(\mathbf{r})F_j(\mathbf{r}')]$ to evaluate quasilocal products when (A′) is satisfied. Since any one-body quantities such as the inner product, electron density, current and so on, are quasilocal, we can evaluate these from $\mathcal{A}[\psi_p^*(\mathbf{r})\psi_{p'}(\mathbf{r}')]$. Generally, we can evaluate matrix elements of a quasilocal operator $X(\mathbf{r}, \mathbf{r}')$ in real space from the 3-component wavefunctions $\psi_p(\mathbf{r})$ in separable form as

$$\langle \psi_p | X | \psi_{p'} \rangle = \int d^3 r\, d^3 r'\, X(\mathbf{r}, \mathbf{r}') \mathcal{A}[\psi_p^*(\mathbf{r})\psi_{p'}(\mathbf{r}')]. \qquad (8)$$

We can read this as a transformation of X to the corresponding operator in the 3-component space.

On the basis of the above prescription, we can define the inner product $\langle \psi_p | \psi_{p'} \rangle$ as $\langle \psi_p | \psi_{p'} \rangle = \sum_{i,j} \alpha_p^{i*} \alpha_{p'}^j O_{ij}$. Here, the overlap matrix O_{ij} is

$$O_{ij} \equiv \langle F_i | F_j \rangle \equiv \int_\Omega d^3 r\, \mathcal{A}[F_i^*(\mathbf{r})F_j(\mathbf{r})]$$

$$= \int_\Omega d^3 r\, F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r}) + \sum_a \int_{|\mathbf{r}| \leq R_a} d^3 r\, F_{1i,a}^*(\mathbf{r})F_{1j,a}(\mathbf{r})$$

$$- \sum_a \int_{|\mathbf{r}| \leq R_a} d^3 r\, F_{2i,a}^*(\mathbf{r})F_{2j,a}(\mathbf{r}). \qquad (9)$$

This can be read as a definition of the inner product in the 3-component space. For a given finite basis set, we can expect that O_{ij} should be positive definite as long as truncation parameters are sufficiently large. The kinetic energy is given by $\rho_{ij} = \sum_p^{\text{occ}} \alpha_p^{i*} \alpha_p^j$ (occ means the sum for occupied states) as $E_k = \sum_{i,j} \rho_{ij} T_{ij}$. Here, the kinetic-energy matrix T_{ij} is given as

$$T_{ij} \equiv \frac{\langle \nabla F_i | \nabla F_j \rangle}{2m_e} \equiv \frac{1}{2m_e} \int_\Omega d^3 r (\nabla_\mathbf{r} \nabla_{\mathbf{r}'} \mathcal{A}[F_i^*(\mathbf{r})F_j(\mathbf{r}')])_{\mathbf{r}=\mathbf{r}'}$$

$$= \frac{1}{2m_e} \int_\Omega d^3 r\, \mathcal{A}[\nabla F_i^*(\mathbf{r})\nabla F_j(\mathbf{r})]$$

$$= \int_\Omega d^3 r\, \frac{\nabla F_{0i}^*(\mathbf{r})\nabla F_{0j}(\mathbf{r})}{2m_e}$$

$$+ \sum_a \int_{|\mathbf{r}| \leq R_a} d^3 r\, \frac{\nabla F_{1i,a}^*(\mathbf{r})\nabla F_{1j,a}(\mathbf{r})}{2m_e}$$

$$- \sum_a \int_{|\mathbf{r}| \leq R_a} d^3 r\, \frac{\nabla F_{2i,a}^*(\mathbf{r})\nabla F_{2j,a}(\mathbf{r})}{2m_e}. \qquad (10)$$

Partial integration gives $T_{ij} = \langle F_i | \frac{-\nabla^2}{2m_e} | F_j \rangle$, since $F_{1i,a}$ and $F_{2i,a}$ have the same value and slope at the MT boundaries. This kinetic energy operator is interpreted as $T = \frac{-\nabla^2}{2m_e} \oplus \{\frac{-\nabla^2}{2m_e}\} \ominus \{\frac{-\nabla^2}{2m_e}\}$ in the 3-component space.

The one-body problem for a given one-particle potential $V(\mathbf{r})$ in real space is translated into a problem in the 3-component space for the Hamiltonian $H = T + V$ under condition (A) or (A′), where $V = V_0 \oplus \{V_{1,a}\} \ominus \{V_{2,a}\}$. Here $V_0(\mathbf{r}) = V(\mathbf{r})$, and $V_{1,a}(\mathbf{r}) = V_{2,a}(\mathbf{r}) = V(\mathbf{r} + \mathbf{R}_a)$ within MTs at $\mathbf{R}_a$. However, we can add any extra potential $\Delta \bar{V}$ simultaneously to both V_0 and $V_{2,a}$ if (A) is completely satisfied.

We have an error because we use Eq. (7) instead of Eq. (3): high-energy contributions contained in the 0th components are not exactly evaluated. However, the error can be small enough to be neglected, as discussed in Appendix A. This error is also related to the question of how can the optimum $\Delta \bar{V}$ be chosen so as to minimize the error. In fact, the success of the PAW[6] is dependent on the choice of $\Delta \bar{V}$, as seen in Sect. 3.2.

The valence electron density n as the 3-component function is given by

$$n = n_0 \oplus n_1 \ominus n_2 = n_0 \oplus \{n_{1,a}\} \ominus \{n_{2,a}\}$$

$$= \sum_{ij} \rho_{ij} F_i^* F_j = \sum_{ij} \rho_{ij} F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r})$$

$$\oplus \left\{ \sum_{ij} \rho_{ij} F_{1i,a}^*(\mathbf{r})F_{1j,a}(\mathbf{r}) \right\} \ominus \left\{ \sum_{ij} \rho_{ij} F_{2i,a}^*(\mathbf{r})F_{2j,a}(\mathbf{r}) \right\}. \qquad (11)$$

We can calculate the Coulomb interaction from $\mathcal{A}[n]$. However, to reduce computational effort, we will also make the Coulomb interaction into a separable form, as shown in Sect. 2.4, with the help of the multipole technique by

Weinert.[26] In Sects. 2.3 and 2.5, we give some preparations for defining the Coulomb interaction in Sect. 2.4.

The total energy should be given as a functional of eigenfunctions in the first-principle calculations, not just as a functional of the coefficients $\{\alpha_p^j\}$. This is important in some cases. For example, it is necessary to know how the change in the basis set affects the total energy when we calculate atomic forces. These are related to the so-called Pulay terms.[27]

2.3 Multipole transformation

In order to define Coulomb interaction in Sect. 2.4, we introduce the multipole transformation ($\mathcal{M}$-transformation) for the 3-component functions. This corresponds to the compensation charges in Ref. 6.

Before defining the $\mathcal{M}$-transformation, we define the gaussian projection $\mathcal{G}_a[f(\mathbf{r})]$ as follows. The projection $\mathcal{G}_a[f(\mathbf{r})]$ is defined for the function $f(\mathbf{r})$ for $|\mathbf{r}| \leq R_a$ as

$$\mathcal{G}_a[f(\mathbf{r})] = \sum_L Q_{aL}[f] G_{aL}(\mathbf{r}), \qquad (12)$$

$$G_{aL}(\mathbf{r}) = \frac{1}{N_{aL}} \exp\left(-\left(\frac{r}{R_{G,a}}\right)^2\right) Y_L(\hat{\mathbf{r}}), \qquad (13)$$

where $Q_{aL}[f] = \int_{|\mathbf{r}| \leq R_a} \mathcal{Y}_L(\mathbf{r}) f(\mathbf{r}) d^3 r$ gives the L-th multipole moment of $f(\mathbf{r})$. Here, $\mathcal{Y}_L(\mathbf{r}) \equiv r^l Y_L(\hat{\mathbf{r}})$. $Y_L(\hat{\mathbf{r}})$ is the real spherical harmonics, where $\hat{\mathbf{r}}$ is the normalized $\mathbf{r}$. N_{aL} is a normalization factor so that $G_{aL}(\mathbf{r})$ has a normalized multipole moment. $R_{G,a}$ in Eq. (13) is chosen to be sufficiently small so that $G_{aL}(\mathbf{r})$ is negligible for $|\mathbf{r}| \geq R_a$ [see Eq. (25) in Ref. 5]. This $\mathcal{G}_a[f(\mathbf{r})]$ is a superposition of gaussians $G_{aL}(\mathbf{r})$ while keeping the multipole moments of $f(\mathbf{r})$. We can take a rather small $R_{G,a}$ without the loss of numerical accuracy; it is possible to take the limit $R_{G,a} \to 0$ because the quantities involved in $G_{aL}(\mathbf{r})$ are evaluated analytically or numerically accurately on a dense radial mesh.

We now define $\mathcal{M}$-transformation for the 3-component density $n = n_0 \oplus n_1 \ominus n_2$ as

$$\mathcal{M}[n] = n_0(\mathbf{r}) + \sum_{a,\mathbf{T},L} Q_{aL}[n_{1,a} - n_{2,a}] G_{aL}(\mathbf{r} - \mathbf{R}_a - \mathbf{T})$$

$$\oplus n_1 \ominus \left\{ n_{2,a}(\mathbf{r}) + \sum_L Q_{aL}[n_{1,a} - n_{2,a}] G_{aL}(\mathbf{r}) \right\}. \qquad (14)$$

Thus, $\mathcal{M}[n]$ adds the same gaussians to both the 0th and 2nd components. $\mathbf{T}$ is the translational vectors of Ω. With this transformation, the multipole moments of the 1st and 2nd components become the same. Note that the $\mathcal{M}$-transformation is not a physically meaningful transformation because $\mathcal{A}[\mathcal{M}[n]] = \mathcal{A}[n]$. With this transformation, the interstitial electrostatic potential calculated from the 0th component of Eq. (14) should be the same as that calculated from $\mathcal{A}[n]$.

2.4 Coulomb interaction

In principle, we can define the Coulomb interaction between $n(\mathbf{r}) = n_0 \oplus n_1 \ominus n_2$ and $m(\mathbf{r}) = m_0 \oplus m_1 \ominus m_2$ from the densities $\mathcal{A}[n]$ and $\mathcal{A}[m]$. We can use $\mathcal{A}[\bar{n}]$ instead of $\mathcal{A}[n]$, where we define $\bar{n}$ as the $\mathcal{M}$-transformed n, i.e., $\bar{n} \equiv \mathcal{M}[n]$, $\bar{m} \equiv \mathcal{A}[m]$ as well. By the nature of $\mathcal{M}$-transformation, we have $\mathcal{A}[\bar{n}] = \mathcal{A}[n]$. Thus, the Coulomb interaction $(n|v|m)_{\text{original}}$ is given as

$(n|v|m)_{\text{original}}$
$$= \sum_{\mathbf{T}} \int_\Omega d^3 r\, d^3 r'\, \mathcal{A}[\bar{n}(\mathbf{r})] v(\mathbf{r} - \mathbf{r}' + \mathbf{T}) \mathcal{A}[\bar{m}(\mathbf{r}')]. \qquad (15)$$

Here, $v(\mathbf{r}) = e^2/|\mathbf{r}|$; $\sum_{\mathbf{T}}$ implicitly includes the division by the number of cells. Equation (15) cannot be easily evaluated because $v(\mathbf{r} - \mathbf{r}' + \mathbf{T})$ contains the cross terms that connect the 0th component with other components.

Thus, we use the approximation

$$(n|v|m) \equiv \mathcal{M}[n] \cdot v \cdot \mathcal{M}[m] = \bar{n} \cdot v \cdot \bar{m}, \qquad (16)$$

instead of Eq. (15), where the dot operator for the 3-component functions is given as

$$\bar{n} \cdot v \cdot \bar{m} \equiv \bar{n}_0 \bullet v \bullet \bar{m}_0 + \bar{n}_1 \circ v \circ \bar{m}_1 - \bar{n}_2 \circ v \circ \bar{m}_2 \qquad (17)$$

$$\bar{n}_0 \bullet v \bullet \bar{m}_0 \equiv \sum_{\mathbf{T}} \int_\Omega d^3 r\, d^3 r'$$
$$\times \bar{n}_0(\mathbf{r}) v(\mathbf{r} - \mathbf{r}' + \mathbf{T}) \bar{m}_0(\mathbf{r}'), \qquad (18)$$

$$\bar{n}_1 \circ v \circ \bar{m}_1 \equiv \sum_a \int_{|\mathbf{r}| \leq R_a} d^3 r \int_{|\mathbf{r}'| \leq R_a} d^3 r'$$
$$\times \bar{n}_{1,a}(\mathbf{r}) v(\mathbf{r} - \mathbf{r}') \bar{m}_{1,a}(\mathbf{r}'), \qquad (19)$$

$$\bar{n}_2 \circ v \circ \bar{m}_2 \equiv \sum_a \int_{|\mathbf{r}| \leq R_a} d^3 r \int_{|\mathbf{r}'| \leq R_a} d^3 r'$$
$$\times \bar{n}_{2,a}(\mathbf{r}) v(\mathbf{r} - \mathbf{r}') \bar{m}_{2,a}(\mathbf{r}'). \qquad (20)$$

Note that $X \bullet Y$ means the integral over Ω, whereas $X \circ Y$ means the integrals within MTs.

Let us evaluate the difference between Eq. (15) and Eq. (16). This can be evaluated with the identity in Appendix A as

$(n|v|m)_{\text{original}} - (n|v|m)$
$$= \sum_a \int_{|\mathbf{r}| \leq R_a} d^3 r \int_{|\mathbf{r}'| \leq R_a} d^3 r' ((\bar{n}_0(\mathbf{r}) - \bar{n}_2(\mathbf{r}))$$
$$\times v(\mathbf{r} - \mathbf{r}')(\bar{m}_1(\mathbf{r}') - \bar{m}_2(\mathbf{r}'))$$
$$+ (\bar{n}_1(\mathbf{r}) - \bar{n}_2(\mathbf{r})) v(\mathbf{r} - \mathbf{r}')(\bar{m}_0(\mathbf{r}') - \bar{m}_2(\mathbf{r}'))). \qquad (21)$$

This is essentially the same as Eq. (13) in Ref. 10. In Eq. (21), the difference consists of contributions of MT sites without terms connecting different MT sites. This is because $\bar{n}_{1,a}(\mathbf{r})$ and $\bar{n}_{2,a}(\mathbf{r})$ have the same multipole moments. Since $\bar{n}_0(\mathbf{r}') - \bar{n}_2(\mathbf{r})$ is a high-l or highly oscillating part, and $\bar{n}_{1,a}(\mathbf{r}) - \bar{n}_{2,a}(\mathbf{r})$ has zero multipole moments and zero at MT boundaries, we expect that the separable form of Eq. (16) should be justified. We can check this by changing the truncation parameters $l_{\text{max},a}$ and $k_{\text{max},a}$.

From Eq. (16), we have the expression of the Coulomb interaction as

$$(F_i^* F_j | v | F_{i'}^* F_{j'}) = \mathcal{M}[F_i^* F_j] \cdot v \cdot \mathcal{M}[F_{i'}^* F_{j'}]. \qquad (22)$$

Here, $F_i^* F_j$ is the diagonal product defined in Eq. (6) at $\mathbf{r} = \mathbf{r}'$. In the calculations such as these arising in the GW approximations,[15] we have to evaluate this as accurately as possible so that the exchange-pair cancellation is kept well.

2.5 Frozen core approximation

We use the frozen core approximation; the charge density due to the cores is evaluated by a superposition of rigid cores as follows.[5] Then we need to treat the spillout of the core density outside of MTs explicitly.

First, we perform a self-consistent atomic calculation under the spherical approximation without a spin polarization to obtain its core density $n_a^c(\mathbf{r})$. Then we make a fitted density $n_{\text{sH},a}^c(\mathbf{r})$ given by a linear combination of several smHankel functions so that $n_{\text{sH},a}^c(\mathbf{r})$ reproduces $n_a^c(\mathbf{r})$ for $|\mathbf{r}| > R_a$ numerically accurately (we can use other types of analytic functions such as gaussians instead of smHankel functions).

Thus, we have the expression of the sum of the all the core electron densities with adding the contributions of nucleus $-Z_a\delta(\mathbf{r})$:

$$n^{\text{Zc}} = \sum_{a,\mathbf{T}} n_{\text{sH},a}^c(\mathbf{r} - \mathbf{R}_a - \mathbf{T})$$
$$\oplus \{n_a^c(\mathbf{r}) - Z_a\delta(\mathbf{r})\} \ominus \{n_{\text{sH},a}^c(\mathbf{r})\}. \quad (23)$$

Applying the $\mathcal{M}$-transformation to n^{Zc} gives

$$\mathcal{M}[n^{\text{Zc}}] = \sum_{a,\mathbf{T}} \left(n_{\text{sH},a}^c(\mathbf{r} - \mathbf{R}_a - \mathbf{T}) \right.$$
$$\left. + \sum_L Q_{aL}^{\text{Zc}} G_{aL}(\mathbf{r} - \mathbf{R}_a - \mathbf{T}) \right) \oplus \{n_a^c(\mathbf{r}) - Z_a\delta(\mathbf{r})\}$$
$$\ominus \left\{ n_{\text{sH},a}^c(\mathbf{r}) + \sum_L Q_{aL}^{\text{Zc}} G_{aL}(\mathbf{r}) \right\}, \quad (24)$$

$$Q_{aL}^{\text{Zc}} = Q_{aL}[n_1^{\text{Zc}} - n_2^{\text{Zc}}]$$
$$= Q_{aL}[n_a^c(\mathbf{r}) - Z_a\delta(\mathbf{r}) - n_{\text{sH},a}^c(\mathbf{r})]. \quad (25)$$

2.6 Total energy in density functional

Let us give the total energy E_{total} for the DFT, and construct the Kohn–Sham equation from it. With the kinetic energy $E_k = \frac{1}{2m_e}\sum_{ij}\rho_{ij}\langle\nabla F_i|\nabla F_j\rangle$ from Eq. (10), the total energy is given as

$$E_{\text{total}} = E_k^{\text{core}} + E_k + E_{\text{es}} + E_{\text{xc}}, \quad (26)$$

where E_k^{core} is the kinetic energy of frozen cores as a constant. E_{es} and E_{xc} are the electrostatic and exchange–correlation energies, respectively. E_{es} is given as the electrostatic energy for the total density $n^{\text{Zcv}} = n^{\text{Zc}} + n$, which are given in Eqs. (11) and (23).

Based on the definition given in Eq. (22), we have

$$E_{\text{es}} = \frac{1}{2}(n^{\text{Zcv}}|v|n^{\text{Zcv}}) = \frac{1}{2}\mathcal{M}[n^{\text{Zcv}}] \cdot v \cdot \mathcal{M}[n^{\text{Zcv}}], \quad (27)$$

where a constant due to the self-interaction of a nucleus is implicitly removed. The components of $\bar{n}^{\text{Zcv}}(\mathbf{r}) = \mathcal{M}[n^{\text{Zcv}}]$ are given as

$$\bar{n}_0^{\text{Zcv}}(\mathbf{r}) = n_0^{\text{Zc}}(\mathbf{r}) + \sum_{a,L,\mathbf{T}}(Q_{aL}^{\text{Zc}} + Q_{aL}^{\text{v}})G_{aL}(\mathbf{r} - \mathbf{R}_a - \mathbf{T}) + n_0(\mathbf{r}),$$
$$\quad (28)$$

$$\bar{n}_{1,a}^{\text{Zcv}}(\mathbf{r}) = n_{1,a}^{\text{Zc}}(\mathbf{r}) + n_{1,a}(\mathbf{r}), \quad (29)$$

$$\bar{n}_{2,a}^{\text{Zcv}}(\mathbf{r}) = n_{2,a}^{\text{Zc}}(\mathbf{r}) + \sum_L(Q_{aL}^{\text{Zc}} + Q_{aL}^{\text{v}})G_{aL}(\mathbf{r}) + n_{2,a}(\mathbf{r}), \quad (30)$$

where $Q_{aL}^{\text{v}} = Q_{aL}[n_{1,a} - n_{2,a}]$. We expand $F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r})$ of n_0 in $\{e^{i\mathbf{Gr}}\}$ (to obtain coefficients, $F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r})$ is tabulated on a real-space mesh, then it is Fourier-transformed). The cutoff of $\mathbf{G}$ is specified by $E_{\text{MAX}}^{\text{rmesh}}$. Then, the 0th component in Eq. (28) is represented by the sum of analytic functions. Thus, we can finally calculate $\frac{1}{2}\bar{n}_0^{\text{Zcv}}(\mathbf{r}) \bullet v \bullet \bar{n}_0^{\text{Zcv}}(\mathbf{r})$ in E_{es}. Thus the terms between gaussians located at different MT sites are evaluated

with the Ewald sum treatment. The term related to MTs in E_{es} is $\frac{1}{2}\bar{n}_1^{\text{Zcv}} \circ w \circ \bar{n}_1^{\text{Zcv}} - \frac{1}{2}\bar{n}_2^{\text{Zcv}} \circ w \circ \bar{n}_2^{\text{Zcv}}$, which is calculated on a radial mesh accurately.

The exchange correlation term can be defined as

$$E_{\text{xc}}[n^{\text{Zcv}}] = E_{\text{xc}}[n_0^{\text{Zcv}}] + \sum_a E_{\text{xc}}[n_{1,a}^{\text{Zcv}}] - \sum_a E_{\text{xc}}[n_{2,a}^{\text{Zcv}}]. \quad (31)$$

The functional derivative of $E_{\text{xc}}[n^{\text{Zcv}}]$ with respect to each component of n^{Zcv} gives

$$v^{\text{xc}} = v_0^{\text{xc}}(\mathbf{r}) \oplus \{v_{1,a}^{\text{xc}}(\mathbf{r})\} \ominus \{v_{2,a}^{\text{xc}}(\mathbf{r})\}. \quad (32)$$

To determine the ground state, E_{total} should be minimized under the orthogonality of eigenfunctions with the constraints (A′) and (B). This ends up with $\delta\psi_p^* \cdot (H - \epsilon_p) \cdot \psi_p = 0$ for the variation $\delta\psi_p^*$ that satisfies (A′) and (B). Here, the operator $H = T + V$ is given as

$$T = \frac{-\nabla^2}{2m_e} \oplus \left\{\frac{-\nabla^2}{2m_e}\right\} \ominus \left\{\frac{-\nabla^2}{2m_e}\right\} \quad (33)$$

$$V = \bar{n}_0^{\text{Zcv}} \bullet v \bullet + v_0^{\text{xc}}$$

$$\oplus \left\{\sum_L \mathcal{Q}_{aL}^{\text{v}}\mathcal{Y}_L(\mathbf{r}) + \bar{n}_{1,a}^{\text{Zcv}} \circ w \circ + v_{1,a}^{\text{xc}}\right\}$$

$$\ominus \left\{\sum_L \mathcal{Q}_{aL}^{\text{v}}\mathcal{Y}_L(\mathbf{r}) + \bar{n}_{2,a}^{\text{Zcv}} \circ w \circ + v_{2,a}^{\text{xc}}\right\}, \quad (34)$$

$$\mathcal{Q}_{aL}^{\text{v}} \equiv \frac{\partial E_{\text{es}}}{\partial Q_{aL}^{\text{v}}} = \bar{n}_0^{\text{Zcv}} \bullet v \bullet G_{aL}(\mathbf{r}' - \mathbf{R}_a)$$
$$- \bar{n}_{2,a}^{\text{Zcv}} \circ w \circ G_{aL}(\mathbf{r}'), \quad (35)$$

where $\bar{n}_0 \bullet v\bullet$ means an integral on a variable, resulting in a function of $\mathbf{r}$.

When the basis set $\{F_j(\mathbf{r})\}$ satisfying (A′) and (B) is fixed, we just need to consider the variation with respect to α_p^{i*} in Eq. (2). Then, we have

$$\sum_j (H_{ij} - \epsilon_p O_{ij})\alpha_p^j = 0, \quad (36)$$

where $H_{ij} = \langle F_i|H|F_j\rangle = \langle F_i|\frac{-\Delta}{2m} + V|F_j\rangle = T_{ij} + V_{ij}$. $V_{ij} = \langle F_i|V|F_j\rangle = V \cdot F_i^*F_j$. Then, the total energy minimization results in the eigenvalue problem. The matrix elements O_{ij}, T_{ij} and V_{ij} are given in Appendix C.

The formula for evaluating atomic forces ise given in Appendix B. It is directly evaluated from the variation on the total energy. This procedure is considerably more simplified than that given in Refs. [5] and [9].

3. PMT Method

Let us give the PMT method based on the 3-component formalism in Sect. 2. We first need to specify a basis set $\{F_i\}$. In the PMT, $\{F_i\}$ is classified into three types of subsets as follows:

(a) APW. We augment the PW by the method explained afterwards.
(b) MTO. We augment the atom-centered smHankel functions.
(c) Local orbital (Lo).[28] We use this to represent some degree of freedom in MTs, such as semicore states. The envelope function of Lo is zero overall.

The smHankel function, as the envelope function of MTO, is first introduced by Methfessel.[5,23] The spherical smHankel

function $h_0(\mathbf{r})$ (for $l = 0$) is defined by the Helmholtz equation with a gaussian source term $g_0(\mathbf{r}) = C \exp(-r^2/R_{\mathrm{SM}}^2)$ [see Eq. (5) in Ref. 5] instead of δ-function;

$$(\nabla^2 + \epsilon)h_0(\mathbf{r}) = -4\pi g_0(\mathbf{r}), \qquad (37)$$

where $C = 1/(\sqrt{\pi} R_{\mathrm{SM}})^3$ is the normalization constant. $\epsilon = -\kappa^2$ is the negative energy for specifying the asymptotic damping behavior of $h_0(\mathbf{r})$. In the limit $R_{\mathrm{SM}} \to 0$ where $g_0(\mathbf{r})$ becomes δ-function (as a point charge), $h_0(\mathbf{r})$ becomes the Hankel function $h_0(\mathbf{r}) = \exp(-\kappa r)/r$. Since the source term is smeared with the radius R_{SM}, we have no divergent behavior at $r = 0$ anymore; the smHankel bends over at $\sim R_{\mathrm{SM}}$ (see Fig. 1 in Ref. 5). From $h_0(\mathbf{r})$, we can make $h_L(\mathbf{r}) \equiv \mathcal{Y}_L(-\nabla)h_0(\mathbf{r})$ for any L. (Recall $\mathcal{Y}_L(\mathbf{r}) = r^l Y_L(\hat{\mathbf{r}})$.) $\mathcal{Y}_L(-\nabla)$ means to substitute $\mathbf{r}$ in $\mathcal{Y}_L(\mathbf{r})$ with $-\nabla$. See Ref. 23 for details.

To augment the PW, that is, to determine the 2nd component from PW as the 0th component, we expand the PW within MTs into the Laguerre polynomial.[8] Any function $f(\mathbf{r})$ (PW in this case) is expanded within a MT $|\mathbf{r} - \mathbf{R}_a| \le R_a$ as

$$f(\mathbf{r}) = \sum_{k,l} C_{akL}[f] P_{akL}(\mathbf{r} - \mathbf{R}_a), \qquad (38)$$

$$P_{akL}(\mathbf{r}) = p_{akl}(r) Y_L(\hat{\mathbf{r}}), \qquad (39)$$

where $k = 0, 1, 2, \ldots$ denotes the order of the polynomial $p_{akl}(r)$. In the case that $f(\mathbf{r})$ is a PW, the coefficients for the function $C_{akL}[f]$ are given analytically.[23]

When we use smHankel centered at $\mathbf{R}_a$ as the envelope function $f(\mathbf{r})$, we have the head part, which is $f(\mathbf{r}) = h_L(\mathbf{r} - \mathbf{R}_a)$ for $|\mathbf{r} - \mathbf{R}_a| \le R_a$, and the tail part, which is in other MT sites $|\mathbf{r} - \mathbf{R}_{a'}| \le R_{a'}$. As for the tail part, we use the expansion of Eq. (38) as in the case of PW. On the other hand, we use the head part as it is;[25] this can be taken into account in the formalism if the set $\{P_{akL}(\mathbf{r})\}$ contains not only the Laguerre polynomials but also $h_L(\mathbf{r})$ as its members.

After specifying $\{P_{akL}(\mathbf{r})\}$, we can determine the corresponding $\{\widetilde{P}_{akL}(\mathbf{r})\}$ as a linear combination of $\phi_{al}(r)Y_L(\hat{\mathbf{r}})$ and $\dot{\phi}_{al}(r)Y_L(\hat{\mathbf{r}})$, where the partial waves $\phi_{al}(r)$ and their energy derivatives $\dot{\phi}_{al}(r)$ are given as the solutions of the radial Schrödinger equation for the spherically-averaged potential of $V_{1,a}$ in Eq. (34), where the energies E_{al} for solving the equation are given as the center of gravities of the occupied states of the partial density of states of the al component; thus, $\phi(r)$ and $\dot{\phi}(r)$ are not with the subscripts aL but with al. This prescription for determining $\{\widetilde{P}_{akL}(\mathbf{r})\}$ can be taken as a quasi-minimization procedure, from the view point of total energy minimization.

As for the al with Lo, we need another partial wave $\phi_{al}^{\mathrm{Lo}}(r)$ corresponding to Lo. When Lo is to describe a deeper level, we can set the energy for solving the radial Schrödinger equation E_{al}^{Lo} at the center of gravity; then we set E_{al} at the Fermi energy instead of the prescription in the previous paragraph.

The number of bases is simply specified by the cutoff energy of the APW for (a). However, the specification of MTOs (b) is not so simple. We use multiple MTOs for each aL to reduce the number of bases while maintaining the computational accuracy.[8] Since $h_L(\mathbf{r})$ as the envelope functions are specified by the parameters R_{SM} and ϵ, we have to specify them for all MTOs. Reference 5 showed their

optimization so as to minimize the total energy. However, as shown in figures in Ref. 5, such nonlinear optimization is too complicated. Thus, it is necessary to give a method of setting the parameters in a simple manner as follows. As for R_{SM}, we can use the condition $R_{\mathrm{SM}} = R_a/2$ for all MTOs. Then the envelope functions outside of MTs well coincide with the conventional Hankel function. Even with this simple setting of R_{SM} without optimization, numerical accuracy can be kept well; we can check the convergence of calculations with the number of APWs. We also see that the dependence of the total energy on ϵ's are rather weak in the PMT method. The dependence becomes weaker when we use larger number of APWs; hence, we do not need to carefully choose the parameters ϵ. Thus, the serious problem of the full-potential LMTO method, i.e., "how to choose MTO parameters", is essentially removed in the PMT method. This is numerically detailed in the paper that gives results for diatomic molecules from H_2 through Kr_2.[14]

We use one more approximation. We drop components higher than $l_{\max,a}$ in the 1st and 2nd components in Eq. (11), even though we have angular momentum components up to $2 \times l_{\max,a}$ in the products of the components. This is because it is meaningless to take such higher components since we have already made truncations for eigenfunctions. In addition, note that this approximation does not affect O_{ij} or T_{ij} because only spherically-symmetric products determine them.

3.1 Problems in the PMT method

Let us examine three problems in the PMT methods, and ways to manage them.

The first problem is the positive definiteness of O_{ij}. Since the last term in Eq. (9) can give a negative contribution, there is a possibility that O_{ij} cannot be positive definite. In principle, we can expect almost zero eigenvalues on the matrix $\int_{|\mathbf{r}| \le R_a} d^3 r(F_{0i}^*(\mathbf{r})F_{0j}(\mathbf{r}) - F_{2i,a}^*(\mathbf{r})F_{2j,a}(\mathbf{r}))$ for all MTs if the truncation parameters are large enough. This guarantees the positive definiteness of O_{ij}. In practice, we typically use $k_{\max,a} \sim 5$ and $l_{\max,a} \sim 4$; they can give satisfactory results while keeping the positive definiteness of O_{ij}, as seen in Refs. 8 and 14.

The second is the undefiniteness of the second component ψ_{2p}. This is clear if (A) is satisfied because ψ_{2p} within MTs is not uniquely determined since it is canceled completely by ψ_{0p} within MTs. Since we use (A') in practice, this undefinitness gives a numerical instability. To illustrate this, let us consider a linear combination of basis functions where only their 0th and 2nd components within MT are nonzero. This is a null vector that has no physical meanings; it gives zero when we apply a Hamiltonian and an overlap matrix to it. This is a kind of ghost. Apparently, this occurs because the 3-component space is not a complete metric space in a mathematical sense. When we enlarge the number of bases, this null vector can cause numerical problems. It can become an origin of an uncontrollable eigenvalue (e.g., 0 divided by 0) or it can attach to some eigenfunctions and deform them easily. In fact, we observed unconverged cases when the 2nd component of electron density becomes too large. Within our current implementation of the PMT, we should use limited number of basis so as to avoid this problem. However, in Refs. 8 and 14, we can see enough stability on the total

energy convergence before such problems occur when we increase the number of bases.

It will be possible to remove the above such undefiniteness in some ways. For example, we can minimize the total energy while adding a fixing term $+\lambda \sum_p \int_{|\mathbf{r}| \le R_a} d^3 r \, \psi^*_{2p,a}(\mathbf{r})(1 - \tilde{P})\psi_{2p,a}(\mathbf{r})$, where λ is a Lagrange multiplier and $\tilde{P}$ is a projector to the space spanned by some pseudo partial waves corresponding to true atomic partial waves. If λ is infinite, 2nd components are only spanned by pseudo partial waves. However, we should avoid a large λ so as not to deteriorate the total energy minimization.

The third problem is the orthogonality to the cores. In the frozen core approximation in Sect. 2.5, we take account of the spillout of the core electron density from MTs; this allows us to use a small MT radius. However, when we use quite small MTs, we observed a problem of orthogonality of wavefunctions to the cores, resulting in unconvergence. In such a case, we need to introduce local orbitals to represent cores so as to maintain the orthogonality. It may be possible to enforce the orthogonality with a projector, as described in Ref. 6.

3.2 Comparison with PAW

Here, we will compare the PMT method with the PAW method[6,10] on the basis of the 3-component formalism.

In the PAW method, we perform all-electron (AE) calculations for a spherical atom as reference in advance. Then the main problem is how to solve the one-body problem for a given one-body potential $V(\mathbf{r})$ in real space. As in Sect. 2.2, the problem is translated into the problem in the 3-component space for $V = V_0 \oplus V_1 \ominus V_2$. For simplicity, we omit the index a in the following.

The basis set in the PAW is given as follows. We first prepare AE partial waves $\{\phi_i(\mathbf{r})\}$ (e.g., two for each aL in Ref. 10), as solutions of the radial Schödinger equation for V_1 at some reference energies $\{\epsilon_i\}$ (in this section, the index i is for the partial wave). Then we set up the corresponding pseudo partial waves $\{\tilde{\phi}_i(\mathbf{r})\}$. The eigenfunction ψ in the PAW can be represented in the 3-component space; for the given 0th-component $\bar{\psi}$ (this is called the *pseudo wavefunction*), we have ψ with projectors $\{\tilde{p}_i\}$ as

$$\psi = \bar{\psi} \oplus \sum_i |\phi_i\rangle\langle\tilde{p}_i|\bar{\psi}\rangle \ominus \sum_i |\tilde{\phi}_i\rangle\langle\tilde{p}_i|\bar{\psi}\rangle. \qquad (40)$$

Here, $\tilde{p}_i$ should satisfy $\langle\tilde{p}_i|\tilde{\phi}_j\rangle = \delta_{ij}$. The minimization of the total energy of the one-body problem $E = \sum_j^{\text{occupied}} \psi_j^* \cdot (T + V) \cdot \psi_j$ with respect to $\bar{\psi}_j$ is given by

$$\left(\frac{-\nabla^2}{2m} + V_0(\mathbf{r}) - \epsilon_j + \sum_{ii'} |\tilde{p}_i\rangle(dH_{ii'} - \epsilon_j dO_{ii'})\langle\tilde{p}_{i'}| \right)\bar{\psi}_j = 0, \qquad (41)$$

$$dH_{ii'} = \langle\phi_i| \frac{-\nabla^2}{2m} + V_1|\phi_{i'}\rangle - \langle\tilde{\phi}_i| \frac{-\nabla^2}{2m} + V_2|\tilde{\phi}_{i'}\rangle, \qquad (42)$$

$$dO_{ii'} = \langle\phi_i|\phi_{i'}\rangle - \langle\tilde{\phi}_i|\tilde{\phi}_{i'}\rangle. \qquad (43)$$

If we use an infinite number of partial waves that make a complete set, Eq. (41) reproduces the original one-body problem in real space.

Let us consider a case where $\psi_j = \bar{\psi}_j \oplus \phi_j \ominus \tilde{\phi}_j$ is the solution of Eq. (41) with the eigenvalue ϵ_j, where $\bar{\psi}_j$ within MT coincides with $\tilde{\phi}_j$. This is given by Eq. (40) from $\bar{\psi}_j$.

When we make a truncation of the number of partial waves, $\{\tilde{p}_i\}$ should satisfy

$$\left(\frac{-\nabla^2}{2m} + V_0(\mathbf{r}) - \epsilon_j \right)|\tilde{\phi}_j\rangle + \sum_i |\tilde{p}_i\rangle(dH_{ij} - \epsilon_j dO_{ij}) = 0, \qquad (44)$$

in order to satisfy Eq. (41). This determines $\{\tilde{p}_i\}$; this is one of the main ideas in the PAW method. In practice, considering the numerical stability, we determine $\tilde{p}_i$ so that Eq. (44) is approximately satisfied.[6]

Another important idea in the PAW method is the introduction of the pseudopotential. This is how to determine V_0 within MT $(= V_2)$. This is because the result strongly depends on the pseudopotential when the number of partial waves is small. In principle, the pseudopotential should be determined so that $\bar{\psi}_j$ contains a high-energy part (high angular momentum l or highly oscillating part) of the wavefunctions, which is missing in the 1st and 2nd components owing to the truncation of the number of partial waves.

Note that the truncation can cause the ghost state problem in the PAW method. To illustrate this, consider a case in which the s-wave part in MT is described only by the two partial waves $2s$ and $3s$. Then the PAW procedure maps $\bar{\psi}$ with zero node to ψ with one node, $\bar{\psi}$ with one node to ψ with two nodes. Problem is that $\bar{\psi}$ with two nodes, which is orthogonal to $\{\bar{\psi}_i\}$ for $2s$ and $3s$, can not be mapped to ψ with three nodes due to the truncation. Thus it is possible that such a function causes a ghost state; we have to design the pseudopotential so that such $\bar{\psi}$ should be kept in a sufficiently high energy region (to push $\bar{\psi}$ far away from the Fermi energy, it may be better to use a relatively strongly repulsive pseudopotential). Reference 10 claims that there is no ghost state for all kinds of atoms. However, it is not easy to check the convergence within the framework of the PAW method.

In the PAW method with PWs proposed in Ref. 10, many PWs are required compared with LAPW. Roughly speaking, the energy cutoff of PWs are ~15 Ry in LAPW, and ~30 Ry in PAW.[10,29] This is because the PAW method, as is the case of pseudopotential methods, needs to uniquely determine pseudo partial waves (0th component) within MT. This is in contrast with the LAPW (and the PMT) method, where 0th component within MT is irrelevant because the 2nd components have enough degree of freedom to efficiently cancel its contribution. However, by sacrificing the cutoff energy, the PAW takes a robust convergence that comes from the absence of the null vector problem discussed in Sect. 3.1.

As a theoretical possibility, we can imagine a method of using smHankels together with the PWs in the basis set for the one-body problem in the PAW method. However, it is not very clear whether or not it becomes an efficient method. To reduce the number of bases of PWs, it is necessary to make the smHankels span high-energy parts of pseudo wavefunctions. Thus, we have to tailor smHankel so that it fits the pseudo wavefunctions not only in interstitial region, but also within MT. This can be not straightforward.

4. Summary

We have reformulated the PMT method on the basis of the 3-component formalism, which is a generalized version of the additive augmentation given by Soler and Williams. The

3-component formalism allows the inclusion of any kinds of bases not necessarily given by a projector as in the PAW method. This fits the procedure for giving the Kohn–Sham equation for a mixed basis method such as the PMT method from the total energy minimization scheme; this results in the transparent derivation of atomic forces. We believe that the formalism shown here could give a basis for future development. Our results for molecules from H_2 through Kr_2 with several new developments on the PMT method are given in Ref. 14.

Acknowledgements

Appendix A: Error Due to the Separable Form

To evaluate the matrix element of a quasilocal operator $X(\mathbf{r}, \mathbf{r}')$, we use the separable form $0X0' + 1X1' - 2X2'$ instead of $(0 + 1 - 2)X(0' + 1' - 2')$ under condition (A′) (see Sect. 2.1). Here, $0, 1, 2$, as well as $0', 1', 2'$, means the three components of a eigenfunction as a 3-component function defined in Sect. 2.1.

We have an error because of the separable form. Here, we reorganize the discussion to evaluate the error[6,11] to fit the formalism in this paper. The error can be evaluated with the identity

$$(0 + 1 - 2)X(0' + 1' - 2') - (0X0' + 1X1' - 2X2')$$
$$= (0 - 2)X(1' - 2') + (1 - 2)X(0' - 2'). \quad \text{(A·1)}$$

Let us examine the error as the right-hand side of Eq. (A·1) under the assumption that X is nearly spherical. Remember that $(0 - 2)$ is completely zero if the condition (A) is satisfied. When the condition (A′) is satisfied instead, i.e., when we introduce the finite truncation parameters $l_{\max,a}$ and $k_{\max,a}$ [see after Eq. (4)], we can expect that $(0 - 2)$ should contain a high-energy remnant (high angular momentum l or highly oscillating remnant) with a small amplitude. The remnant $(0 - 2)$ for each L is largest at the MT boundaries. In contrast, when (A′) is satisfied, $(1' - 2')$ is the low-energy part that converges quickly on the truncation parameters. The value and slope of $(1' - 2')$ are zero at MT boundaries. Thus, we can expect that the product $(0 - 2)(1' - 2')$ should be small and nearly orthogonal, i.e., $\delta n_a(\mathbf{r}) = (0 - 2)_a(1' - 2')_a$ should satisfy $\int_a d^3 r \, \delta n_a(\mathbf{r}) Y_L(\hat{\mathbf{r}}) \approx 0$ for a low L. Here, suffix a means quantities within MT at $\mathbf{R}_a$. Based on these considerations, we expect that the error affects hardly the total energy. This can be checked by changing the truncation parameters within the PMT method.

This logic is applicable not only to the products of the eigenfunctions, but also to the electron density for the Coulomb interaction with some modifications.

Appendix B: Atomic Force

First, we define the Harris energy E_{Harris},[9,30] which is the total energy of a functional of the density; this gives a reasonable estimate of the total energy even when the density is a little away from the density of self-consistency. During the iteration cycle to attain the convergence of self-consistency, the input density n^{in} generates the one-particle potential V. Here, V is given by Eq. (34) where we use n^{in} for n. Then the output density n^{out} is calculated from eigenfunctions of the eigenvalue problem of V. Now, E_{Harris} in the frozen core approximation as a functional of $\bar{n}_i^{\text{in}}$ is defined by:[9]

$$E_{\text{Harris}} = E_k^{\text{core}} + E_B - V[n^{\text{Zc}} + n^{\text{in}}, \mathbf{R}_a] \cdot n^{\text{in}}$$
$$+ E_{\text{es}}[n^{\text{Zc}} + n^{\text{in}}, \mathbf{R}_a] + E_{\text{xc}}[n^{\text{Zc}} + n^{\text{in}}], \quad \text{(B·1)}$$

$$E_B = \sum_p^{\text{occupied}} \alpha_p^{i*} \langle F_i | H^{\text{in}} | F_j \rangle \alpha_p^j, \quad \text{(B·2)}$$

where E_B is the band energy. α_i^p is the eigenvector of $\langle F_i | H^{\text{in}} | F_j \rangle = \langle F_i | \frac{-\Delta}{2m} + V[n^{\text{Zc}} + n^{\text{in}}, \mathbf{R}_a] | F_j \rangle$. Thus, we have $E_B = \sum_p^{\text{occupied}} \epsilon_p$, where ϵ_p are eigenvalues. The $\mathbf{R}_a$ dependence explicitly shown in Eq. (B·1) is through the $\mathcal{M}$-transformation and $\mathcal{A}$-mapping; even when $n^{\text{Zc}} + n^{\text{in}}$ is fixed as a 3-component function, the $\mathbf{R}_a$ dependence is introduced in Eq. (34) through Eqs. (28) and (35). In addition, we have the $\mathbf{R}_a$ dependence through $n^{\text{Zc}} + n^{\text{in}}$.

Atomic forces are given as the change in the total energy for atomic displacement $\delta\mathbf{R}_a$. Here, let us consider the change in E_{Harris}, written as δE_{Harris}. To obtain δE_{Harris}, we use the derivative chain rule where we treat E_{Harris} as a function of $\mathbf{R}_a$ through $\{F_i(\mathbf{r}), n^{\text{in}}, V, \mathbf{R}_a\}$; V means $V[n^{\text{Zc}} + n^{\text{in}}, \mathbf{R}_a]$ in Eqs. (B·1) and (B·2). Remember that there is an $\mathbf{R}_a$ dependence through n^{Zc}. Here, we assume the partial waves ($\{\phi_{al}(r), \dot{\phi}_{al}(r), \phi_{al}^{\text{Lo}}(r)\}$ in the case of the PMT method) are not dependent on atomic positions, as in Ref. 9.

Let us evaluate δE_{Harris}. As for $E_B = \sum_p^{\text{occupied}} \epsilon_p$ as a functional of $\{F_i(\mathbf{r}), V\}$, the perturbation theory on Eq. (36) gives

$$\delta E_B = \sum_p^{\text{occupied}} \delta\epsilon_p = \sum_p^{\text{occupied}} \sum_i \sum_j \alpha_{p*}^i (\delta H_{ij} - \epsilon_p \delta O_{ij}) \alpha_p^j$$
$$= \delta V \cdot n^{\text{out}} + \delta E_B^{\text{Puley}}, \quad \text{(B·3)}$$

$$\delta E_B^{\text{Puley}} = \sum_p^{\text{occupied}} \sum_i \sum_j \alpha_p^{i*} (\delta H_{ij}^F - \epsilon_p \delta O_{ij}^F) \alpha_p^j, \quad \text{(B·4)}$$

where we have used $\delta(V \cdot F_i^* F_j) = \delta V \cdot F_i^* F_j + V \cdot \delta(F_i^* F_j)$. $\delta E_B^{\text{Puley}}$ is calculated from $\delta F_{0i}(\mathbf{r})$ and δC_{akL}^i, which are given as a functional of $\delta\mathbf{R}_a$.

Since $E_{\text{es}} + E_{\text{xc}}$ is a functional of $\{n^{\text{in}}, \mathbf{R}_a\}$, we have

$$\delta E_{\text{Harris}} = \delta E_B - \delta(V \cdot n^{\text{in}}) + \delta(E_{\text{es}} + E_{\text{xc}})$$
$$= \delta V \cdot (n^{\text{out}} - n^{\text{in}}) + \delta E_B^{\text{Puley}} + \frac{\partial(E_{\text{es}} + E_{\text{xc}})}{\partial\mathbf{R}_a}\bigg|_{n^{\text{in}}} \delta\mathbf{R}_a.$$
$$\text{(B·5)}$$

There are three terms on the right-hand side of Eq. (B·5). The first term appears because E_{Harris} is for non self-consistent density ($n^{\text{out}} \neq n^{\text{in}}$).

To calculate the first term, we need to know δn^{in} that determines δV. When self-consistency is attained, that is, $n^{\text{in}} = n^{\text{out}}$, $\delta\mathbf{R}_a$ uniquely determines $\delta n^{\text{in}} = \delta n^{\text{out}}$. However, this is not true when $n^{\text{in}} \neq n^{\text{out}}$. In this case, there is no unique way to determine δn^{in} for a given $\delta\mathbf{R}_a$. Thus, we need an extra assumption to determine it. As a reasonable and convenient choice, we use $\delta n^{\text{in}} = 0$ in the sense of the 3-component representation. Physically, this means that $n_{1,a}(\mathbf{r}) - n_{2,a}(\mathbf{r})$ together with the frozen core centered at $\mathbf{R}_a$ moves rigidly to $\mathbf{R}_a + \delta\mathbf{R}_a$. Then we can calculate the corresponding δV from the change $\delta\bar{n}_0^{\text{Zcv}}$ in Eq. (34). $\delta\bar{n}_0^{\text{Zcv}}$ is

evaluated from Eq. (28). Note that $n_0^{Zc}(\mathbf{r})$ contains an $\mathbf{R}_a$ dependence as given in Eq. (23).

Appendix C: Onsite Matrix

Here, we summarize the expressions of the one-center matrix for O_{ij}, T_{ij}, and V_{ij}. These are essentially the same as that shown in Ref. 5. With the help of Eqs. (4) and (5), Eqs. (9), (10), and (34) are reduced to

$$O_{ij} = \int_\Omega d^3 r \, F_{0i}^*(\mathbf{r}) F_{0j}(\mathbf{r}) + \sum_{akk'L} C_{akL}^{*i} \sigma_{akk'L} C_{akL}^j \quad \text{(C·1)}$$

$$T_{ij} = \frac{1}{2m_e} \int_\Omega d^3 r \, \nabla F_{0i}^*(\mathbf{r}) \nabla F_{0j}(\mathbf{r})$$
$$+ \sum_{akk'L} C_{akL}^{*i} \tau_{akk'L} C_{ak'L}^j, \quad \text{(C·2)}$$

$$V_{ij} = \int_\Omega d^3 r \, F_{0i}^*(\mathbf{r}) V_0(\mathbf{r}) F_{0j}(\mathbf{r})$$
$$+ \sum_{akk'LL'} C_{akL}^{*i} \pi_{akk'LL'} C_{ak'L'}^j, \quad \text{(C·3)}$$

where

$$\sigma_{akk'l} = \int_{|\mathbf{r}| \le R_a} d^3 r (\widetilde{P}_{akL}(\mathbf{r}) \widetilde{P}_{ak'L}(\mathbf{r})$$
$$- P_{akL}(\mathbf{r}) P_{ak'L}(\mathbf{r})), \quad \text{(C·4)}$$

$$\tau_{akk'l} = \frac{1}{2m_e} \int_{|\mathbf{r}| \le R_a} d^3 r (\nabla \widetilde{P}_{akL}(\mathbf{r}) \nabla \widetilde{P}_{ak'L}(\mathbf{r})$$
$$- \nabla P_{akL}(\mathbf{r}) \nabla P q_{ak'L}(\mathbf{r})), \quad \text{(C·5)}$$

$$\pi_{akk'LL'} = \sum_M Q_{kk'LL'M} \mathcal{Q}_{aM}^v$$
$$+ (\bar{n}_{1,a}^{Zcv} \circ w + v_{1,a}^{xc}) \circ \widetilde{P}_{akL}(\mathbf{r}') \widetilde{P}_{ak'L}(\mathbf{r}')$$
$$- (\bar{n}_{2,a}^{Zcv} \circ w + v_{2,a}^{xc}) \circ P_{akL}(\mathbf{r}') P_{ak'L}(\mathbf{r}'), \quad \text{(C·6)}$$

$$Q_{kk'LL'M} = \int_{|\mathbf{r}| \le R_a} d^3 r (\widetilde{P}_{akL}(\mathbf{r}) \widetilde{P}_{ak'Ll}(\mathbf{r})$$
$$- P_{akL}(\mathbf{r}) P_{ak'Ll}(\mathbf{r})) \mathcal{Y}_M(\mathbf{r}). \quad \text{(C·7)}$$

Note that $\sigma_{akk'l}$ and $\tau_{akk'l}$ are dependent only on l of $L = (l, m)$. In Ref. 5, this $\pi_{akk'LL'}$ is further divided as $\pi_{akk'LL'}^{mesh} + \pi_{akk'LL'}^{local}$. $\mathcal{Q}_{aM}^v$ is given by Eq. (35).

Appendix D: Scalar Relativistic Approximation in the Augmentation

Generally, it is valid to take the scalar relativistic (SR) approximation (e.g., see Ref. 2) if we can safely replace the non-relativistic (NR) wavefunctions with the SR wavefunctions within MTs. The SR wavefunctions contain major and minor components. The major component should be smoothly connected to the NR wavefunction in the interstitial region, where the minority parts are negligible. All physical quantities within MT should be evaluated through the SR wavefunctions. In the followings, we explain how the above idea can be implemented in the 3-component augmentation for bilinear products.

First, we modify the 1st component of the basis. We use two component wavefunctions $\{g_{1i,aL}(\mathbf{r}), f_{1i,aL}(\mathbf{r})\}$ instead of $F_{1i,a}(\mathbf{r})$, where the SR approximation gives $f_{1i,a}(\mathbf{r}) = \frac{1}{2m_e c} \frac{dg_{1i,a}(\mathbf{r})}{dr}$, where c is the speed of light. For the given F_{0i} and F_{2i} (they are the same as those in the NR case), we ask

that the major components $g_{1i,a}(\mathbf{r})$ satisfy the boundary conditions for value and slope at MT boundaries.

In order to calculate the contributions of the 1st components within the SR approximation instead of the NR approximation, we make the replacement $F_{1i,a}^*(\mathbf{r}) F_{1j,a}(\mathbf{r}') \to g_{1i,a}^*(\mathbf{r}) g_{1j,a}(\mathbf{r}') + (\frac{1}{2m_e c})^2 f_{1i,a}^*(\mathbf{r}) f_{1j,a}(\mathbf{r}')$. With this replacement, we can evaluate the density n, the matrix O_{ij}, and so on. This ends up with the total energy in the SR approximation.

Finally, we see that changes are in the replacement Eqs. (C·4)–(C·7), where products $\widetilde{P}_{akL}(\mathbf{r}) \widetilde{P}_{ak'L}(\mathbf{r})$ (and those with ∇) should be interpreted not only from the products of the majority wavefunctions, but also from those of the minority wavefunctions. This also occurs for the density $n_{1,a}$ included in Eq. (C·6).

In such a way, we can include the SR effect in the 3-component formalism. In a similar manner, we can include the spin–orbit coupling in the 1st component, which results in spin off-diagonal contributions.[31]

1) O. K. Andersen, Phys. Rev. B **12**, 3060 (1975).
2) R. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, New York, 2000).
3) D. J. Singh, *Planewaves, Pseudopotential, and the LAPW Method* (Kluwer Academic, Boston, MA, 1994).
4) S. Blügel and G. Bihlmayer (2006) [http://www2.fz-juelich.de/nic-series/volume31/bluegel.pdf].
5) M. Methfessel, M. van Schilfgaarde, and R. A. Casali, in *Lecture Notes in Physics*, ed. H. Dreysse (Springer, Berlin, 2000) Vol. 535.
6) P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
7) D. Singh, Phys. Rev. B **43**, 6388 (1991).
8) T. Kotani and M. van Schilfgaarde, Phys. Rev. B **81**, 125117 (2010).
9) M. Methfessel and M. van Schilfgaarde, Phys. Rev. B **48**, 4937 (1993).
10) G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).
11) J. M. Soler and A. R. Williams, Phys. Rev. B **40**, 1560 (1989).
12) J. M. Soler and A. R. Williams, Phys. Rev. B **42**, 9728 (1990).
13) J. M. Soler and A. R. Williams, Phys. Rev. B **47**, 6784 (1993).
14) T. Kotani and H. Kino, J. Phys. Soc. Jpn. **82**, 124714 (2013).
15) T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Phys. Rev. B **76**, 165106 (2007).
16) M. van Schilfgaarde, T. Kotani, and S. V. Faleev, Phys. Rev. B **74**, 245125 (2006).
17) S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
18) M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
19) T. Kotani, J. Phys. Soc. Jpn. **83**, 094711 (2014).
20) M. J. Han, H. Kino, and T. Kotani, Phys. Rev. B **90**, 035127 (2014).
21) H. Nagara, T. Ishikawa, and T. Kotani, High Pressure Res. **34**, 215 (2014).
22) J. Klimeš, M. Kaltak, and G. Kresse, Phys. Rev. B **90**, 075125 (2014).
23) E. Bott, M. Methfessel, W. Krabs, and P. C. Schmidt, J. Math. Phys. (N.Y.) **39**, 3393 (1998).
24) K. Iwashita, T. Oguchi, and T. Jo, Phys. Rev. B **54**, 1159 (1996).
25) M. van Schilfgaarde, private communication. The head part of the smooth Hankel functions is difficult to expand in the series of polynomials accurately.
26) M. Weinert, J. Math. Phys. (N.Y.) **22**, 2433 (1981).
27) P. Pulay, Mol. Phys. **17**, 197 (1969).
28) E. Sjöstedt, L. Nordström, and D. J. Singh, Solid State Commun. **114**, 15 (2000).
29) C. Filippi, D. J. Singh, and C. J. Umrigar, Phys. Rev. B **50**, 14947 (1994).
30) J. Harris, Phys. Rev. B **31**, 1770 (1985).
31) A. N. Chantis, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **96**, 086405 (2006).