Contracted Plane Wave satisfying periodic gauge

Takao Kotani
Department of Applied Mathematics and Physics, Tottori university, Tottori 680-8552, Japan
(Dated: October 9, 2020)

We introduce the contracted plane waves (CPWs), which satisfy the periodic gauge in the Brillouin-zone torus as in the case of usual Wannier functions. CPWs are very simply given as the sum of plane waves. We will be able to use CPWs instead of the Wannier functions for the interpolation of physical quantities given as the function of wave vectors in BZ. Furthermore, it will be easy to complement the set of Gaussian bases by CPWs.

PACS numbers: 71.15.Ap, 71.15.-m, 31.15.-p

I. INTRODUCTION

In the first-principles methods such as the pseudopotential (PP), LAPW, and PAW, we use plane waves (PWs) or augmented plane waves (APWs) for the bases to expand eigenfunctions [1]. Even the linearized Muffin-tin orbital (MTO) method have recently evolved to be the APW plus MTO method (the PMT method) [2], where we can perform accurate and robust calculations with the basis set of the highly-localized MTOs (without material-dependent parameters) complemented by the set of the low-energy-cutoff ($\lesssim 4$ Ry) APWs [3, 4]. In contrast to these methods utilizing PWs/APWs, it is not so easy to obtain high-energy states accurately in the methods with Gaussians as bases [5], because of the difficulty to enlarge the number of the Gaussian bases systematically. Hereafter, discussion is seemingly for PWs, however, also for APWs with trivial modifications.

Although the set of PWs is very useful, the set is often inconvenient for kinds of applications. Especially, PWs are not suitable for the interpolation in the Brillouin zone (InBZ) for variety of quantities. InBZ means evaluating matrix $F(q)$ at any $q$ point in BZ from $\{ F(q) \mid q \in \text{mesh points in the BZ} \}$. As an example, let us consider the case to perform the GW calculations in the PW basis [3, 7]. Then we treat the non-local one-particle effective Hamiltonian $H_q(G, G') = \frac{1}{2} \int d^d r \int d^d r' \exp(-i(q + G + G')r)H(r, r') \exp(i(q + G')r)$. Here $G$ denotes the reciprocal vectors of a crystal; $N$ denotes the number of primitive cells in the Born–von Karman boundary condition. Because the GW calculations are expensive, we can calculate $H_q(G, G')$ only at limited number of $q$ in BZ. Thus we need to obtain $H_q(G, G')$ at any $q$ via InBZ. InBZ is very important, for example, to calculate effective mass and/or the Fermi surfaces in the GW calculations. InBZ is essential to perform stable QSGW calculations [8–10]. InBZ is a key methodology even when we evaluate other kinds of quantities such as electron-electron interaction, electron-phonon interactions, topological quantities, and so on.

The difficulty of the InBZ is due to the non-periodicity of the PWs $\exp(i(q + G)r)$ with respect to $q$: when $q$ changes across BZ to be $q + G_{\text{shift}}$, $\exp(i(q + G)r)$ changes to $\exp(iG_{\text{shift}}r) \exp(i(q + G)r)$. That is, $\exp(i(q + G)r)$ do not satisfy the periodic gauge [11], $\exp(i(q + G)r) \neq \exp(iG_{\text{shift}}r) \exp(i(q + G)r)$. This ends up with $H_q(G, G') \neq H_{q + G_{\text{shift}}}(G, G')$. This causes difficulty to use PWs for InBZ.

To avoid this difficulty, the so-called Wannier interpolation (WI) [7] is introduced as a method for InBZ. After we construct a set of the Wannier functions by the procedure of maximally localized Wannier function [12–14], we re-expand $H_q(G, G')$ by the Wannier functions instead of PWs as $H_q(i, i')$, where $i, i'$ are the the indexes of the Wannier functions. Since the Wannier functions satisfy the periodic gauge condition [11], we can easily interpolate $H_q(i, i')$ for any $q$. However, the construction of the maximally-localized Wannier functions usually used in the WI are not so simple [15]. We need to choose initial conditions and choose energy windows. It is not so easy to make the method automatic without examination by human. This causes a difficulty when we apply the method to the material’s informatics where we have to analyze thousands of possible materials in a work. In order to avoid the difficulty of WI, Wang et al. suggested to use the first-principles method with the atomic-like localized bases satisfying the periodic gauge [16]. However, this suggestion do not remove a problem in WI; for supercell with huge vacuum region, we have to fill the region by the bases of PWs if we need to describe scattering states well. The localized orbitals as well as the Wannier functions can hardly fill the vacuum region systematically.

Here we will introduce new bases named as the contracted plane waves (CPWs). CPWs satisfy the periodic gauge, thus are represented by the Bloch sum of the non-orthogonalized localized bases. CPWs are generated easily as the linear combinations of PWs. Thus we will be easily make InBZ. As another application, CPWs can be used as bases for the first-principles calculations. Especially, we expect CPWs can be easily included as a part of bases in the Gaussian-based packages as Crystal [8].
II. CONSTRUCTION OF THE SET OF CPWS

A. Definition of CPW

Bases in the set of CPWs \( \{P_{qn}(r)\} \) are defined as

\[
P_{qn}(r) = \frac{1}{\sqrt{N_n}} \sum_G C_n(q + G) \exp(i(q + G)r), \tag{1}
\]

\[
C_n(q + G) = 
\exp\left(-\frac{1}{2\alpha^2} \sum_{ij} (q_i + G_i + \vec{G}_{ni}) A_{ij} (q_j + G_j + \vec{G}_{nj})\right) \tag{2}
\]

where \( q \) is the wavevectors in the BZ, \( q_i \) denotes the x,y,z components for \( q \); \( G_i \) and \( \vec{G}_{ni} \) are for \( G \) and \( \vec{G}_n \), as well. \( \vec{G}_n \) is a vector in a set \( \Omega = \{ \vec{G}_n | n = 1, 2, 3, \ldots N_{\text{max}} \} \); we show how to choose \( \Omega \) and \( \alpha \) in Sec. \[ II \]. In Eq. (2), we take sum for all the reciprocal vectors \( G \). \( \sqrt{N_n} \) are normalization factors so that \[ \int d^3r (P_{qn}(r))^2 = 1 \]. Since the symmetric matrix \( A_{ij} \) is determined for the Bravais lattice as follows, the set \( \{P_{qn}(r)\} \) is specified by the parameters \( \alpha \) and \( \Omega \).

\( A_{ij} \) is an invariant symmetric matrix under the symmetry of the Bravais lattice like the dielectric constants. While \( A_{ij} \) is essentially trivial except a constant factor in the case of simple lattice, let us consider case of general Bravais lattice. We ask that the spheroid given by \[ \sum_{i,j} q_i A_{ij} q_j = 1 \] should give an optimum fitting of the 1st Brillouin zone (BZ). We have a few possible options to determine \( A_{ij} \). One is that the spheroid can be given as the largest spheroid inside the 1st BZ, one another is that we can determine \( A_{ij} \) to reproduce the ‘moment of inertia’ of the 1st BZ. In either way, we can obtain three orthonormalized vectors \( n_1, n_2, \) and \( n_3 \) with corresponding eigenvalues \( c_1, c_2, \) and \( c_3 \) for such \( A_{ij} \). Thus we can interpret that \( A_{ij} \) specify an approximation of the 1st BZ by a cuboid where as we scale the size of \( A_{ij} \) so that \[ c_1 c_2 c_3 = |A_{ij}| = \left(\frac{2\alpha}{\sqrt{\Omega}}\right)^2 \].

We can easily see that the CPWs satisfy the periodicity of the usual Bloch functions as

\[
P_{qn}(r + T) = P_{qn}(r) \exp(iqT), \tag{3}
\]

\[
P_{qn+Gn}(r) = P_{qn}(r), \tag{4}
\]

where \( T \) is the translation vectors. We can define \( P_{qn}(r) \) for \( q_n \) even for the augmented plane waves (APWs) if we use APWs instead of \( \exp(i(q + G)r) \). It is virtually possible to take infinite sum for all \( \vec{G} \) numerically, because of the truncation due to the Gaussian factor in Eq. (2). Eq. (2) shows that a PW whose \( q + G \) nearest to \( \vec{G}_n \) has largest \( C_n(q + G) \). Thus we may say that \( P_{qn}(r) \) is similar with \( \exp(iG_n r) \) when \( \alpha \) is small enough. As we noted in Sec. [ I ] one of the usage of CPWs is for the BZ. Because \( P_{qn}(r) \) satisfies the periodic gauge condition Eq. (4), quantities such as \[ \langle P_{qn}^\dagger H P_{qn} \rangle \], is periodic for \( q \) in the BZ. This allows us to make the BZ.

From Eq. (1), we have real-space localized functions \( P_{qn}(r) \) as

\[
P_{qn}(r) = \sum_q P_{qn}(r) = \frac{1}{\sqrt{N_n}} (2\pi \alpha)^{3/2} \times \exp(-\frac{\alpha^2}{2} \sum_{ij} r_j A_{ij}^{-1} r_j + iG_n r). \tag{5}
\]

When we use real \( \alpha \), the center of \( P_{qn}(r) \) is located at \( r = 0 \). Note that bases in the set \( \{P_{qn}(r)\} \) are not orthogonalized. Thus we need overlap matrix \( O_q(n, n') = \langle P_{qn}(r) | P_{qn'}(r) \rangle \) to handle the set. CPWs are nothing but the Bloch sum of the Gaussians with oscillations as shown in Eq. (5).

To use CPWs in the first-principles calculations, especially for the BZ, we have following requirements for the set \( \{P_{qn}(r)\} \).

1. Each \( P_{qn}(r) \) should be smoothly changing as a function of \( q \) in the BZ.

2. The low-energy Hilbert space spanned by \[ \{\exp(i(q + G)r) | G \in \Omega_S(q)\} \] should be contained well in the Hilbert space spanned by \{\( P_{qn}(r) \),\} where \( \Omega_S(q) \) is a set of \( G \) near \( q + G = 0 \).

3. Linear-dependency of bases are kept. That is, the eigenvalues of overlap matrix \( O_q(n, n') \) should be not too small for double precision calculations.

Under the assumption that \[ \{\exp(i(q + G)r) | G \in \Omega_S(q)\} \] is good enough to expand eigenfunctions, the condition (2) assures that the space spanned by \{\( P_{qn}(r) \)\} is good enough, too. In fact, we expect not so large \( \Omega_S \) is required to obtain accurate bands in GW calculations [3]. Together with the condition (1) and (3), we can use \( P_{qn}(r) \) as bases to expand the one-body Hamiltonian for the inter-polation.

Although our Hilbert space spanned by the set \{\( P_{qn}(r) \)\} is not satisfying the translational symmetry, we will see that the translational symmetry can be recovered well as shown in Sec. [ III ].

B. Parameters to specify a set of CPWs

To specify the set \( \{P_{qn}(r)\} \), we need \( \Omega \). \( \vec{G}_n \) in \( \Omega \) is given by

\[
\vec{G}_n = \frac{1}{\beta} G(n), \tag{6}
\]

where \( G(n) \) is the reciprocal \( G \) vectors satisfying \[ |G(n)| < G_{\text{MAX}} \]. Thus the number of bases in the set \( \{P_{qn}(r)\} \) is given by \( G_{\text{MAX}} \). \( \beta \) is a scaling factor, a little larger than unity. The value of \( \beta \) is examined in Sec. [ III ]. This procedure of Eq. (6) gives a little denser mesh of \( \vec{G}_n \) than the mesh of \( G \). As we see in Sec. [ III ] the scaling procedure Eq. (6) is essential to reproduce eigenfunctions around the BZ boundaries accurately.
In addition, we need to determine the parameter $\alpha$. The condition (1) in Sec.IIA requires that $\alpha$ is large enough so that $C_n(q+G)$ is a smooth function of $q$, while the condition (2) requires $\alpha$ is small enough so that $C_n(q+G)$ chooses one of $q+G$ dominantly. To let these requirements balanced, we use a balancing condition between the damping factor in Eq. (2) measured by the unit of the $G$-lattice spacing, and the damping factor of Eq. (5) measured by the realspace-lattice spacing. Because we have approximated the 1st BZ by the cuboids given by $A_{ij}$ (real-space cuboid by $A_{ij}$ as well), we have

$$\frac{-4}{2\alpha^2} \approx -\frac{\alpha^2 (\pi)^2}{2}. \quad (7)$$

Thus we have $\alpha = \sqrt{2/\pi} \approx 0.8$. We use this value in the test calculations shown in Sec.III.

With the three parameters $G_{\text{MAX}}, \alpha$, and $\beta$, we can give the set $\{P_n(r)\}$ for given primitive cell vectors of a crystal structure. In Sec.III we will evaluate quality of the set when we are changing these parameters.

Let us summarize the algorism to specify $\{P_n(r)\}$. For given $G_{\text{MAX}}$, we first make a set $G(n)|n = 1, \ldots, N_{\text{MAX}}; |G(n)| < G_{\text{MAX}}$. Then, we have $\bar{\Omega}$ by Eq. (6). Then, Eq. (1) and Eq. (2) yields $\{P_n(r)\}$, whereas $A_{ij}$ is given in advance so that the 1st BZ is approximated by the cuboid specified by $A_{ij}$. The number of $G$ for the sum in Eq. (2) should be large enough so that the sum converges well.

There is a possibility to use another algorism to determine $\alpha$; furthermore $\alpha$ can be $n$-dependent. In fact, we have tested some other possibilities such as $G_n = \frac{G(n)}{\left|G(n)\right|\pi^2}G(n)$ instead of Eq. (6). However, we did not observe meaningful improvements in the tests shown in Sec.III Thus we focus on the algorism presented here.

### III. Numerical Tests and Determination of Parameters of CPWS

Here we perform test calculation for the fcc empty lattice. In the followings, we consider a case of the fcc lattice for Si. The size of primitive cell is $a_t$ where $a = 5.43\text{Å}$. We show that the set $\{P_n\}$ spans the low energy part of the Hilbert space spanned by PWs very well, as long as we take adequate choice of $(\alpha, \beta)$.

We first look into eigenvalues. We solve the following eigenvalue problem to determine eigenvalue $\lambda_{q}^{\text{m}}$ for empty lattice:

$$\sum_{n'} \left( <P_{q \nu}|\frac{-\nabla^2}{2m^*}P_{q \nu'}>| \lambda_{q}^{\text{m}} (P_{q \nu}|P_{q \nu'}) \right) \epsilon_{q n'}^{m} = 0. \quad (8)$$

In Fig. we plot $\lambda_{q}^{\text{m}}$ for the fcc empty lattice with $\alpha = 0.8$ and with $\beta = 1, 1.6, 2.0$. We use $G_{\text{MAX}} = 4$, which determines the number of bases is 56. We show exact eigenvalues of the empty lattice by solid lines together.

We see disagreements at the BZ boundaries for the case of no scaling, $\beta = 1$. However, we see good agreement with the exact ones for $\beta = 1.6$ and $\beta = 2.0$ in the whole BZ. We show details of agreement afterwards in Fig.4. In the middle panel of $\beta = 2.0$, we see a branch from $\Gamma$ to $X$ are bending at $32\text{eV}$ and getting to be $23\text{eV}$ at $X$ point. This is reasonable because of Eq. (2), when we are changing $q, G$ giving the largest $\left| q+G-G_n \right|$ switches. For larger $\beta$, we have better agreement at low energies, but we have more bumpy bands at higher energies. Since larger $\beta$ gives denser bases for lower energy, larger $\beta$ shows better agreement at low energy but worse at high energy.

Let us examine the effect of parameter $\alpha$. In Fig. we show how the bands change for changing $\alpha$. The middle panel of Fig. is the same as that in Fig. As we expect from Eq. (2), we have better smoothness of energy bands in the BZ for larger $\alpha$.

Another check is the ability to reproduce PWs as superpositions of $\{P_n\}$. For this purpose, we calculate the square of the amplitude of projection to the space of PWs, $Q(q+G) = \sum_n |\langle P_n|\exp(i(q+G)r)\rangle|^2$. We show a plot, eigenvalue $\frac{|\langle P_n|\exp(i(q+G)\nu)\rangle|^2}{\epsilon_{q n}^{m}}$ (y-axis) vs. $Q(q+G)$ (x-axis), in Fig. Lines from multiple branches of energy bands are plotted. This shows most of all $Q(q+G)$ below $\sim 30\text{eV}$ is almost at one-hundred percent. This means that such low energy part is well-expanded by $\{P_n\}$.

#### A. Determination of parameters

Based on the condition in Sec.IIA and observations in Sec.III, we discuss how to determine parameters optimally. For given $G_{\text{MAX}}$, we have to determine $(\alpha, \beta)$.

For this purpose, we plot maximum error of eigenvalues at low energies (eigenvalues below 20eV) and the smallest size of the eigenvalues of the overlap matrix in Fig. for given $(\alpha, \beta)$. Each line is for each $\alpha$. Along the line, we have different symbols for different $\beta$. The top panel is for $G_{\text{MAX}} = 4$. For example, we can see that the error is $\sim 0.02\text{eV}$ for $(\alpha = 0.8, \beta = 1.6)$. Fig.4 shows that enlarging $\beta$ can efficiently reduce error at low energy. However, because of round-off error in numerical calculations, it is safer to use $(\alpha, \beta)$ which gives not too small eigenvalues of the overlap matrix.

Thus optimum $(\alpha, \beta)$ may be chosen so that it is at right-bottom in the panels. The discussion around Eq. suggested $\alpha = 0.8$. We can see that behavior of results shown in Fig. looks stable enough around $\alpha = 0.8$. Thus we claim that $\alpha = 0.8$ is not a bad choice. Then we determine $\beta$ showing not too small eigenvalue of the overlap matrix. If we set the smallest eigenvalue is $10^{-7}$ (on the red vertical line), we can see $\beta \sim 1.07$ from the line of $\alpha = 0.8$ in the case of bottom panel $G_{\text{MAX}} = 8$. Thus we suggest a prescription to determine $\alpha$ and $\beta$ for given $G_{\text{MAX}}$: use $\alpha = 0.8$, and determined $\beta$ so that the smallest eigenvalue is not too small. This determination can be done by test calculations or a table prepared in
advance. We think this can be easily implemented.

B. CPWs in the supercell

Although the procedure to give a set of CPWs do not depends on the cell size, we have one another choice to give the set of CPW for a supercell in the following manner, based on the fact that the supercell is made from small cells. The primitive vectors of the supercell \( \mathbf{A}_i \) are given as \( \mathbf{A}_i = \sum_{j} n_{ij} \mathbf{a}_j \) by the \( 3 \times 3 \)-integer matrix \( n_{ij} \), where \( \mathbf{a}_j \) is the primitive vectors of a small cell such as the fcc lattice. The volume of the supercell cell is \( N = N_{S} X S_{MAX} \). By definition of this construction, the set of CPWs to practical methods. Although the set of CPWs are the oscillating Gaussians in real space as shown in Eq. (5), CPWs can be relatively easily used in the Gaussian-based methods such as Crystal [5]. With CPWs, we can obtain high-energy states easily without being bothered with the choices of the Gaussian basis sets. The idea of CPWs might be a key to bridge the Gaussian bases and the PW bases. In addition, CPWs can directly used in the usual first-principles methods such as PP, LAPW, and PP methods. The real-space representation of CPWs may work for reducing the number of the bases when we treat a supercell with large vacuum region.

IV. DISCUSSION AND SUMMARY

The set of CPWs will be very useful to make interpolation in the BZ. For the matrix as a function of \( \mathbf{q} \) expanded in PWs/APWs, we can easily re-expand them in the set of CPWs since the transformation matrix between PWs and CPWs is explicitly given. Then the interpolation becomes easy because of the periodicity of CPWs in the BZ. This procedure via CPWs allows automatic interpolation which was difficult in the Wannier interpolation.

Since CPWs are the oscillating Gaussians in real space as shown in Eq. (5), CPWs can be relatively easily used in the Gaussian-based methods such as Crystal [5]. With CPWs, we can obtain high-energy states easily without being bothered with the choices of the Gaussian basis sets. The idea of CPWs might be a key to bridge the Gaussian bases and the PW bases. In addition, CPWs can directly used in the usual first-principles methods such as PP, LAPW, and PP methods. The real-space representation of CPWs may work for reducing the number of the bases when we treat a supercell with large vacuum region.

We did not yet get aware serious problems to implement CPWs to practical methods. Although the set of CPWs does not satisfy the translational symmetry completely, it will cause little problem since the low-energy part of the Hilbert space spanned by the PWs are well-reproduced. Determination of parameters to specify a set of CPWs is not difficult, thus it can be automatic without tuning by human.

ACKNOWLEDGMENTS

This is supported by JSPS KAKENHI Grant Number 17K05499. T. Kotani thanks to discussion with Prof. Hiroyuki Sakakibara. We also thank the computing time provided by Research Institute for Information Technology (Kyushu University).

[1] R. Martin, Electronic Structure: Basic Theory and Practical Methods (Cambridge University Press, New York, 2000).
[2] T. Kotani and M. van Schilfgaarde, Phys. Rev. B 81, 125117 (2010)
[3] T. Kotani, H. Kino, and H. Akai, J. Phys. Soc. Jpn. 84, 034702 (2015)
[4] A first-principles method package, ecalj is available from https://github.com/tkotani/ecalj/.
[5] The homepage of Crystal is at http://www.crystal.unito.it/.
[6] M. S. Hybertsen and S. G. Louie, Phys. Rev. B 34, 5390 (1986)
[7] D. R. Hamann and D. Vanderbilt, Physical Review B 79, 045109 (2009)
[8] T. Kotani, J. Phys. Soc. Jpn. 83, 094711 [11 Pages] (2014)
[9] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Phys. Rev. B 76, 165106 (2007)
[10] H. Sakakibara, T. Kotani, M. Obata, and T. Oda, Phys. Rev. B 101, 205120 (2020)
[11] D. Vanderbilt, Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators (Cambridge University Press, 2018).
[12] N. Marzari and D. Vanderbilt, Physical review B 56, 12847 (1997)
[13] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B 65, 035109 (2001)
[14] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Reviews of Modern Physics 84, 1419 (2012)
[15] C. Wang, S. Zhao, X. Guo, X. Ren, B.-L. Gu, Y. Xu, and W. Duan, 21, 093001 (2019)
FIG. 1. We plot \( \epsilon_m^m \) in Eq. (8) for the fcc empty lattice, that is, the band structure of empty lattice calculated with the set of CPWs \( \{ P_{q,n} \} \). We plot the exact empty bands \( \epsilon_{q,\text{Exact}} = \frac{\pi^2 q + G^2}{2m} \) with purple solid lines together. At low energy, we see that \( \epsilon_m^m \) clearly on the solid lines. Three panels are for \( \beta = 1.0, 1.6 \) and 2.0, while \( G_{\text{MAX}} = 4.0 \) Ry and \( \alpha = 0.8 \).
FIG. 2. Similar plots in Fig. 1. Three panels are for $\alpha = 0.6$, 0.8, and 1.0, while $G_{\text{MAX}} = 4.0$ Ry and $\alpha = 0.8$. Middle panel is the same as that in Fig. 1.
FIG. 3. Reproductivity of PWs. For \( \frac{k^2(q + G)^2}{2m} \) (y-axis) along X-Γ-L, we plot the projection weight \( Q(q + G) \). Lines for all branches are plotted. In this case, \( Q(q + G) \) below \( \sim 30 \) eV are almost one-hundred percent.
FIG. 4. Maximum error of eigenvalues and minimum value
of overlap matrix for varieties of \((\alpha, \beta)\). The top panel is for
\(G_{\text{MAX}} = 4\) which gives 58 number of bases as shown in its
title. Other panels as well. Horizontal and Vertical red lines
are guide for eye; right-bottom area suggests preferable \((\alpha, \beta)\).