Journal of the Physical Society of Japan **83**, 094711 (2014)
http://dx.doi.org/10.7566/JPSJ.83.094711

Quasiparticle Self-Consistent *GW* Method Based on the Augmented Plane-Wave and Muffin-Tin Orbital Method

Takao Kotani

Department of Applied Mathematics and Physics, Tottori University, Tottori 680-8552, Japan

(Received May 7, 2014; accepted June 12, 2014; published online August 13, 2014)

We have developed the quasiparticle self-consistent *GW* (QSGW) method based on a recently developed mixed basis all-electron full-potential method (PMT method), which uses the augmented plane waves (APWs) and the highly localized muffin-tin orbitals (MTOs) simultaneously. We call this PMT-QSGW. Because of the two types of augmented bases, we have an efficient description of one-particle eigenfunctions in materials with a small number of basis functions. In QSGW, we have to treat a static nonlocal exchange–correlation potential, which is generated from the self-energy. We expand the potential in the highly localized MTOs. This allows us to carry out stable interpolation of the self-energy in the whole Brillouin zone. In addition, we have improved the offset-Γ method for the Brillouin zone integration, so that we take into account the anisotropy of the screened Coulomb interaction in the calculation of the self-energy. For GaAs and cubic SiO_2, we checked the convergence of calculated band gaps on cutoff parameters. PMT-QSGW is implemented in a first-principles electronic structure package ecalj, which is freely available from github.

1. Introduction

The quasiparticle self-consistent *GW* method (QSGW) is a self-consistent perturbation method within the *GW* approximation. QSGW find out an optimum static one-body Hamiltonian $\hat{H}^0$ describing the independent-particle picture [or the quasiparticle (QP) picture]. In other words, QSGW divides the full many-body Hamiltonian $\hat{H}$ into $\hat{H} = \hat{H}^0 + (\hat{H} - \hat{H}^0)$. Then $(\hat{H} - \hat{H}^0)$ is chosen so that it virtually does not affect the determination of QPs. That is, we extract $\hat{H}^0$ as a kernel of $\hat{H}$. Note that $(\hat{H} - \hat{H}^0)$ should contain not only the bare Coulomb interaction but also quadratic term, which is missing in conventional model Hamiltonians. Since we evaluate $(\hat{H} - \hat{H}^0)$ in the *GW* approximation in QSGW, $\hat{H}^0$ (or the QPs, equivalently) is determined while taking into account the charge fluctuation in the random phase approximation (RPA) self-consistently. QSGW is conceptually completely different from the fully self-consistent *GW* method,[1–5] which tries to calculate the full one-body Green's function self-consistently.

QSGW was first introduced by Faleev, van Schilfgaarde, and Kotani.[6] It was implemented on the basis of the the all-electron full-potential linearized muffin-tin orbital (FP-LMTO) package[7] organized by van Schilfgaarde, in combination with the *GW* package developed by Kotani for Ref. 8, where he started from a detailed analysis of a *GW* package developed by Aryasetiawan[9–11] based on the LMTO in the atomic sphere approximation. We refer to the implementation as FP-LMTO-QSGW in the following. QSGW is now widely accepted as a possible candidate to go beyond the ability of current first-principles methods.[12,13] QSGW is also implemented in other first-principles electronic structure packages in different manners.[14–19] For example, Bruneval has calculated the ionization energies of atoms in QSGW.[17,18]

In QSGW, we have to treat the static nonlocal exchange–correlation potential $V^{\rm xc}(\mathbf{r}, \mathbf{r}')$ (the spin index is omitted for simplicity here). It is given by removing the energy dependence from the self-energy $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$ in the manner shown in Eq. (7). We determine eigenvalues and eigenfunctions with $V^{\rm xc}(\mathbf{r}, \mathbf{r}')$, with which we evaluate not only the diagonal but also the off-diagonal elements. The importance of the off-diagonal elements is seen especially in the dispersion crossing. We see that the conventional one-shot *GW* with only the diagonal self-energy cannot give a band gap for Ge as shown Fig. 6 of Ref. 20. This is because the connectivity of the dispersion in GGA/LDA cannot be altered. In contrast to this case of one-shot *GW*, the connectivity is correctly altered when we include the off-diagonal elements (or fully include the nonlocality).

To plot the energy-band dispersion in the whole Brillouin zone (BZ), we have to know nonlocal potential $V^{\rm xc}_{\mathbf{k}}(\mathbf{r}, \mathbf{r}')$ at any $\mathbf{k}$ point in the BZ by interpolation, where $\mathbf{r}$ and $\mathbf{r}'$ are within the primitive cell. This interpolation is also needed for the offset-Γ method in Sect. 3.2, and useful for calculating physical quantities which requires integrations in the BZ. For the interpolation, we inevitably require the real-space representation $V^{\rm xc}(\mathbf{r}, \mathbf{R}' + \mathbf{r}')$, where $\mathbf{R}'$ specifies the origins of primitive cells. With the inverse Fourier transformation, we obtain $V^{\rm xc}_{\mathbf{k}}(\mathbf{r}, \mathbf{r}')$ at any $\mathbf{k}$. The fact that MTOs are the localized functions enables us to carry out such interpolation in FP-LMTO-QSGW.[21]

Even though FP-LMTO-QSGW has been successfully applied to many cases,[22–27] it still has difficulties. The main difficulty originates from the FP-LMTO method, the applicability of which is limited to systems that can be described only by MTOs. Because of this, we have to fill empty regions with empty spheres (ESs) in cases such as loosely packed systems and slab models. Because of its ambiguity, we may need to repeat many calculations to check the numerical convergence. Therefore, it is not easy to apply QSGW to, for example, surfaces. In addition, it is not so easy to enlarge the basis set systematically as in the case of the LAPW method. Furthermore, the interpolation of $V^{\rm xc}_{\mathbf{k}}(\mathbf{r}, \mathbf{r}')$ is not stable enough because we need to use MTOs which are not localized well: used MTOs are specified by the damping factor $\propto \exp(-\kappa r)$ where we use not so small κ as $\kappa^{-2} \sim 0.1$ (bohr)$^{-2}$. This requires us to use a very complicated interpolation procedure.[21]

To overcome the difficulty in the FP-LMTO in DFT, we have recently introduced a new all-electron full-potential first-principles method of electronic structure calculations in the GGA/LDA (one-body problem solver).[28,29] This method

is named the linearized augmented plane wave and muffin-tin orbital method (PMT method), which is a mixed basis method for augmented waves, APWs and MTOs. Within our knowledge, there is no other mixed basis method of augmented waves. We can use a procedure to set the parameters of MTOs almost automatically, as given in Ref. 29 (see Fig. 1 and Table I of Ref. 29). The important point is that a serious difficulty in the FP-LMTO, i.e., how to set the parameters, has now been overcome. In the conventional FP-LMTO, we need to repeat many calculations to determine reasonable parameters of MTOs. In contrast, we can check the convergence only by changing the number of APWs. By using this procedure, we find that highly localized MTOs [$\propto \exp(-\kappa r)$ where $\kappa^{-2} = 1$–2 $(\text{bohr})^{-2}$] in combination with APWs whose cutoff energy $\sim$4 Ry can give good convergence of total energy in the GGA/LDA;[29] we successfully obtained atomization energy for homonuclear diatomic molecules, which is converged higher than the chemical accuracy ($\sim$1 kcal/mol). Thus, we can expand eigenfunctions with highly localized atom-centered MTOs and low-energy APWs.

In this paper, we show how to implement the QSGW method in the PMT method, that is, the PMT-QSGW method. After we explain the QSGW theory in Sect. 2, we explain the implementation of PMT-QSGW in Sect. 3. In particular in Sect. 3.2, we show a new improvement in the offset-Γ method, so that the method takes into the anisotropy of the screened Coulomb interaction accurately; in Sect. 3.3, we explain the interpolation of $V^{\text{xc}}_{\mathbf{k}}(\mathbf{r},\mathbf{r}')$. Finally, in Sect. 4, we show detailed numerical tests for the band gap (at the Γ point) for GaAs and cubic SiO_2 (β-cristobalite). We show how the QSGW band gap depends on the cutoff parameters.

2. Theory of the QSGW Method

Here we summarize the QSGW method. We treat the following many-body Hamiltonian for an electronic system. With the field operator $\hat{\psi}_\sigma(\mathbf{r})$, spin index σ, external potential $V^{\text{ext}}_\sigma(\mathbf{r})$, and Coulomb interaction $v(\mathbf{r},\mathbf{r}') = \frac{e^2}{|\mathbf{r}-\mathbf{r}'|}$, the Hamiltonian is written as

$$\hat{H} = \hat{H}^{\text{k}} + \hat{V}^{\text{ee}} + \hat{V}^{\text{ext}}, \tag{1}$$

$$\hat{H}^{\text{k}} = \sum_\sigma \int d\mathbf{r}\, \hat{\psi}^\dagger_\sigma(\mathbf{r})\left(-\frac{\nabla^2}{2m}\right)\hat{\psi}_\sigma(\mathbf{r}), \tag{2}$$

$$\hat{V}^{\text{ext}} = \sum_\sigma \int d\mathbf{r}\, V^{\text{ext}}_\sigma(\mathbf{r})\hat{n}_\sigma(\mathbf{r}), \tag{3}$$

$$\hat{V}^{\text{ee}} = \frac{e^2}{2}\sum_{\sigma\sigma'}\int d^3r\, d^3r'\, v(\mathbf{r},\mathbf{r}')\hat{\psi}^\dagger_\sigma(\mathbf{r})\hat{\psi}^\dagger_{\sigma'}(\mathbf{r}')\hat{\psi}_{\sigma'}(\mathbf{r}')\hat{\psi}_\sigma(\mathbf{r}). \tag{4}$$

Here we omit the classical electrostatic nucleus-nucleus energy for simplicity. We explicitly show the electron mass m and charge e^2 in the following formulas but with $\hbar = 1$. $V^{\text{ext}}(\mathbf{r})$ mainly contains those coming from nuclei, in addition to perturbations such as external magnetic fields. Hats on symbols indicate the second quantized quantities [for example, $\hat{V}^{\text{ext}}$ and $V^{\text{ext}}_\sigma(\mathbf{r})$ denote the same physical quantities in different representations]. In the following, we omit the spin index σ and often $\mathbf{r}$ for simplicity.

Let us consider how to obtain the best one-body Hamiltonian H^0 for describing QPs for a given $\hat{H}$. If we have the self-energy $\Sigma(\mathbf{r},\mathbf{r}',\omega)$ for a given $\hat{H}$, we can determine the QP energies and eigenfunctions as the solutions of

$$H(\epsilon_i)|\Psi_i(\mathbf{r})\rangle = \epsilon_i|\Psi_i(\mathbf{r})\rangle, \tag{5}$$

at least near the Fermi energy, where the one-particle dynamical effective Hamiltonian $H(\omega)$ is

$$H(\omega) = -\frac{\nabla^2}{2m} + V^{\text{ext}} + V^{\text{H}} + \Sigma(\omega). \tag{6}$$

Here V^{ext} is the external potential from the nuclei and V^{H} is the Hartree potential. If $H(\omega)$ were ω-independent and Hermitian, we could have directly identified this as H^0. Apparently, this is not true; however, on the basis of the Landau–Silin's QP theory (or the independent-particle picture), we can still expect that the physical properties will be successfully evaluated using the eigenvalues and eigenfunctions of QPs $\{\epsilon_i, \Psi_i(\mathbf{r})\}$. This means a physical picture that primary excitations are specified by adding electrons or holes for these orbitals; then they interact with each other by the screened Coulomb interaction. A theoretical inconvenience is that the set of QPs is not a complete set since $\Sigma(\omega)$ is energy-dependent and non-Hermitian. If we can use a static Hermitian one-body potential $V^{\text{xc}}(\mathbf{r},\mathbf{r}')$ in place of $\Sigma(\omega)$, such a problem will not occur. Then the set $\{\Psi_i(\mathbf{r})\}$ is an orthonormal complete set, and physical quantities can be represented in the Fock space of the set. In other words, we divide the full many-body Hamiltonian $\hat{H}$ into $\hat{H} = \hat{H}^0 + (\hat{H} - \hat{H}^0)$, where the many-body contribution due to $(\hat{H} - \hat{H}^0)$ should not change the QPs given by $\hat{H}^0$.

Following the above discussion, we need the following two methods should be combined for obtaining such $\hat{H}^0$ for a given $\hat{H}$.

(i) A method of calculating $\Sigma(\omega)$ (and also V^{H}) for a given division of $\hat{H} = \hat{H}^0 + (\hat{H} - \hat{H}^0)$.

(ii) A method of determining V^{xc} as a good substitution of a given $\Sigma(\omega)$.

If both methods (i) and (ii) are specified, we can close a self-consistent cycle. That is, we have the cycle $\hat{H}^0 \to \{V^{\text{H}}, \Sigma(\omega)\} \to \hat{H}^0 \to \cdots$; this cycle is repeated until converged. In QSGW, we use the GW approximation for (i). As for (ii), we use mapping from $\Sigma(\omega)$ to a static Hermitian potential $V^{\text{xc}}(\mathbf{r},\mathbf{r}')$ as

$$\begin{aligned} V^{\text{xc}} &= \frac{1}{2}\int_{-\infty}^{\infty} d\omega\, \mathcal{R}[\Sigma(\omega)]\delta(\omega - H^0) + \text{c.c.} \\ &= \sum_{ij}|\Psi_i\rangle\langle\Psi_i|\frac{\mathcal{R}[\Sigma(\varepsilon_i) + \Sigma(\varepsilon_j)]}{2}|\Psi_j\rangle\langle\Psi_j|, \end{aligned} \tag{7}$$

where $\mathcal{R}[X] = \frac{X+X^\dagger}{2}$ means taking the Hermitian part of X. Equation (7) is given so as to reproduce $\{\epsilon_i, \Psi_i(\mathbf{r})\}$ given by Eq. (5) as much as possible (this is not a unique choice[21]). If necessary, we can derive Eq. (7) from an approximate minimization of the difference $G^{-1} - (G^0)^{-1}$, where the difference is given as $\text{Tr}[(G^{-1} - (G^0)^{-1})\delta((G^0)^{-1})(G^{-1} - (G^0)^{-1})] + \text{c.c.}$[12] The GW approximation together with Eq. (7) results in a fundamental equation of QSGW.

Let us summarize the computational steps of the QSGW calculation. We start from a trial one-particle static Hamiltonian written as

$$H^0 = \frac{-\nabla^2}{2m} + V^{\text{eff}}(\mathbf{r},\mathbf{r}'). \tag{8}$$

This H^0 is just the initial condition for the iteration cycle and does not affect the final result. The GW method is applied to the division $\hat{H} = \hat{H}^0 + (\hat{H} - \hat{H}^0)$. Its steps are following (I)–(V).

(I) We have the non-interacting Green's function G^0 for a given H^0. It is

$$G^0(\mathbf{r}, \mathbf{r}', \omega) = \sum_i \frac{\Psi_i(\mathbf{r})\Psi_i^*(\mathbf{r}')}{\omega - \varepsilon_i \pm i\delta}, \tag{9}$$

where $\{\varepsilon_i, \Psi_i\}$ are eigenvalues and eigenfunctions of H^0; $+i\delta$ is for unoccupied states, and $-i\delta$ for occupied states.

(II) Calculate the dynamical screened Coulomb interaction W as

$$W = \epsilon^{-1}v = (1 - v\Pi)^{-1}v, \tag{10}$$

where we use the proper polarization $\Pi = -iG^0 \times G^0$.

(III) Calculate the self-energy $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$ as

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi}\int d\omega' \, G^0(\mathbf{r}, \mathbf{r}', \omega - \omega') \times W(\mathbf{r}, \mathbf{r}', \omega')e^{-i\delta\omega'}. \tag{11}$$

(IV) Simultaneously, we can calculate V^{H} for electron density from G^0. Together with $V^{\mathrm{ext}} + V^{\mathrm{H}}$, we have $H(\omega) = \frac{-\nabla^2}{2m} + V^{\mathrm{ext}} + V^{\mathrm{H}} + \Sigma(\omega)$.

(IV) From $\Sigma(\mathbf{r}, \mathbf{r}', \omega)$, we obtain V^{xc} through Eq. (7).

(V) For V^{xc}, we obtain a new H^0 as $H^0 = \frac{-\nabla^2}{2m} + V^{\mathrm{ext}} + V^{\mathrm{H}} + V^{\mathrm{xc}}$. From this, we start again from (I).

We repeat these steps until converged. In (I), we assume a non-interacting ground state by filling electrons up to the Fermi energy. The self-consistency ensures that this ground state is stable for the GW approximation.

We emphasize the ability of the nonlocality of the one-particle potential $V^{\mathrm{eff}}(\mathbf{r}, \mathbf{r}')$ to describe QPs, in comparison with ability of the local potential used in GGA/LDA. The nonlocality may be classified into two kinds of nonlocality. One is the onsite nonlocality which can be partially described by U of LDA+U. We may add such an onsite nonlocal potential to the exchange–correlation term to enhance the size of the orbital magnetic moment;[30] note that the local potential cannot break the time-reversal symmetry (resulting in too small orbital moment). The other is the offsite nonlocality that is important for giving the difference in eigenvalues between bonding and anti-bonding orbitals. For example, we can imagine a nonlocal potential that behaves as a projector to push down only eigenvalues of a bonding orbital. A local potential can hardly give this effect.

The QSGW method can be justified from the view point of the $GW\Gamma$ theory which takes into account the vertex Γ; Ishii, Maebashi, and Takada[31] analyzed the effects of the vertex Γ appeared in the calculations of the self-energy and of the polarization function. They claimed that the effect of Γ is virtually canceled out. We see an illustration of their claim in the renormalization factor of QPs, Z, contained in G; in the calculation of $\Sigma = G \times W \times \Gamma$, this Z is canceled out by the vertex Γ since the vertex is reduced to be $1/Z$ at $\mathbf{q}, \omega \to 0$. That is, we see the cancellation $Z \times 1/Z$ in $G \times \Gamma$.[21] This cancellation is generalized by the Ward identity, and they concluded that we should use $\Sigma = G_0 \times W$ rather than $\Sigma = G \times W$ if we neglect the vertex correction ($\Gamma = 1$). In the context of QSGW, we interpret their theory as "If we have a good H^0 for describing QPs, we can calculate good self-energy to determine QPs by $\Sigma = G_0 \times W$". Although they did not discuss how to obtain H^0, we think that QSGW is a possible candidate for determining such H^0. As for the polarization function, they also discussed about not using $\Pi = -iG \times G$ but $\Pi = -iG_0 \times G_0$ for the proper polarization. This is reasonable because $\Pi = -iG \times G$ contains the QP electron–hole excitations with too small weight $Z_{\mathrm{occupied}} \times Z_{\mathrm{unoccupied}}$. This discussion on the proper polarization is consistent with the fact that the first-principles calculations of dielectric functions with $\Pi = -iG_0 \times G_0$ show good agreement with experiments[32] (and the agreement is improved by taking into account the two-body correlations in the Bethe–Salpeter equation). On the other hand, numerical calculations by Bechstedt et al.[33] showed that the poorness of $\Pi = -iG \times G$ is corrected if we include the contribution of Γ to Π. This is consistent with our claim here. The discussions here support the QSGW rather than the full self-consistent GW methods.[1–5]

Let us consider two effects that are missing in the QSGW method. One is the effect that is missing in the GW method utilized in method (i) [or, almost equivalently, how to improve $W(\omega)$ in step (II)]. QSGW, for example, tends to give a slightly larger band gap than the experimental one,[21] which is traced back to a slightly strong $W(\omega)$ (slightly small screening effect) in the RPA. Thus, we need better $W(\omega)$ beyond the RPA. Along this line, some works have been performed until now: $W(\omega)$ including pair excitations,[15] $W(\omega)$ including phonons,[34] and $W(\omega)$ including a vertex correction.[31] The other missing effects are, for example, the contribution to the self-energy due to low-energy excitations such as magnetic fluctuations and phonons. Note that QSGW gives QPs, where charge fluctuation is already taken into account in the RPA self-consistently. Thus we expect that the main missing contribution comes from the low-energy excitations. If such contribution to the self-energy is taken into account, the QP dispersion near the Fermi energy can be deformed; a kink-like structure (mass enhancement) is added just near the Fermi energy[35] on top of the QP dispersion of QSGW as long as the effects due to such fluctuations are not very large. From the opposite viewpoint, this means that QSGW describes the overall feature of energy bands including the Fermi surface except such mass enhancement near the Fermi energy. Such low-energy part of the self-energy may be calculated with $W(\omega = 0)$ (neglecting energy dependence) in the many-body perturbation theory, although we need to avoid the double-counting problem of Feynman diagrams intrinsic in the first-principles many-body perturbation theory.[36] Not much research has been performed along this line, in contrast to the first-principles method combined with the dynamical mean field theory.[37]

Let us discuss the total energy in QSGW. Formally, the total energy can be given by an adiabatic connection, usually specified by the parameter λ changing from zero through unity as $\hat{H}^\lambda = \hat{H}^0 + (\lambda\hat{V}^{\mathrm{ee}} - \hat{V}^{\mathrm{eff}}_\lambda)$; this path starts from $\hat{H}^0$ at $\lambda = 0$ and ends with $\hat{H}$ at $\lambda = 1$ (note that $\hat{H}^0$ and $\hat{V}^{\mathrm{eff}}_{\lambda=1}$ are the second-quantized expressions of H^0 and V^{eff}, respectively). Along the path, $\hat{V}^{\mathrm{eff}}_\lambda$ is supposed to be chosen so that QSGW applied to $\hat{H}^\lambda$ gives $\hat{H}^0$ for any λ. Then the total energy is given as

$$E = E^0 + \int_0^1 d\lambda \frac{dE^\lambda}{d\lambda}$$
$$= E^0 + \int_0^1 d\lambda \langle 0_\lambda | \hat{V}^{\rm ee} | 0_\lambda \rangle - \int_0^1 d\lambda \langle 0_\lambda | \frac{\partial \hat{V}^{\rm eff}_\lambda}{\partial \lambda} | 0_\lambda \rangle, \quad (12)$$

where $|0_\lambda\rangle$ is the ground state for $\hat{H}^\lambda$. Equation (12) is an exact formula without approximation. As the lowest-order approximation, we replace $|0_\lambda\rangle$ with $|0_{\lambda=0}\rangle$. Then we have the Hartree–Fock energy calculated from the eigenfunctions of H^0. As a more accurate approximation, we evaluate $\langle 0_\lambda | \hat{V}^{\rm ee} | 0_\lambda \rangle$ in the random phase approximation (RPA); we apply it to $\hat{H}^0$ whose ground state is $|0\rangle$, with the interaction of $\lambda \hat{V}^{\rm ee}$. This gives the polarization function $\Pi(1 - \lambda v \Pi)^{-1}$, where $\Pi(\omega)$ is the polarization function of the non-interacting ground state $|0\rangle$. Then we have the RPA total energy

$$E^{\rm RPA} = E^{\rm k}_0 + E^{\rm ext}_0 + E^{\rm H}_0 + E^{\rm x}_0 + E^{\rm c}, \quad (13)$$

$$E^{\rm c} = \frac{-i}{2} {\rm Tr}[\log(1 - v\Pi) + v\Pi]. \quad (14)$$

The derivative of $E^{\rm RPA}$ with respect to the number of occupations for the orbital $\{\epsilon_i, \Psi_i\}$ gives

$$\frac{\partial E^{\rm RPA}}{\partial n_i} = \langle \Psi_i | -\frac{\nabla^2}{2m} + V^{\rm ext} + V^{\rm H} + \Sigma(\epsilon_i) | \Psi_i \rangle. \quad (15)$$

Since $\langle \Psi_i | \Sigma(\epsilon_i) | \Psi_i \rangle$ are the diagonal elements of Eq. (7), this change (derivative) in the RPA energy equals the QP energy given by the QSGW. In addition, the minimization of the right-hand side of Eq. (15) as a functional of Ψ_i gives Eq. (5) if we can neglect Ψ_i contained in $\Sigma(\epsilon_i)$. These show that the QSGW is related to the "RPA" total energy. We need to be cautious about the meaning of the QP energy ϵ_i. It is not the change of the total energy for one electron added/removed, but the derivative with respect to the number of occupancy. This is common to the case of Koopman–Slater–Janak's theorem. This is related to the localization–delocalization problem,[38] where we need to know how the eigenvalue ϵ_i changes as a function of fractional occupancy. In principle, we can calculate ionization energies by integration of ϵ_i.[17,18]

Originally the QSGW is proposed to treat solids; however, today we have requirement to treat molecules on the surface for such problems as catalysis. In the case of molecules (zero-dimensional systems), there are not only continuous eigenvalues but also discrete ones in H^0. Even in this case, Eq. (5) is the equation for determining eigenstates of the system. However, it is not trivial whether we can extract the independent-particle (or the QP) picture in the manner of QSGW. Only a limited number of publications on the QSGW applied to molecules are currently available,[17,18] and not so much has been clarified yet.

3. Implementation

In Sect. 3.1, we will give an overview of the method to perform the QSGW calculation. We made some improvements to the method in Refs. 8 and 21, where we took some ideas from another *GW* implementation given by Friedrich, Blügel, and Schindlmayr.[39]

In Sect. 3.2, we show a new improvement in the offset-Γ method, which is made in order to treat the $\mathbf{k} \to 0$ divergence of the integrand for the self-energy calculation. This improvement can correctly capture the anisotropy of the screened Coulomb interaction, although the previous offset-Γ method in FP-LMTO-QSGW[21] can be problematic for treating anisotropic systems.

In Sect. 3.3, we explain the interpolation procedure of $V^{\rm xc}_{\mathbf{k}}(\mathbf{r}, \mathbf{r}')$. The procedure is simplified in comparison with that used in FP-LMTO-QSGW.

3.1 Overview

In the PMT method,[28] the valence eigenfunctions for a given H^0 are represented in the linear combinations of the Bloch-summed MTOs $\chi^{\mathbf{k}}_{\mathbf{R}Lj}(\mathbf{r})$ and the APWs $\chi^{\mathbf{k}}_{\mathbf{G}}(\mathbf{r})$;

$$\Psi_{\mathbf{k}n}(\mathbf{r}) = \sum_{\mathbf{R}Lj} z^{\mathbf{k}n}_{\mathbf{R}Lj} \chi^{\mathbf{k}}_{\mathbf{R}Lj}(\mathbf{r}) + \sum_{\mathbf{G}} z^{\mathbf{k}n}_{\mathbf{G}} \chi^{\mathbf{k}}_{\mathbf{G}}(\mathbf{r}), \quad (16)$$

where we use indexes of the wave vector $\mathbf{k}$, band index n, and reciprocal lattice vector $\mathbf{G}$. The MTOs in the primitive cell are specified by the index of MT site $\mathbf{R}$, angular momentum $L = (l, m)$, and j for radial functions. As for core eigenfunctions, we calculate them under the condition that they are restricted within MTs. Then we consider the contributions of the cores only to the exchange part defined in Eq. (23) in the following. (In other words, we apply "core1" treatment in Ref. 21 for all cores.)

Let us explain the basis sets for eigenfunctions in the PMT method. In Ref. 28, we have tested a variety of basis sets of MTOs with APWs, whose numbers are specified by the APW cutoff energy $E^{\rm APW}_{\rm MAX}$. Then we propose a simple and systematic procedure for choosing the MTO basis sets in Ref. 29. By this procedure, we can perform stable and accurate calculations, where we use a large set of MTOs (two or three MTOs per L for valence electrons) together with APWs with low-cutoff energy, typically, ∼4 Ry. Owing to the APWs, we can include only highly localized MTOs. For the damping factors $\propto \exp(-\kappa r)$ contained in MTOs, we use $\kappa^2 = 1.0$ and 2.0 $({\rm bohr})^{-2}$. In Ref. 29, we have shown that it is not necessary to optimize the κ parameters if we use a sufficiently large $E^{\rm APW}_{\rm MAX}$ (∼4 Ry), as shown in Fig. 1 of Ref. 29. Other parameters for specifying MTOs are also fixed in a simple manner. The smoothing radii of the smooth Hankel functions, which are the envelope functions of MTOs, are set to be one-half of the MT radii. Thus MTOs are chosen essentially automatically, and the convergence is controlled only by $E^{\rm APW}_{\rm MAX}$. In addition, we do not need to use ESs because APWs are substituted for the MTO basis of ESs. We have shown that such a basis set works well in practice for determining the atomization energies of homonuclear dimers from H_2 through Kr_2 with the convergence of ∼1 kcal/mol or less in the DF calculation in the PBE exchange correlation functional in a large supercell.[28] Note that such supercell calculations are tough tests for augmented wave methods (FP-LAPW requires a very high $E^{\rm APW}_{\rm MAX}$ because of a small MT radius; it is not easy to apply FP-LMTO because of the difficulty in filling space by ESs). In comparison with methods using only the localized basis set such as `Gaussian` in quantum chemistry, the PMT method is advantageous in the point that it can describe scattering states (higher than zero level) accurately.

To perform the *GW* calculation, we first re-expand $\Psi_{\mathbf{k}n}(\mathbf{r})$ in Eq. (16) as the sum of the augmentation parts in MTs and the PW parts in the interstitial region:

$$\Psi_{\mathbf{k}n}(\mathbf{r}) = \sum_{\mathbf{R}u} \alpha^{\mathbf{k}n}_{\mathbf{R}u} \varphi^{\mathbf{k}}_{\mathbf{R}u}(\mathbf{r}) + \sum_{\mathbf{G}} \beta^{\mathbf{k}n}_{\mathbf{G}} P^{\mathbf{k}}_{\mathbf{G}}(\mathbf{r}), \tag{17}$$

where the interstitial plane wave (IPW) is defined as

$$P^{\mathbf{k}}_{\mathbf{G}}(\mathbf{r}) = \begin{cases} 0 & \text{if } \mathbf{r} \in \text{any MT} \\ \exp(i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}) & \text{otherwise} \end{cases} \tag{18}$$

and $\varphi^{\mathbf{k}}_{Ru}(\mathbf{r})$ are Bloch sums of the atomic functions $\varphi_{Ru}(\mathbf{r})$ defined within the MT at R,

$$\varphi^{\mathbf{k}}_{Ru}(\mathbf{r}) \equiv \sum_{\mathbf{T}} \varphi_{Ru}(\mathbf{r}-\mathbf{R}-\mathbf{T}) \exp(i\mathbf{k}\cdot\mathbf{T}). \tag{19}$$

$\mathbf{T}$ and $\mathbf{G}$ are lattice translation vectors in real and reciprocal spaces, respectively.

In the *GW* calculation, we need not only the basis set for eigenfunctions, but also the basis set for expanding the product of eigenfunctions. The basis is called the mixed product basis (MPB) $\{M^{\mathbf{k}}_I(\mathbf{r})\}$ first introduced in Ref. 8 by Kotani. The MPB consists of the product basis (PB) within MTs[10] and the IPW in the interstitial region. Since $\{M^{\mathbf{k}}_I(\mathbf{r})\}$ contains IPWs which are not orthogonal, we define dual for $\{M^{\mathbf{k}}_I(\mathbf{r})\}$ as

$$|\tilde{M}^{\mathbf{k}}_I\rangle \equiv \sum_{I'} |M^{\mathbf{k}}_{I'}\rangle (O^{\mathbf{k}})^{-1}_{I'I}, \tag{20}$$

$$O^{\mathbf{k}}_{I'I} = \langle M^{\mathbf{k}}_{I'} | M^{\mathbf{k}}_I \rangle. \tag{21}$$

From $v^{\mathbf{k}}_{IJ} = \langle M^{\mathbf{k}}_I | v | M^{\mathbf{k}}_J \rangle$, we calculate the eigenfunction for the generalized eigenvalue problem defined by $\sum_J (v^{\mathbf{k}}_{IJ} - v^{\mathbf{k}}_\mu O^{\mathbf{k}}_{IJ}) w^{\mathbf{k}}_{\mu J} = 0$, where $v_\mu(\mathbf{k})$ are the eigenvalues of the Coulomb interaction matrix. Then we have the Coulomb interaction represented by matrix elements as

$$v(\mathbf{k}) = \sum_\mu |E^{\mathbf{k}}_\mu\rangle v_\mu(\mathbf{k}) \langle E^{\mathbf{k}}_\mu|, \tag{22}$$

where we define a new MPB $|E^{\mathbf{k}}_\mu(\mathbf{r})\rangle = \sum_J |M^{\mathbf{k}}_J\rangle w^{\mathbf{k}}_{\mu J}$, which is orthonormal and is diagonal to the Coulomb interaction $v(\mathbf{k})$. For the all-electron full-potential *GW* approximation, Eq. (22) is introduced in Ref. 39. This corresponds to the representation in the plane wave expansion $v(\mathbf{k}+\mathbf{G}, \mathbf{k}+\mathbf{G}') = \frac{4\pi\delta_{\mathbf{G}\mathbf{G}'}}{|\mathbf{k}+\mathbf{G}|^2}$. $\mu = 1$ corresponds to the largest eigenvalue of v_μ, and $v_{\mu=1}$ is $\sim \frac{4\pi e^2}{|\mathbf{k}|^2}$, which is related to the divergent term discussed in Sect. 3.2.

With the definition of $\langle A|B\rangle = \int d^3r\, A^*(\mathbf{r}) B(\mathbf{r})$, the exchange part of $\Sigma(\omega)$ is written as

$$\Sigma^{\mathrm{x}}_{nm}(\mathbf{q}) = \langle \Psi_{\mathbf{q}n} | \Sigma_{\mathrm{x}} | \Psi_{\mathbf{q}m} \rangle = -\sum_{\mathbf{k}}^{\mathrm{BZ}} \sum_{n'}^{\mathrm{occ}} \langle \Psi_{\mathbf{q}n} | \Psi_{\mathbf{q}-\mathbf{k}n'} E^{\mathbf{k}}_\mu \rangle v_\mu(\mathbf{k}) \langle E^{\mathbf{k}}_\mu \Psi_{\mathbf{q}-\mathbf{k}n'} | \Psi_{\mathbf{q}m} \rangle. \tag{23}$$

The screened Coulomb interaction $W(\omega)$ is calculated using Eq. (10), where the polarization function $\Pi(\omega)$ is written as

$$\Pi_{\mu\nu}(\mathbf{q},\omega) = \sum_{\mathbf{k}}^{\mathrm{BZ}} \sum_{n}^{\mathrm{occ}} \sum_{n'}^{\mathrm{unocc}} \frac{\langle E^{\mathbf{q}}_\mu \Psi_{\mathbf{k}n} | \Psi_{\mathbf{q}+\mathbf{k}n'} \rangle \langle \Psi_{\mathbf{q}+\mathbf{k}n'} | \Psi_{\mathbf{k}n} E^{\mathbf{q}}_\nu \rangle}{\omega - (\varepsilon_{\mathbf{q}+\mathbf{k}n'} - \varepsilon_{\mathbf{k}n}) + i\delta} + \sum_{\mathbf{k}}^{\mathrm{BZ}} \sum_{n}^{\mathrm{unocc}} \sum_{n'}^{\mathrm{occ}} \frac{\langle E^{\mathbf{q}}_\mu \Psi_{\mathbf{k}n} | \Psi_{\mathbf{q}+\mathbf{k}n'} \rangle \langle \Psi_{\mathbf{q}+\mathbf{k}n'} | \Psi_{\mathbf{k}n} E^{\mathbf{q}}_\nu \rangle}{-\omega - (\varepsilon_{\mathbf{k}n} - \varepsilon_{\mathbf{q}+\mathbf{k}n'}) + i\delta}. \tag{24}$$

When time-reversal symmetry is assumed, $\Pi(\omega)$ can be simplified to read

$$\Pi_{\mu\nu}(\mathbf{q},\omega) = \sum_{\mathbf{k}}^{\mathrm{BZ}} \sum_{n}^{\mathrm{occ}} \sum_{n'}^{\mathrm{unocc}} \langle E^{\mathbf{q}}_\mu \Psi_{\mathbf{k}n} | \Psi_{\mathbf{q}+\mathbf{k}n'} \rangle \langle \Psi_{\mathbf{q}+\mathbf{k}n'} | \Psi_{\mathbf{k}n} E^{\mathbf{q}}_\nu \rangle \times \left(\frac{1}{\omega - \varepsilon_{\mathbf{q}+\mathbf{k}n'} + \varepsilon_{\mathbf{k}n} + i\delta} - \frac{1}{\omega + \varepsilon_{\mathbf{q}+\mathbf{k}n'} - \varepsilon_{\mathbf{k}n} - i\delta} \right). \tag{25}$$

To evaluate Eq. (24) or (25), we first accumulate the imaginary parts (anti-Hermitian part) of $\Pi_{\mu\nu}(\mathbf{q},\omega)$ along bins of histograms on the real axis ω by the tetrahedron technique,[40] and then determine the real part via the Hilbert transformation. The bins are dense near the Fermi energy and coarser at higher energy as described in Ref. 21. This procedure is not only more efficient but also safer than the methods of calculating the real part directly. We also use the extended irreducible zone (EIBZ) symmetrization procedure described in Ref. 39.

The correlation part of the screened Coulomb interaction $W^{\mathrm{c}}(\omega) = W(\omega) - v$, which is calculated from v and $\Pi(\omega)$, is given as

$$W^{\mathrm{c}}(\mathbf{k},\omega) = \sum_{\mu\nu} |E^{\mathbf{k}}_\mu\rangle W^{\mathrm{c}}_{\mu\nu}(\mathbf{k},\omega) \langle E^{\mathbf{k}}_\mu|. \tag{26}$$

With this $W^{\mathrm{c}}(\mathbf{k},\omega)$, we have the correlation part of the self-energy as

$$\Sigma^{\mathrm{c}}_{n,n'}(\mathbf{q},\omega) = \sum_{\mathbf{k},m} \int_{-\infty}^{\infty} d\omega' \sum_{\mu,\nu} \frac{\langle \Psi_{\mathbf{q}n} | \Psi_{\mathbf{q}-\mathbf{k}m} E^{\mathbf{k}}_\mu \rangle W^{\mathrm{c}}_{\mu\nu}(\mathbf{k},\omega') \langle E^{\mathbf{k}}_\nu \Psi_{\mathbf{q}-\mathbf{k}m} | \Psi_{\mathbf{q}n'} \rangle e^{-i\delta\omega'}}{\omega - \omega' - \epsilon_{\mathbf{q}-\mathbf{k}m} \pm i\delta}. \tag{27}$$

Here, we use $-i\delta$ for occupied states of $\mathbf{q}-\mathbf{k}m$, and $+i\delta$ for unoccupied states. In QSGW, we have to calculate the Hermitian part of $\Sigma_{nn'}(\mathbf{q}, \epsilon_{\mathbf{q}n})$, to obtain $V^{\mathrm{xc}}_{\mathbf{q}}$ using Eq. (7).

There are two key points to handle the *GW* procedure given above. The first key point, given in Sect. 3.2, is the improved offset-Γ method, which treats the divergence of $W^{\mathrm{c}}(\mathbf{k}\to 0,\omega)$ in Eq. (27). For this purpose, we define the non-divergent effective interaction $\overline{W^{\mathrm{c}}}(\mathbf{k}=0,\omega)$ instead of $W^{\mathrm{c}}(\mathbf{k}=0,\omega)$. Then we can take a simple discrete sum for both expressions of Eqs. (23) and (27).

The second point in Sect. 3.3 is how to perform an interpolation to give $V^{\mathrm{xc}}_{\mathbf{q}}$ at any $\mathbf{q}$ in the whole BZ, from $V^{\mathrm{xc}}_{\mathbf{q}}$ calculated only at limited numbers of $\mathbf{q}$ points. This is required in the offset-Γ method shown in Sect. 3.2, that is, we have to calculate eigenfunctions at some $\mathbf{q}$ points near $\mathbf{q}=0$. For the interpolation, we expand the static nonlocal potential V^{xc} in Eq. (7) in highly localized MTOs in real space. Thus such MTOs are used for two purposes: one as the basis of the eigenfunctions; and two as the basis of expanding V^{xc}. The interpolation procedure of $V^{\mathrm{xc}}_{\mathbf{k}}(\mathbf{r},\mathbf{r}')$ becomes stabler and simpler than the complicated interpolation procedure in Ref. 21. This is because we now use highly localized MTOs. In the planewave-based QSGW method by Hamann and Vanderbilt,[14] they expand V^{xc} in the maximally localized Wannier functions instead of MTOs.

In practical implementation, the LDA or GGA exchange–correlation potential $V_{\rm LDA}^{\rm xc}$ is used to perform efficient numerical calculations. That is, it is used in order to generate core eigenfunctions as well as radial functions within MTs (in this paper, we use the subscript LDA even when we use GGA. "LDA/GGA" means LDA or GGA). The difference $V^{\rm xc} - V_{\rm LDA}^{\rm xc}$ is used for the interpolation in the BZ (explained in Sect. 3.3), because this difference is numerically small as long as $V_{\rm LDA}^{\rm xc}$ roughly gives an approximation to $V^{\rm xc}$. This procedure utilizing $V_{\rm LDA}^{\rm xc}$ to perform efficient numerical calculations give a very weak dependence to the final numerical results in practice as seen in Sect. 4, although the results formally does not depend on the LDA/GGA exchange–correlation functions.

3.2 *Improved offset-Γ method*

The offset-Γ method, originally invented for Ref. 8 by Kotani (it is described in Ref. 21), is a key to perform accurate *GW* calculation. It is for the integration of $\mathbf{k}$ in Eqs. (23) and (27), where we have the integrands that diverge at $\mathbf{k} \to 0$. It works well for highly symmetric systems; however, it can be problematic to apply to less symmetric systems, because the anisotropic divergence of the integrands may not be treated accurately. Here we show an improved offset-Γ method, which treats the anisotropy of $W(\mathbf{k}, \omega)$ accurately. In the following, we use expression $W(\mathbf{k})$ for simplicity (omit subscripts and ω) instead of $W_{\mu\nu}(\mathbf{k}, \omega)$, since we are concerned with the $\mathbf{k}$ integral.

Let us give a formula for calculating $\int_{\rm BZ} f(\mathbf{k})\, d^3k$ using a discrete sum on $\mathbf{k}$-mesh, where $f(\mathbf{k}) = G(\mathbf{q} - \mathbf{k}) \times W(\mathbf{k})$. As the $\mathbf{k}$-mesh, we use

$$\mathbf{k}(i_1, i_2, i_3) = 2\pi\left(\frac{i_1}{N_1}\mathbf{b}_1 + \frac{i_2}{N_2}\mathbf{b}_2 + \frac{i_3}{N_3}\mathbf{b}_3\right),$$

where $\mathbf{b}_1, \mathbf{b}_2$, and $\mathbf{b}_3$ are the primitive reciprocal vectors [the same as the Eq. (47) in Ref. 21]. The 1st BZ is divided into $N = N_1 \times N_2 \times N_3$ microcells $(i_1 = 0, 1, \ldots, N_1 - 1,$ and also the same for i_2 and $i_3)$. The microcell including the Γ point is called the Γ cell.[41] The main problem is how to evaluate the contribution of the Γ cell. The divergent part of $f(\mathbf{k})$ behaves $\approx$ (analytic function of $\mathbf{k}$)/$(\mathbf{k}^{\rm T}\mathbf{L}\mathbf{k})$, where $\mathbf{k}^{\rm T}$ denotes the transpose of $\mathbf{k}$; $\mathbf{L}$ is a 3×3 Hermitian matrix.[39] We neglect an odd part of $\mathbf{k}$ in the above (analytic function of $\mathbf{k}$) because it has no contribution to the integral around $\mathbf{k} = 0$. Thus it is sufficient to consider the integral for $f(\mathbf{k})$ whose divergent parts behave as $f(\mathbf{k}) = \sum_L \frac{f_L Y_L(\hat{\mathbf{k}})}{|\mathbf{k}|^2}$ at $\mathbf{k} \to 0$, where l of $L \equiv (l, m)$ is restricted to be an even number. We evaluate the integral using the formula

$$\int_{\rm BZ} f(\mathbf{k})\, d^3k \approx \frac{1}{N}\sum^{\mathbf{k}\neq 0} f(\mathbf{k}) + \sum_L f_L w_L + \frac{1}{N}\tilde{f}, \qquad (28)$$

which is introduced in Ref. 41. Here the weight w_L is determined in a manner as follows, so as to take into account the contributions of the divergent part of $f(\mathbf{k})$ at $\mathbf{k} \to 0$ in the Γ cell. $\tilde{f}$ is the constant part of $f(\mathbf{k})$ at $\mathbf{k} \to 0$.

To determine w_L, we can use the following procedure instead of that given in Ref. 41. We first introduce the auxiliary function

$$F_L(\mathbf{k}) = \sum_{\mathbf{G}} \frac{\exp(-\alpha|\mathbf{k} - \mathbf{G}|^2) Y_L(\widehat{\mathbf{k} - \mathbf{G}})}{|\mathbf{k} - \mathbf{G}|^2}. \qquad (29)$$

This is a generalization of an auxiliary function used in the offset-Γ method (then we only used F_{00}[21]). We usually take the $\alpha \to 0$ limit, or a sufficiently small α instead. Let us apply Eq. (28) to $F_L(\mathbf{k})$. Then we can evaluate the left-hand side of Eq. (28) exactly [the exact values are zero except for $L = (0, 0)$]. On the other hand, the first and third terms on the right-hand side of Eq. (28) can be evaluated numerically. In addition, we know that $f_{L'}$ for $F_L(\mathbf{k})$ is unity for $L' = L$, and zero otherwise. Thus we can determine w_L in Eq. (28) so that Eq. (28) is exactly satisfied for $F_L(\mathbf{k})$ for any L.

Let us apply Eq. (28) to $f(\mathbf{k}) = G(\mathbf{q} - \mathbf{k}) \times W(\mathbf{k})$. Then we perform an approximation taking only the most divergent term in $W(\mathbf{k})$ in addition to its analytic part. That is, we use

$$W_{\mu\nu}(\mathbf{k}) \sim \widetilde{W}_{\mu\nu}(\mathbf{0}) + \frac{4\pi}{\mathbf{k}^{\rm T}\mathbf{L}\mathbf{k}}\delta_{1\mu}\delta_{1\nu} \qquad (30)$$

at $\mathbf{k} \to 0$. $\widetilde{W}_{\mu\nu}(\mathbf{0}) = 0$ for $\mu = 1$ or $\nu = 1$. See Eq. (36) in Ref. 39 to know what is neglected in the approximation of Eq. (30).

Then we finally obtain

$$\int_{\rm BZ} d^3k\, G(\mathbf{q} - \mathbf{k}) W(\mathbf{k}) \approx \overline{\sum G(\mathbf{q} - \mathbf{k}) W(\mathbf{k})}, \qquad (31)$$

where its right-hand side is defined as

$$\overline{\sum G(\mathbf{q} - \mathbf{k}) W(\mathbf{k})} \equiv \frac{1}{N}\sum_{\mathbf{k}\neq 0} G(\mathbf{q} - \mathbf{k}) W(\mathbf{k}) + \frac{1}{N} G(\mathbf{q})\overline{W}(\mathbf{0}), \qquad (32)$$

$$\overline{W}(\mathbf{0}) \equiv N\sum w_L W_L + \widetilde{W}(\mathbf{0}). \qquad (33)$$

Here $\overline{W}(\mathbf{0})$ can be taken as a kind of average of W in the Γ cell. With this $\overline{W}(\mathbf{0})$, we can evaluate integrals just as the sum on the discrete $\mathbf{k}$-mesh. When the matrix $\mathbf{L}$ is given (a method of calculating $\mathbf{L}$ is given in the next paragraph), the non-analytic (but non-divergent) function $\mathbf{k}^{\rm T}\mathbf{L}\mathbf{k}/|\mathbf{k}^2|$ is expanded in the spherical harmonics. Then W_L is calculated for a given $\mathbf{L}$ in the manner shown in Ref. 39. We can evaluate the accuracy of integrals with a discrete $\mathbf{k}$-mesh in combination with the approximation of Eq. (30) by calculations while changing the size of the $\mathbf{k}$-mesh.

The remaining problem is how to calculate the matrix $\mathbf{L}$ in Eq. (30); there are two possible ways. One is the $\mathbf{k}\cdot\mathbf{p}$ method (perturbation) used in Ref. 39; the other is the numerical method to calculate $\mathbf{L}$ at some $\mathbf{k}$ points near $\mathbf{k} = 0$. Here we use the latter method. Because of the point-group symmetry of the system, $\mathbf{L}$ can be expressed by the linear combination of invariant tensors μ_{ij}^g for the symmetry of the unit cell,

$$L_{ij}(\omega) = \sum_{g=1}^{N_g} a_g(\omega)\mu_{ij}^g, \qquad (34)$$

where g is the index of the invariant tensor. The number of g's N_g, can be from one (cubic symmetry) through six (no symmetry). It is possible to determine the coefficient $a_g(\omega)$ from the dielectric functions $\hat{\mathbf{k}}_{0i}^{\rm T}\mathbf{L}\hat{\mathbf{k}}_{0i}$ calculated at $\{\mathbf{k}_{0i}\}$ points around $\mathbf{k} = 0$, where $\{\mathbf{k}_{0i}; i = 1, N_g\}$ is a set of the offset-Γ points. The offset-Γ points are chosen so that the conversion matrix from $\hat{\mathbf{k}}_{0i}^{\rm T}\mathbf{L}(\omega)\hat{\mathbf{k}}^{0i}$ to $a_g(\omega)$ is not numerically degenerated. The length $|\mathbf{k}^{0i}|$ can be chosen to be sufficiently enough, but avoiding numerical error as the average of $W(\mathbf{k})$ in the Γ cell. The improved offset-Γ method

shown here can be applicable even to metal cases, as long as $\hat{\mathbf{k}}_{0i}^{\mathrm{T}}\mathbf{L}(\omega)\hat{\mathbf{k}}_{0i}$ contains the contribution of intraband transition.

3.3 *Interpolation of the self-energy in the Brillouin zone*

Here we show an interpolation procedure for giving $V_{\mathbf{k}}^{\mathrm{xc}}$ at any $\mathbf{k}$, from $V_{\mathbf{k}}^{\mathrm{xc}}$ calculated only at the regular mesh points $\mathbf{k}(i_1, i_2, i_3)$. This interpolation is used for the offset-Γ method that requires $W(\omega)$ at $\{\mathbf{k}_{0i}\}$; to calculate this $W(\omega)$, we need eigenfunctions and eigenvalues not only at the regular mesh points $\mathbf{k}(i_1, i_2, i_3)$ but also at $\mathbf{k}(i_1, i_2, i_3) + \mathbf{k}_{0i}$. This interpolation is also useful for plotting energy bands. A key point of the interpolation is that V^{xc} is expanded in real space in highly localized MTOs as follows.

At the end of step (IV) in Sect. 2, we obtain the matrix elements $\langle\Psi_{\mathbf{k}n}|\Delta V_{\mathbf{k}}^{\mathrm{xc}}|\Psi_{\mathbf{k}m}\rangle$ on the regular mesh points of $\mathbf{k}$, where $\Delta V_{\mathbf{k}}^{\mathrm{xc}} = V_{\mathbf{k}}^{\mathrm{xc}} - V_{\mathbf{k}}^{\mathrm{xc,LDA}}$. Then it is converted to the representation in the APW and MTO bases as

$$\langle\chi_a^{\mathbf{k}}|\Delta V_{\mathbf{k}}^{\mathrm{xc}}|\chi_b^{\mathbf{k}}\rangle = \sum_{n,m}(z^{-1})_{an}^{*}\langle\Psi_{\mathbf{k}n}|\Delta V_{\mathbf{k}}^{\mathrm{xc}}|\Psi_{\mathbf{k}m}\rangle z_{bm}^{-1}, \qquad (35)$$

where we use the simplified basis index a, which is the index for specifying a basis ($\mathbf{R}Lj$ for MTO or $\mathbf{G}$ for APW). Thus $\chi_a^{\mathbf{k}}$ denotes the APWs or MTOs in Eq. (16); z_{na} ($\mathbf{k}$ is omitted for simplicity) denotes the coefficients of the eigenfunctions at $\mathbf{k}$, that is, $z_{\mathbf{R}Lj}^{\mathbf{k}n}$ and $z_{\mathbf{G}}^{\mathbf{k}n}$ in Eq. (16) together. This z_{an} is identified as a conversion matrix that connects eigenfunctions (band index n) and the APW and MTO bases (basis index a).

To obtain real-space representation of ΔV^{xc}, we need a representation expanded in the basis that consist of the Bloch-summed localized orbitals, which are periodic for $\mathbf{k}$ in the BZ. However, this is not the case for the APWs in Eq. (35). To overcome this problem, we use an approximation in which we only take the matrix elements related to MTOs, that is, the elements $\langle\chi_a^{\mathbf{k}}|\Delta V_{\mathbf{k}}^{\mathrm{xc}}|\chi_b^{\mathbf{k}}\rangle$ where a and b specify MTOs. This means that the part of ΔV^{xc} related to APWs is projected onto the basis of MTOs. This approximation can be reasonable as long as the main part of ΔV^{xc} can be well expanded in MTOs, although we need numerical tests to confirm the accuracy as shown in Sect. 4. Then we obtain a real-space representation of ΔV^{xc} expanded in MTOs from the MTO part of $\langle\chi_a^{\mathbf{k}}|\Delta V_{\mathbf{k}}^{\mathrm{xc}}|\chi_b^{\mathbf{k}}\rangle$ by Fourier transformation. Then we can have interpolated ΔV^{xc} at any $\mathbf{k}$ by inverse Fourier transformation. Since we use highly localized MTOs, this interpolation is more stable than the previous one in FP-LMTO-QSGW.[21] The complicated interpolation procedure given in Sect. II-G in Ref. 21 is no longer necessary.

To reduce the computational time, we calculate the matrix elements $\langle\Psi_{\mathbf{k}n}|\Delta V_{\mathbf{k}}^{\mathrm{xc}}|\Psi_{\mathbf{k}m}\rangle$ only up to the states whose eigenvalues are less than $E_{\mathrm{MAX}}^{\Sigma}$. Then the high energy parts of the matrix elements are assumed to be diagonal, where their values are given by a simple average of calculated diagonal elements.

4. Numerical Test

Here we show the results of test calculations for PMT-QSGW applied to two examples, GaAs and cubic SiO_2 (β-cristobalite, denoted as SiO2c hereafter). The latter has large interstitial regions; it has the same structure as Si but oxygen atoms are located in the middle of Si–Si bonds. We use the lattice constants 5.653 Å for GaAs and 7.165 Å for SiO2c.

Table I. MTOs used for GaAs and SiO2c (β-cristobalite). These are specified by the principal quantum number and angular momentum. The MTO's envelope functions are the smooth Hankel functions, which are specified by two parameters, that is, the damping factor κ and the smoothing radius R_{sm}. We set the parameters in the manner shown in Ref. 28 (we currently use $\kappa^2 = 1.0$ and $\kappa^2 = 2.0$ for all kinds of atoms). R_{sm} is given to be one-half of the MT radius R_{MT}, which is shown in the unit of bohr radius. Empty spheres (ESs) are located in the middle of the interstitial region (two ESs per primitive cell in both GaAs and SiO2c). ESs are used only in cases specified by "vwn,es" and "pbe,es" in II and IV. The unit of κ^2 is in $(\mathrm{bohr})^{-2}$.

	Valence	R_{MT}
GaAs		
Ga	3d(lo), 4s4p4d4f($\kappa^2 = 1.0$), 4s4p4d($\kappa^2 = 2.0$)	2.19
As	3d(lo), 4s4p4d4f($\kappa^2 = 1.0$), 4s4p4d($\kappa^2 = 2.0$)	2.30
(ES)	1s2p3d($\kappa^2 = 1.0$), 1s2p($\kappa^2 = 2.0$)	2.80
SiO2c (two Si and four O in a primitive cell)		
Si	4s4p4d4f($\kappa^2 = 1.0$), 4s4p4d($\kappa^2 = 2.0$)	2.19
O	2s3p4d($\kappa^2 = 1.0$), 2s3p4d($\kappa^2 = 2.0$)	2.30
(ES)	1s2p3d4f($\kappa^2 = 1.0$)	2.80

Table II. Band gap (eV) in LDA/GGA for checking the convergence on the basis set. For sets of MTOs shown in Table I, we tabulate the calculated band gaps for $E_{\mathrm{MAX}}^{\mathrm{APW}}$. Note that "vwn,es" and "pbe,es" means cases in which ESs are included. We can see that band gaps converge well with a small number of APWs; this is consistent with the case of atomization energies.[29] In the GGA case with ESs, the convergence behavior becomes slightly unstable (not converged for 6.0 Ry for SiO2c), because of numerical instability due to the poor linear-dependency of the basis set. The number n_{APW}, determined by $E_{\mathrm{MAX}}^{\mathrm{APW}}$, means number of APWs at $\mathbf{k} = 0$. The number of MTOs except ESs are 60 for GaAs and 168 for SiO2c.

$E_{\mathrm{MAX}}^{\mathrm{APW}}$ (Ry)	vwn	vwn,es	pbe	pbe,es	n_{APW}
GaAs					
0.0	0.425	0.308	0.665	0.541	0
1.0	0.322	0.295	0.558	0.528	1
2.0	0.294	0.294	0.526	0.529	15
3.0	0.294	0.294	0.528	0.532	27
4.0	0.294	0.294	0.530	0.535	51
5.0	0.294	0.294	0.530	0.536	59
6.0	0.294	0.294	0.530	0.538	65
SiO2c					
0.0	8.560	6.131	8.592	6.186	0
1.0	5.406	5.434	5.663	5.563	15
2.0	5.437	5.445	5.670	5.622	27
3.0	5.442	5.445	5.665	5.652	59
4.0	5.444	5.446	5.665	5.668	65
5.0	5.446	5.447	5.668	5.658	113
6.0	5.446	5.445	5.669	—	169

We perform calculations with different settings to show the convergence properties of band gaps. We use a simple and systematic procedure to determine sets of MTOs and APWs, as explained after Eq. (16). We use the MTOs shown in Table I. As for Ga(3d) and As(3d), we use the local orbitals shown as "lo".[42]

Before showing the band gaps calculated by QSGW, let us show those calculated by LDA/GGA in Table II. We can check the convergence behavior by changing the APW cutoff energy $E_{\mathrm{MAX}}^{\mathrm{APW}}$. For the functional of LDA, we employ the VWN exchange–correlation functional;[43] for GGA, we employ the PBE.[44] The labels "vwn,es" and "pbe,es" mean cases in which ESs are included. The convergence behavior is satisfactory, as in the case of total energy for homonuclear dimers.[29] We see better convergence behavior for "vwn,es" and "pbe,es" than "vwn" and "pbe", since we have larger

number of bases. For example, "vwn,es" for GaAs showing 0.295 eV for $E^{\rm APW}_{\rm MAX} = 1$ Ry is essentially the same as the converged value of 0.294 eV, while "vwn" requires $E^{\rm APW}_{\rm MAX} \gtrsim 2$ Ry to have a similar convergence. For SiO2c, the convergence is slightly slower because SiO2c has large interstitial region, e.g., the band gap of 5.437 eV for "vwn" at $E^{\rm APW}_{\rm MAX} \gtrsim 2$ Ry shows ~0.01 eV difference from the converged value of 5.445 eV (we took the case of "vwn,es" at $E^{\rm APW}_{\rm MAX} = 6$ Ry). Within this small error, we can determine the band gap even without ESs. This confirms our expectation that the missing part of the Hilbert space spanned by highly localized MTOs [large damping factors $\kappa^2 = 1.0$ and 2.0 $(\text{bohr})^{-2}$] is complemented by APWs with such a very low $E^{\rm APW}_{\rm MAX}$. The wave number of the cutoff roughly corresponds to the distance between nearest-neighbor atoms. We observe slight instability (we need many iterations) in the calculations when we use $E^{\rm APW}_{\rm MAX} \gtrsim 5$ Ry in the case of "pbe,es", since GGA requires better numerical accuracy for calculating derivative of density. This is because of the over-completeness problem of the basis set, that is, we lose linear-independency of basis functions for a large $E^{\rm APW}_{\rm MAX}$. We conclude that Table II gives a satisfactory convergence behavior within this limitation.

Let us summarize the settings (and parameters) for performing the PMT-QSGW calculations. These can be classified as follows.

(A) IPW cutoff $|\mathbf{q}+\mathbf{G}|^{\Psi}_{\rm Max}$ to give allowed $P^{\mathbf{q}}_{\mathbf{G}}(\mathbf{r})$ in the expansion of eigenfunctions, Eq. (17).
(B) Settings of the MPB. We have parameters for specifying product basis (PB) within MTs. IPWs belonging to the MPB are given by the cutoff $|\mathbf{q}+\mathbf{G}|^{W}_{\rm Max}$. The sets of PB are shown in Table III.
(C) Cutoff energy for self-energy. As we explained in Sect. 3.3, we calculate $\langle i|\Delta V^{\rm xc}|j\rangle$ only for $\epsilon_i \le E^{\Sigma}_{\rm MAX}$ and $\epsilon_j \le E^{\Sigma}_{\rm MAX}$, where $E^{\Sigma}_{\rm MAX}$ is measured from the top of the valence. See the bottom of Sect. 3.3.
(D) Energy-axis parameters for GW.
These are to accumulate the imaginary part of $W(\omega)$. See the explanation after Eq. (25). We use an energy mesh (bin width) whose bin width is 0.005 Ry at $\omega = 0$ and quadratically coarser at larger ω (Sect. II-D of Ref. 21); the bin width becomes doubles at 0.04 Ry. For integration along the imaginary axis, we use ten points in the imaginary axis of ω. The parameters are good enough to yield reasonable results as seen in Ref. 21.
(E) $E^{\rm APW}_{\rm MAX}$.
(F) Use ESs or not.
(G) LDA or GGA. Theoretically, results do not depends on this choice. However, this choice slightly affect to results in practice because of reasons shown at the bottom of Sect. 3.1.

Here (E), (F), and (G) are settings common with the LDA/GGA-level calculations.

In Table IV, we show the band gaps for GaAs and SiO2c calculated by PMT-QSGW for changing the settings of (A)–(G) except (D). We calculate the self-energy only at **k**-mesh points, which are $4\times4\times4$ and $2\times2\times2$ in the 1st BZ for GaAs and SiO2c, respectively (we use sufficiently large **k**-mesh for electron density, $10\times10\times10$ for GaAs, and $6\times6\times6$ for SiO2c). No spin orbit coupling is included. In the calculation of the polarization function of Eq. (25), we take all occupied and unoccupied states (no cutoff energy for unoccupied states). The top line data labeled as "REF", which show the gaps of 1.939 eV(GaAs) and 11.16 eV(SiO2c), are treated as bases for the following comparisons with other cases. In the cases of "REF", we use $E^{\Sigma}_{\rm MAX}$ = "all", which means taking all the matrix elements of $\Delta V^{\rm xc}$ (that is, $E^{\Sigma}_{\rm MAX}$ is infinity). Empty spaces in Table IV mean that we use the same settings as with "REF". For example, the line next to "REF" for GaAs denotes a case with the same settings as "REF" except for changes in $|\mathbf{q}+\mathbf{G}|^{\Psi}_{\rm Max} = 6.0$ and $|\mathbf{q}+\mathbf{G}|^{W}_{\rm Max} = 4.0\,\text{bohr}^{-1}$.

We read the following points from Table IV. Here we take ~0.1 eV as our target numerical accuracy in the PMT-QSGW method. This is because it seems difficult to attain a numerical error within ~0.1 eV. It is not so meaningful to discuss small differences.

1. In the first section (sections are divided by horizontal lines) just below "REF", we can see the dependence on $|\mathbf{q}+\mathbf{G}|^{\Psi}_{\rm Max}$ and $|\mathbf{q}+\mathbf{G}|^{W}_{\rm Max}$. We see that the "REF" setting, $(|\mathbf{q}+\mathbf{G}|^{\Psi}_{\rm Max}, |\mathbf{q}+\mathbf{G}|^{W}_{\rm Max}) = (4.0, 3.0)$ $(\text{bohr})^{-1}$, show a convergence of ~0.01 eV even in the case of SiO2c (~ 0.001 eV for GaAs) for these parameters. We have shown a similar check in Ref. 21.
2. In our test cases of the PB in Table III, we can estimate numerical errors caused by choosing PB. As for PB in GaAs, 1.939 eV given by PB1 (REF) shows good agreement with 1.946 eV by PB1*l*, which is the largest PB among those we used here. For SiO2c, there is little

Table III. The used product basis (PB) is constructed from the products of atomic bases. The names of PB are PB0, PB0t, and so on, shown in the left-end column. After all the products are generated, linearly-dependent ones are removed with the use of the overlap matrix of the products. See Ref. 21 for details. $l_{\rm cut}$ means the allowed maximum l of the PB. $n_{\rm PB}$ shows the total number of PBs in each MT.

		Products	$l_{\rm cut}$	tol	$n_{\rm PB}$
GaAs					
PB0	Ga	$\phi(4s,4p,3d,4d)\times\phi(4s,4p,3d,4d,4f)$	4	10^{-3}	97
	As	$\phi(4s,4p,3d,4d)\times\phi(4s,4p,3d,4d,4f)$	4	10^{-3}	106
PB0*t*	Ga	PB0	4	10^{-5}	119
	As	PB0	4	10^{-5}	126
PB0*l*	Ga	PB0	6	10^{-3}	119
	As	PB0	6	10^{-3}	128
PB1	Ga	$\phi(4s,4p,3d,4d)\times\phi(4s,4p,3d,4d,4f)$ $\phi(4s,4p,3d,4d)\times\dot\phi(4s,4p,3d,4d,4f)$	4	10^{-3}	115
	As	$\phi(4s,4p,3d,4d)\times\phi(4s,4p,3d,4d,4f)$ $\phi(4s,4p,3d,4d)\times\dot\phi(4s,4p,3d,4d,4f)$	4	10^{-3}	115
	(ES)	$\phi(1s,2p,3d)\times\phi(1s,2p,3d,4f)$	2	10^{-3}	22
PB1*l*	Ga	PB1	6	10^{-5}	175
	As	PB1	6	10^{-5}	178
SiO2c					
PB0	Si	$\phi(3s,3p,3d)\times\phi(3s,3p,3d,4f)$	4	10^{-3}	75
	O	$\phi(2s,2p,3d)\times\phi(2s,2p,3d,4f)$	4	10^{-3}	67
PB0s	Si	PB0	4	10^{-3}	76
	O	PB0	2	10^{-3}	31
PB1	Si	$\phi(3s,3p,3d)\times\phi(3s,3p,3d,4f)$ $\phi(3s,3p,3d)\times\dot\phi(3s,3p,3d,4f)$	4	10^{-3}	76
	O	$\phi(2s,2p,3d)\times\phi(2s,2p,3d,4f)$ $\phi(2s,2p,3d)\times\dot\phi(2s,2p,3d,4f)$	2	10^{-3}	31
	(ES)	$\phi(1s,2p,3d)\times\phi(1s,2p,3d,4f)$	2	10^{-3}	22
PB1*l*	Si	PB1	4	10^{-5}	76
	O	PB1	4	10^{-5}	70

Table IV. Band gaps for GaAs and cubic SiO2c (β-cristobalite SiO_2) in the PMT-QSGW method in different cutoffs/settings. The numbers of used **k** points in the 1st BZ are $4 \times 4 \times 4$ for GaAs, and $2 \times 2 \times 2$ for SiO2c. There is no spin–orbit coupling. The first line named "REF" is treated as a base for comparison with others in this table. There are empty spaces; an empty space means using the same setting as that of "REF". The column $|\mathbf{q}+\mathbf{G}|^{\Psi,W}_{\mathrm{Max}}$ shows the used $|\mathbf{q}+\mathbf{G}|^{\Psi}_{\mathrm{Max}}$ and $|\mathbf{q}+\mathbf{G}|^{W}_{\mathrm{Max}}$. Lines marked by $*$ show values with largest bases. Lines marked by $**$ is the case used in Fig. 1. Each line corresponds to a QSGW calculation for a given setting. Lines are divided into sections by horizontal lines (guide of eyes).

GaAs XC	$\lvert\mathbf{q}+\mathbf{G}\rvert^{\Psi,W}_{\mathrm{Max}}$ (1/bohr)	PB	$E^{\Sigma}_{\mathrm{MAX}}$ (Ry)	$E^{\mathrm{APW}}_{\mathrm{MAX}}$ (Ry)	Band gap (eV)
REF: vwn	4.0, 3.0	PB1	all	3.0	1.939
	6.0, 4.0				1.939
	3.5, 3.0				1.940
	3.0, 2.5				1.934
		PB0			1.956
		PB0t			1.938
		PB0l			1.967
		PB1l			1.946
				2.0	1.931
				4.0	1.950
				5.0	1.959
				6.0	1.969
			3.0		1.942**
			3.0	6.0	1.980
vwn,es					1.945
vwn,es				6.0	1.982*
vwn,es			3.0		1.903
vwn,es			3.0	6.0	1.940
pbe				2.0	1.973
pbe					1.981
pbe				4.0	1.992
pbe				5.0	2.001
pbe				6.0	2.010
pbe,es					1.969
pbe,es			3.0		1.940
pbe,es				6.0	2.002*

SiO2c XC	$\lvert\mathbf{q}+\mathbf{G}\rvert^{\Psi,W}_{\mathrm{Max}}$ (1/bohr)	PB	$E^{\Sigma}_{\mathrm{MAX}}$ (Ry)	$E^{\mathrm{APW}}_{\mathrm{MAX}}$ (Ry)	Band gap (eV)
REF: vwn	4.0, 3.0	PB1	all	3.0	11.16
	8.0, 6.0				11.28
	6.0, 4.0				11.27
	3.5, 3.0				11.15
	3.0, 2.5				10.76
		PB0			11.20
		PB0s			11.17
		PB1l			11.19
		PB1l		6.0	11.21
				6.0	11.18
			3.0		10.38**
			6.0		10.78
			9.0		10.99
vwn,es					10.49
vwn,es				4.0	10.47
vwn,es				6.0	10.41*
vwn,es			3.0		10.09
pbe					11.31
pbe				6.0	11.33
pbe,es					10.57
pbe,es				5.0	10.54*

dependence on the choice of the PB used here. In particular, in the case of PB0s, we use a set of PBs on oxygen only with $l_{\mathrm{cut}} = 2$. This choice reduces the computational time considerably for larger systems.

3. The band gap gradually increases when we increase $E^{\mathrm{APW}}_{\mathrm{MAX}}$ in GaAs. The band gap monotonically increases from 1.939 eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 3.0$ Ry to 1.969 eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 6.0$ Ry for "vwn" (we see similar changes for "vwn,es" where 1.945 eV increases to 1.982 eV). Thus we cannot see a convergence behavior within this range of $E^{\mathrm{APW}}_{\mathrm{MAX}}$. This 1.969 eV can be taken as the best value for "vwn" among the performed calculations since we use the largest number of APWs in this case. Because of the over-completeness problem of basis sets in the PMT method, it is not easy to enlarge the number of APWs. In addition, eigenfunctions at high energy are not accurate (we do not include local orbital for high energy bands). Thus we inevitably take this behavior as a limitation of our current implementation of PMT-QSGW. Recall that such a slow convergence on the number of unoccupied bands (= number of APWs in our case) is also observed in Ref. 45.
   We observe similar levels of small changes in the case of SiO2c. The band gap of SiO2c changes from 11.16 eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 3.0$ Ry (REF), to 11.18 eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 6.0$ Ry. We also see a similar change for "vwn,es" from 10.49 eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 3.0$ Ry to 10.41 eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 6.0$ Ry.
4. Let us discuss other points for GaAs.
   We see that using $E^{\Sigma}_{\mathrm{MAX}} = 3.0$ Ry (marked by $**$) gives little difference from REF (1.942–1.939 eV). Thus we can use $E^{\Sigma}_{\mathrm{MAX}} = 3.0$ Ry to reduce computational effort. We take that "vwn,es" is better setting than "vwn" because we include the MTOs of ESs as bases. We see the difference between "vwn" and "vwn,es" is very small ($1.945 - 1.939 = 0.006$ eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 3.0$ Ry; $1.982 - 1.969 = 0.013$ eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 6.0$ Ry). Thus we do not need to use ESs for GaAs.
   When we use "pbe" (explained at the bottom of Sect. 3.1), the result changes slightly. The best value of 2.002 eV (marked by $*$) shows small difference of 0.02 eV from that in "vwn,es" of 1.982 eV (marked by $*$).
   In conclusion, except for the non-converging behavior on $E^{\mathrm{APW}}_{\mathrm{MAX}}$, the analysis presented here shows that it might be safer to estimate numerical errors as $\sim$0.1 eV.
5. Let us discuss other points for SiO2c.
   Look in the "vwn,es" section, where we include MTOs at ESs. The band gap of 10.49 eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 3.0$ Ry changes to 10.41 eV at $E^{\mathrm{APW}}_{\mathrm{MAX}} = 6.0$ Ry. The difference of $10.49 - 10.41 = 0.08$ eV is not so large, although we do not claim that this shows convergence on $E^{\mathrm{APW}}_{\mathrm{MAX}}$.

We see a large difference of $11.16 - 10.49 = 0.67$ eV between the value in "vwn,es" and 11.16 eV for "REF". This can be taken as an error because of no ESs. The error decreases to $10.38 - 10.09 = 0.29$ eV if we use $E^{\Sigma}_{\rm MAX} = 3.0$ Ry (compare 10.38 eV marked by $**$ and 10.09 eV in "vwn,es"). Thus more than half of the error of 0.67 eV is due to the high-energy part of $\Delta V^{\rm xc}$. In other words, "Without ESs" shows a numerical accuracy of $\sim$0.29 eV under the constraint $E^{\Sigma}_{\rm MAX} = 3.0$ Ry. It seems that using a larger $E^{\Sigma}_{\rm MAX}$ is an unbalanced setting (when we use a large $E^{\Sigma}_{\rm MAX}$, we need to include MTOs for ESs, especially as the basis to expand $\Delta V^{\rm xc}$). The above value of 10.38 eV is in good agreement with the value of 10.49 eV of $E^{\Sigma}_{\rm MAX}$ = "all" in "vwn,es" because of a fortunate error cancellation. That is, their difference can be divided as $10.38 - 10.49$ eV $= (10.38 - 10.09) + (10.09 - 10.49)$ eV. Here $(10.38 - 10.09) = 0.29$ eV is the error which we already discussed above; $(10.09 - 10.49)$ eV is the error between $E^{\Sigma}_{\rm MAX} = 3.0$ Ry and $E^{\Sigma}_{\rm MAX}$ = "all" in "vwn,es".

See the difference between "vwn,es" and "pbe,es". It gives an extra numerical error of $\sim$0.1 eV.

Let us summarize the key findings. When we include ESs, the convergence behaviors for band gap are satisfactory (convergence within $\sim$0.1 eV) except for $E^{\rm APW}_{\rm MAX}$. We can observe the stability of the band gap when we enlarge $E^{\rm APW}_{\rm MAX}$. However, a systematic convergence check on $E^{\rm APW}_{\rm MAX}$ is difficult to carry out because of the limited ability describing high-energy bands (over-completeness problem of a basis set in the PMT method). Constraint on $E^{\Sigma}_{\rm MAX}$ is required to maintain numerical accuracy when we do not use ESs. If we use $E^{\Sigma}_{\rm MAX} = 3.0$ Ry, we have a smaller difference $\sim$0.29 eV for the band gap of SiO2c between cases with/without ESs.

Considering the balance of computational effort and the discussion in the above paragraph, we think that a procedure such as "$E^{\Sigma}_{\rm MAX} = E^{\rm APW}_{\rm MAX} = 3.0$ Ry without ESs" is useful for application to a variety of materials. In particular, the cutoff of $E^{\Sigma}_{\rm MAX}$ is needed to maintain numerical accuracy in the case without ESs. Then we take this procedure as an approximation to true results of the fundamental equation of QSGW. Note that it is not so meaningful to pursue fully converged results in QSGW in practical applications, since it is inevitable for QSGW to give some differences from experimental values. In fact, Refs. 12 and 21 show that QSGW tends to give systematic overestimation of band gaps. The calculated value of 10.41 eV for SiO2c ("vwn,es", $E^{\rm APW}_{\rm MAX} = 6.0$ Ry) in Table IV is larger than the experimental value of $\sim$8.9 eV,[46] thus not directly compared with experiments (another QSGW calculation by Shaltat et al.[16] gives band gap 8.8 eV by QSGW for SiO2c. The difference from our value of 10.41 eV suggests numerical difficulty in achieving convergence). In papers such as Refs. 22 and 47, we empirically correct this discrepancy from experiments by a hybrid method such as $(1 - \alpha) \times$ QSGW $+ \alpha \times$ LDA ($\alpha \sim 0.2$). Thus, from a practical viewpoint, it will be better to take the parameter α as a combined correction on the theoretical and numerical errors due to the approximation. Or we may need to invent a better fundamental equation to go beyond QSGW, which is numerically stable while keeping the advantages of QSGW.

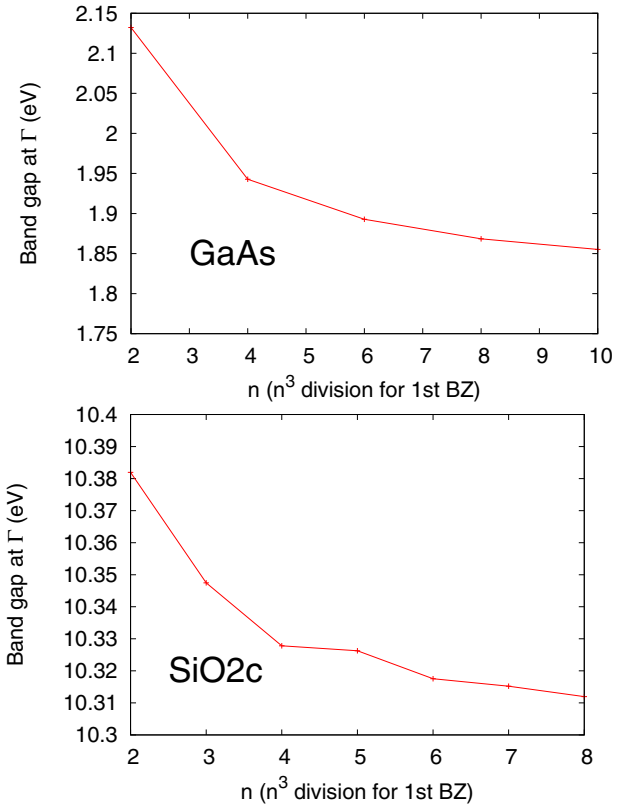

Fig. 1. (Color online) Dependence of band gap on the number of **k** points in the 1st BZ for self-energy calculation. The integer n of the x-axis means that the number of divisions of BZ is $n \times n \times n$. The y-axis is for the band gap.

In Fig. 1, we show the convergence check concerning the number of **k** points for self-energy calculation in the 1st BZ (the **k** point mesh for electron density is fixed). The integer n of the x-axis means that the number of used **k** points is $n \times n \times n$. As for GaAs, we see a smooth convergence on the number of **k** points. In the $4 \times 4 \times 4$ calculation, we see $\sim$0.1 eV overestimation in comparison with the value at $10 \times 10 \times 10$. We need to choose the number of **k** points to have the best accuracy within the allowed computational resources. As for SiO2c, pay attention to the energy scale of the y-axis. The difference of the band gap between $n = 2$ and $n = 8$ is small, only $\sim$0.04 eV. In our analysis, the unsmooth behavior of this plot is due to the cutoff of $E^{\Sigma}_{\rm MAX}$; see the dependence on $E^{\Sigma}_{\rm MAX}$ in Table IV. Energy bands near $E^{\Sigma}_{\rm MAX}$ are taken into account or not with a slight change of **k** point.

In Fig. 2, the energy dispersion curve for QSGW obtained with the largest number of **k** point cases in Fig. 1 is shown, in order to show the difference from LDA/GGA.

5. Summary

We have developed a new method, the PMT-QSGW method of performing the QSGW calculation based on the PMT method. PMT-QSGW have the advantages of robustness, ease to use, and accuracy over FP-LMTO-QSGW. It also does not require the tuning of the parameters for MTOs. Owing to the use of APWs, we can use highly localized MTOs with low-energy APWs ($\sim$4 Ry). Thus we employ a simplified procedure of interpolation to the self-energy instead of a previous complicated one in FP-LMTO-QSGW.

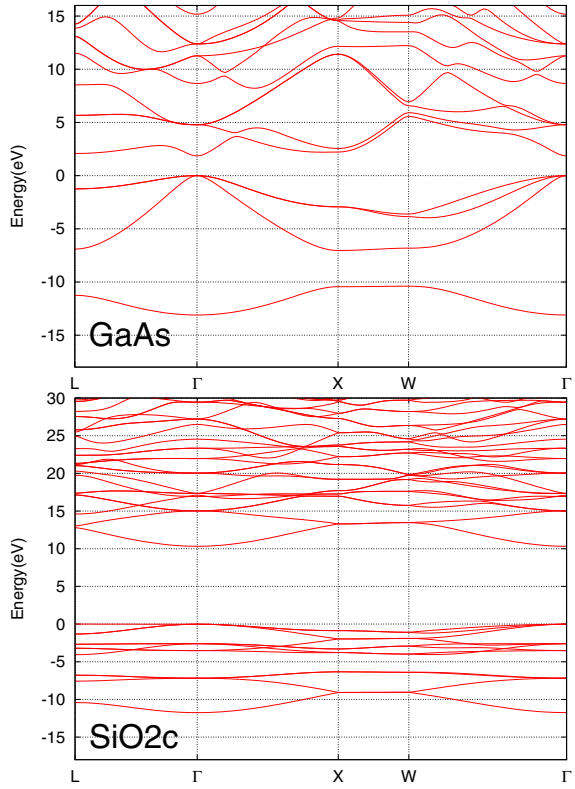

Fig. 2. (Color online) Band plots for GaAs, corresponding to the case of $10 \times 10 \times 10$ in Fig. 1, and for SiO2c corresponding to the case of $8 \times 8 \times 8$.

We have shown a convergence check on the band gaps of two typical cases, GaAs and cubic SiO_2. We analyzed how their band gaps depend on the cutoff parameters and computational settings. Then we observe the performance and limitations of PMT-QSGW. The results shown in this paper can be reproduced by the PMT-QSGW method implemented in the ecalj package, which is freely available from github.[48)]

The PMT-QSGW method with the highly localized MTOs and low-energy APWs is advantageous for theoretical treatment. The techniques developed here can be useful even when we develop a method going beyond QSGW.

Acknowledgement

I wish to thank to Dr. H. Kino for discussions, coding, and advises regarding this manuscript. This work was partly supported by the Advanced Low Carbon Technology Research and Development Program (ALCA) of Japan Science and Technology Agency (JST), and by Grant-in-Aid for Scientific Research 23104510. We also acknowledge computing time provided by Computing System for Research in Kyushu University.

1) B. Holm and U. v. Barth, Phys. Rev. B **57**, 2108 (1998).
2) W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **89**, 126401 (2002).
3) A. Stan, N. E. Dahlen, and R. v. Leeuwen, Europhys. Lett. **76**, 298 (2006).
4) C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen, Phys. Rev. B **81**, 085103 (2010).
5) F. Caruso, P. Rinke, X. Ren, A. Rubio, and M. Scheffler, Phys. Rev. B **88**, 075105 (2013).
6) S. V. Faleev, M. v. Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
7) M. Methfessel, M. van Schilfgaarde, and R. A. Casali, in *Lecture Notes in Physics*, ed. H. Dreysse (Springer, Berlin, 2000) Vol. 535.
8) T. Kotani and M. v. Schilfgaarde, Solid State Commun. **121**, 461 (2002).
9) F. Aryasetiawan and O. Gunnarsson, Phys. Rev. Lett. **74**, 3221 (1995).
10) F. Aryasetiawan and O. Gunnarsson, Phys. Rev. B **49**, 16214 (1994).
11) F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
12) M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
13) T. Kotani, M. van Schilfgaarde, and S. V. Faleev, Phys. Rev. B **76**, 165106 (2007).
14) D. R. Hamann and D. Vanderbilt, Phys. Rev. B **79**, 045109 (2009).
15) M. Shishkin, M. Marsman, and G. Kresse, Phys. Rev. Lett. **99**, 246403 (2007).
16) R. Shaltaf, G.-M. Rignanese, X. Gonze, F. Giustino, and A. Pasquarello, Phys. Rev. Lett. **100**, 186401 (2008).
17) F. Bruneval, Phys. Rev. Lett. **103**, 176403 (2009).
18) F. Bruneval, J. Chem. Phys. **136**, 194107 (2012).
19) S.-H. Ke, Phys. Rev. B **84**, 205415 (2011).
20) M. van Schilfgaarde, T. Kotani, and S. V. Faleev, Phys. Rev. B **74**, 245125 (2006).
21) T. Kotani and M. van Schilfgaarde, Phys. Rev. B **76**, 165106 (2007).
22) A. N. Chantis, M. v. Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **96**, 086405 (2006).
23) P. Lukashev, W. R. L. Lambrecht, T. Kotani, and M. van Schilfgaarde, Phys. Rev. B **76**, 195202 (2007).
24) T. Kotani and H. Kino, J. Phys.: Condens. Matter **21**, 266002 (2009).
25) N. E. Christensen, A. Svane, R. Laskowski, B. Palanivel, P. Modak, A. N. Chantis, M. van Schilfgaarde, and T. Kotani, Phys. Rev. B **81**, 045203 (2010).
26) A. Svane, N. E. Christensen, I. Gorczyca, M. van Schilfgaarde, A. N. Chantis, and T. Kotani, Phys. Rev. B **82**, 115102 (2010).
27) L.-y. Huang and W. R. L. Lambrecht, Phys. Rev. B **88**, 165203 (2013).
28) T. Kotani and M. van Schilfgaarde, Phys. Rev. B **81**, 125117 (2010).
29) T. Kotani and H. Kino, J. Phys. Soc. Jpn. **82**, 124714 (2013).
30) I. V. Solovyev, A. I. Liechtenstein, and K. Terakura, Phys. Rev. Lett. **80**, 5758 (1998).
31) S. Ishii, H. Maebashi, and Y. Takada, arXiv:1003.3342.
32) B. Arnaud and M. Alouani, Phys. Rev. B **63**, 085208 (2001).
33) F. Bechstedt, K. Tenelsen, B. Adolph, and R. D. Sole, Phys. Rev. Lett. **78**, 1528 (1997).
34) S. Botti and M. A. L. Marques, Phys. Rev. Lett. **110**, 226404 (2013).
35) X. Deng, M. Ferrero, J. Mravlje, M. Aichhorn, and A. Georges, Phys. Rev. B **85**, 125137 (2012).
36) M. Springer, F. Aryasetiawan, and K. Karlsson, Phys. Rev. Lett. **80**, 2389 (1998).
37) A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996).
38) A. J. Cohen, P. Mori-Sanchez, and W. Yang, Science **321**, 792 (2008).
39) C. Friedrich, S. Blügel, and A. Schindlmayr, Phys. Rev. B **81**, 125102 (2010).
40) J. Rath and A. J. Freeman, Phys. Rev. B **11**, 2109 (1975).
41) C. Freysoldt, P. Eggert, P. Rinke, A. Schindlmayr, R. Godby, and M. Scheffler, Comput. Phys. Commun. **176**, 1 (2007).
42) E. Sjöstedt, L. Nordström, and D. J. Singh, Solid State Commun. **114**, 15 (2000).
43) L. W. S. H. Vosko and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
44) J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
45) C. Friedrich, M. C. Mller, and S. Blügel, Phys. Rev. B **83**, 081101 (2011).
46) T. DiStefano and D. Eastman, Solid State Commun. **9**, 2259 (1971).
47) T. Kotani and M. van Schilfgaarde, Phys. Rev. B **81**, 125201 (2010).
48) The first-principles electronic structure suite based on the PMT method, ecalj package, is freely available from https://github.com/tkotani/ecalj.