Full versus quasi-particle self consistency in vertex corrected GW approaches

Andrey L. Kutepov*
Condensed Matter Physics and Materials Science Department,
Brookhaven National Laboratory, Upton, NY 11973

Using seven semiconductors/insulators with band gaps covering the range from 1 eV to 10 eV we systematically explore the performance of two different variants of self-consistency associated with famous Hedin’s system of equations: the full self-consistency and the so called quasi-particle approximation to it. The pros and cons of these two variants of self-consistency are sufficiently well documented in literature for the simplest GW approximation to the Hedin’s equations. Our study, therefore, aims primarily at the level of theory beyond GW approximation, i.e. at the level of theory which includes vertex corrections. Whereas quasi-particle self-consistency has certain advantages at GW level (well known fact), the situation becomes quite different when vertex corrections are included. In the variant with full self-consistency, vertex corrections (both for polarizability and for self energy) systematically reduce the calculated band gaps making them closer to the experimental values. In the variant with quasi-particle self-consistency, however, an inclusion of the same diagrams has considerably larger effect and calculated band gaps become severely underestimated. Different effect of vertex corrections in two variants of self-consistency can be related to the Z-factor cancellation which plays positive role in quasi-particle self-consistency at GW level of theory but appears to be destructive for the quasi-particle approximation when higher order diagrams are included. Second result of our study is that we were able to reproduce the results obtained with the Questaal code using our FlapwMBPT code when the same variant of self-consistency (quasi-particle) and the same level of vertex corrections (for polarizability only, static approximation for screened interaction, and Tamm-Dancoff approximation for the Bethe-Salpeter equation) are used.

PACS numbers:

Introduction

Reproducibility of results in computational material science is an important issue. In the field of electronic structure calculations, the issue is considered to be resolved at the level of density functional theory (DFT) calculations. General consensus is that modern electronic structure codes, however employ different basis sets (LAPW, LMTO, PAW et al.), upon convergence demonstrate quite similar electronic structure of materials. When one goes beyond the DFT approximation (for instance if one uses Hedin’s GW approach) the number of setup parameters in the calculation increases. Convergence of the results now depends not only on the occupied one-electron states which have to be represented accurately, but also on a number of excited (unoccupied) states which brings considerable difference in the results if excited states are represented differently or if their number (those which included in the calculation) varies. Besides that, the appearance of two-point bosonic functions (such as polarizability P and screened interaction W) requires efficient basis set to represent them. For example, the so called product basis (PB) set was designed specifically for this purpose. Greater complexity of GW approximation (as compared to DFT) unavoidably brings more differences in implementations which makes the reproducibility of results at the level of GW a more serious issue. Nevertheless, as it was shown in non-self-consistent (G0W0) calculations for 100 molecular systems, the reproducibility of the molecular electronic structure though not perfect is still acceptable in many cases. Namely, by comparing G0W0 approach implemented in three different codes FHI-aims7,9, BerkeleyGW10, and TURBOMOLE10,11, authors of Ref. 6 conclude that at convergence, the HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital) levels agree on the order of 200 meV. In the process of their work, authors of Ref. 6 also have identified two crucial aspects that control the accuracy of the G0W0 quasi-particle energies: the size of the basis set and the treatment of the frequency dependence.

Vertex corrected (diagrammatically) GW calculations for realistic materials now only begin to appear12,17,18,22. Their increased complexity (even comparative to G0W0) as well as their relatively recent introduction to the field make the reproducibility of results an open and important issue. Additional (as compared to GW) setup parameters for the methods which diagrammatically go beyond GW approximation are the following: i) specific set of diagrams beyond GW; ii) details of implementation of these high order diagrams. Thorough investigation of the effects of using different sets of diagrams as well as the details of implementation is far beyond the scope of a single study. The objective of this work is more specific and it was motivated by recent vertex corrected GW calculations of the band gaps in semiconductors performed with the FlapwMBPT code19, and with the Questaal code (B. Cunningham et al., 13).

In both studies, 19 and 13, it is concluded that vertex corrections provide the biggest numerical improve-

*e-mail: akutepov@bnl.gov
ment of the results obtained with fully self-consistent GW (scGW, Ref. [19]) or with quasi-particle self-consistent GW (QSGW, Ref. [13]). Authors of work [13] also include the effect of electron-phonon interaction (for polar semiconductors) but this effect quantitatively is smaller than vertex corrections in most cases. What is important for the present study is the fact that inclusion of electron-phonon interaction always reduces the calculated band gaps as it is evident not only from Ref. [19] but also from earlier works [23–26]. The list of materials studied in Ref. [13] is slightly different (and also longer) than the corresponding list in Ref. [19]. But remarkable tendency of improvements of scGW (or QSGW) results is unmistakable in both studies. Only one noticeable exception is the case of CuCl in Ref. [13] where vertex correction worsens the QSGW result (band gap becomes seriously underestimated). What is surprising (and which is one of the motivations for the present study) is the fact that similar (and good) results were obtained with quite different variants of vertex corrections applied in two studies.

Whereas both works formally are based on exact Hedin’s equations, the details of the applied approximations differ a lot. In the study [19], vertex corrections are used to improve both polarizability and self energy $\Sigma$. Authors of Ref. [13] use vertex correction only for polarizability. Further, all vertex corrections in [19] (for $P$ and for $\Sigma$) use fully frequency dependent screened interaction $W(\nu)$, whereas vertex correction to $P$ in Ref. [13] is evaluated with frequency independent (taken at zero frequency) $W(\nu = 0)$. Also, the Tamm-Dancoff (TD) approximation was used in Ref. [13]. There are also some differences in the basis set (see below). But the most important and dramatic (as it will be shown below) difference consists in using full or quasi-particle Green’s function in the evaluation of diagrams. In Ref. [19] all vertex corrected calculations are performed with full self-consistency applied to Green’s function $G$. At the same time, authors of work [13] use additional (the so called quasi-particle, QP) approximation which is not intended in Hedin’s equations. Validity of quasi-particle approximation is well justified at the level of GW method (without vertex corrections). It is known, that QSGW approach usually is more accurate than fully self-consistent scGW method. The success of QP approximation at the GW level of theory is based on the so called Z-factor cancellation which was clearly explained in the pioneer work on QSGW in Ref. [27]. Briefly, the essence of the trick is that diagrams should be evaluated either with full Green’s function (no QP approximation) and including vertex part or with QP Green’s function but excluding the vertex part. Therefore, if one excludes vertex part (GW level of theory) than it is of advantage to use QP approximation. Alternatively, if one intends to apply vertex corrections, QP approximation for $G$ should not be used and one has to use full $G$ in the evaluation of diagrams instead. Thus, from the point of view of Z-factor cancellation, simultaneous use of vertex corrections and QP approximation for $G$ should be questioned for consistency. Some results on the inadequacy of quasi-particle self-consistency were published a few years ago using Hubbard-dimer model as an example.

The wish to understand deeper the reason why two seemingly different approximations used in [19] and [13] lead to similar results was the main motivation for the present work. In order to accomplish the goal we first answer the question whether one can reproduce the results obtained in Ref. [13] (with the Questaal code) using the same approximations as in Ref. [19] but running different code (FlapwMBPT). Secondly, we extend both studies, [19] and [13], by performing the calculations which are intended to trace step by step the differences in the implementations of vertex corrections between Refs. [19] and [13]. Namely, in the first set of calculations (for each material considered) we start with scGW, then we add vertex correction to polarizability but with $W(\nu = 0)$ in the corresponding diagrams, then we apply vertex correction again to $P$ only but with full $W(\nu)$, and finally we add vertex correction to self energy to mimic the full approximation used in Ref. [19].

Second set of calculations consists of exactly the same steps but all calculations in the second set are supplemented with QP approximation for Green’s function. In this case the variant with vertex correction to $P$ only and with static $W(\nu = 0)$ mimics the level of approximation accepted in Ref. [19]. As authors of work [13] also used the Tamm-Dancoff approximation in their evaluation of vertex correction to polarizability, the calculations with QP self-consistency (and with static $W(\nu = 0)$ approximation) in present work were performed in both ways: with Tamm-Dancoff approximation and without it. In this respect, all calculations with full self-consistency were performed without using the Tamm-Dancoff approximation. Additional steps, i.e. vertex correction to $P$ with full $W(\nu)$ and, finally, with inclusion of vertex correction to $\Sigma$ represent the steps which authors of work [13] mention as possible ways to improve their results but they do not perform these steps. However, as it will be shown below, inclusion of these steps in the calculations with QP approximation for $G$, in fact, worsens the QP-based results considerably and, therefore, cannot be considered as a valid improvement for calculations with QP self-consistency.

As it is shown below, the calculations performed at similar level of approximations result in very similar results when one uses Questaal code or FlapwMBPT code. At the same time, it is also shown below that a few omissions (or rather 'constraints') accepted in Ref. [13], namely:

- Tamm-Dancoff approximation;
- insufficient basis set (number of unoccupied states included in vertex corrections);
- vertex corrections are applied to polarizability only, but not to self energy;
• static approximation for screened interaction with its value taken at zero frequency;

all result in reduction of the calculated band gaps. When the above ‘constraints’ are removed (with QP self-consistency) the obtained band gaps demonstrate dramatic level of underestimation of the corresponding experimental values. When all ‘constraints’ are lifted, the calculations even become unstable for small gap semiconductors (when QP self-consistency is used).

The paper begins with a discussion of the approximations used in this study (Section ‘Methods’). The discussion of convergence issues and of setup parameters is provided next. The principal results obtained in this work are presented in Section ‘Results’. The conclusions are given afterwards.

Methods

All calculations in this study formally are based on the Hedin equations. For convenience, we remind the reader about how Hedin’s equations could be solved self-consistently in practice.

Suppose one has a certain initial approach for Green’s function \( G \) and screened interaction \( W \). Then one calculates the following quantities:

three-point vertex function from the Bethe-Salpeter equation

\[
\begin{align*}
\Gamma^\alpha(123) & = \delta(12)\delta(13) \\
& + \sum_\beta \frac{\delta \Sigma^\alpha(12)}{\delta G^\beta(45)} G^\beta(46) \Gamma^\beta(673) G^\beta(75),
\end{align*}
\] (1)

where \( \alpha \) and \( \beta \) are spin indexes, and the digits in the brackets represent space-Matsubara’s time arguments, polarizability

\[
P(12) = \sum_\alpha G^\alpha(13) \Gamma^\alpha(342) G^\alpha(41),
\] (2)

screened interaction

\[
W(12) = V(12) + V(13) P(34) W(42),
\] (3)

and the self energy

\[
\Sigma^\alpha(12) = -G^\alpha(14) \Gamma^\alpha(425) W(51).
\] (4)

In the equation (3) \( V \) stands for the bare Coulomb interaction. New approximation for the Green function is obtained from Dyson’s equation

\[
G^\alpha(12) = G_0^\alpha(12) + G_0^\alpha(13) \Sigma^\alpha(34) G^\alpha(42),
\] (5)

where \( G_0 \) is the Green function in Hartree approximation. Eqn. (1) comprise one iteration. If convergence is not yet reached one can go back to the equation (1) to start the next iteration with renewed \( G \) and \( W \).

The system of Hedin’s equations formally is exact, but one has to introduce certain approximations when solving \( \Gamma \) for the vertex function \( \Gamma^\alpha(123) \) in order to make the solving of the system manageable in practice. Approximations which we use in this study are dictated by the goals of the work. In order to justify their choice let us summarize the goals again:

• reproduce the results of Ref. [13];

• check the validity of speculations made in Ref. [13] about the possibility to improve QSGW-based results by taking into account frequency dependence of \( W \) when solving the BSE and adding vertex correction to self energy;

• compare full self-consistency with QP self-consistency.

Guided by the goals, we conducted two sets of calculations (for each material studied) which were already sketched in the Introduction section. Starting point for the first set is scGW approximation (fully self-consistent GW), whereas second set of calculations has QSGW approach as a starting point (as in Ref. [13]). In scGW method, vertex function \( \Gamma \) is approximated by its trival part \( \Gamma^\alpha(123) = \delta(12)\delta(13) \). In QSGW, we use additional approximation related to the self energy in Eq. (5), namely we linearize frequency dependence of self energy around zero frequency (see details in Refs. [31,32]). In this respect, our construction of QSGW differs from Ref. [13] where the approximation for self energy in Eq. (5) consists in taking its hermitian part. As it will be shown below, however, numerical results obtained with above two variants of QSGW are pretty much similar for majority of materials. Concerning the vertex part, all vertex corrected calculations included the solution of Bethe-Salpeter equation (BSE) for polarizability with screened interaction \( W \) as a kernel of BSE (\( \frac{\delta \Sigma^\alpha(12)}{\delta G^\alpha(342)} = \delta_{\alpha}\delta(13)\delta(24) W(12) \)). Corresponding diagrammatic representation for the vertex correction to \( P \) is shown in Fig. 1. At this level, we conducted a few variants of calculations in order to explore approximations made by authors of Ref. [13], namely the Tamm-Dancoff approximation and the use of frequency independent \( W(\nu = 0) \). Finally, our the most sophisticated calculations included vertex correction to self energy of second order (second term in Fig. 2). This approximation for self energy corresponds to expansion of vertex function in Eq. (1) up to the first order in \( W \). As one can easily notice, the above described vertex corrected variants assume different approaches for vertex function when it is used in the expression for polarizability (2) and in the expression for self energy (4) and, as a result, they are not conserving in Baym-Kadanoff definition[13] (i.e. corresponding \( P \) and \( \Sigma \)
calculations includes: and with the corresponding abbreviations. First set of approximations used in this study together with the purpose follow from the chosen approximation for Ψ-functional. (Fig. 4) and for self energy (Fig. 2) in sc(GW+G3W2). Diagrammatic representations for irreducible polarizability (Fig. 3) and for self energy (Fig. 2) in sc(GW+G3W2) follow from the chosen approximation for Ψ-functional.

For convenience, we list here all variants of approximations used in this study together with the purpose and with the corresponding abbreviations. First set of calculations includes:

- scGW, which is used primarily to generate initial approximation to start vertex corrected calculations and, by doing this, to reduce number of iterations with vertex corrections. We also compare the scGW results with the ones obtained with QSGW;
- sc(BSE0:P@GW), where the part after the symbol @ stands for diagrammatic representation of self energy, whereas the part before the symbol @ says that polarizability is evaluated from BSE with static screened interaction taken at zero frequency (W(ν = 0)) as the kernel of BSE. The goal of this variant is to assess the quality of static approximation for W in BSE, as well as to compare this variant with the same variant but based on QP self-consistency;
- sc(BSE:P@GW), where one uses full frequency dependent W(ν) in BSE. The goal of this variant is to assess the quality of static approximation for W in BSE (by comparing results with sc(BSE0:P@GW)) and also to compare this variant with the same variant but based on QP self-consistency;
- sc(BSE:P@GW+G3W2), where diagrammatic representation of self energy includes second order (in W) diagram. All diagrams in this variant (for P and for Σ) use full frequency dependent W(ν). The goal of this variant is to assess the effect of inclusion of vertex correction to self energy and also to compare this variant with the same variant but based on QP self-consistency;
- sc(GW+G3W2), which is conserving in Baym-Kadanoff definition. Only diagrammatic definition of self energy (GW+G3W2) is needed to be specified in this case. Diagrammatic representation for P, G2+G4W1, follows if ones uses the same first order vertex function (as for self energy) in Eq. ③.

All calculations in this work were performed using code FlapwMBPT. Technical details of the GW part were described in Refs. 31,32. Detailed account of the implementation of vertex corrected schemes can be found in Refs. 18,21. Figure 5 presents the flowchart of the calculations (for three selected approximations) which gives a general idea how the calculations are organized. The
Σ = .

FIG. 2. Diagrammatic representation of self energy up to the second order in screened interaction W.

Ψ = \frac{-1}{2} + \frac{1}{4}

FIG. 3. Diagrammatic representation of Ψ-functional which includes the simplest non-trivial vertex. First diagram on the right hand side stands for scGW approximation, whereas total expression corresponds to sc(GW+G3W2) approximation.

flowchart in Fig. 5 corresponds to full self-consistency. In the case of quasi-particle self-consistency, the formal change consists only in the fact that instead of Dyson’s equation \((G = G_0 + G_0 \Sigma G)\), a special construction for G is used as it is described in Refs. [31,32]. The diagrammatic (GW and the diagrams beyond GW) parts of the FlapwMBPT code take full advantage of the fact that certain diagrams can more efficiently be evaluated in reciprocal (and frequency) space whereas other diagrams are easier to evaluate in real (and time) space. As a result, GW part of the code scales as \(N_k N_\omega N_b^3\) where \(N_k\) is the number of \(k\)-points in the Brillouin zone, \(N_\omega\) is the number of Matsubara frequencies, and \(N_b\) stands for the size of the basis set. The vertex part of the code scales as \(N_k^2 N_\omega^2 N_b^2\). For comparison, if one uses naive (all in reciprocal space and frequency) implementation then GW part scales as \(N_k^2 N_\omega^2 N_b^4\) (i.e. exactly as the vertex part when the implementation is efficient), and the vertex part scales as \(N_k^3 N_\omega^3 N_b^5\). Besides of efficiency of the implementation, we have to mention two more factors which make the use of the diagrams beyond GW feasible. First is the fact that the higher order diagrams converge much faster than the GW diagram with respect to the basis set size and to the number of \(k\)-points. Second is that the higher order diagrams are very well suited for massive parallelization.

Calculation setups and convergence checks

Let us now specify the setup parameters used in the calculations. First of all, our selection of materials for this study was dictated by the following constraints: i) band gaps of the selected compounds should cover (approximately uniformly) a broad range of energies (1 – 10 eV); ii) selected materials should be taken from the list studied in Ref. [13] (only AlP does not fit in this constraint); iii) selected materials should be sufficiently simple as our vertex corrected calculations which use full frequency dependence of \(W\) are rather time consuming. In order to make presentation more compact, the list of
TABLE I. Setup parameters of the solids studied in this work. Lattice parameters are in Angstroms. $N_{bnd}$ is the number of band states used as a basis set for evaluation of GW part. $N_{bnd}^{VRT}$ represents the corresponding number for the vertex part.

| Space group | Lattice parameter | $N_{bnd}$ | $N_{bnd}^{VRT}$ |
|-------------|------------------|-----------|----------------|
| Si          | 227              | 5.43      | 160            |
| AlP         | 216              | 5.451     | 185            |
| CuCl        | 216              | 5.64      | 260            |
| C           | 227              | 3.57      | 160            |
| MgO         | 225              | 4.217     | 110            |
| NaCl        | 225              | 5.62      | 150            |
| LiCl        | 225              | 5.13      | 120            |

selected compounds, their principal structural parameters and the size of basis sets have been collected in Table I. All calculations have been performed for the electronic temperature $600K$ (Matsubara’s formalism is used throughout the work). Commonly used setup parameter $RK_{max}$ for LAPW-based calculations was set to 8.0 in all calculations of this work. The sampling of the Brillouin zone for GW part (i.e. excluding vertex correction diagrams) was $12 \times 12 \times 12$ in all cases. Evaluation of the diagrams associated with vertex part (i.e. all diagrams in Fig. 1 and the second diagram in Fig. 2) was performed with sampling $3 \times 3 \times 3$. As Table I shows, the number of band states used in the evaluation of vertex part also was considerably smaller than the number of band states included in the evaluation of GW part of the diagrams. The fact, that the diagrams representing the vertex part require smaller basis set and coarser sampling of the Brillouin zone was discussed before in Refs. 18, 19, so that the choice of these two setup parameters for the present study is justified (see for instance Table I in Ref. 19). At the same time, one can notice that the basis set for the BSE part used in Ref. 13 is smaller than ours by almost a factor of two (our basis sets for the vertex part shown in Table I are the sums of valence and conduction bands included). This fact was the reason that we perform our QPsc(BSE0TD@GW) calculations with the basis set (for the vertex part) exactly corresponding to the basis set used in Ref. 13. Therefore, the difference between the band gaps obtained with QPsc(BSE0@GW) and QPsc(BSE0TD@GW) is, in fact, a total effect of the TD approximation and of the basis set mismatch. In most cases, however, the effect of TD approximation was prevailing.

It is well known that LAPW basis set has to be supplemented with sufficient number of high energy local orbitals (HELO) in order to ensure the convergence of calculated band gaps in GW-based approximations (see for instance Refs. 36–40). Therefore, for all studied materials, we extended standard LAPW basis set by including 3–4 (per atom) HELO’s of s-type, 2–3 HELO’s of p- and d-type, 1–2 HELO’s of f-type, and also 1 HELO of g-type. In this respect, our additional basis set (HELO’s) was also larger than additional local orbital basis set used in Ref. 15 (see Table I there) which can be another reason for small differences in results at QSGW level (besides of the different way to introduce the quasi-particle approximation).

In most of our vertex corrected calculations (excluding QPsc(BSE0TD@GW)), BSE is solved iteratively15 which especially is needed when one uses frequency dependent $W(\nu)$. Referring to Fig. 1 one iteration in this case means adding one more term in the infinite sequence of ladder diagrams. In practice, the infinite sequence has to be truncated. In this study we used six terms as a cutoff parameter for the iterative solution of BSE. As we did show before (see Fig. 7 in Ref. 19), this choice of the cutoff means that the contribution of the rest of ladder diagrams (i.e. those which are not included) could be only 1/50–1/100 of the first term contribution. In fact, our checks with eight terms have demonstrated that the addition of two more ladder diagrams changes the calculated band gaps by less than 0.005 eV.

Our QPsc(BSE0TD@GW) approach which serves to reproduce the results obtained in Ref. 13 was implemented similar to as described in Ref. 13. Namely, we solved BSE directly (not iteratively as in all other our approaches) using $k$-dependent product basis for electron-hole pairs. Details of this basis set specific for LAPW implementation can be found in Ref. 41. Tamm-Dancoff approximation, therefore, was implemented by neglecting the anti-resonant part and keeping only the resonant part of the transition space.

As it will be shown later in this work, the approximation of full frequency dependent $W(\nu)$ by static function $W(\nu = 0)$ gives qualitatively (but not quantitatively) correct results for vertex corrections to polarizability. However, as it was discussed before15, similar replacement of $W(\nu)$ by $W(\nu = 0)$ in the vertex correction for self energy gives even qualitatively incorrect results. Namely, vertex correction to the band gaps evaluated with static $W(\nu = 0)$ is positive (band gaps increase) whereas correct evaluation of the corresponding diagrams (with full $W(\nu)$) always reduces the gaps. This observation also collaborates with the increase in the calculated band gaps obtained by A. Gruneis et al. in Ref. 12 where static $W(\nu = 0)$ was used. By this reason, self energy vertex corrections in this work were always evaluated with full frequency dependent $W(\nu)$.

Results

Principal results of this work are presented in Figs. 6 and 7 where the band gaps obtained with full self-consistency (Fig. 6) and with QP self-consistency (Fig. 7) are shown. Let us start our discussion with the full self-consistency. Firstly, we observe that calculations
FIG. 6. Band gaps as obtained with full self-consistency. Positioning of the data for each compound relative to the X-axis corresponds to the experimental band gap of the specific compound (Refs. [29,42–44]). Y-axis represents the deviations of the calculated band gaps from the experimental ones. Calculated results do not include electron-phonon interaction.

without vertex corrections (scGW) severely overestimate the calculated band gaps (by 0.5–1.5 eV). Secondly, principal improvement comes from the vertex correction to polarizability as it is evidenced in sc(BSE:P@GW) calculations. It is important to point out, however, that in order to get quantitatively correct results one has to use full frequency dependent \( W(\nu) \) when solving BSE which is used to obtain vertex correction to polarizability. Using static \( W(\nu = 0) \) in BSE can only give qualitatively correct correction to \( P \), but quantitatively it underestimates the correction by 20–50% (compare sc(BSE0:P@GW) with sc(BSE:P@GW)). Thirdly, vertex correction to self energy always reduces the calculated band gaps (compare sc(BSE:P@GW) with sc(BSE:P@GW+G3W2)). It needs to be pointed out that we always use full frequency dependent \( W(\nu) \) in the evaluation of self energy vertex correction (as we already mentioned before). As it was discussed in Ref. [18], the use of static \( W(\nu = 0) \) results in qualitatively incorrect correction to band gaps: they are increasing instead of decreasing when one uses full \( W(\nu) \). The effect of vertex correction to self energy is smaller than the effect of vertex correction to polarizability but it still is important as the total result (band gaps as obtained in sc(BSE:P@GW+G3W2)) is very close to the experimental band gaps with remaining discrepancy mostly attributed to the electron-phonon interaction which was not included in present study. For instance, the biggest discrepancy in the band gap of 0.5 eV (case of carbon) can be nicely accounted for by considering the corresponding electron-phonon band reduction (\( \sim 0.4 \) eV, [23,45,46]). Finally, Fig. 6 also includes the band gaps evaluated with conserving (in Baym-Kadanoff[13] definition) approach sc(GW+G3W2). This approach has a merit of not only being conserving but also of being more computationally efficient because only one diagram (of first order) in the sequence of Fig. 1 has to be evaluated for polarizability vertex correction. As one can judge from Fig. 6, sc(GW+G3W2) approach is especially useful for small gaps semiconductors where the corresponding band gaps are close to the calculated with sc(BSE:P@GW+G3W2) band gaps. For large gap insulators, however, solving of full BSE for polarizability vertex correction is essential.

Let us discuss now the results obtained with QP self-consistency which are presented in Fig. 7. As one can conclude, our results confirm that at GW level, quasi-particle self-consistency works better than full self-consistency. Especially it is noticeable for small gap (1–3 eV) materials. Next important observation from Fig. 7 is that our QPsc(BSE0TD:P@GW) results are not only in qualitative but also in quantitative agreement with the corresponding results obtained using the Questaal code (see Fig. 6 in Ref. [13]). Similar to Ref. [13], QPsc(BSE0TD:P@GW) approach does a good job in bringing the calculated band gaps in close agreement with experiments (especially if one takes into account the electron-phonon correction as authors of Ref. [13] demonstrate). Similar to the work [13],
there is one notable exception, CuCl, where the calculated with QPsc(BSE0TD:P@GW) band gap is severely underestimated. Thus, we are arriving to an important conclusion that using the same approximation, QPsc(BSE0TD:P@GW), we are able to reproduce the results obtained with the Questaal code in Ref. [13]. It is, therefore, interesting that partial vertex correction (correction to only P, static $W(\nu = 0)$, TD approximation) combined with quasi-partial self-consistency mimics the total result (vertex corrections to both P and $\Sigma$, full $W(\nu)$, no Tamm-Dancoff approximation) obtained with full self-consistency. However, in disagreement with the speculations made by authors of Ref. [13] that results can be further improved by using full $W(\nu)$ and also adding vertex correction to self energy, we see from Fig. 7 that both speculated 'improvements' result in too big reduction of the calculated band gaps making them seriously underestimated. In fact, when self energy vertex correction is taken into account, calculations become unstable for small gap materials. Thus, as it seems, the consideration about Z-factor cancellation done by authors of work [29] works. But, in conjunction with vertex corrections, it works in negative direction essentially explaining the fact that QP self-consistency should not be combined with vertex corrections. One more result, which one can get from Fig. 7 is that TD approximation is of rather poor quality (especially for large gap insulators) when one uses it in the context of band gaps evaluation which involves integration over the Brillouin zone. This finding corroborates with the finding made by authors of Ref. [47] that TD approximation fails for finite momentum transfers. Summarizing our observations of QP self-consistency, we can state that this variant of self-consistency can only be combined with vertex corrections if one makes additional (which are not assumed in Hedin's equations) approximations such as Tamm-Dancoff approximation, static $W(\nu = 0)$, and no Tamm-Dancoff approximation.

Conclusions

The study conducted in the present work resulted in two principal conclusions. The first one, which clearly represents a positive achievement, is that two codes (Questaal and FlapwMBPT) produce similar band gaps in vertex corrected QSGW calculations for a number of materials provided that vertex corrections are evaluated similarly (correction to only polarizability, static $W(\nu = 0)$, Tamm-Dancoff approximation for BSE). The second conclusion is that when one adds diagrams (be-
Beyond GW) in self-consistent calculations one should use full self-consistency approach. Namely, Green’s function has to be properly evaluated from Dyson’s equation without referring to the quasi-particle approximation. As it is shown in this study, combining the vertex corrected calculations with QP self-consistency as it is advocated in Refs. [13–15], can only be successful if vertex corrections are evaluated with a number of restrictive approximations or constraints such as polarizability only correction, static $W(\nu = 0)$ in vertex diagrams, and Tamm-Dancoff approximation when solving BSE. From this point of view, quasi-particle self-consistency combined with vertex-corrected GW approach can be considered as an ad hoc theory where one imposes specific constraints on the vertex part in order to avoid too large (and destructive for the final result) effect. Nevertheless, the approach still can be useful from practical point of view allowing one to quickly estimate the possible effect of vertex corrections before addressing the problem with full vertex and full self-consistency.

The second conclusion, as it seems, is in a contradiction with our previous advocating the combination of QSGW and dynamical mean field theory (QSGW+DMFT, [48, 49]). Formally, the addition of DMFT to QSGW can be considered as a vertex correction and, according to the discussion above, cannot be a valid approximation. However, similar to the approach used in Refs. [13–15], our implementation of QSGW+DMFT also uses ‘constraints’ for the vertex (DMFT) part: i) only one iteration which includes DMFT (one-shot type of DMFT correction performed on top of QSGW); ii) effective interaction in DMFT part is not evaluated from proper DMFT self-consistency condition, [50], but is provided by constrained random-phase approximation (cRPA, [51]). Specifically, the second constraint (using the cRPA) can clearly be considered as an ad hoc part where the setup parameters of cRPA are adjusted in order to get reasonable effective interaction. Thus, the success of our QSGW+DMFT calculations can also be attributed to the use of ‘constraints’ in the vertex part.

Acknowledgments

This work was supported by the U.S. Department of energy, Office of Science, Basic Energy Sciences as a part of the Computational Materials Science Program.
A. L. Kutepov, V. S. Oudovenko, G. Kotliar, Comp. Phys. Comm. 219, 407 (2017).
G. Baym, and L. P. Kadanoff, Phys. Rev. 124, 287 (1961).
C. O. Almladl, U. von Barth and R. van Leeuw, Int. J. of Mod. Phys. B 13, 535 (1999).
The latest publicly available version of the FlapwMBPT code (FlapwMBPT2106) can be downloaded from the website https://github.com/andreykutepov65/FlapwMBPT.
C. Friedrich, A. Schindlmayr, and S. Blügel, T. Kotani, Phys. Rev. B 74, 045104 (2006).
C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B 83, 081101 (2011).
C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B 84, 039906(E) (2011).
H. Jiang, and P. Blaha, Phys. Rev. B 93, 115203 (2016).
D. Nabok, A. Gulans, and C. Draxl, Phys. Rev. B 94, 035118 (2016).
C. Vorwerk, B. Aurich, C. Cocchi, and C. Draxl, Electron. Struct. 1, 037001 (2019).
F. J. Himpsel, J. A. Knapp, J. A. van Vechten, and D. E. Eastman, Phys. Rev. B 20, 624 (1979).
P. Lautenschlager, M. Garriga, S. Logothetidis, and M. Cardona, Phys. Rev. B 35, 9174 (1987).
A. Mang, K. Reimann, St. Rübenacke, and M. Steube, Phys. Rev. B 53, 16283 (1996).
M. Cardona, Solid State Commun. 133, 3 (2005).
B. Monserrat, G. J. Conduit, and R. J. Needs, Phys. Rev. B 90, 184302 (2014).
T. Sander, E. Maggio, and G. Kresse, Phys. Rev. B 92, 045209 (2015).
S. Choi, A. Kutepov, K. Haule, M. van Schilfgaarde, and G. Kotliar, NPJ Quantum Materials 1, 16001 (2016).
S. Choi, P. Semon, B. Kang, A. Kutepov, G. Kotliar, Comp. Phys. Comm. 244, 277 (2019).
S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. 90, 086402 (2003).
F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A. I. Lichtenstein, Phys. Rev. B 70, 195104 (2004).