

Full versus quasiparticle self-consistency in vertex-corrected GW approachesAndrey L. Kutepov ^{*}*Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, New York 11973, USA*

(Received 3 September 2021; revised 20 November 2021; accepted 10 January 2022; published 19 January 2022)

Using seven semiconductors and insulators with band gaps covering the range from 1 to 10 eV we systematically explore the performance of two different variants of self-consistency associated with the famous Hedin system of equations: the full self-consistency and the so-called quasiparticle approximation to it. The pros and cons of these two variants of self-consistency are sufficiently well documented in the literature for the simplest GW approximation to the Hedin 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 quasiparticle self-consistency has certain advantages at the GW level (a 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 quasiparticle self-consistency, however, an inclusion of the same diagrams has a considerably larger effect and calculated band gaps become severely underestimated. Different effects of vertex corrections in two variants of self-consistency can be related to the Z -factor cancellation which plays a positive role in quasiparticle self-consistency at the GW level of theory but appears to be destructive for the quasiparticle approximation when higher-order diagrams are included. The 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 (quasiparticle) 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.

DOI: [10.1103/PhysRevB.105.045124](https://doi.org/10.1103/PhysRevB.105.045124)**I. 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 [1]. The general consensus is that modern electronic structure codes, however, employ different basis sets [linearized augmented plane wave (LAPW), linear muffin tin orbitals method (LMTO), projector augmented wave (PAW), etc.], and upon convergence they demonstrate quite similar electronic structure of materials. When one goes beyond the DFT approximation (for instance if one uses Hedin's GW approach [2]) 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 are 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 sets to represent them. For example, the so-called product basis (PB) set [3,4] 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 [5], the reproducibility of the molecular electronic structure, though not perfect, is still acceptable in many cases. Namely, by comparing the G0W0 approach implemented in the three different codes FHI-AIMS [6,7], BERKELEYGW [8], and TURBOMOLE [9,10], the authors of Ref. [5] concluded that at convergence, the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) levels agree on the order of 200 meV. In the process of their work, the authors of Ref. [5] also identified two crucial aspects that control the accuracy of the G0W0 quasiparticle 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 appear [11–22]. 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 the GW approximation are the following: (i) specific sets of diagrams beyond GW and (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 are far beyond the scope of a single study. The objective of this

^{*} akutepov@bnl.gov

work is more specific and it was motivated by recent vertex-corrected GW calculations of the band gaps in semiconductors performed with the FLAPWMBPT code [18] and with the QUESTAAL code [12].

In both studies [12,18], it is concluded that vertex corrections provide the biggest numerical improvement of the results obtained with fully self-consistent GW (scGW [18]) or with quasiparticle self-consistent GW (QSGW [12]). The authors of work [12] 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. [12] but also from earlier works [23–26]. The list of materials studied in Ref. [12] is slightly different (and also longer) than the corresponding list in Ref. [18]. But the 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. [12] where vertex correction worsens the QSGW result (the 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 equations [2], the details of the applied approximations differ a lot. In Ref. [18], vertex corrections are used to improve both polarizability and self-energy Σ . The authors of Ref. [12] use vertex correction only for polarizability. Further, all vertex corrections in Ref. [18] (for P and for Σ) use fully frequency-dependent screened interaction $W(\nu)$, whereas vertex correction to P in Ref. [12] is evaluated with frequency-independent (taken at zero frequency) $W(\nu = 0)$. Also, the Tamm-Dancoff (TD) approximation [27] was used in Ref. [12]. 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 a full or quasiparticle Green's function in the evaluation of diagrams. In Ref. [18] all vertex-corrected calculations are performed with full self-consistency applied to Green's function G . At the same time, the authors of Ref. [12] use additional [the so-called quasiparticle (QP)] approximation which is not intended in Hedin's equations. The validity of the quasiparticle approximation is well justified at the level of the GW method (without vertex corrections). It is known that the QSGW approach usually is more accurate than the fully self-consistent scGW method [18,28]. The success of the 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. [29]. Briefly, the essence of the trick is that diagrams should be evaluated either with full Green's function (no QP approximation) and including the vertex part or with the QP Green's function but excluding the vertex part. Therefore, if one excludes the vertex part (GW level of theory) then it is of advantage to use the QP approximation. Alternatively, if one intends to apply vertex corrections, the 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, the simultaneous

use of vertex corrections and QP approximation for G should be questioned for consistency. Some results on the inadequacy of quasiparticle self-consistency were published a few years ago using a Hubbard-dimer model as an example [30].

The wish to understand deeper the reason why two seemingly different approximations used in Refs. [12,18] 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. [12] (with the QUESTAAL code) using the same approximations as in Ref. [12] but running a different code (FLAPWMBPT). Second, we extend both studies [12,18] by performing the calculations which are intended to trace step by step the differences in the results obtained with full self-consistency and with quasiparticle self-consistency when approximate forms of vertex corrections are applied. The reason for doing this is the following: whereas full vertex correction to QSGW [i.e., for both polarizability and self-energy and with full $W(\nu)$] should be expected to be destructive (because of the unphysical $1/Z$ factor [29]), a partial vertex correction to QSGW (like the one applied in Ref. [12]) can be useful. Therefore, experimenting with approximate forms of vertex correction and with both forms of self-consistency can be instructive for future prospects of vertex-corrected GW calculations. In order to accomplish the above goal, we conduct two sets of calculations. 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. [18].

The second set of calculations consists of exactly the same steps but all calculations in the second set are supplemented with the QP approximation for the 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. [12]. As the authors of Ref. [12] also used the Tamm-Dancoff approximation [27] in their evaluation of vertex correction to polarizability, the calculations with QP self-consistency [and with static $W(\nu = 0)$ approximation] in the 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 Σ , represent the steps which the authors of Ref. [12] 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 the 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 a similar level of approximations result in very similar results when one uses the QUESTAAL code or FLAPWMBPT code. At the same time, it is also shown below that a few omissions (or rather "constraints") accepted in Ref. [12], namely, (a) the Tamm-Dancoff approximation, (b) insufficient basis set (number of unoccupied states included in vertex corrections), (c)

vertex corrections that are applied to polarizability only, but not to self-energy, and (d) a static approximation for screened interaction with its value taken at zero frequency, when removed, all result in reduction of the calculated band gaps. Therefore, with the above constraints removed (with QP self-consistency), the obtained band gaps demonstrate a 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 (Methods). The discussion of convergence issues and of setup parameters is provided next. The principal results obtained in this work are presented in the Results section. The conclusions are given afterwards.

II. METHODS

All calculations in this study formally are based on the Hedin equations [2]. 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 (integration over arguments that appear on the right-hand side but not on the left is implied): the three-point vertex function from the Bethe-Salpeter equation,

$$\Gamma^\alpha(123) = \delta(12)\delta(13) + \sum_\beta \frac{\delta \Sigma^\alpha(12)}{\delta G^\beta(45)} G^\beta(46) \Gamma^\beta(673) G^\beta(75), \quad (1)$$

where α and β are spin indices, and the digits in the brackets represent space–Matsubara's time arguments; polarizability,

$$P(12) = \sum_\alpha G^\alpha(13) \Gamma^\alpha(342) G^\alpha(41); \quad (2)$$

screened interaction,

$$W(12) = V(12) + V(13)P(34)W(42); \quad (3)$$

and the self-energy,

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

In Eq. (3), V stands for the bare Coulomb interaction. A new approximation for the Green's function is obtained from Dyson's equation,

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

where G_0 is the Green's function in Hartree approximation. G_0 depends on the Hartree potential which depends on the electronic density (i.e., on full G) from the previous iteration and, therefore, G_0 is updated in the course of iterations. Equations (1)–(5) comprise one iteration. If convergence is not yet reached one can go back to Eq. (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 Eq. (1) for the vertex function $\Gamma^\alpha(123)$ in order to make the solving of the system manageable in practice. Approximations which

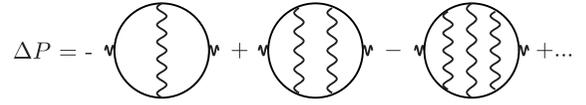


FIG. 1. Ladder sequence of diagrams for the vertex correction to polarizability.

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:

(i) Reproduce the results of Ref. [12].

(ii) Check the validity of speculations made in Ref. [12] about the possibility to improve QSGW-based results by taking into account the frequency dependence of W when solving the Bethe-Salpeter equation (BSE) and adding vertex correction to self energy.

(iii) 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. The starting point for the first set is the scGW approximation, whereas the second set of calculations has the QSGW approach as a starting point (as in Ref. [12]). In the scGW method, vertex function (1) is approximated by its trivial part $\Gamma^\alpha(123) = \delta(12)\delta(13)$. In QSGW, we use an additional approximation related to the self-energy in Eq. (5); namely, we linearize the frequency dependence of self-energy around zero frequency (see details in Refs. [31,32] or in Appendix D). In this respect, our construction of QSGW differs from Ref. [12] 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 the above two variants of QSGW are pretty much similar for the majority of materials. Concerning the vertex part, all vertex-corrected calculations included the solution of the BSE for polarizability with screened interaction W as a kernel of the BSE [$\frac{\delta \Sigma^\alpha(12)}{\delta G^\beta(34)} \approx \delta_{\alpha\beta} \delta(13)\delta(24)W(12)$]. A 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. [12], namely, the Tamm-Dancoff approximation and the use of frequency-independent $W(\nu = 0)$. Finally, our 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 the 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 the vertex function when it is used in the expression for polarizability, Eq. (2), and in the expression for self-energy, Eq. (4), and, as a

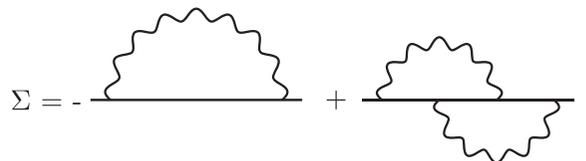


FIG. 2. Diagrammatic representation of self-energy up to the second order in screened interaction W .

$$\Psi = -\frac{1}{2} \text{ (circle with wavy line) } + \frac{1}{4} \text{ (circle with wavy line and vertical zigzag) }$$

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

result, they are not conserving in the Baym-Kadanoff definition [33] (i.e., corresponding P and Σ cannot be obtained from the same functional). In order to check the effect of using a scheme which is conserving, we also included vertex-corrected calculations (with full self-consistency only, not QP) where the vertex correction to polarizability consists of the first only term shown in Fig. 1 and the vertex correction to self-energy consists of the second term in Fig. 2. This approach, sc(GW + G3W2), as well as scGW, can also be defined using the Ψ -functional formalism of Almladh *et al.* [34] and, therefore, is conserving in the sense of Baym and Kadanoff [33]. The corresponding Ψ functional which includes vertex corrections is shown in Fig. 3. In Fig. 3, the first diagram corresponds to the GW approximation, whereas the sum of the first and the second diagrams represents the sc(GW + G3W2) approximation. Diagrammatic representations for irreducible polarizability (Fig. 4) and for self-energy (Fig. 2) in sc(GW + G3W2) follow from the chosen approximation for the Ψ functional.

For convenience, we list here all variants of approximations used in this study together with the purpose and with the corresponding abbreviations. The first set of calculations includes scGW, sc(BSE0 : P@GW), sc(BSE : P@GW), sc(BSE : P@GW + G3W2), and sc(GW + G3W2).

scGW. This is used primarily to generate the initial approximation to start vertex-corrected calculations and, by doing this, to reduce the number of iterations with vertex corrections. We also compare the scGW results with the ones obtained with QSGW.

sc(BSE0 : P@GW). The part after the symbol @ stands for diagrammatic representation of self-energy, whereas the part before the symbol @ says that polarizability is evaluated from the BSE with static screened interaction taken at zero frequency [$W(\nu = 0)$] 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). One uses fully frequency-dependent $W(\nu)$ in the BSE. The goal of this variant is to assess the

$$P = \text{ (circle with wavy line) } - \text{ (circle with wavy line and vertical zigzag) }$$

FIG. 4. Diagrammatic representation of irreducible polarizability in the simplest conserving vertex-corrected scheme sc(GW + G3W2).

quality of the static approximation for W in the 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). Diagrammatic representation of self-energy includes a second-order (in W) diagram. All diagrams in this variant (for P and for Σ) use fully frequency-dependent $W(\nu)$. 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). This variant is conserving in the Baym-Kadanoff definition. Only a diagrammatic definition of self-energy (GW + G3W2) is needed to be specified in this case. Diagrammatic representation for P , G2 + G4W1, follows if one uses the same first-order vertex function (as for self-energy) in Eq. (2). All diagrams in this variant (for P and for Σ) also use fully frequency-dependent $W(\nu)$. The goal of this variant is to assess the effect of applying the conserving approximation.

The second set of calculations includes QSGW, QS(BSE0TD : P@GW), QS(BSE0 : P@GW), QS(BSE:P@GW), and QS(BSE:P@GW+G3W2).

QSGW. QSGW is used to compare the results with the ones obtained in Ref. [12] at the same level of theory.

QS(BSE0TD : P@GW). The part in brackets before the symbol @ says that polarizability is evaluated from the BSE with static screened interaction taken at zero frequency [$W(\nu = 0)$] as the kernel of the BSE. Plus, the Tamm-Dancoff approximation is assumed. The basis size for the BSE is taken exactly as in Ref. [12]. Initial symbol “QS” stand for quasiparticle self-consistency. The goal of this variant is to compare the results with the ones obtained in Ref. [12] at the same level of theory and to assess the effect of the Tamm-Dancoff approximation.

QS(BSE0 : P@GW). This is the same as sc(BSE0 : P@GW) but with QP self-consistency instead of full. The goal of this variant is to assess the effect of the Tamm-Dancoff approximation and of the differences in size of the basis set for the BSE.

QS(BSE : P@GW). This is the same as sc(BSE : P@GW) but with QP self-consistency instead of full. The goal of this variant is to assess the quality of the static approximation for W in the BSE [by comparing results with QS(BSE0 : P@GW) in the case of QP self-consistency].

QS(BSE : P@GW + G3W2). This is the same as sc(BSE : P@GW + G3W2) but with QP self-consistency instead of full. The goal of this variant is to assess the effect of inclusion of vertex correction to self-energy when one uses QP self-consistency.

All calculations in this work were performed using the code FLAPWMBPT [35]. Technical details of the GW part were described in Refs. [31,32]. Detailed accounts of the implementation of vertex-corrected schemes can be found in Refs. [17–20]. Figure 5 presents the flowchart of the calculations (for three selected approximations) which gives a general idea of how the calculations are organized. The flowchart in Fig. 5 corresponds to full self-consistency. In the case of quasiparticle self-consistency, the formal change consists only in the fact that instead of Dyson’s equation ($G =$

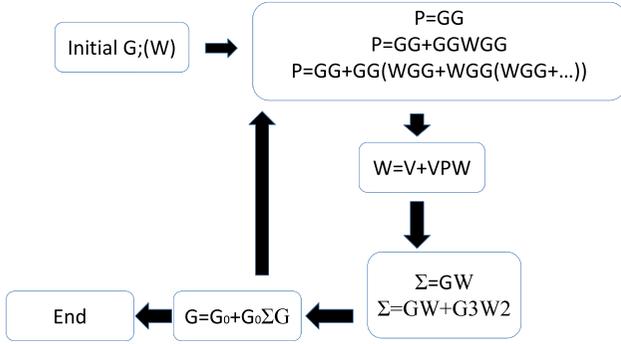


FIG. 5. Flowchart of scGW, sc(GW + G3W2), and sc(BSE : P@GW + G3W2) calculations. All equations are presented using symbolic notations. In the expressions for polarizability, the first equation corresponds to scGW, the second equation is used in sc(GW + G3W2), and the third one in sc(BSE : P@GW + G3W2). In the expressions for self-energy, the first equation corresponds to scGW, and the second one to both sc(GW + G3W2) and sc(BSE : P@GW + G3W2). G_0 stands for the Green's function in Hartree approximation. Any calculation begins with self-consistent DFT iterations where the basis set is formed and the initial approach for G is generated. Iterations of the scGW method use this initial Green's function as an input in order to start. During scGW iterations, G is updated and screened interaction W is generated. Both G and W serve as inputs to start iterations of sc(GW + G3W2) or sc(BSE : P@GW + G3W2) approaches. sc(BSE : P@GW + G3W2), being computationally most demanding, can be run after a few iterations of sc(GW + G3W2), which can save computer time.

$G_0 + G_0 \Sigma G$, 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, the GW part of the code scales as $N_k N_\omega N_b^3$ where N_k is the number of \mathbf{k} points in the Brillouin zone, N_ω 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 $N_k^2 N_\omega^2 N_b^4$. For comparison, if one uses naive (all in reciprocal space and frequency) implementation then the 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 \mathbf{k} points [17,18]. Second is that the higher-order diagrams are very well suited for massive parallelization.

III. 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)

TABLE I. Setup parameters of the solids studied in this work. Lattice parameters are in angstroms. N_{bnd}^{GW} is the number of band states used as a basis set for evaluation of the GW part. N_{bnd}^{VRT} represents the corresponding number for the vertex part.

Solid	Space group	Lattice parameter	N_{bnd}^{GW}	N_{bnd}^{VRT}
Si	227	5.43	160	20
AIP	216	5.451	185	14
CuCl	216	5.64	260	24
C	227	3.57	160	14
MgO	225	4.217	110	16
NaCl	225	5.62	150	16
LiCl	225	5.13	120	16

a broad range of energies (1–10 eV); (ii) selected materials should be taken from the list studied in Ref. [12] (only AIP does not fit in this constraint); and (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 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 600 K (Matsubara's formalism is used throughout the work). Whereas specific electronic temperature is not very critical for materials with a band gap, its value establishes certain minimal values for the numbers of Matsubara's time and frequency points which one can use in the calculation. Figure 6 shows the dependence of the MgO band gap as a function of the corresponding numbers for two temperatures, 300 and 600 K. As it was explained in Ref. [31], we use inhomogeneous grids of points on both the imaginary time interval $[0 : 1/T]$ (T is temperature) and Matsubara's frequency interval $[0 : \infty]$. The functions are evaluated and stored only at these inhomogeneous grids of points. In our calculations we used the same number of time or frequency points. This choice is reasonable because it ensures that accuracy of representation of functions on the time or frequency interval is similar. As one can see from Fig. 6, the essential difference between 300 and 600 K is that stable calculation can be performed at a somewhat smaller minimal number

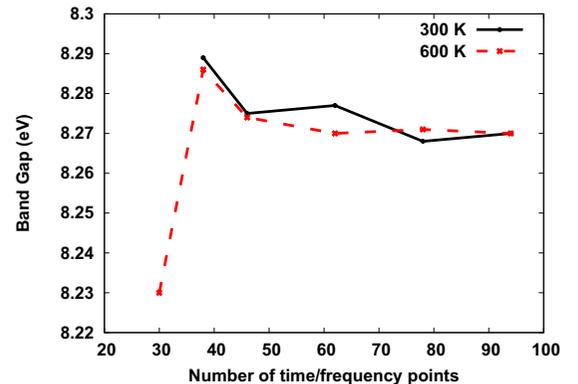


FIG. 6. Convergence of the MgO band gap obtained in the sc(GW + G3W2) approximation with respect to the number of time or frequency points used. See the text for more details.

of time or frequency points when the temperature is higher. Also, general convergence is achieved faster at 600 K. In order to obtain all below-presented results we used 62 time or frequency points. As one can judge from Fig. 6 it ensures the convergence of the calculated band gap at about 0.002 eV. The 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 the GW part (i.e., excluding vertex correction diagrams) was $12 \times 12 \times 12$ in all cases. Evaluation of the diagrams associated with the vertex part (i.e., all diagrams in Fig. 1 and the second diagram in Fig. 2) was performed with $3 \times 3 \times 3$ sampling. As Table I shows, the number of band states used in the evaluation of the vertex part also was considerably smaller than the number of band states included in the evaluation of the GW part of the diagrams. The fact, that the diagrams representing the vertex part require a smaller basis set and coarser sampling of the Brillouin zone was discussed before in Refs. [17,18], so the choice of these two setup parameters for the present study is justified (see, for instance, Table I in Ref. [18]). At the same time, one can notice that the basis set for the BSE part used in Ref. [12] is smaller than ours by almost a factor of 2 (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 performed our QS(BSE0TD@GW) calculations with the basis set (for the vertex part) exactly corresponding to the basis set used in Ref. [12]. Therefore, the difference between the band gaps obtained with QS(BSE0@GW) and QS(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 the TD approximation was prevailing.

It is well known that the LAPW basis set has to be supplemented with a sufficient number of high-energy local orbitals (HELOs) 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 the standard LAPW basis set by including three or four (per atom) HELOs of s type, two or three HELOs of p and d type, one or two HELOs of f type, and also one HELO of g type. In this respect, our additional basis set (HELOs) was also larger than the additional local orbital basis set used in Ref. [12] (see Table I there) which can be another reason for small differences in results at the QSGW level (besides the different way to introduce the quasiparticle approximation).

In most of our vertex-corrected calculations [excluding QS(BSE0TD@GW)], the BSE is solved iteratively [17] 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 the BSE. As we did show before (see Fig. 7 in Ref. [18]), 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 to 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 QS(BSE0TD@GW) approach which serves to reproduce the results obtained in Ref. [12] was implemented similar

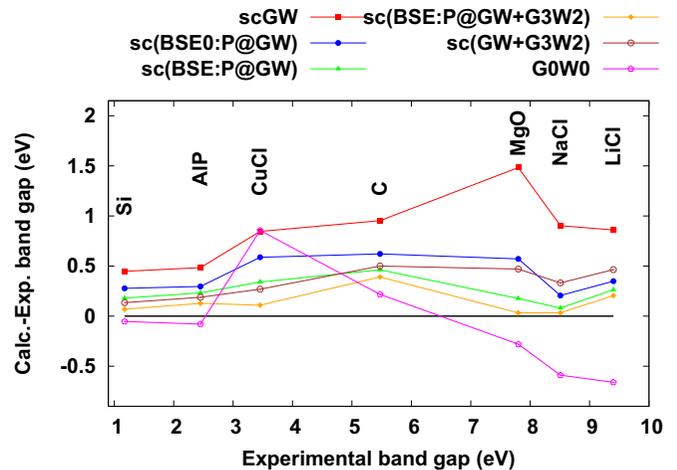


FIG. 7. Band gaps as obtained with full self-consistency. Positioning of the data for each compound relative to the x axis corresponds to the experimental band gap of the specific compound [29,44–46]. The y axis represents the deviations of the calculated band gaps from the experimental ones. Calculated results do not include electron-phonon interaction. G0W0 results are shown for the purpose of reference.

to that described in Ref. [12]. Namely, we solved the BSE directly (not iteratively as in all other our approaches) using a \mathbf{k} -dependent product basis for electron-hole pairs. Details of this basis set specific to the LAPW implementation can be found in Ref. [41]. The Tamm-Dancoff approximation, therefore, was implemented by neglecting the antiresonant part and keeping only the resonant part of the transition space [41].

As it will be shown later in this work, the approximation of fully frequency-dependent $W(\nu)$ by the static function $W(\nu = 0)$ gives qualitatively (but not quantitatively) correct results for vertex corrections to polarizability. However, as it was discussed before [17], a similar replacement of $W(\nu)$ by $W(\nu = 0)$ in the vertex correction for self-energy gives even qualitatively incorrect results. Namely, the 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 Grüneis *et al.* in Ref. [11], where static $W(\nu = 0)$ was used. By this reason, self-energy vertex corrections in this work were always evaluated with fully frequency-dependent $W(\nu)$.

IV. RESULTS

Principal results of this work are presented in Figs. 7 and 8 where the band gaps obtained with full self-consistency (Fig. 7) and with QP self-consistency (Fig. 8) are shown. For the purpose of reference, we also provide quasiparticle energies for high-symmetry points in the Brillouin zone for two materials, Si and NaCl, in Appendix E. Let us start our discussion with the full self-consistency. First, we observe that calculations without vertex corrections (scGW) severely overestimate the calculated band gaps (by 0.5–1.5 eV). Second,

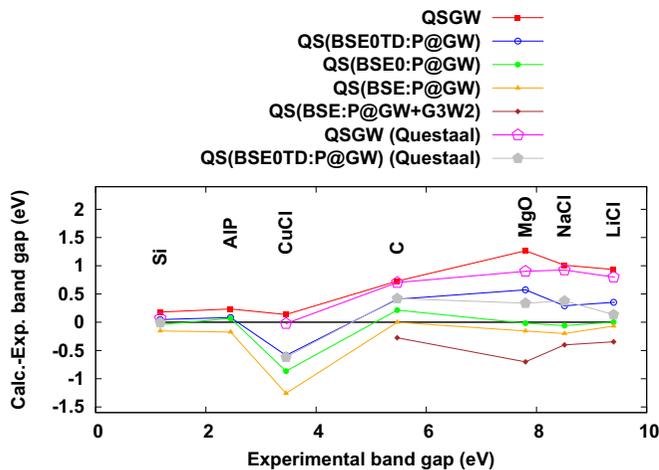


FIG. 8. Band gaps as obtained with quasiparticle self-consistency. Positioning of the data for each compound relative to the x axis corresponds to the experimental band gap of the specific compound [29,44–46]. The y axis represents the deviations of the calculated band gaps from the experimental ones. Calculated results do not include electron-phonon interaction. QS(BSE:P@GW+G3W2) band gaps are presented only for materials where the self-consistency iterations are stable.

principal improvement comes from the vertex correction to polarizability as it is evidenced in $sc(\text{BSE} : \text{P@GW})$ calculations. It is important to point out, however, that in order to get quantitatively correct results one has to use fully frequency-dependent $W(\nu)$ when solving the BSE which is used to obtain vertex correction to polarizability. Using static $W(\nu = 0)$ in BSE can only give a qualitatively correct correction to P , but quantitatively it underestimates the correction by 20–50 % [compare $sc(\text{BSE0} : \text{P@GW})$ with $sc(\text{BSE} : \text{P@GW})$]. Third, vertex correction to self-energy always reduces the calculated band gaps [compare $sc(\text{BSE} : \text{P@GW})$ with $sc(\text{BSE} : \text{P@GW} + \text{G3W2})$]. It needs to be pointed out that we always use fully frequency-dependent $W(\nu)$ in the evaluation of the self-energy vertex correction (as we already mentioned before). As it was discussed in Ref. [17], the use of static $W(\nu = 0)$ results in a 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 $sc(\text{BSE} : \text{P@GW} + \text{G3W2})$] is very close to the experimental band gaps with the remaining discrepancy mostly attributed to the electron-phonon interaction which was not included in the present study. For instance, the biggest discrepancy in the band gap of 0.5 eV (the case of carbon) can be nicely accounted for by considering the corresponding electron-phonon band reduction (~ 0.4 eV [23,42,43]). Finally, Fig. 7 also includes the band gaps evaluated with conserving (in Baym-Kadanoff [33] definition) approach $sc(\text{GW} + \text{G3W2})$. This approach has the 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. 7, the $sc(\text{GW} + \text{G3W2})$ approach is especially use-

ful for small-gap semiconductors where the corresponding band gaps are close to the band gaps calculated with $sc(\text{BSE} : \text{P@GW} + \text{G3W2})$. For large-gap insulators, however, solving of the full BSE for polarizability vertex correction is essential.

Let us discuss now the results obtained with QP self-consistency which are presented in Fig. 8. As one can conclude, our results confirm that at the GW level, quasiparticle self-consistency works better than full self-consistency. It is especially noticeable for small-gap (1–3 eV) materials. The next important observation from Fig. 8 is that our QS(BSE0TD : P@GW) results are not only in qualitative but also in quantitative agreement with the corresponding results obtained using the QUESTAAL code. Namely, the agreement for the smaller-gap materials (up to carbon) is perfect and the deviation of FLAPWMBPT gaps from the gaps obtained using QUESTAAL code does not exceed 0.2 eV for larger-gap insulators (i.e., it is at the level of the mismatch in the comparative test for 100 molecules [5]). Similar to Ref. [12], the QS(BSE0TD : P@GW) approach implemented in the FLAPWMBPT code 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 the authors of Ref. [12] demonstrate). Similar to the QUESTAAL results, there is one notable exception, CuCl, where the band gap calculated with QS(BSE0TD : P@GW) is severely underestimated. Thus, we arrive at an important conclusion that using the same approximation, QS(BSE0TD : P@GW), we are able to reproduce the results obtained with the QUESTAAL code in Ref. [12]. It is, therefore, interesting that partial vertex correction [correction to only P , static $W(\nu = 0)$, TD approximation] combined with quasiparticle self-consistency mimics the total result [vertex corrections to both P and Σ , full $W(\nu)$, no Tamm-Dancoff approximation] obtained with full self-consistency. However, in disagreement with the speculations made by the authors of Ref. [12] that results can be further improved by using full $W(\nu)$ and also adding vertex correction to self-energy, we see from Fig. 8 that both speculated “improvements” result in too big a 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 the authors of Ref. [29] works. But, in conjunction with vertex corrections, it works in the 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. 8, is that the TD approximation is of rather poor quality (especially for large-gap insulators) when one uses it in the context of band-gap evaluation which involves integration over the Brillouin zone. This finding corroborates the finding made by the authors of Ref. [47] that the 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 the Tamm-Dancoff approximation, static $W(\nu = 0)$, and no self-energy vertex correction. As it is now confirmed empirically by two different codes (QUESTAAL and FLAPWMBPT), such approximations still allow to improve QSGW band gaps but

TABLE II. Computational cost (average wall time in seconds per one iteration) of methods used in this study. The range corresponds to the shortest and longest times among the materials studied. Only timings for the approaches with full self-consistency are given. The switch from full self-consistency to the quasiparticle one changes the time very little. All calculations were performed using 192 MPI (Message passing interface) processes. Therefore, in order to get full CPU time per 1 MPI process (approximately) one has to multiply the given number with 192.

Approach	Shortest time	Longest time
scGW	540	1470
sc(BSE0:P@GW)	670	1710
sc(BSE:P@GW)	6110	24380
sc(BSE:P@GW+G3W2)	7920	27110
sc(GW+G3W2)	2030	4920

at the same time allow one to avoid the destructive effect of Z-factor cancellation when full vertex corrections are used in connection with QP self-consistency.

Table II shows how computational time (per one iteration) depends on the approximation in use. As one can see, evaluation of the vertex correction to polarizability with static W increases the scGW time only by about 20%. This is because the vertex corrections are evaluated on coarse grids of \mathbf{k} points and with considerable reduction in the basis set size (see the section Calculation Setups and Convergence Checks). However, when we consider fully frequency-dependent W , the time increases considerably (more than by a factor of 10 as compared to scGW), even with the above-mentioned reductions in the basis set size and in the number of \mathbf{k} points. Vertex correction to self-energy takes (very roughly) about the same time as one iteration of BSE (with fully frequency-dependent W) and, therefore, is a few times cheaper than the full BSE solution. In general, the computation time per iteration varies considerably from one approximation to another. Therefore, in order to reduce the total computational time we used a restart option and iterated the most time-consuming approaches starting from less expensive variants. For instance, for each material we always used scGW to start with. It takes 12–20 iterations to converge scGW for the materials studied. Converged scGW provides initial G and W for vertex-corrected schemes and allows to reduce number of iterations with vertex corrections by about factor of 2 as compared to if one starts with vertex-corrected iterations from local density approximation (LDA). In order to reduce the number of iterations in the two most computationally expensive approaches, sc(BSE:P@GW) and sc(BSE:P@GW+G3W2), even more, we began them from converged sc(BSE0:P@GW) and sc(GW + G3W2), correspondingly.

V. 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 $W(\nu = 0)$, and Tamm-Dancoff approximation for the BSE]. The second conclusion is that when one adds diagrams (beyond GW) in self-consistent calculations one should use a full self-consistency approach. Namely, the Green’s function has to be properly evaluated from Dyson’s equation without referring to the quasiparticle approximation. As it is shown in this study, combining the vertex-corrected calculations with QP self-consistency as it is advocated in Refs. [12–14] can only be successful if vertex corrections are evaluated with a number of restrictive approximations or “constraints” such as a polarizability-only correction, static $W(\nu = 0)$ in vertex diagrams, and the Tamm-Dancoff approximation when solving the BSE. From this point of view, quasiparticle self-consistency combined with a 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) an effect. Nevertheless, the approach still can be useful from a 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 contradiction with our previous advocating for 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. [12–14], our implementation of QSGW + DMFT also uses “constraints” for the vertex (DMFT) part: (i) only one iteration which includes DMFT is used (a one-shot type of DMFT correction performed on top of QSGW) and (ii) an effective interaction in the DMFT part is not evaluated from a 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 the cRPA are adjusted in order to get reasonably 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.

APPENDIX A: SELF-CONSISTENT GW METHOD

In this Appendix we outline the steps which are performed when one uses the scGW approximation. The details were published before in Refs. [31,32,52,53]. In the following, we provide the most important steps of the algorithm, giving references for more details. In references, we differentiate, if necessary, between the fully relativistic formulation of the scGW approach [31] and its non(scalar)-relativistic approach [32].

(1) Perform self-consistent LDA calculation. The basis of band states is obtained here [52,53].

- (2) Construct the product basis set (Sec. 4 in Ref. [32]).
- (3) Calculate the matrix elements of the bare Coulomb interaction in \mathbf{q} space using the product basis (Sec. II A in Ref. [31]).
- (4) Calculate the effective (or “pseudo”) bare Coulomb interaction in \mathbf{R} space (Sec. II G in Ref. [31]).
- (5) Perform initial approximation for the Green’s function (using LDA one-electron energies) in $(\mathbf{k}; \tau)$ representation (Eq. (17) in Ref. [31]).
- (6) Transform the Green’s function from $(\mathbf{k}; \tau)$ to $(\mathbf{R}; \tau)$ representation (Sec. II C in Ref. [31]).
- (7) Calculate the polarizability in $(\mathbf{R}; \tau)$ variables (Sec. II D in Ref. [31] or Sec. 6 in Ref. [32]).
- (8) Transform the polarizability from $(\mathbf{R}; \tau)$ to $(\mathbf{q}; \nu)$ representation (Sec. II E in Ref. [31]).
- (9) Calculate the screened interaction W in $(\mathbf{q}; \nu)$ representation (Sec. II F in Ref. [31]).
- (10) Transform the screened interaction from $(\mathbf{q}; \nu)$ to $(\mathbf{R}; \tau)$ representation (Sec. II G in Ref. [31]).
- (11) Calculate the exchange (static) part of the self-energy in (\mathbf{R}) representation (Sec. II H in Ref. [31]).
- (12) Transform the exchange part of the self-energy from (\mathbf{R}) to (\mathbf{k}) representation (Sec. II H in Ref. [31]).
- (13) Solve the effective Hartree-Fock eigenvalue problem including the core levels, and obtain the exchange part of the Green’s function in $(\mathbf{k}; \tau)$ representation (Sec. II I in Ref. [31]).
- (14) Calculate the correlation (dynamic) part of the self-energy in $(\mathbf{R}; \tau)$ representation (Sec. II H in Ref. [31] or Sec. 8 in Ref. [32]).
- (15) Transform the correlation part of the self-energy from $(\mathbf{R}; \tau)$ to $(\mathbf{k}; \omega)$ representation (Sec. II H and Appendix A in Ref. [31]).
- (16) Calculate the correlation part of the Green’s function in $(\mathbf{k}; \omega)$ representation (Sec. II J in Ref. [31]).
- (17) Transform the correlation part of the Green’s function from $(\mathbf{k}; \omega)$ to $(\mathbf{k}; \tau)$ representation (Appendix B in Ref. [31]).
- (18) Add the exchange and correlation parts of the Green’s function to obtain the full Green’s function in $(\mathbf{k}; \tau)$ representation.
- (19) Go to step 6.

APPENDIX B: VERTEX CORRECTION TO POLARIZABILITY

In this section we provide the steps which one follows when solving the Bethe-Salpeter equation for the vertex correction to polarizability. Our implementation of the BSE uses full frequency dependence of screened interaction W opposite to a common approximation [54,55] where one uses static (frequency-independent and taken at zero frequency) screened interaction W . As a result, the BSE is solved iteratively in this study. Each iteration adds one more diagram from an infinite sequence shown in Fig. 1 into the vertex correction to polarizability ΔP . Therefore, the first-order correction to polarizability which is embodied, for instance, in the sc(GW + G3W2) approach represents just one iteration of the current scheme. Here we give the steps of iterations with some comments on where one can find more detailed information. A full (and rather lengthy) account of the implementation

was published in Ref. [17], which includes the details of the basis sets, \mathbf{k} dependencies, and handling of time-to-frequency and frequency-to-time transformations. In this brief account, space arguments of all functions are represented by digits. Integration over repeated space arguments (if they are only on the right-hand sides of equations) is assumed. Below we use auxiliary functions K_0 , K , ΔK , and $\Delta\Gamma$, which are defined by the corresponding equations.

- (1) Before the iterations we evaluate the quantity K_0 ,

$$K_0(123; \omega, \nu) = -G(13; \omega)G(32; \omega - \nu), \quad (\text{B1})$$

and assign $\Delta K = 0$. ω and ν are fermionic and bosonic Matsubara frequencies, correspondingly. Also we transform $K_0(123; \tau, \nu) = \frac{1}{\beta} \sum_{\omega} e^{-i\omega\tau} K_0(123; \omega, \nu)$ where τ is Matsubara’s time and $\beta = 1/T$. Diagrammatically, the quantity K_0 represents two Green’s function lines joined at one point (the rightmost part of each term in Fig. 1). More details are in Appendices A2 and A3 of Ref. [17].

During each iteration we perform the following steps [Eqs. (B2)–(B6)]:

- (2) Form the full quantity K from K_0 and correction ΔK :

$$K(123; \tau, \nu) = K_0(123; \tau, \nu) + \Delta K(123; \tau, \nu). \quad (\text{B2})$$

- (3) Evaluate the nontrivial part of the three-point vertex function:

$$\Delta\Gamma(123; \tau, \nu) = W(21; \tau)K(123; \tau, \nu). \quad (\text{B3})$$

Diagrammatically, the quantity $\Delta\Gamma$ represents the sum of all terms shown in Fig. 1 with each term lacking two leftmost Green’s function lines. More details are in Appendix A4 of Ref. [17].

- (4) Transform the nontrivial part of the vertex function to a double-frequency representation:

$$\Delta\Gamma(123; \omega, \nu) = \int d\tau e^{i\omega\tau} \Delta\Gamma(123; \tau, \nu). \quad (\text{B4})$$

More details are in Appendix A4 of Ref. [17].

- (5) Evaluate the improved correction to the quantity K :

$$\Delta K(123; \omega, \nu) = -G(14; \omega)\Delta\Gamma(453; \omega, \nu)G(52; \omega - \nu). \quad (\text{B5})$$

Diagrammatically, the quantity ΔK represents the sum of all terms shown in Fig. 1 with each term having two leftmost Green’s function lines disconnected. More details are in Appendix A5 of Ref. [17].

- (6) Transform ΔK to a one-time or one-frequency representation:

$$\Delta K(123; \tau, \nu) = \frac{1}{\beta} \sum_{\omega} e^{-i\omega\tau} \Delta K(123; \omega, \nu). \quad (\text{B6})$$

More details are in Appendix A2 of Ref. [17].

- (7) Go to step 2 (if necessary). The above steps are repeated a specific number of times (iterations).

- (8) At the end of the iterations we evaluate the vertex correction to polarizability:

$$\Delta P(12; \nu) = -\Delta K(112; \tau = 0, \nu). \quad (\text{B7})$$

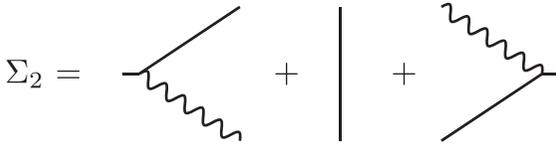


FIG. 9. Scheme of the evaluation of the second-order diagram for self-energy.

Diagrammatically, this step represents the connecting of two leftmost Green's function lines in Fig. 1. More details are in Appendix B of Ref. [17].

If one uses static approximation for W then only a simplified form of the quantities K_0 , K , and $\Delta\Gamma$ is needed, namely, with $\tau = 0$ in the arguments: $K(112; \tau = 0, \nu)$. This makes the iterations above considerably less time consuming. For weakly correlated semiconductors (which are the object of the present study), iterations (B2)–(B6) converge very fast (see, for instance, Fig. 7 in Ref. [18]).

APPENDIX C: VERTEX CORRECTION TO SELF-ENERGY

We consider the diagram of the second order represented by the second term in Fig. 2 which we, for convenience of presentation below, have divided into three terms as shown in Fig. 9. The meaning of “+” symbols in Fig. 9 is that three parts are connected to each other (not summed up) in the course of the evaluation process. The diagram has the following rendition in the formal math language:

$$\Sigma_2(12) = W(14)G(13)G(34)G(42)W(32), \quad (C1)$$

with digits representing space-time arguments and summation or integration over repeated indices on the right-hand side is assumed. Let us introduce the notations for convenience:

$$K(562) = G(52)W(62), \quad (C2)$$

which diagrammatically is the rightmost term in Fig. 9.

$$T(562) = G(65)K(562), \quad (C3)$$

which represents the combined second and third terms in Fig. 9.

$$S(342) = W(35)G(46)T(562), \quad (C4)$$

which represents all three terms combined but with leftmost ends of G and W lines disconnected.

Then the final result is the connection of the leftmost (open) lines of quantity S :

$$\Sigma_2(12; \omega) = S(112; \tau = 0; \omega). \quad (C5)$$

The fact that screened interaction W is represented by a sum of bare Coulomb interaction V and frequency-dependent screening part \tilde{W} makes it convenient to separate the evaluation of vertex correction to self-energy into three terms. For convenience, let us call the term with both interaction lines represented by \tilde{W} as dynamic, terms with one interaction line represented by V and another interaction line by \tilde{W} as semi-dynamic, and the term with both interaction lines represented by V as static. We begin the presentation of formulas with the dynamic case.

Let us write down the steps of the evaluation of the dynamic part of the self-energy with explicit time and frequency arguments. Digits in Eqs. (C6)–(C12), therefore, represent space arguments only. We begin with calculating the quantity K using a double-frequency representation:

$$K(562; \omega'; \omega) = G(52; \omega')\tilde{W}(62; \omega - \omega'). \quad (C6)$$

In order to proceed with the evaluation of quantity T , we have to transform quantity K into a one-time or one-frequency representation:

$$K(562; \tau; \omega) = \frac{1}{\beta} \sum_{\omega'} e^{-i\omega'\tau} K(562; \omega', \omega). \quad (C7)$$

Next goes the evaluation of quantity T in a one-time or one-frequency representation,

$$T(562; \tau, \omega) = G(65; -\tau)K(562; \tau, \omega), \quad (C8)$$

and the transform of it back to the double-frequency form (but with one frequency of bosonic type):

$$T(562; \nu; \omega) = \int e^{i\nu\tau} T(562; \tau, \omega). \quad (C9)$$

Evaluation of the quantity S is performed in double-frequency form,

$$S(342; \nu, \omega) = \tilde{W}(35; \nu)G(46; \omega - \nu)T(562; \nu, \omega), \quad (C10)$$

with subsequent transform to a one-time or one-frequency representation:

$$S(342; \tau; \omega) = \frac{1}{\beta} \sum_{\nu} e^{-i\nu\tau} S(342; \nu, \omega). \quad (C11)$$

Finally the correction to self-energy is obtained by connecting the leftmost line ends in Fig. 9:

$$\Sigma_2(12; \omega) = S(112; \tau = 0; \omega). \quad (C12)$$

Now we are going to repeat the above steps with explicit showing of the basis sets used to represent the dependence on the space arguments. In order to make the reading of the following easier, the notations have been collected here:

- (a) $\lambda, \lambda', \lambda''$ are the band indices.
- (b) \mathbf{k}, \mathbf{q} are points in the Brillouin zone.
- (c) i, j are product basis (PB) indices. When index i or j is used together with vector \mathbf{q} in the Brillouin zone (the corresponding PB function is $\Pi_i^{\mathbf{q}}$), the index runs over all PB (muffin tins plus interstitial). When it is used together with atomic index \mathbf{t} (the corresponding PB function in this case is $\Pi_i^{\mathbf{t}}$), it runs over the part of the full PB belonging to the given atom.
- (d) ω, ω' represent the fermionic Matsubara frequency.
- (e) ν is the bosonic Matsubara frequency.
- (f) τ, τ' represent Matsubara's time.
- (g) $\epsilon_{\lambda}^{\mathbf{k}}$ are band energies.
- (h) $\Psi_{\lambda}^{\mathbf{k}}$ are band wave functions.
- (i) β is the inverse temperature.
- (j) \mathbf{R} are vectors of translations in real space.
- (k) \mathbf{t}, \mathbf{t}' are coordinates (or indices) of atoms in the unit cell.

(l) L, L' are indices combining orbital moment l , its projection m , and other quantum numbers distinguishing the orbitals ϕ_{tL} for a given atom \mathbf{t} (L indices also distinguish between ϕ and $\dot{\phi}$).

(m) $N_{\mathbf{k}}$ is the full number of \mathbf{k} points in the Brillouin zone.

(n) \mathbf{r}, \mathbf{r}' are the points on the regular real space mesh in the unit cell. Below we use \mathbf{r} and \mathbf{r}' instead of digits for space arguments.

(o) \mathbf{G} are reciprocal lattice vectors.

(p) \mathbf{G}_i is the reciprocal lattice vector associated with the reduced product basis index i .

(q) Ω_0 is the primitive cell volume.

Now, let us use basis sets to represent space arguments. For the sake of efficiency, in some formulas below we use reciprocal space representation where quantities depend on momentum (\mathbf{k} -point) and band states. In other formulas we use (in a somewhat generalized sense) the real-space representation. The real-space representation is characterized by using unit cell indices instead of momentum and by using LAPW orbitals in muffin tin (MT) spheres and a real-space grid in the interstitial region instead of band states. Therefore, similar to the transform between time and frequency representations, we use transformations between reciprocal and real spaces. With this in mind, let us proceed. Formula (C6) in the basis of band states and product basis looks as the following (we do not show spin index for simplicity):

$$K_{\lambda i}^{\mathbf{k}'}(\mathbf{k}\lambda'; \omega'; \omega) = \sum_{j\lambda''} \langle \Psi_{\lambda'}^{\mathbf{k}} | \Psi_{\lambda''}^{\mathbf{k}'} \Pi_j^{\mathbf{k}-\mathbf{k}'} \rangle^* \times G_{\lambda\lambda''}^{\mathbf{k}'}(\omega') W_{ij}^{\mathbf{k}-\mathbf{k}'}(\omega - \omega'). \quad (\text{C13})$$

Next, we transform the quantity K to real space. The transform to the real space is different depending on where the space arguments \mathbf{r} and \mathbf{r}' are.

Mt-Mt (muffin tin-muffin tin):

$$K_{tL;i}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}'} e^{i\mathbf{k}'\mathbf{R}} \sum_{\lambda} Z_{tL;\lambda}^{\mathbf{k}'} K_{\lambda i}^{\mathbf{k}'}(\mathbf{k}\lambda'; \tau; \omega), \quad (\text{C14})$$

where the relation $\Psi_{\lambda}^{\mathbf{k}}(\mathbf{r})|_{\mathbf{t}} = \sum_L Z_{tL;\lambda}^{\mathbf{k}} \phi_{tL}(\mathbf{r})$ was used.

Int-Mt (interstitial-muffin tin):

$$K_{\mathbf{r};i}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}'} e^{i\mathbf{k}'\mathbf{R}} \frac{e^{i\mathbf{k}'\mathbf{r}}}{\sqrt{\Omega_0}} \sum_{\mathbf{G}} e^{i\mathbf{G}\mathbf{r}} \times \sum_{\lambda} A_{\mathbf{G};\lambda}^{\mathbf{k}'} K_{\lambda i}^{\mathbf{k}'}(\mathbf{k}\lambda'; \tau; \omega), \quad (\text{C15})$$

where the relation $\Psi_{\lambda}^{\mathbf{k}}(\mathbf{r})|_{\text{Int}} = \sum_{\mathbf{G}} A_{\mathbf{G};\lambda}^{\mathbf{k}} \frac{e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}}{\sqrt{\Omega_0}}$ was used.

Mt-Int (muffin tin-interstitial):

$$K_{tL;\mathbf{r}'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}'} e^{i\mathbf{k}'\mathbf{R}} \sum_{\lambda} Z_{tL;\lambda}^{\mathbf{k}'} \frac{e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}'}}{\sqrt{\Omega_0}} \times \sum_i K_{\lambda i}^{\mathbf{k}'}(\mathbf{k}\lambda'; \tau; \omega) e^{i\mathbf{G}_i\mathbf{r}'}. \quad (\text{C16})$$

Int-Int (interstitial-interstitial):

$$K_{\mathbf{r};\mathbf{r}'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}'} e^{i\mathbf{k}'\mathbf{R}} \frac{e^{i\mathbf{k}'\mathbf{r}}}{\sqrt{\Omega_0}} \sum_{\mathbf{G}} e^{i\mathbf{G}\mathbf{r}} \sum_{\lambda} A_{\mathbf{G};\lambda}^{\mathbf{k}'} \frac{e^{i(\mathbf{k}-\mathbf{k}')\mathbf{r}'}}{\sqrt{\Omega_0}} \times \sum_i K_{\lambda i}^{\mathbf{k}'}(\mathbf{k}\lambda'; \tau; \omega) e^{i\mathbf{G}_i\mathbf{r}'}. \quad (\text{C17})$$

Augmentation of the Green's function line (to form quantity T) and subsequent transform to the reciprocal space is again different dependent on where the arguments are.

Mt-Mt:

$$T_{tL;t'L'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) = - \sum_{L_1 L_2} \sum_{L_3 i'} G_{tL_1;t'L_2}^{*\mathbf{R}}(\beta - \tau) \times \langle \phi_{L_1}^{\mathbf{t}} | \phi_{L_3}^{\mathbf{t}} \Pi_i^{\mathbf{t}} \rangle K_{tL_3;t'i'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) \times \langle \phi_{L'}^{\mathbf{t}'} | \phi_{L_2}^{\mathbf{t}'} \Pi_{i'}^{\mathbf{t}'} \rangle, \quad (\text{C18})$$

$$T_{tL;\lambda''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \sum_{\mathbf{t}'L'} \left\{ \sum_{\mathbf{R}} e^{-i\mathbf{q}\mathbf{R}} T_{tL;t'L'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) \right\} Z_{\mathbf{t}'L';\lambda''}^{*\mathbf{k}-\mathbf{q}}. \quad (\text{C19})$$

Mt-Int:

$$T_{tL;\mathbf{r}'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) = - \sum_{L_1 L_3} G_{tL_1;\mathbf{r}'}^{*\mathbf{R}}(\beta - \tau) \times \langle \phi_{L_1}^{\mathbf{t}} | \phi_{L_3}^{\mathbf{t}} \Pi_i^{\mathbf{t}} \rangle K_{tL_3;\mathbf{r}'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega). \quad (\text{C20})$$

In order to transform this expression to the reciprocal space we first perform a Fourier transform (after $\mathbf{R} \rightarrow \mathbf{q}$):

$$T_{tL;\mathbf{r}'}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \sum_{\mathbf{G}''} T_{tL;\mathbf{G}''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) \frac{e^{i(\mathbf{k}-\mathbf{q}+\mathbf{G}'')\mathbf{r}'}}{\sqrt{\Omega_0}}, \quad (\text{C21})$$

with

$$T_{tL;\mathbf{G}''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \frac{\Omega_0}{N_{\mathbf{r}'}} \sum_{\mathbf{r}'} T_{tL;\mathbf{r}'}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) \frac{e^{-i(\mathbf{k}-\mathbf{q}+\mathbf{G}'')\mathbf{r}'}}{\sqrt{\Omega_0}}. \quad (\text{C22})$$

After that

$$T_{tL;\lambda''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \int_{\Omega_{\text{Int}}} d\mathbf{r}' T_{tL;\mathbf{r}'}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) \Psi_{\lambda''}^{*\mathbf{k}-\mathbf{q}}(\mathbf{r}') = \sum_{\mathbf{G}''} T_{tL;\mathbf{G}''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) \int_{\Omega_{\text{Int}}} d\mathbf{r}' \Psi_{\lambda''}^{*\mathbf{k}-\mathbf{q}}(\mathbf{r}') \frac{e^{i(\mathbf{k}-\mathbf{q}+\mathbf{G}'')\mathbf{r}'}}{\sqrt{\Omega_0}} = \sum_{\mathbf{G}''} T_{tL;\mathbf{G}''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) X_{\mathbf{G}'';\lambda''}^{\mathbf{k}-\mathbf{q}}, \quad (\text{C23})$$

with

$$X_{\mathbf{G};\lambda}^{\mathbf{k}} = \int_{\Omega_{\text{Int}}} d\mathbf{r} \Psi_{\lambda}^{*\mathbf{k}}(\mathbf{r}) \frac{e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}}{\sqrt{\Omega_0}} = \sum_{\mathbf{G}'} A_{\mathbf{G}';\lambda}^{*\mathbf{k}} \int_{\Omega_{\text{Int}}} d\mathbf{r} \frac{e^{-i(\mathbf{k}+\mathbf{G}')\mathbf{r}}}{\sqrt{\Omega_0}} \frac{e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}}}{\sqrt{\Omega_0}}. \quad (\text{C24})$$

Int-Mt:

$$T_{\mathbf{r};\nu'L}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) = - \sum_{L_2'} G_{\mathbf{r};\nu'L_2}^{\mathbf{R}}(\beta - \tau) \times K_{\mathbf{r};\nu'L}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) \langle \phi_{L_2'}^{\nu'} | \phi_{L_2}^{\nu'} \Pi_{\nu'}^{\nu'} \rangle. \quad (\text{C25})$$

In this case we first represent

$$T_{\mathbf{G}'';\nu'L}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \sum_{\mathbf{G}''} \frac{e^{i(\mathbf{q}+\mathbf{G}'')\mathbf{r}}}{\sqrt{\Omega_0}} T_{\mathbf{G}'';\nu'L}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega), \quad (\text{C26})$$

with

$$T_{\mathbf{G}'';\nu'L}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \frac{\Omega_0}{N_{\mathbf{r}}} \sum_{\mathbf{r}} \frac{e^{-i(\mathbf{q}+\mathbf{G}'')\mathbf{r}}}{\sqrt{\Omega_0}} \left\{ \sum_{\mathbf{R}} e^{-i\mathbf{q}\mathbf{R}} T_{\mathbf{r};\nu'L}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) \right\}. \quad (\text{C27})$$

After that

$$T_{\mathbf{G}'';\lambda''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \sum_{\mathbf{G}''} \int_{\Omega_{\text{Int}}} d\mathbf{r} \frac{e^{-i(\mathbf{q}+\mathbf{G}'')\mathbf{r}}}{\sqrt{\Omega_0}} \frac{e^{i(\mathbf{q}+\mathbf{G}'')\mathbf{r}}}{\sqrt{\Omega_0}} \times \sum_{\nu'L} T_{\mathbf{G}'';\nu'L}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) Z_{\nu'L;\lambda''}^{*\mathbf{k}-\mathbf{q}} = \sum_{\mathbf{G}''} S_{\mathbf{G}\mathbf{G}''}^{\mathbf{q}} \sum_{\nu'L} T_{\mathbf{G}'';\nu'L}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) Z_{\nu'L;\lambda''}^{*\mathbf{k}-\mathbf{q}}. \quad (\text{C28})$$

Int-Int:

$$T_{\mathbf{r};\mathbf{r}'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega) = -G_{\mathbf{r};\mathbf{r}'}^{\mathbf{R}}(\beta - \tau) K_{\mathbf{r};\mathbf{r}'}^{\mathbf{R}}(\mathbf{k}\lambda'; \tau; \omega). \quad (\text{C29})$$

First we represent

$$T_{\mathbf{r};\mathbf{r}'}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \sum_{\mathbf{G}''\mathbf{G}'''} \frac{e^{i(\mathbf{q}+\mathbf{G}'')\mathbf{r}}}{\sqrt{\Omega_0}} T_{\mathbf{G}'';\mathbf{G}'''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) \frac{e^{i(\mathbf{k}-\mathbf{q}+\mathbf{G}''')\mathbf{r}'}}{\sqrt{\Omega_0}}, \quad (\text{C30})$$

with

$$T_{\mathbf{G}'';\mathbf{G}'''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \frac{\Omega_0}{N_{\mathbf{r}}^2} \sum_{\mathbf{r}\mathbf{r}'} e^{-i(\mathbf{q}+\mathbf{G}'')\mathbf{r}} T_{\mathbf{r};\mathbf{r}'}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) e^{-i(\mathbf{k}-\mathbf{q}+\mathbf{G}''')\mathbf{r}'}. \quad (\text{C31})$$

After that

$$T_{\mathbf{G}'';\lambda''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) = \sum_{\mathbf{G}''} \sum_{\eta'\mathbf{G}'''} S_{\mathbf{G}''\mathbf{G}'''}^{\mathbf{q}} T_{\mathbf{G}'';\mathbf{G}'''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau; \omega) X_{\mathbf{G}'';\lambda''}^{\mathbf{k}-\mathbf{q}}. \quad (\text{C32})$$

The evaluation of quantity S is accomplished in reciprocal space:

$$S_{i\lambda''}^{\mathbf{q}}(\mathbf{k}\lambda'; \nu; \omega) = \sum_{j\lambda''} W_{ij}^{\mathbf{q}}(\nu) G_{\lambda\lambda''}^{\mathbf{k}-\mathbf{q}}(\omega - \nu) T_{j;\lambda''}^{\mathbf{q}}(\mathbf{k}\lambda'; \nu, \omega). \quad (\text{C33})$$

Finally, the dynamic part of the correction to self-energy is evaluated as the following:

$$\Sigma_{\lambda\lambda''}^{\mathbf{k}}(\omega) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{q}} \sum_{i\lambda''} \langle \Psi_{\lambda}^{\mathbf{k}} | \Psi_{\lambda''}^{\mathbf{k}-\mathbf{q}} \Pi_i^{\mathbf{q}} \rangle \times \frac{1}{\beta} \sum_{\nu} S_{i\lambda''}^{\mathbf{q}}(\mathbf{k}\lambda'; \nu, \omega). \quad (\text{C34})$$

TABLE III. Quasiparticle energies of Si at high-symmetry points in the Brillouin zone. For fully self-consistent methods they represent the positions of the peaks in corresponding \mathbf{k} -resolved spectral functions.

\mathbf{k} point	Γ	Γ	L	L	L	X	X
scGW	0	3.85	-1.33	2.68	4.66	-3.22	1.62
sc(BSE0:P@GW)	0	3.67	-1.32	2.50	4.47	-3.19	1.46
sc(BSE:P@GW)	0	3.62	-1.29	2.46	4.41	-3.11	1.39
sc(BSE:P@GW+G3W2)	0	3.54	-1.25	2.36	4.36	-3.10	1.31
sc(GW+G3W2)	0	3.63	-1.28	2.43	4.44	-3.12	1.36
QSGW	0	3.58	-1.28	2.44	4.40	-3.10	1.45
QS(BSE0TD:P@GW)	0	3.46	-1.27	2.30	4.23	-3.06	1.31
QS(BSE0:P@GW)	0	3.33	-1.25	2.21	4.12	-3.03	1.22
QS(BSE:P@GW)	0	3.00	-1.20	1.96	3.87	-2.88	1.02

Let us consider the static part now. In this case, the mathematical expression of the whole diagram is simpler:

$$\Sigma_2(12; \tau) = G(13; \tau) V(14) G(34; -\tau) G(42; \tau) V(32). \quad (\text{C35})$$

Again, we introduce the following notations for convenience:

$$K(342; \tau) = G(42; \tau) V(32), \quad (\text{C36})$$

$$T(342; \tau) = G(34; -\tau) K(342; \tau), \quad (\text{C37})$$

and

$$S(112; \tau) = G(13; \tau) V(14) T(342; \tau). \quad (\text{C38})$$

The use of basis sets and reciprocal and real spaces is totally identical to the dynamic case, only the time and/or frequency role is different. Therefore, the evaluation proceeds similar to the dynamic case above but with simplified time or frequency dependence of the quantities involved. The final result for the static part of the self-energy correction is

$$\Sigma_{2,\lambda\lambda''}^{\mathbf{k}}(\tau) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{q}} \sum_{i\lambda''} \langle \Psi_{\lambda}^{\mathbf{k}} | \Psi_{\lambda''}^{\mathbf{k}-\mathbf{q}} \Pi_i^{\mathbf{q}} \rangle S_{i\lambda''}^{\mathbf{q}}(\mathbf{k}\lambda'; \tau). \quad (\text{C39})$$

Let us consider now the semidynamic part of the self-energy correction. First we note that in this case

$$\Sigma_2(12; \tau) = \tilde{\Sigma}(12; \tau) + \tilde{\Sigma}(21; \tau), \quad (\text{C40})$$

where

$$\tilde{\Sigma}(12; \tau) = G(13; \tau) \tilde{W}(14; \tau - \tau_4) \times G(34; -\tau_4) G(42; \tau_4) V(32). \quad (\text{C41})$$

Using quantity T from the static case, we have

$$\tilde{\Sigma}(12; \tau) = G(13; \tau) \tilde{W}(14; \tau - \tau_4) T(342; \tau_4). \quad (\text{C42})$$

Again, the use of basis sets is the same as in the dynamic case, so we do not repeat the corresponding formulas.

APPENDIX D: QUASIPARTICLE APPROXIMATION

Our method is exclusively based on imaginary axis data, which makes it different from earlier proposed methods (see, for example, Ref. [29]).

TABLE IV. Quasiparticle energies of NaCl at high-symmetry points in the Brillouin zone. For fully self-consistent methods they represent the positions of the peaks in corresponding \mathbf{k} -resolved spectral functions.

\mathbf{k} point	Γ	Γ	L	L	L	X	X	X
scGW	0	9.41	-2.35	-0.35	11.57	-2.07	-0.77	12.20
sc(BSE0:P@GW)	0	8.71	-2.32	-0.34	10.90	-2.05	-0.75	11.52
sc(BSE:P@GW)	0	8.59	-2.32	-0.34	10.75	-2.05	-0.75	11.41
sc(BSE:P@GW+G3W2)	0	8.54	-2.36	-0.33	10.72	-2.08	-0.77	11.39
sc(GW+G3W2)	0	8.84	-2.35	-0.33	11.01	-2.08	-0.77	11.67
QSGW	0	9.52	-2.37	-0.35	11.69	-2.09	-0.77	12.33
QS(BSE0TD:P@GW)	0	8.80	-2.35	-0.34	10.98	-2.07	-0.76	11.72
QS(BSE0:P@GW)	0	8.45	-2.35	-0.34	10.76	-2.08	-0.76	11.29
QS(BSE:P@GW)	0	8.31	-2.35	-0.34	10.48	-2.07	-0.75	11.16
QS(BSE:P@GW+G3W2)	0	8.11	-2.45	-0.36	10.32	-2.17	-0.80	11.01

We proceed as follows. In the Dyson equation for the Green's function

$$G_{\lambda\lambda'}^{-1}(\mathbf{k}; \omega) = (i\omega + \mu - \varepsilon_{\lambda}^{\mathbf{k}})\delta_{\lambda\lambda'} - \Sigma_{\lambda\lambda'}^c(\mathbf{k}; \omega), \quad (\text{D1})$$

with the band indices (λ, λ') corresponding to the effective exchange Hamiltonian [31], we approximate the frequency dependence of the self-energy by a linear function

$$\Sigma_{\lambda\lambda'}^c(\mathbf{k}; \omega) = \Sigma_{\lambda\lambda'}^c(\mathbf{k}; \omega = 0) + \frac{\partial \Sigma_{\lambda\lambda'}^c(\mathbf{k}; \omega)}{\partial (i\omega)} \Big|_{\omega=0} (i\omega). \quad (\text{D2})$$

With this approximation, the Dyson equation can be simplified:

$$G_{\lambda\lambda'}^{-1}(\mathbf{k}; \omega) = Z_{\lambda\lambda'}^{-1}(\mathbf{k})(i\omega) + (\mu - \varepsilon_{\lambda}^{\mathbf{k}})\delta_{\lambda\lambda'} - \Sigma_{\lambda\lambda'}^c(\mathbf{k}; 0), \quad (\text{D3})$$

where we have introduced a renormalization factor Z matrix,

$$Z_{\lambda\lambda'}^{-1}(\mathbf{k}) = \delta_{\lambda\lambda'} - \frac{\partial \Sigma_{\lambda\lambda'}^c(\mathbf{k}; \omega)}{\partial (i\omega)} \Big|_{\omega=0}. \quad (\text{D4})$$

Representing the Z factor as a symmetrical product,

$$Z_{\lambda\lambda'}^{-1}(\mathbf{k}) = \sum_{\lambda''} Z_{\lambda\lambda''}^{-1/2}(\mathbf{k}) Z_{\lambda''\lambda'}^{-1/2}(\mathbf{k}), \quad (\text{D5})$$

we reduce the Dyson equation to the following form:

$$\begin{aligned} & \sum_{\lambda''\lambda'''} Z_{\lambda\lambda''}^{1/2}(\mathbf{k}) G_{\lambda''\lambda'''}^{-1}(\mathbf{k}; \omega) Z_{\lambda'''\lambda'}^{1/2}(\mathbf{k}) \\ &= i\omega\delta_{\lambda\lambda'} + \sum_{\lambda''\lambda'''} Z_{\lambda\lambda''}^{1/2}(\mathbf{k}) [(\mu - \varepsilon_{\lambda''}^{\mathbf{k}})\delta_{\lambda''\lambda'''} \\ & \quad - \Sigma_{\lambda''\lambda'''}^c(\mathbf{k}; 0)] Z_{\lambda'''\lambda'}^{1/2}(\mathbf{k}). \end{aligned} \quad (\text{D6})$$

The second term on the right-hand side of this equation is a Hermitian matrix. It can be diagonalized:

$$\begin{aligned} \mu\delta_{\lambda\lambda'} - H_{\lambda\lambda'}^{\mathbf{k}} &= \sum_{\lambda''\lambda'''} Z_{\lambda\lambda''}^{1/2}(\mathbf{k}) [(\mu - \varepsilon_{\lambda''}^{\mathbf{k}})\delta_{\lambda''\lambda'''} \\ & \quad - \Sigma_{\lambda''\lambda'''}^c(\mathbf{k}; 0)] Z_{\lambda'''\lambda'}^{1/2}(\mathbf{k}) \\ &= \sum_i Q_{\lambda i}^{\mathbf{k}} E_i^{\mathbf{k}} Q_{i\lambda'}^{\dagger \mathbf{k}}, \end{aligned} \quad (\text{D7})$$

where $E_i^{\mathbf{k}}$ are the effective eigenvalues. After the diagonalization, we can rewrite Eq. (D6) as

$$\sum_{\lambda''\lambda'''} Z_{\lambda\lambda''}^{1/2}(\mathbf{k}) G_{\lambda''\lambda'''}^{-1}(\mathbf{k}; \omega) Z_{\lambda'''\lambda'}^{1/2}(\mathbf{k}) = \sum_i Q_{\lambda i}^{\mathbf{k}} [i\omega + \mu - E_i^{\mathbf{k}}] Q_{i\lambda'}^{\dagger \mathbf{k}}, \quad (\text{D8})$$

or, for the Green's function,

$$G_{\lambda\lambda'}^{\mathbf{k}}(\omega) = \sum_i \frac{(Z^{1/2} Q)_{\lambda i}^{\mathbf{k}} (Q^{\dagger} Z^{1/2})_{i\lambda'}^{\mathbf{k}}}{i\omega + \mu - E_i^{\mathbf{k}}}. \quad (\text{D9})$$

This expression differs from the full GW Green's function by a linear approximation for the frequency-dependent self-energy.

At this point, we construct the quasiparticle Green's function by setting $Z_{\lambda\lambda'}^{\mathbf{k}} = \delta_{\lambda\lambda'}$ (in order to satisfy the Z -factor cancellation condition [29]) in the above equation to obtain

$$G_{\lambda\lambda'}^{\mathbf{k}}(\omega) = \sum_i \frac{Q_{\lambda i}^{\mathbf{k}} Q_{i\lambda'}^{\dagger \mathbf{k}}}{i\omega + \mu - E_i^{\mathbf{k}}}. \quad (\text{D10})$$

APPENDIX E: EFFECTIVE QUASIPARTICLE ENERGIES AT \mathbf{k} POINTS OF HIGH SYMMETRY

In this Appendix we present (for the purpose of reference) the effective quasiparticle one-electron energies for two materials: Si (Table III) and NaCl (Table IV). Silicon represents an example of a small-gap semiconductor whereas NaCl represents a large-gap insulator.

[1] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso *et al.*, *Science* **351**, 1415 (2016).

[2] L. Hedin, *Phys. Rev.* **139**, A796 (1965).

[3] F. Aryasetiawan and O. Gunnarsson, *Phys. Rev. B* **49**, 16214 (1994).

- [4] T. Kotani and M. van Schilfhaarde, *Solid State Commun.* **121**, 461 (2002).
- [5] M. J. van Setten, F. Caruso, S. Sharifzadeh, X. Ren, M. Scheffler, F. Liu, J. Lischner, L. Lin, J. R. Deslippe, S. G. Louie, C. Yang, F. Weigend, J. B. Neaton, F. Evers, and P. Rinke, *J. Chem. Theor. Comput.* **11**, 5665 (2015).
- [6] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, *New J. Phys.* **14**, 053020 (2012).
- [7] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, *Comput. Phys. Commun.* **180**, 2175 (2009).
- [8] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, *Comput. Phys. Commun.* **183**, 1269 (2012).
- [9] M. J. van Setten, F. Weigend, and F. Evers, *J. Chem. Theor. Comput.* **9**, 232 (2013).
- [10] TURBOMOLE V6.6, 2014; a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007; TURBOMOLE GmbH, since 2007, <http://www.turbomole.com>.
- [11] A. Grüneis, G. Kresse, Y. Hinuma, and F. Oba, *Phys. Rev. Lett.* **112**, 096401 (2014).
- [12] B. Cunningham, M. Grüning, D. Pashov, and M. van Schilfhaarde, [arXiv:2106.05759](https://arxiv.org/abs/2106.05759).
- [13] S. K. Radha, W. R. L. Lambrecht, B. Cunningham, M. Grüning, D. Pashov, and M. van Schilfhaarde, *Phys. Rev. B* **104**, 115120 (2021).
- [14] S. Acharya, D. Pashov, B. Cunningham, A. N. Rudenko, M. Rösner, M. Grüning, M. van Schilfhaarde, and M. I. Katsnelson, *Phys. Rev. B* **104**, 155109 (2021).
- [15] M. C. T. D. Müller, S. Blügel, and C. Friedrich, *Phys. Rev. B* **100**, 045130 (2019).
- [16] A. L. Kutepov, *Phys. Rev. Materials* **5**, 083805 (2021).
- [17] A. L. Kutepov, *Phys. Rev. B* **94**, 155101 (2016).
- [18] A. L. Kutepov, *Phys. Rev. B* **95**, 195120 (2017).
- [19] A. L. Kutepov and G. Kotliar, *Phys. Rev. B* **96**, 035108 (2017).
- [20] A. L. Kutepov, *Phys. Rev. B* **104**, 085109 (2021).
- [21] A. L. Kutepov, *J. Phys.: Condens. Matter* **33**, 485601 (2021).
- [22] D. Nabok, S. Blügel, and C. Friedrich, *npj Comput. Mater.* **7**, 178 (2021).
- [23] B. Monserrat and R. J. Needs, *Phys. Rev. B* **89**, 214304 (2014).
- [24] G. Antonius, S. Ponce, P. Boulanger, M. Cote, and X. Gonze, *Phys. Rev. Lett.* **112**, 215501 (2014).
- [25] M. Cardona, M. L. W. Thewalt, *Rev. Mod. Phys.* **77**, 1173 (2005).
- [26] B. Monserrat, *Phys. Rev. B* **93**, 100301(R) (2016).
- [27] M. Grüning, A. Marini, and X. Gonze, *Nano Lett.* **9**, 2820 (2009).
- [28] M. Grumet, P. Liu, M. Kaltak, J. Klimes, and G. Kresse, *Phys. Rev. B* **98**, 155143 (2018).
- [29] T. Kotani, M. van Schilfhaarde, and S. V. Faleev, *Phys. Rev. B* **76**, 165106 (2007).
- [30] A. L. Kutepov, *J. Phys.: Condens. Matter* **27**, 315603 (2015).
- [31] A. Kutepov, K. Haule, S. Y. Savrasov, and G. Kotliar, *Phys. Rev. B* **85**, 155129 (2012).
- [32] A. L. Kutepov, V. S. Oudovenko, and G. Kotliar, *Comput. Phys. Commun.* **219**, 407 (2017).
- [33] G. Baym and L. P. Kadanoff, *Phys. Rev.* **124**, 287 (1961).
- [34] C.-O. Almbladh, U. von Barth, and R. van Leeuwen, *Int. J. Mod. Phys. B* **13**, 535 (1999).
- [35] The latest publicly available version of the FLAPWMBPT code (FlapwMBPT2106) can be downloaded from <https://github.com/andreykutepov65/FlapwMBPT>.
- [36] C. Friedrich, A. Schindlmayr, S. Blügel, and T. Kotani, *Phys. Rev. B* **74**, 045104 (2006).
- [37] C. Friedrich, M. C. Müller, and S. Blügel, *Phys. Rev. B* **83**, 081101(R) (2011).
- [38] C. Friedrich, M. C. Müller, and S. Blügel, *Phys. Rev. B* **84**, 039906(E) (2011).
- [39] H. Jiang and P. Blaha, *Phys. Rev. B* **93**, 115203 (2016).
- [40] D. Nabok, A. Gulans, and C. Draxl, *Phys. Rev. B* **94**, 035118 (2016).
- [41] C. Vorwerk, B. Aurich, C. Cocchi, and C. Draxl, *Electron. Struct.* **1**, 037001 (2019).
- [42] M. Cardona, *Solid State Commun.* **133**, 3 (2005).
- [43] B. Monserrat, G. J. Conduit, and R. J. Needs, *Phys. Rev. B* **90**, 184302 (2014).
- [44] F. J. Himpsel, J. A. Knapp, J. A. VanVechten, and D. E. Eastman, *Phys. Rev. B* **20**, 624 (1979).
- [45] P. Lautenschlager, M. Garriga, S. Logothetidis, and M. Cardona, *Phys. Rev. B* **35**, 9174 (1987).
- [46] A. Mang, K. Reimann, S. Rübenacke, and M. Steube, *Phys. Rev. B* **53**, 16283 (1996).
- [47] T. Sander, E. Maggio, and G. Kresse, *Phys. Rev. B* **92**, 045209 (2015).
- [48] S. Choi, A. Kutepov, K. Haule, M. van Schilfhaarde, and G. Kotliar, *npj Quantum Mater.* **1**, 16001 (2016).
- [49] S. Choi, P. Semon, B. Kang, A. Kutepov, and G. Kotliar, *Comput. Phys. Commun.* **244**, 277 (2019).
- [50] S. Biermann, F. Aryasetiawan, and A. Georges, *Phys. Rev. Lett.* **90**, 086402 (2003).
- [51] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A. I. Lichtenstein, *Phys. Rev. B* **70**, 195104 (2004).
- [52] A. L. Kutepov and S. G. Kutepova, *J. Phys.: Condens. Matter* **15**, 2607 (2003).
- [53] A. L. Kutepov, *Phys. Rev. B* **103**, 165101 (2021).
- [54] S. Albrecht, L. Reining, R. Del Sole, and G. Onida, *Phys. Rev. Lett.* **80**, 4510 (1998).
- [55] F. Fuchs, C. Rödl, A. Schleife, and F. Bechstedt, *Phys. Rev. B* **78**, 085103 (2008).