Effect of propagator renormalization on the band gap of solids

Sergei Iskakov,1 Alexander A. Rusakov,2 Dominika Zgid,2,1,3 and Emanuel Gull1,3

1Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA
2Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA
3Center for Computational Quantum Physics, Flatiron Institute, New York, NY 10010, USA
(Dated: December 19, 2018)

We present momentum-resolved spectral functions and band gaps from bare and self-consistent second order perturbation theory for insulating periodic solids. We establish that, for systems with large gap sizes, both bare and self-consistent perturbation theory yield reasonable gaps. However, smaller gap sizes require a self-consistent adjustment of the propagator. In contrast to results obtained within a quasi-particle formalism used on top of bare second order perturbation theory, no unphysical behaviour of the band gap is observed. We establish that the renormalization or ‘screening’ of interaction terms in addition to propagators is not required to obtain accurate gaps in insulators. Our implementation of a fully self-consistent, \( \Phi \)-derivable and thermodynamically consistent finite temperature diagrammatic perturbation theory forms a framework on which embedding theories such as the dynamical mean field theory or self-energy embedding theories can be implemented.

A truly ab-initio quantitative many-body description of weakly correlated systems beyond density functional theory (DFT) [1, 2] remains challenging despite enormous theoretical [3] and computational [4–9] advances achieved during last thirty years. While such calculations are important for explaining the behavior of semiconductors, they are also necessary as a first step for many strongly correlated computational techniques such as some derivatives [10–14] of the dynamical mean field theory (DMFT) [15–17], where the first calculation step consists of evaluating the system by a perturbative weak coupling method.

Three types of perturbation theories exist: bare (non-self-consistent) perturbation theory based on the expansion in interaction terms of the original Hamiltonian, where neither propagators nor interactions are renormalized; self-consistent perturbation theories where propagators but not the interactions are self-consistently renormalized; and perturbation theories with both propagators and interactions renormalized.

Second order bare non-self-consistent perturbation theory is known as Møller - Plesset second order (MP2) [18–27] when applied to real materials. The self-consistent second order Green’s function perturbation theory (GF2) [28–33] renormalizes propagators but not interactions. While fully self-consistent GW [3] renormalizes both propagators and interactions, its non-self-consistent variants such as \( G_0 W_0 \) [4, 34] renormalize interactions without renormalizing propagators.

The differences in the treatment of propagators and interactions between all these types of perturbation theories are crucial since approaches that do not renormalize the interactions are expected to fail in 3D metallic systems [27, 35]. This breakdown is not expected to occur in insulators.

For weakly correlated materials, most results so far have been obtained within zero-temperature approaches such as MP2 [23, 24, 27] and GW [36, 37]. Only recently, finite-temperature results for solids have started to appear [38–41]. It was demonstrated that MP2 gaps are wildly inaccurate for semiconductors with band gaps smaller than 6 eV [27], leading to a breakdown of the band gap estimation for silicon and silicon carbide. In contrast, zero-temperature, non-self-consistent GW [3] has been very successful in predicting band gaps for semiconductors. This success is usually attributed to the renormalization of the interactions by an infinite series of ‘bubble’ (RPA [3]) diagrams, which renders this method convergent in the metallic limit.

Since there is no obvious reason for the breakdown of the perturbative series in semiconductors and band insulators, it is interesting to explore how the renormalization with self-consistent propagators but unrenormalized interactions affects the band gaps and compare these results to the ones obtained by bare MP2. Moreover, since the MP2 band gap is evaluated using approximate band energies for the lowest unoccupied and highest occupied bands [27] reminiscent of the formulas usually employed to solve the quasi-particle (QP) equations in GW [42], it interesting to compare these MP2 band gaps to the ones evaluated without solving the QP equations.

The renormalization of interactions and propagators is commonly discussed in many-body textbooks [35, 43], usually at the example of weakly interacting or uniform systems. Due to both the computational cost and implementation difficulties, their effect in realistic solids is difficult to explore.

In this paper, we focus on quantifying the effects of the renormalization of propagators and self-energies on the value of band gaps in simple 3D solids. This is possible due to the implementation of a fully self-consistent finite temperature second order perturbation theory (GF2) that was so far not available for realistic 3D systems.

We investigate the physics of 3D solids by choosing a
finite basis set on each atom, and choosing a finite lattice of atoms periodic in all three spatial directions. This yields a periodic electronic structure Hamiltonian best expressed in a Bloch basis in reciprocal space, resulting in the Coulomb integrals

\[
V_{ijkl}^{\text{k}_1\text{k}_2\text{k}_3\text{k}_4} = \int \frac{\delta^{(3)}(\text{k}_1,\tau)\delta^{(3)}(\text{k}_2,\tau')\delta^{(3)}(\text{k}_3,\tau')\delta^{(3)}(\text{k}_4,\tau)}{|\tau - \tau'|} d\tau d\tau',
\]

where \(\phi\) are basis functions in reciprocal space. translational invariance is guaranteed by the momentum conservation \(\text{k}_1 + \text{k}_2 = \text{k}_3 + \text{k}_4\). These integrals can be decomposed [44–47] into a product of two low-rank tensors, \(V_{ijkl}^{\text{k}_1\text{k}_2\text{k}_3\text{k}_4} = \sum_{Q} \tilde{V}_{\text{k}_1,\text{l}\text{k}_4,\text{j}\text{k}_2,\text{k}_3} Q\tilde{V}_{\text{k}_2,\text{k}_4,\text{k}_3,\text{k}_1} Q\), where \(Q\) is an auxiliary basis index.

The second-order self-energy is then evaluated in reciprocal space in an imaginary time formalism,

\[
\Sigma^{(2)}_{ij}(\tau) = -(2\tilde{V}_{\text{q}_1,\text{q}_2,\text{k}_1\text{k}_2} Q\tilde{V}_{\text{q}_1,\text{q}_2,\text{k}_3\text{k}_4} - \tilde{V}_{\text{q}_1,\text{q}_2,\text{k}_1\text{k}_2} Q\tilde{V}_{\text{q}_1,\text{q}_2,\text{k}_3\text{k}_4}) \times \tilde{V}_{\text{q}_1,\text{q}_2,\text{k}_1\text{k}_2} Q\tilde{V}_{\text{q}_1,\text{q}_2,\text{k}_3\text{k}_4} Q G_{p\text{q}}^{\text{k}_1}(\tau) G_{\text{k}\text{l}}^{\text{k}_2}(\tau) G_{\text{n}\text{m}}^{\text{k}_3}(\tau) \delta_{\text{k}+\text{k}_3,\text{k}_1+\text{k}_2},
\]

with \(\tau\) denoting imaginary time \(0 < \tau < \beta\) and \(\beta = 1/T\) the inverse of the physical temperature \(T\). We assume Einstein summation over internal indices. In bare second-order perturbation theory, the Hartree-Fock Green’s function is employed in Eq. 2 and the Dyson equation is evaluated only once to yield the interacting Green’s function.

In renormalized perturbation theories such as GF2, to achieve self-consistency in both the density matrix (or, correspondingly, the Fock and the matrix frequency independent term of the self-energy \(\Sigma_{\infty}\)) and the dynamical self-energy \(\Sigma(\tau)\), we employ a modification of the iterative procedure described in Ref. 28 for molecular systems. First, we adjust the chemical potential to find the correct particle number of the Hartree-Fock (HF) solution. We then obtain the HF propagator, calculate the second-order self-energy, and recompute the interacting Green’s function and density matrix using the Dyson equation, adjusting the chemical potential until the correct density is found and subsequent \(\Sigma_{\infty}\) is evaluated. This propagator is then used for the next second-order self-energy evaluation, until convergence is achieved in all quantities.

Energies, entropies, free energies, and specific heats are then computed using standard thermodynamic formulae [31, 35, 48, 49], see also supplemental materials [50].

Self-consistent finite temperature perturbation theory renormalizes the propagators. In addition, quantities obtained in diagrammatic approximations by thermodynamic integration are in general dependent on the integration path [51, 52]; i.e. the integration of a quantity such as the energy or the entropy may differ if it is obtained by integration from zero \(T\), infinite \(T\), infinite chemical potential or via coupling constant integration. Only so-called ‘\(\Phi\)-derivable’ [53], self-consistent methods, such as the self-consistent GF2 method investigated here or the fully self-consistent GW approximation, avoid this problem and are intrinsically thermodynamically consistent.

Standard finite temperature perturbation theories are formulated on the imaginary axis, thus, static quantities such as the density matrix, the energy, the entropy, the static magnetic susceptibility or the specific heat are directly accessible. In contrast, real-frequency-dependent quantities such as the spectral function, the gap, the optical conductivity, or the dynamical magnetic susceptibility require analytical continuation to the real axis in order to be compared to experiment. This analytical continuation is ill conditioned and leads to an amplification of uncertainties, even for data known up to numerical precision, in particular at high temperature and high frequency. The problem is intrinsic to the finite-temperature field-theory formalism on the imaginary axis and can only be overcome by reformulating the method in frequency space or real time. In the present work, we used the maximum entropy method [54, 55]. Other methods, such as the spectral method [56], Padé [57], stochastic analytical continuation [58–60], or Consistent Constraints [61] could be explored, as could the continuation of the self-energy to obtain spectra and gaps via the quasi-particle equation [62].

![Figure 1. Thermodynamics of solid LiH. Clockwise from the top: Internal energy, Helmholtz energy, Specific heat, and entropy as a function of temperature. Data evaluated on a periodic 4x4x4 lattice in the poh-TZVP [63] basis.](image-url)
Figure 2. Momentum-resolved spectral function for 3d neon obtained at $\beta = 100 \text{ Ha}^{-1}$. Left panel: comparison between self-consistent GF2 (black lines) and bare second-order perturbation theory (orange lines) on a periodic $4 \times 4 \times 4$ lattice. Thick lines: directly evaluated $k$-points. Thin lines: interpolation. Results are listed in the 6-311+G* basis set. Middle panel: Comparison of the momentum resolved GF2 spectral function along the $\Delta$ direction on a periodic $4 \times 4 \times 4$ (solid black) and $3 \times 3 \times 3$ (dashed green) lattice. Right panel: GF2 on a periodic $4 \times 4 \times 4$ lattice, in the basis set 6-311+G* (black) and aug-cc-pVDZ (purple).

 dermination of thermodynamic properties for solid LiH in Fig. 1. Smooth curves have been obtained by Chebyshev interpolation on a temperature grid.

As these insulators have such vastly different band gaps, a different amount of Green’s function renormalization is expected to be necessary, allowing us to examine how the iterative nature of GF2 changes the results in comparison to bare perturbation theory and Hartree-Fock. Moreover, since evaluating band gaps using either the QP equation as used in Ref. 27 or analytical continuation [54] may give different results, we compare our values for band gaps (obtained using analytical continuation) to the ones available in the literature, where QP equations were used with the bare perturbation theory.

Our GF2 implementation for periodic systems uses a compact Chebyshev polynomial [74] representation of Green’s functions that converges exponentially. This representation is superior to the one reported in Ref. [75]. In the current implementation we do not use a spline representation in frequency space [76]. Our implementation relies on the open source ALPS library [77]. Additionally, we use periodic density-fitted integrals in Gaussian orbitals, evaluated using the open source pySCF package [78, 79]. Our calculations result in a set of imaginary time Green’s function orbitals [82, 83]. Thick lines denote points in the Brillouin zone that coincide with our momentum grid. Thin lines denote interpolated values. All spectral functions are plotted as a function of frequency in eV.

In this system, the band gap and the spectral functions are converged with respect to the momentum discretization. This is illustrated in the middle panel of Fig. 2, where we show results on a $3 \times 3 \times 3$ and on a $4 \times 4 \times 4$ lattice along the $\Delta$ direction in momentum space. Thick lines denote values on the respective momentum grids (black for $4 \times 4 \times 4$, green for $3 \times 3 \times 3$), thin lines are obtained by interpolation. The X point is absent on the smaller grid. Data has been obtained in the 6-311+G* basis set.

Our calculations also demonstrate that, in the frequency window shown, the spectral function is relatively insensitive to the choice of the basis set, see right panel of Fig. 2. However, it should be stressed that while our results do not show significant differences between aug-cc-pVDZ [84] and 6-311+G* basis sets, they may not be converged with respect to the basis set size, since both bases are small. Converging our calculations with respect to the basis set size would require a systematic increase of the cardinal number $X$ in the series of aug-cc-pVXZ basis sets. This is exceedingly difficult in ordinary solid state calculations since regular Gaussian basis sets such as aug-cc-pVXZ become linearly dependent for higher cardinal numbers.

Hartree-Fock gaps can be extracted directly from the eigenvalues. The determination of the gap in correlated methods via the spectral function leaves some arbitrariness, as finite temperature Green’s functions are broadened by self-energy, temperature, and analytical continuation artifacts. For GF2, we chose to define the band gap as the peak-to-peak distance of the k-space peaks closest to the Fermi energy from above and below. For solid neon, GF2 in both the aug-cc-pVDZ and the 6-311+G* basis yields a band gap of 20.55 eV at the $\Gamma$ point. We
attribute the difference between between the GF2 gap (20.55 eV) and the experimental band gap (21.7 eV) to the basis set effect and/or analytical continuation effects.

Similarly to solid neon, solid LiF has a wide experimental band gap of 14.2 eV [66]. For the 4×4×4 k-point grid, we observe a band gap of 13.03 eV in self-consistent GF2, as extracted from the peak-to-peak distance of the spectral function. The difference of the spectral function between the first iteration and self-consistent GF2 is negligible, see left panel of Fig. 3. For the information about k-grid convergence of the unit cell energy and band gap see Tab. I and Fig. 1 in the supplement. [50]

Both LiH and diamond have moderate band gaps. Therefore, a significant change of the band gap between the first iteration of GF2 and the fully self-consistent result is expected. Indeed, both the middle and the right panel of Fig. 3 confirm that during the first iteration of GF2 the starting band gaps are much wider and they are getting contracted once the self-consistent iterations progress. Here, to evaluate spectral functions during the first iteration of GF2, we use the analytical continuation of \( G(i\omega) = \left[ (i\omega + \mu)1 - F_{HF} - \Sigma^{(2)} \right]^{-1} \), where \( \Sigma^{(2)} \) was obtained in the first iteration of the GF2 method, i.e., using \( G_{HF}(i\omega) = \left[ (i\omega + \mu)1 - F_{HF} \right]^{-1} \) as the propagator in Eq. 2. The Fock matrix \( F_{HF} \) comes from a preceding HF calculation. Consequently, the first iteration of GF2 lacks two types of renormalization: first, the renormalization coming from the self-consistently updated \( \Sigma_{\infty} \) and then consequently updated Fock matrix and, second, the one from the fully self-consistent evaluation of \( \Sigma^{(2)} \) which in subsequent iterations is evaluated with renormalized propagators.

For diamond in the pob-TZVP basis, we observe an indirect band gap of 4.8 eV between \( \Gamma \) and the point halfway between the \( \Gamma \) and \( X \) points. The direct band gap at the \( \Gamma \)-point is about 6.6 eV. The positions and values of the direct and indirect band gaps are in good agreement with previous experimental and theoretical results. [70–73] It is worth emphasizing that the band gap for diamond obtained from self-consistent GF2 is 4.8 eV, while the MP2 band gap obtained using a QP formalism listed in Ref. 27 is 1.9 eV. This underestimation of band gaps smaller than 6eV is very noticeable (as listed in Ref. 27) for bare second order perturbation theory coupled with band gap evaluation based on the QP formalism. This deficiency seems to be avoided when the fully self-consistent GF2 is employed, as we observe both in the case of LiH and of diamond.

For LiH, we plot the convergence of the band gap and unit cell energy for different k-grids in Fig. 4 and Tab. I. The data clearly show that while the unit cell energy converges rather quickly, the convergence of the band gap is substantially slower.

![Figure 3](image3.png)

**Figure 3.** Momentum-resolved spectral function of solid LiF (left panel), LiH (middle panel), and diamond (right panel) obtained from GF2 (solid black lines) and bare second-order perturbation theory (dashed orange lines) on a periodic 4×4×4 lattice. Results are listed in the pob-TZVP basis. Inset: zoom to the upper gap edge as indicated in the main panel.

![Figure 4](image4.png)

**Figure 4.** Gap size in solid LiH for different k-grids. Results are listed in pob-TZVP basis.

| Size     | \( E(\text{HF}) \) | \( e_{g}(\text{HF}) \) | \( E(\text{GF2}) \) | \( e_{g}(\text{GF2}) \) |
|----------|--------------------|------------------------|--------------------|------------------------|
| 3×3×3    | -8.0659            | 11.82                  | -8.1097            | 10.08                  |
| 4×4×4    | -8.0629            | 11.44                  | -8.1079            | 7.73                   |
| 5×5×5    | -8.0618            | 11.31                  | -8.1076            | 6.66                   |
| 6×6×6    | -8.0612            | 11.24                  | -8.1075            | 5.93                   |

**Table 1.** HF and GF2 total energies and band gaps of LiH for different system sizes in pob-TZVP basis.

In conclusion, we computed the momentum-resolved spectral functions and band gaps for Ne, LiH, LiF, and diamond. In all of these cases, we found reasonable agreement with experimental values. Wide-gap insulators were found to be fully converged w.r.t the k-point
grid, whereas larger momentum grids are needed to converge the band gap (but not the total energy) of LiH. We show that thermodynamic consistency is obeyed in our calculations, opening the door for systematic thermodynamic calculations of the electronic system of real materials.

Our study reveals three major insights. First, the fact that reasonable gap sizes are obtained within GF2 implies that a renormalization of interactions, as done e.g. in GW, is not necessary to obtain gap values in the insulators analyzed. Second, the fact that our results from bare second order perturbation theory yield reasonable gap values that differ substantially from the published MP2 values hints at a breakdown of the QP formalism for gap extraction. Finally, our results show that controlled self-consistent diagrammatic many-body calculations in standardized Gaussian basis sets are now routinely possible.

S.I. and E.G. acknowledge support by the Simons Foundation via the Simons Collaboration on the Many-Electron Problem. D.Z. and A.R. are supported by NSF-CHE-1453894. E.G and D.Z thank the Simons Foundation for sabbatical support. The Flatiron Institute is a division of the Simons Foundation. S.I. thanks Miguel A. Morales for fruitful discussion. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

[1] W. Kohn, Rev. Mod. Phys. 71, 1253 (1999).
[2] W. Kohn and L. J. Sham, Phys. Rev. 140, A1133 (1965).
[3] L. Hedin, Phys. Rev. 139, A796 (1965).
[4] M. S. Hybertsen and S. G. Louie, Phys. Rev. B 34, 5390 (1986).
[5] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. 74, 601 (2002).
[6] M. Govoni and G. Galli, Journal of Chemical Theory and Computation 11, 2680 (2015).
[7] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, Computer Physics Communications 183, 1269 (2012).
[8] M. Shishkin and G. Kresse, Phys. Rev. B 74, 035101 (2006).
[9] P. Bruneval and M. A. L. Marques, Journal of Chemical Theory and Computation 9, 324 (2013).
[10] S. Biermann, F. Aryasetiawan, and A. Georges, in Physics of Spin in Solids: Materials, Methods and Applications, edited by S. Halilov (Springer Netherlands, Dordrecht, 2005) pp. 43–65.
[11] F. Nilsson, L. Boehnke, P. Werner, and F. Aryasetiawan, Phys. Rev. Materials 1, 043003 (2017).
[12] A. A. Kananenka, E. Gull, and D. Zgid, Phys. Rev. B 91, 121111 (2015).
[13] D. Zgid and E. Gull, New Journal of Physics 19, 023047 (2017).
