Self-Consistent Calculations of Quasiparticle States in Metals and Semiconductors

Wolf-Dieter Schöne and Adolfo G. Eguiluz

Department of Physics and Astronomy, The University of Tennessee, Knoxville, Tennessee 37996-1200
and Solid State Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831-6030

(Received 18 May 1998)

We solve the Dyson equation for the one-particle Green's function in a periodic solid self-consistently within the shielded-interaction (or *GW*) approximation. The effect of self-consistency (propagator renormalization) is found to be substantial. We illustrate this finding via calculations of quasiparticle states in K and Si. In the case of Si we show that the current standard of *ab initio* quasiparticle theory, the *GW* self-energy diagram, fails to account for the observed value of the absolute band gap—if properly, i.e., self-consistently, evaluated. [S0031-9007(98)06874-4]

PACS numbers: 71.20.Mq, 71.10.–w, 71.15.Mb

An *ab initio* investigation of the spectrum of electronic excitations in "real-life" materials requires tackling a formidable many-body problem for which only partial benchmarks are available. In the case of semiconductors, as is well known, the energy difference between the lowest unoccupied and the highest occupied Kohn-Sham eigenvalues [1] [typically obtained within the local-density approximation (LDA) [1]] disagrees with the measured value of the absolute band gap by 50%–100%. Thus, a large volume of work has been devoted in recent years to the study of the gap within the framework of quasiparticle (QP) theory [2–7]. The so-called *GW* approximation [8,9] yields results in apparent *quantitative* agreement with experiment (error < 0.1 eV), for the gap and for the valence and conduction bands of semiconductors [2,4–9]. Thus, currently the *GW* one-electron band structure is the standard [5] against which the quality of other approximate schemes is routinely judged.

However, all of the published calculations for realistic models of semiconductors share one limitation [10]: the neglect of full self-consistency—the requirement that the propagators must be dressed self-consistently with the self-energy [11]. Partial tests of self-consistency [2,4,5] have commonly been taken as indicators of its presumed quantitative irrelevance. Otherwise, the complications of a full self-consistent treatment have motivated the adoption of shortcuts, such as the introduction of chemical potential shifts, designed to circumvent the problem that lack of self-consistency leads to violation of conservation laws [12].

In the case of metals, the impact of the *GW* self-energy insertion on the width of the occupied band of Na and K [13], and on the exchange splitting of the magnetic bands of Ni [14], has been calculated—again, without an attempt at self-consistency. Overall, the agreement with experiment is not nearly as good as in the case of semiconductors.

In this Letter we elucidate the importance of self-consistency in the QP band structure of periodic solids. To this end we implement an efficient method for the self-consistent solution of the Dyson equation. The self-energy is evaluated within the shielded-interaction approximation (SIA) of Baym and Kadanoff [11]; this approximation, which is diagrammatically equivalent to Hedin's popular *GW* approximation [8], has the virtue that it fulfills all microscopic conservation laws [11]. We show that the true *GW* band gap of Si disagrees with experiment—in fact, the *GW* overestimates the absolute gap by as much as the LDA underestimates it. Similarly, the *GW* bandwidth of K is too wide (by ~ 1 eV).

The starting point of a QP calculation is the Dyson equation for the one-particle Green's function $G(1, 1')$, which we write down as [11]

$$G^{-1}(1, 1') = G^{-1}_{\mathrm{LDA}}(1, 1') - \tilde{\Sigma}(1, 1'), \qquad (1)$$

where $G_{\mathrm{LDA}}(1, 1')$ is the Green's function for "free" propagation in the LDA band structure, and the labels $1, 1'$ denote space-time points; the time variables are Matsubara times $0 \leq \tau, \tau' \leq \beta\hbar$ [15]. All correlations beyond the LDA are contained in the self-energy $\tilde{\Sigma}(1, 1')$, which is a functional of G—thus the self-consistent nature of the problem. We emphasize that $\tilde{\Sigma}(1, 1')$ differs from the "free space" self-energy $\Sigma(1, 1')$ introduced in the Baym-Kadanoff treatment; technically, we account for the "tadpole" diagrams, and avoid double-counting the interactions already built into $G_{\mathrm{LDA}}(1, 1')$, via the definition $\tilde{\Sigma}(1, 1') = \Sigma(1, 1') - [V_{\mathrm{H}}(1) + V_{\mathrm{xc}}(1)]\delta(1 - 1')$, where $V_{\mathrm{H}}(1)$ and $V_{\mathrm{xc}}(1)$ are, respectively, the Hartree and exchange-correlation (XC) potentials entering the Kohn-Sham equation in LDA [1]. In more physical language, we note that a solid must, first of all, bind—and this the LDA does well; this essential feature is already contained in $G_{\mathrm{LDA}}(1, 1')$. Thus, the expectation is that this propagator provides a suitable starting point (a "good guess") for a self-consistent solution of Eq. (1). In the SIA [11], whose physical content is the dynamical screening of the Coulomb interaction, we have that

$$\Sigma(1, 1') = \delta(1 - 1') \int d\bar{3}\, \nu(1 - \bar{3}) G(\bar{3}, \bar{3}^{+}) + \Sigma_{\mathrm{xc}}(1, 1'), \qquad (2)$$

 0031-9007/98/81(8)/1662(4)$15.00

where the XC self-energy is given by

$$\Sigma_{\rm xc}(1,1') = -V_S(1,1')G(1,1'), \tag{3}$$

in terms of the shielded (or screened) interaction $V_S(1,1')$, the solution of the integral equation

$$V_S(1,1') = \nu(1-1') + \int d\bar{1} \int d\bar{2}\, \nu(1-\bar{1}) \times P(\bar{1},\bar{2})V_S(\bar{2},1'), \tag{4}$$

where ν is the bare Coulomb interaction, and $P(1,1') = 2G(1,1')G(1',1)$ is the polarizability.

We note that Hedin's GW approximation [8,9] is formally the same as the SIA. In Hedin's approach our $V_S(1,1')$ is denoted $W(1,1')$ (i.e., $\Sigma_{\rm xc} = -G^*W$); we adhere to the notation of Ref. [11] in order to emphasize that our work is centered on the *exact* fulfillment of conservation laws, which is the cornerstone of the method of Baym and Kadanoff, who showed that the solution of the above system of equations is conserving—*if obtained self-consistently*.

We express the frequency-Fourier transform of all two-point functions according to

$$G_{j,j'}(\vec{q};i\omega_n) = \int d^2x \int d^3x'\, \phi^*_{\vec{q}j}(\vec{x})G(\vec{x},\vec{x}';i\omega_n) \times \phi_{\vec{q}j'}(\vec{x}'), \tag{5}$$

where $\omega_n = (2n+1)\pi/\beta\hbar$, n being an integer (positive, negative, or zero) and $\beta = 1/k_{\rm B}T$; for bosonlike quantities, such as the polarizability, $\omega_n = 2n\pi/\beta\hbar$. In Eq. (5) the $\{\phi_{\vec{q}j}(\vec{x})\}$ are Kohn-Sham (Bloch) states, $\vec{q}$ is a wave vector in the first Brillouin zone (BZ), and j is a band index. In this representation, Eq. (1) becomes

$$[G^{-1}(\vec{q};i\omega_n)]_{j,j'} = [G^{-1}_{\rm LDA}(\vec{q};i\omega_n)]_{j,j'} - \tilde{\Sigma}_{j,j'}(\vec{q};i\omega_n), \tag{6}$$

which is the form of the Dyson equation we actually solve. Now, the XC self-energy $\Sigma^{\rm xc}_{j,j'}(\vec{q};i\omega_n)$ involves convolutions over wave vectors and Matsubara frequencies (in addition to a sum over four band indices, weighted by integrals involving three Bloch states [16]). The convolutions are evaluated straightforwardly; in particular, since the sampling of the BZ is done for imaginary ω's, the $\vec{q}$ convolution requires relatively coarse meshes. Although the polarizability $P_{i,j}(\vec{q};i\omega_n)$ can also be expressed as a convolution over q's and ω's, the latter convolution converges poorly. (Note that $P \sim G^*G$, and $G \sim \omega_n^{-1}$ for large ω_n; by contrast, $\Sigma_{\rm xc} \sim V_S^*G$ and $V_S \sim \omega_n^{-2}$.) Thus, we obtain $P_{i,j}(\vec{q};i\omega_n)$ via a prior evaluation on the τ axis,

$$P(\vec{q},\tau) = \frac{2}{\Omega}\sum_{\vec{k}} G(\vec{k},-\tau)G(\vec{k}+\vec{q},\tau), \tag{7}$$

where Ω is the volume of the macrocrystal, and for brevity we omit reference to band indices, as we also do in Eq. (8); full expressions are given in [16]. We generate the Green's function for τ times for a given iteration of the solution of Eq. (6) for $G_{j,j'}(\vec{q};i\omega_n)$ according to

$$G(\vec{q},\tau) = \frac{1}{\beta\hbar}\sum_{i\omega_n} e^{-i\omega_n\tau}G(\vec{q},i\omega_n) \approx \frac{1}{\beta\hbar}\sum_{i|\omega_n|\le i\omega_{\rm max}} e^{-i\omega_n\tau}\{G(\vec{q},i\omega_n) - G_x(\vec{q},i\omega_n)\} - e^{-[\omega_q+\Sigma_x(\vec{q})]\tau}\left(1 - \frac{1}{e^{\beta\hbar[\omega_{\vec{q}}+\Sigma_x(\vec{q})]}+1}\right), \tag{8}$$

where we have added and subtracted an auxiliary Green's function, G_x, defined according to Eq. (1), but with the Fock self-energy Σ_x replacing Σ. Since Σ_x is ω independent, the sum involving G_x is performed in closed form, the result being the last term in Eq. (8). The remaining sum in Eq. (8) requires the use of a cutoff $\omega_{\rm max}$; however, since $G \to G_x$ for large ω_n (correlation becomes inoperative in this limit), this sum converges rapidly [16,17]. From the knowledge of $G_{j,j'}(\vec{k};\tau)$ we obtain $P_{j,j'}(\vec{q};\tau)$ according to Eq. (7), and the required coefficients $P_{j,j'}(\vec{q};i\omega_n)$ are obtained via fast-Fourier-transform techniques.

We start out with a well-converged LDA ground state [18]. The solution of Eq. (6) was iterated until for two consecutive iterations the largest change in any element of the diagonal of the $\tilde{\Sigma}$ matrix, for all $\vec{q}$'s and $i\omega_n$'s, was $<1\%$ (typically, this involves six iterations; with this criterion, the Si band gap has converged to within 1%). We consider first the case of K [19], the emphasis being placed on the impact of correlation on the occupied bandwidth. Figure 1 shows the density of one-particle states (DOS), given by the trace (sum over bands and over the BZ) of the spectral function $A_{j,j}(\vec{q};\omega) = -\pi^{-1}{\rm Im}\,G_{j,j}(\vec{q};\omega)$, where ω is real. The analytic continuation of $G_{j,j}(\vec{q};i\omega_n)$ to real ω's was performed via Padé approximants [20]. Since $A_{j,j}(\vec{q};\omega)$ contains sharp features, a dense $\vec{q}$ mesh is required in the evaluation of the trace, while, as implied above, Eq. (6) is solved on a coarser mesh [19,21,22]. A three-dimensional interpolation scheme was used to generate $A_{j,j}(\vec{q};\omega)$ for the dense mesh [16].

The results shown in Fig. 1 correspond to the LDA (no dynamical correlations), to the first evaluation of the self-energy, $\tilde{\Sigma}[G_{\rm LDA}]$, and to the self-consistent $\tilde{\Sigma}[G]$, G being the solution of Eq. (6) for the same $\tilde{\Sigma}$ ("same" within the tolerance stated above). In each case the zero of energy is the chemical potential μ, obtained by solving the implicit equation $n = 2\Omega^{-1}\sum_{\vec{k}}(\beta\hbar)^{-1}\sum_{i\omega_m} e^{i\omega_m\delta}G(\vec{k};i\omega_m)$ for each iteration of Eq. (6). (A convergence algorithm similar to the one outlined above for P was used here as well.) It is apparent that $\tilde{\Sigma}[G_{\rm LDA}]$ induces a narrowing of the band relative to the LDA (Table I), which agrees

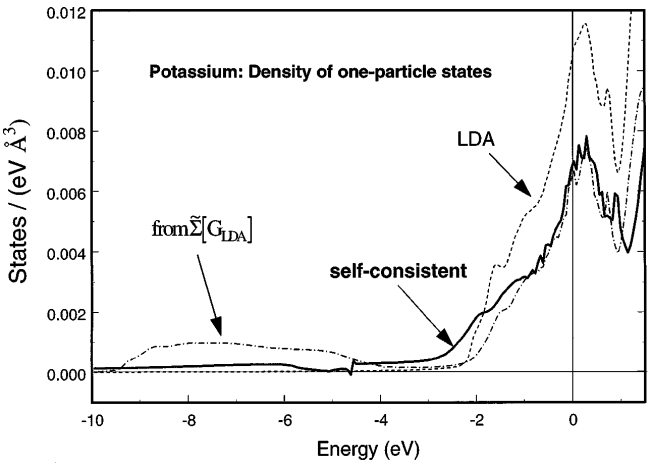

FIG. 1. Calculated DOS for K. Shown are the LDA result (dashes), and the results obtained from $\tilde{\Sigma}[G_{\mathrm{LDA}}]$ (dot-dashes) and from the self-consistent solution of Eq. (6) (solid line). The zero of energy is the corresponding chemical potential.

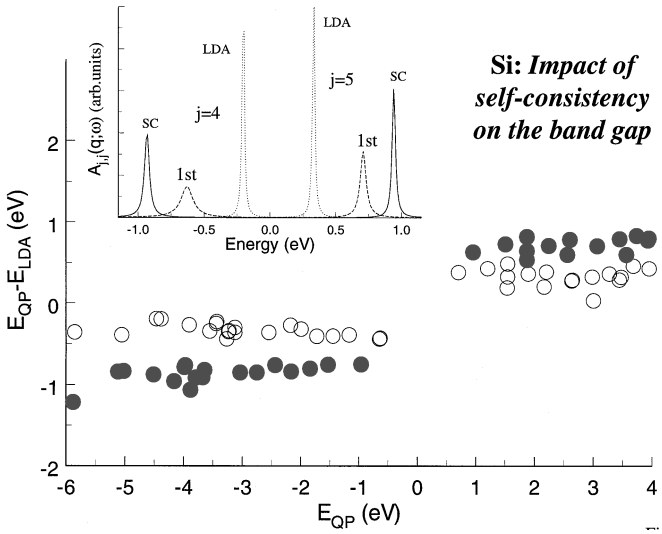

FIG. 2. Shift of the energy of QP states in Si near the absolute gap relative to the corresponding LDA eigenvalues. The empty circles correspond to $\tilde{\Sigma}[G_{\mathrm{LDA}}]$; the solid circles correspond to the self-consistent solution of Eq. (6). Inset: Spectral function $A_{j,j}(\vec{q};\omega)$ for the states at the bottom (fourth band, Γ point) and top (fifth band, $\vec{q} \sim 0.8\Gamma X$) of the absolute gap, respectively. Shown are the LDA band states (dots), and the QP peaks for the first (dashes) and self-consistent (solid line) evaluation of the self-energy.

qualitatively with experiment [23]. However, the self-consistent treatment of the correlations built into $\tilde{\Sigma}[G]$ yields a bandwidth which is *significantly wider* (Table I) [10,22].

This result is intimately connected with the reduction in the weight of the satellites which develop in $A_{j,j}(\vec{q};\omega)$ as a feedback of the plasmon resonance onto the one-particle spectrum. The holelike satellite structure below the band is quite apparent in the DOS obtained from $\tilde{\Sigma}[G_{\mathrm{LDA}}]$; this structure is strongly suppressed in the self-consistent calculation. A concomitant result is the increase of the weight of the QP state at the Fermi surface Z_{k_F} [15]; see Table I. This transfer of spectral weight from the satellite to the QP is controlled by (or is consistent with) sum rules obeyed by $A_{j,j}(\vec{q};\omega)$ [15]. Specifically, we find that the zero- and first-frequency moment sum rules are fulfilled to better than 1% [16,22].

We turn next to our results for QP states near the absolute band gap of Si [19]. Figure 2 shows the shift of the energy of the QP states, E_{QP}, relative to the Kohn-Sham LDA eigenvalue E_{LDA}, for the corresponding band and wave vector [E_{QP} is obtained from the position of the QP peak in $A_{j,j}(\vec{q};\omega)$—see inset]. The empty circles in Fig. 2 represent the shift obtained when the self-energy is $\tilde{\Sigma}[G_{\mathrm{LDA}}]$. At this non-self-consistent level our results agree well with previous work (Table II); the small difference with the $T = 0$ K results of Refs. [6,7] (in which P was also evaluated without plasmon-pole approximations) is attributed to residual effects of finite temperature [16,19]. The dominant feature of Fig. 2 is the correlation-induced widening of the gap relative to the LDA, by an amount which happens to agree well with experiment [24] (Table II). However, the self-consistent treatment of QP propagation built into $\tilde{\Sigma}[G]$ (solid circles) yields a much wider gap (Fig. 2 and Table II). Indeed, the "true" GW band gap ($\sim$1.9 eV) deviates from the experimental value (1.17 eV) by an amount comparable with, but of the opposite sign, the error built into the gap obtained from LDA eigenvalues (0.53 eV). This trend

TABLE I. Potassium: Calculated values of the occupied bandwidth (in eV) and QP renormalization factor at the Fermi surface Z_{k_F}. The experimental value of the bandwidth is also given.

	Occupied bandwidth	Z_k
LDA	2.21	1.0
From $\tilde{\Sigma}[G_{\mathrm{LDA}}]$	2.04	0.60
From self-consistent solution of Eq. (6), $\tilde{\Sigma}[G]$	2.64	0.72
Experiment [23]	1.60	⋯

TABLE II. Silicon: Occupied bandwidth, direct gap at the Γ point, and absolute gap (energies in eV)—comparison with previous GW calculations, and with experiment.

	Occupied bandwidth	Direct gap	Absolute gap
Ref. [3]	12.04	3.35	1.29
Ref. [6]	12.35	3.33	1.17
Ref. [7]	⋯	3.29	1.29
Ref. [8]	11.57	3.23	1.19
LDA	11.93	2.57	0.53
Experiment [24]	12.5 ± 0.6	3.4	1.17
Present work; from $\tilde{\Sigma}[G_{\mathrm{LDA}}]$	11.65	3.27	1.34
Present work; self-consistent	13.10	4.02	1.91

is easily visualized in the inset in Fig. 2, which shows the spectral function of the QP states at both edges of the gap, for the LDA, and for the first (1st) and self-consistent evaluations of $\tilde{\Sigma}$. Similar conclusions apply for the occupied bandwidth and the direct gap at Γ (Table II).

As for the physics behind the overestimation of the Si band gap (and the widening of the band of K as well), a mechanism left out in Eq. (3) is as follows. The renormalized polarizability P, obtained from propagators dressed self-consistently with the self-energy, does not give rise to a plasmon for small q's, unlike the "bare" polarizability $P_{\mathrm{LDA}} \sim G^{*}_{\mathrm{LDA}} G_{\mathrm{LDA}}$ [16]. Thus, correlations beyond Eq. (3) must play a role. For example, at least to low order in perturbation theory, it has been shown that the competition between self-energy insertions and vertex corrections produces a polarizability that does yield a plasmon for small q's [25]. A conserving scheme which includes renormalized vertices, and which may be suitable for the present problem, has been proposed recently for strongly correlated electrons [26].

In summary, we have presented *ab initio*, conserving [11] calculations of QP states for archetypes of metallic and semiconducting behavior, K and Si. We have shown that self-consistency in the treatment of correlation at the level of the *GW* approximation yields QP energies (gaps, bandwidths) which disagree with experiment—i.e., "the *GW* diagram is not enough." Competing correlations (e.g., vertex corrections) must be incorporated in the QP band structure. Our work makes it clear that those effects must be incorporated fully self-consistently. Whether an approximate (if self-consistent) inclusion of additional correlations will restore agreement with experiment to the level that has been generally accepted to be true for the *GW* band structure (0.1 eV) is a matter of conjecture at the present time.

W.-D. S. was partially supported by the Deutsche Forschungsegemeinschaft. A. G. E. was supported by National Science Foundation Grant No. DMR-9634502 and the National Energy Research Supercomputer Center. Oak Ridge National Laboratory is managed by Lockheed Martin Energy Research Corp., for the Division of Materials Sciences, U.S. DOE under Contract No. DE-AC05-96OR22464.

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965). For reviews of density-functional theory, see, e.g., *Density Functional Theory of Many-Fermion Systems*, edited by S. B. Trickey (Academic, New York, 1990)
[2] M. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[3] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **35**, 4170 (1987).
[4] F. Bechstedt *et al.*, Phys. Rev. B **49**, 7357 (1994).
[5] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **56**, R7065 (1997); **57**, 6485 (1998).
[6] H. N. Rojas, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **74**, 1827 (1995).
[7] A. Fleszar and W. Hanke, Phys. Rev. B **56**, 10 228 (1997).
[8] L. Hedin, Phys. Rev. **139**, A796 (1965).
[9] For recent reviews, see S. G. Louie, Surf. Sci. **299/300**, 346 (1994); F. Aryasetiawan and O. Gunnarsson, cond-mat/9712013.
[10] Full self-consistency for electrons in jellium has been implemented by B. Holm and U. von Barth, Phys. Rev. B **57**, 2108 (1998), and by the present authors [22].
[11] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).
[12] A. Schindlmayr, Phys. Rev. B **56**, 3528 (1997).
[13] M. P. Surh, J. E. Northrup, and S. G. Louie, Phys. Rev. B **38**, 5976 (1988).
[14] F. Aryasetiawan, Phys. Rev. B **46**, 13 051 (1992).
[15] G. D. Mahan, *Many-Particle Physics* (Plenum, New York, 1990), 2nd ed.
[16] W.-D. Schöne, J. A. Gaspar, and A. G. Eguiluz (to be published).
[17] Our convergence algorithm is in the spirit of procedures introduced in Hubbard-model calculations, by, e.g., J. W. Serene and D. W. Hess, Phys. Rev. B **44**, R3391 (1991); N. E. Bickers and S. R. White, Phys. Rev. B **43**, 8044 (1991).
[18] M. Bockstedte, A. Kley, J. Neugebauer, and M. Scheffler, Comput. Phys. Commun. **107**, 187 (1997).
[19] Numerical parameters used in the calculations for K are $T = 1100$ K, $\omega_{\max} = 80$ eV (256 Matsubara frequencies), $8 \times 8 \times 8$ Monkhorst-Pack mesh sampling of the BZ, six bands are used in evaluating P. For Si: $T = 2320$ K, $\omega_{\max} = 161$ eV, $9 \times 9 \times 9$ mesh, 26 bands; convergence tests involving these parameters are presented in [16].
[20] H. J. Vidberg and J. W. Serene, J. Low Temp. Phys. **29**, 179 (1977).
[21] "Conventional" $T = 0$ K methods [14] use coarse meshes in the evaluation of Σ, but at the expense of introducing numerical smearing in the eV range in the evaluation of P.
[22] A. G. Eguiluz and W.-D. Schöne, Mol. Phys. **94**, 87 (1998).
[23] B. S. Itchkawitz, I.-W. Lyo, and E. W. Plummer, Phys. Rev. B **41**, 8075 (1990).
[24] *Zahlenwerte und Funktionen aus Naturwissenschaften und Technik*, Landolt-Bornstein, New Series, Vol. III, Pt. 17a (Springer, New York, 1982).
[25] B. E. Sernelius, Phys. Rev. B **36**, 1080 (1987).
[26] Y. M. Vilk and A.-M. S. Tremblay, Europhys. Lett. **33**, 159 (1996).