All-Electron Self-Consistent *GW* Approximation: Application to Si, MnO, and NiO

Sergey V. Faleev,[1] Mark van Schilfgaarde,[2] and Takao Kotani[3]
[1]*Sandia National Laboratories, Livermore, California 94551, USA*
[2]*Arizona State University, Tempe, Arizona, 85284, USA*
[3]*Department of Physics, Osaka University, Toyonaka 560, Japan*
(Received 18 October 2003; published 17 September 2004)

We present a new kind of self-consistent *GW* approximation based on the all-electron, full-potential linear muffin-tin orbital method. By iterating the eigenfunctions of the *GW* Hamiltonian, self-consistency in both the charge density and the quasiparticle spectrum is achieved. We explain why this form of self-consistency should be preferred to the conventional one. Some results for Si (a representative semiconductor) are presented. Finally we consider many details in the electronic structure of the antiferromagnetic insulators MnO and NiO. Excellent agreement with experiment is shown for many properties, suggesting that a Landau quasiparticle (energy band) picture provides a reasonable description of electronic structure even in these correlated materials.

DOI: 10.1103/PhysRevLett.93.126406 PACS numbers: 71.10.–w, 71.15.Ap, 71.20.–b, 75.50.Ee

The *GW* approximation (GWA) of Hedin [1] is generally believed to accurately predict excited-state properties, and, in particular, improve on the Kohn-Sham band structure, for example, in local density approximation (LDA), whose limitations are well known, e.g., to underestimate band gaps in semiconductors and insulators. Usually GWA is computed as a one-shot calculation starting from the LDA eigenfunctions and eigenvalues; the self-energy Σ is approximated as $\Sigma = iG^{\mathrm{LDA}}W^{\mathrm{LDA}}$, where G^{LDA} is a bare Green function constructed from LDA eigenfunctions, and W^{LDA} is the screened Coulomb interaction constructed from G^{LDA} in the random phase approximation (RPA). However, establishing the validity of the one-shot approach has been seriously hampered by the fact that nearly all calculations to date make further approximations, e.g., computing Σ from valence electrons only, the plasmon-pole approximation, and the pseudopotential (PP) approximation to deal with the core. Only recently, when reliable all-electron implementations have begun to appear, has it been shown that the one-shot GWA with PP leads to systematic errors [2–4]. There is general agreement among the all-electron calculations (see Table I) that the Γ-X transition in Si is underestimated when $\Sigma \approx iG^{\mathrm{LDA}}W^{\mathrm{LDA}}$. And we have shown previously [2] that the tendency for $\Sigma = iG^{\mathrm{LDA}}W^{\mathrm{LDA}}$ to underestimate gaps is almost universal in semiconductors. This is reasonable because small gaps used to make G increase the screening and underestimate gap corrections. G constructed from quasiparticles (QP) with a wider gap (e.g., a self-consistent G) reduces the screening and therefore generates *GW* with a wider gap.

However, there are many possible ways to achieve self-consistency. The theoretically simplest (and internally consistent) is the fully self-consistent scheme (SCGW), which is derived through the Luttinger-Ward functional with the exchange-correlation energy approximated as the sum of RPA ring diagrams. Then W is evaluated as $W = v(1 - vP)^{-1}$ with the proper part of the polarization function $P = -iG \times G$. However, such a construction may not give reasonable W [5], resulting in a poor G, for the following reason. If Σ is energy dependent, G can be partitioned into a QP part and a residual satellite part. The QP part consists of terms whose energy dependence varies as $Z_i/(\omega - \epsilon_i \pm i\Gamma_i)$, where ϵ_i, Γ_i, and Z_i are, respectively, the QP energies, inverse lifetimes, and renormalization factors ($Z_i < 1$, typically between 0.7 and 1). The QP parts are thus weighted by factors Z; the residual weights $1 - Z$ go into the plasmon-related satellite parts, high in energy. Thus $P = -iG \times G$ contains contributions from the particle-hole pair excitations as does $-iG^{\mathrm{LDA}} \times G^{\mathrm{LDA}}$, but reduced by the products of two Z factors, one from occupied and the other from unoccupied states. However, this construction of P is not consistent with Landau's quasiparticle theory, which insists

TABLE I. Minimum energy gap E_g and selected energy eigenvalues for Si, relative to Γ'_{25v} (eV). Three all-electron methods are shown: linearized augmented-plane-wave (LAPW) and projector-augmented-wave (PAW) approaches, and linear muffin-tin orbital (this work). The PAW calculation included valence electrons only. The last row compares the Ge valence bandwidth. The results of this work differ slightly from Ref. [2] because a large basis set (50 orbitals/atom) was employed in the present work.

	PAW [3]	LAPW [4]		This work		
	$(GW)^{\mathrm{LDA}}$	$(GW)^{\mathrm{LDA}}$	SCGW	$(GW)^{\mathrm{LDA}}$	SCGW	Exp.
E_g	0.92	0.85	1.03	0.84	1.14	1.17
X_{1c}	1.01			0.98	1.28	1.32
L_{1c}	2.05			2.03	2.24	2.04
Γ_{15c}	3.09	3.12	3.48	3.06	3.40	3.40
Γ_{1v}		−12.1	−13.5	−12.1	−12.3	−12.5
Γ_{1v}(Ge)		−13.1	−14.8	−12.9	−13.1	−12.6

that one-particle excitations remain meaningful (at least near the Fermi energy). Based on the theory, we should instead evaluate the QP contributions to P without Z factors, as they dominate the static screening $W(\omega \to 0)$. Inclusion of Z can lead to $W(0)$ being underscreened; moreover $W(\omega)$ does not satisfy the f sum rule [5]. Consequently, $W(0)$ will be overestimated resulting in a tendency to overestimate bandwidths, as is well known in the extreme (Hartree-Fock) case, where there is no screening of W.

Indeed Holm and von Barth found that SCGW overestimates the bandwidth in the homogeneous electron gas [6]. Very recently Ku presented a SCGW calculation which similarly overestimates the valence bandwidth in Si and Ge; see Table I [4]. Another practical justification for the argument that a bare G without Z should be used when we construct P as $-iG \times G$, is that W^{LDA} is already known to be rather good if we add some enlargement of band gap by hand to correct for errors in the LDA ϵ_i [7].

For these reasons, we do not adopt the full SCGW scheme but construct two kinds of constrained self-consistent GW methods. For a set of trial eigenfunctions and quasiparticle energies $\{\psi_{\mathbf{q}n}, \epsilon_{\mathbf{q}n}\}$, we can calculate the one-particle Green function G, and in turn the self-energy $\Sigma^{\mathbf{q}}_{nn'}(\omega)$ in the GWA. Then we generate an energy-independent, Hermitian $\Sigma^{\mathbf{q}}_{nn'}$ in one of two ways:

$$\Sigma^{\mathbf{q}}_{nn'} = \Sigma^{\mathbf{q}}_{nn'}(E_{\mathrm{F}}) + \delta_{nn'}\mathrm{Re}[\Sigma^{\mathbf{q}}_{nn}(\varepsilon_{\mathbf{q}n}) - \Sigma^{\mathbf{q}}_{nn}(E_{\mathrm{F}})], \quad (1)$$

$$\Sigma^{\mathbf{q}}_{nn'} = \mathrm{Re}[\Sigma^{\mathbf{q}}_{nn'}(\varepsilon_{\mathbf{q}n}) + \Sigma^{\mathbf{q}}_{nn'}(\varepsilon_{\mathbf{q}n'})]/2, \quad (2)$$

where Re means that we take only the Hermitian parts and E_{F} is set to the middle of the band gap. From $\Sigma^{\mathbf{q}}_{nn'}$ we can construct a new density $n(\mathbf{r})$ and corresponding Hartree potential and proceed to a new set of $\{\psi_{\mathbf{q}n}, \epsilon_{\mathbf{q}n}\}$, iterating to self-consistency. Our method is not related to the LDA (though in practice the LDA is used to make a starting guess for Σ and the augmented-wave basis set). Schemes (1) and (2) differ in the treatment of the off-diagonal parts of $\Sigma_{nn'}$, but both restrict the potential to be nonlocal, Hermitian, and ω independent. Thus the problem in the full SCGW is avoided; also the numerical computation becomes rather stable. In Ref. [4], off-diagonal matrix elements of $\Sigma^{\mathbf{q}}_{nn'}$ were completely neglected. However, these are important for MnO and NiO: eigenfunctions and the density cannot be changed from LDA if we neglect them. We find that converged QP energies differ in these two schemes by small amounts (less than 0.02 eV for Si and $\sim$0.1 eV for NiO), which is within the resolution of the method ($\sim$ 0.1 eV).

Our implementation is based on the method of Ref. [2]. W is expanded in a mixed basis which consists of two contributions: local atom-centered functions (product basis) confined to muffin-tin spheres and plane waves with the overlap to the local functions projected out. The former can include any of the core states: thus the valence and core states can be treated on an equal footing and the contribution of the latter to Σ included [8]. We calculate the full energy dependence of W without the plasmon-pole approximation. This approach shares some features in common with both the full-potential, all-electron plane-wave based methods [3,4] and the product-basis method [9], combining the advantages of each, e.g., efficient treatment of localized valence electrons.

Results for Si are shown in Table I. Agreement between the three all-electron methods is generally excellent. The $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ gaps are $\sim$0.3 eV smaller than experiment; the SCGW gaps fall much closer. As we will show elsewhere, most properties of weakly correlated systems calculated with the present SCGW method (fundamental and higher gaps, valence bandwidths, effective mass, position of deep d levels) are in excellent agreement with experiment, with small systematic residual errors.

Turning to the TM oxides, we first consider MnO because it is less correlated. Figure 1 compares the SCGW energy bands and corresponding density of states (DOS) to the LDA and the $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ gap. The conduction band at Γ is evidently a dispersive band of sp character. Above this, fall the t_{2g} bands ($\sim$ 6–9 eV); still higher at $\sim$10 eV is a narrow e_g band, whose width is $\sim$3 eV. Thus, the itinerant and d bands are well separated. The minimum gap is 3.5 eV, in good agreement with the bremsstrahlung isochromat spectroscopy (BIS) gap [10] (3.9 $\pm$ 0.4 eV). The BIS spectrum also shows a peak at $\sim$6.8 eV, which probably corresponds to a convolution of the peaks of t_{2g} symmetry seen in the DOS at 6.6 and 7.3 eV. These bands are in stark contrast to the LDA, which shows the t_{2g} and e_g bands overlapping and hybridizing with the sp band at 1 to 4 eV.

LDA and SCGW valence bands are more similar (Fig. 1). In the LDA there is a narrow upper e_g band at

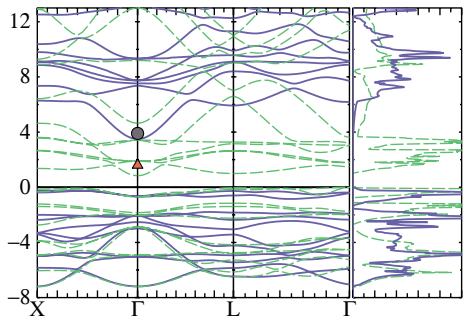

FIG. 1 (color online). SCGW energy bands and DOS of MnO. Solid lines: SCGW bands; dashed lines: LDA bands. The valence-band maximum (VBM) is set to energy 0. Circle and triangle at Γ: BIS and $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ gaps. The right panel shows the corresponding DOS. Peaks at -0.5 eV (-0.1 eV in LDA) and -5 eV are the nearly dispersionless e_g bands. Peaks at -2.2 eV (-1.2 eV) and 6.6 and 7.3 eV (1.7 and 1.9 eV) derive from Mn t_{2g} states.

0.1 eV below the VBM, and another one at VBM − 5 eV. Both weakly hybridize with the O $2p$ band. The SCGW pushes the upper e_g band down to VBM − 0.5 eV, so that the VBM takes more O $2p$ character, and the band at VBM − 5 eV takes more Mn d character. The splitting Δ_v between the upper e_g level and the t_{2g} level widens from 1.0 eV (LDA) to 1.7 eV (SCGW), in good agreement with a photoemission measurement of 1.9 eV [10]. An approximately similar picture emerges from a model GW calculation of Massidda *et al.* [11], the most important difference being that the model GW d conduction bands fall ∼1 eV lower than ours.

Figure 2 compares the SCGW energy bands for NiO along the [110] and [100] lines to the angle-resolved photoemission spectroscopy data of Shen *et al.* [12] for the valence bands, and to the LDA and $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ conduction bands. As can be seen, self-consistency has a dramatic effect on the TM oxides. The right panel shows the DOS for both LDA and SCGW, and Fig. 3 shows the total DOS resolved into components. Also shown in the top panel are BIS data [14].

Several features are of interest as follows:

(i) The conduction-band minimum falls at the Γ point; the VBM falls at the point $(1/2, 1/2, 1/2)$ (not shown). The calculated minimum gap and magnetic moment are dramatically improved (see Table II).

(ii) The SCGW conduction bands are a mixture of a dispersive band composed of sp approximately equally weighted on the Ni and O sites, and a nearly dispersionless e_g state (see the discussion of EELS below). Peaks in the BIS spectrum labeled "1," "2," and "3" closely coincide to those in the SCGW total DOS, apart from a constant shift of 0.8 eV.

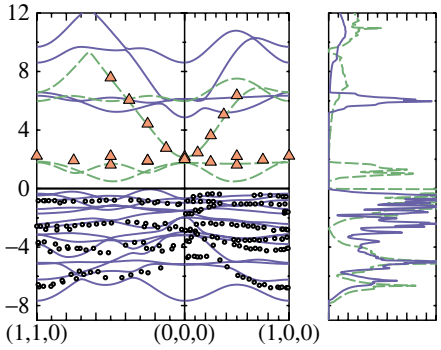

FIG. 2 (color online). SCGW energy bands and DOS of NiO. Valence-band maximum is set at energy 0. Solid lines: SCGW bands; dashed lines: LDA bands (only conduction bands are shown). Circles show photoemission data of Ref. [12], using a Fermi level of VBM + 0.8 eV [13]. Triangles show the lowest $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ conduction bands, without the Z factor. The $G^{\mathrm{LDA}}W^{\mathrm{LDA}}$ sp band is similar to the LDA; the dispersionless e_g band shows a modest improvement relative to the LDA. The right panel shows the corresponding total DOS.

(iii) The SCGW valence bands are in very good agreement with experiment: indeed they agree as well with the Shen data as the latter agrees with an independent experiment by Kuhlenbeck *et al.* [15] (not shown).

(iv) There is an increased dispersion in the valence bands relative to the LDA at the VBM because the nearly dispersionless Ni t_{2g} levels are pushed down. Thus the VBM acquires somewhat more O $2p$ character. This supports the generally accepted view that the LDA too heavily favors the Mott-Hubbard picture.

The bottom panel of Fig. 3 compares the EELS spectrum from the O $1s$ core, calculated as described in Ref. [16]. Calculated data were convolved by a Gaussian of 0.5 eV width (which was the resolution reported) to compare to experimental data reported by Dudarev *et al.* [17], The calculated results were shifted to align the spectra with the DOS; Dudarev *et al.*'s data were shifted by 526 eV to align the peaks with the SCGW results. (Peaks labeled 1, 2, and 3 should correspond to the peaks with the same labels in the BIS spectrum; indeed the measured EELS peaks and BIS peaks almost perfectly align if the EELS data is further shifted by 0.8 eV.)

Apart from the 0.8 eV shift, the EELS data are in excellent agreement with the SCGW results. Spacings between the three peaks agree to within ∼0.1 eV, and the spectral weight under each peak (estimated by numerical integration) also agrees well. This establishes that the SCGW relative positions of the sp and Ni d bands are correctly predicted. This is a significant result, because

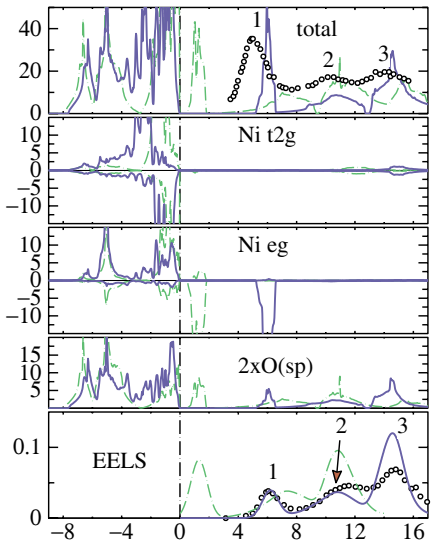

FIG. 3 (color online). DOS and electron-energy-loss spectroscopy (EELS) of NiO. Solid lines: SCGW data; dash-dotted lines: LDA data. Top panel: total DOS, together with BIS data of Ref. [14] (circles). Panels 2 and 3 show the Ni t_{2g} and e_g partial DOS, with positive DOS showing majority spin and negative showing minority spin. Panel 4 shows the O sp partial DOS; panel 5 compares the calculated and measured [17] EELS spectra from the O $1s$ level.

TABLE II. Magnetic moments and minimum gaps in MnO and NiO.

	Moment			Band gap		
Compound	LDA	SCGW	Expt.	LDA	SCGW	Expt.
MnO	4.48	4.76	4.6	0.78	3.5	3.9 ± 0.4
NiO	1.28	1.72	1.9	0.45	4.8	4.3

the relative positions of the sp and d bands is a rather delicate quantity [8]. In contrast, the LDA overestimates the spacing between peaks 1 and 2 by ~0.8 eV and underestimates the spacing between peaks 2 and 3 by ~0.7 eV. Moreover, it overestimates the spectral weight of the first peak by a factor of ~2. This result is also significant, because the EELS spectra largely reflect the O $2p$ partial DOS. Without coupling between the Ni e_g level, the itinerant band would adopt a simple parabolic form; thus the amplitude of the first peak is a reflection of the hybridization between the Ni e_g and the itinerant band. The fact that SCGW gets the correct weight for this peak establishes that it accurately estimates this coupling, while the LDA overestimates it by a factor of 2.

Many of the results found here confirm many conclusions drawn in a model GW calculation [18], as well as various LDA + Hubbard U calculations [17,19,20], both of which may be viewed as model approaches to the present theory. Some significant differences do arise. The relative positions of different bands and the energy gaps depend rather sensitively on the choice of parameters in the model approaches. For example, in Ref. [20], the O-derived sp conduction band appears to fall at 2.8 eV, ~2 eV below the middle of the e_g level when U is assumed to be 5 (somewhat lower than the constrained LDA estimate, $U \sim 8$ eV). Massidda *et al.*'s model GW calculation[18] shows the sp band ~1 eV above the e_g.

To what extent does the Landau QP picture based on the preceding SCGW results fail to describe the true electronic structure of MnO and NiO? We have shown that a great deal is correctly described, including many details of the valence and conduction bands. The main discrepancy is with x-ray photoemission measurements. For optics, the peak in $\mathrm{Im}(\epsilon)$ corresponding to the gap in NiO is about VBM + 4 eV, whereas this peak is at VBM + 5 eV in the BIS data. But $\mathrm{Im}(\epsilon)$ is directly related to the excitonic process or correlated motion of electron-hole pairs, which can shift $\mathrm{Im}(\epsilon)$ downward. The difference between the two experiments can be due to this correlation. So the poles of the true Green function (which are reflected in the DOS) should correspond to the unoccupied d position at VBM + 5 eV as is shown in BIS. Peak 1 in the SCGW DOS falls slightly higher than experiment, at ~5.8 eV. If we include correlation beyond RPA, e.g., inclusion of ladder diagrams, screening will increase and the band gap in the Green function may be reduced. A model calculation [21] estimated the reduction to be ~1 eV in NiO. Thus, it would seem that the RPA explains quite well the important experimental data, apart from a slight tendency to underestimate screening of W [22]. We have not yet attempted to include excitonic effects, so we cannot say to what extent photoemission data can be explained within the RPA, though estimates in a model context were reasonably successful [19]. Thus, we believe that the band picture [23] for NiO is a reasonable starting point for the description of the electronic structure of NiO, much better than previously thought, and in many respects more appropriate than the ligand-field picture.

This work was supported by ONR Contract No. N00014-02-1-1025 and by BES Division of Materials Sciences, Contract No. DE-AC04-94AL85000.

[1] L. Hedin, Phys. Rev. **139**, A796 (1965).
[2] T. Kotani and M. van Schilfgaarde, Solid State Commun. **121**, 461 (2002).
[3] S. Lebegue *et al.*, Phys. Rev. B **67**, 155208 (2003).
[4] W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **89**, 126401 (2002).
[5] D. Tamme *et al.*, Phys. Rev. Lett. **83**, 241 (1999).
[6] B. Holm and U. von Barth, Phys. Rev. B **57**, 2108 (1998).
[7] B. Arnaud and M. Alouani, Phys. Rev. B **63**, 085208 (2001).
[8] For reliable results it was essential to use well-converged basis sets. Local orbitals were included for both the Ni $3p$ and $4d$ channels. Without the Ni $4d$ orbital, e_g falls ~1–2 eV higher in energy. Treating the Ni $3p$ in an exchange-only approximation and neglecting the $3p$-valence hybridization, the gap widens by ~0.5 eV. Approximating the $3p$ at the LDA level induces still further errors [2,4].
[9] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[10] J. van Elp *et al.*, Phys. Rev. B **44**, 1530 (1991).
[11] S. Massidda *et al.*, Phys. Rev. Lett. **74**, 2323 (1995).
[12] Z.-X. Shen *et al.*, Phys. Rev. B **44**, 3604 (1991).
[13] K. Shih (private communication).
[14] G. A. Sawatzky and J. W. Allen, Phys. Rev. Lett. **53**, 2339 (1984).
[15] H. Kuhlenbeck *et al.*, Phys. Rev. B **43**, 1969 (1991).
[16] A. T. Paxton *et al.*, J. Phys. C **12**, 729 (2000).
[17] S. L. Dudarev *et al.*, Phys. Rev. B **57**, 1505 (1998).
[18] S. Massidda *et al.*, Phys. Rev. B **55**, 13 494 (1997).
[19] V. I. Anisimov *et al.*, Phys. Rev. B **48**, 16 929 (1993).
[20] O. Bengone *et al.*, Phys. Rev. B **62**, 16 392 (2000).
[21] M. Takahashi and J. Igarashi, Phys. Rev. B **54**, 13 566 (1996).
[22] Interestingly, the semiconductors also demonstrate a slight but universal tendency to overestimate the band gaps; see, e.g., Ku's results for Ge [4]. Not surprisingly, the electron-hole pair correlation is rather stronger in a correlated material such as NiO.
[23] K. Terakura *et al.*, Phys. Rev. B **30**, 4734 (1984).