Effect of self-consistency on quasiparticles in solids

Fabien Bruneval*
Laboratoire des Solides Irradiés, UMR 7642, CNRS-CEA/DSM, École Polytechnique, F-91128 Palaiseau, France and European Theoretical Spectroscopy Facility (ETSF)

Nathalie Vast
Laboratoire des Solides Irradiés, UMR 7642, CNRS-CEA/DSM, École Polytechnique, F-91128 Palaiseau, France

Lucia Reining
Laboratoire des Solides Irradiés, UMR 7642, CNRS-CEA/DSM, École Polytechnique, F-91128 Palaiseau, France and European Theoretical Spectroscopy Facility (ETSF)

(Received 27 January 2006; revised manuscript received 15 May 2006; published 6 July 2006)

We have evaluated the self-energy of solids within different self-consistent approximations, from bare and screened exchange to screened exchange plus Coulomb hole (COHSEX) and *GW* approximations. Our calculations for silicon, aluminum, and argon assess the quality of the local-density approximation (LDA) valence wave functions and density with respect to their *GW* counterparts. The LDA conduction wave functions are shown to be of significantly poorer quality, in particular for not highly symmetric points of the Brillouin zone. Comparing the different approaches we show that screened exchange alone is not reliable and we propose to combine a self-consistent COHSEX with a subsequent perturbative *GW* calculation that gives a reasonable estimate of the self-consistent *GW* results, at a much lower computational cost. The calculations are based on the pseudopotential method; they confirm its reliability.

DOI: 10.1103/PhysRevB.74.045102 PACS number(s): 71.10.−w, 71.15.Ap, 71.20.−b

I. INTRODUCTION

Hedin's *GW* approximation[1] has become the method of choice to determine the band structure of reasonably correlated solids. The success has been broad, in particular for the determination of the band gaps of semiconductors.[2] This is important, because the widely spread Kohn-Sham (KS) approach[3] of density-functional theory[4] (DFT) systematically gives a poor estimate of this quantity, for both fundamental and practical reasons.[5] KS eigenvalues should in fact not be interpreted as band structures. Available approximations within DFT, such as the local density approximation (LDA), may further worsen the results. Even though in practice KS LDA often yields adequate band dispersions for valence states, it is desirable to turn to theoretically better justified methods.

On the contrary, the energies in the many-body perturbation theory (MBPT), of which *GW* is one of the prominent approximations, have a physical meaning: they can be interpreted as quasiparticle (QP) energies.[6,7] The price to pay is that the key quantity of the theory is much more complex than the ground-state density $\rho(\mathbf{r})$ of DFT. In MBPT, the one-particle Green's function $G(\mathbf{r}_1 t_1, \mathbf{r}_2 t_2)$ has to be considered explicitly. Even within the *GW* approximation of the MBPT framework, the calculations are much more involved and additional technical approximations are mandatory.

In practice, the *GW* method refers not only to an approximation within MBPT, but also to the technical recipe used to solve the equations, first introduced in pioneering *ab initio* works of the mid-1980's.[8,9] In particular, the LDA wave functions and *GW* QP wave functions are assumed to be equal. This allows one to evaluate the *GW* band structure as a first-order perturbation with respect to the LDA one. The overlap between LDA and *GW* wave functions has been claimed to be greater than 99.9% for simple bulk systems.[10] Moreover, the standard procedure calculates the Green's function constructed from LDA eigenvalues and wave functions. This "one-shot" method, denoted LDA+G_0W_0 in the following, avoids the calculation of self-consistent *GW* quantities. It is of practical interest, because full self-consistency would be out of reach for realistic systems: for instance, as the *GW* self-energy is not Hermitian, it would be required to consider the nonorthogonal left and right eigenvectors of the Hamiltonian. It is also of theoretical interest, since a recent work on the homogeneous electron gas[11] showed that self-consistency in *GW* worsens the agreement between the *GW* spectral function and the exact one. The self-consistency issue is currently a much debated point.[12–16]

Another point plagues the *GW* community nowadays. While the first applications of *GW* to solids were based on the pseudopotential technique that separates core and valence electrons, recent results[14,16,17] show that all-electron methods give different results, with in general smaller band gaps. Furthermore, the difference between all-electron and pseudopotential methods is exacerbated, when performing self-consistent calculations.[13]

From the theoretical point of view, the dynamical part of *GW* has been found responsible for the poor performance of self-consistency in the homogeneous electron gas.[11] The present work concentrates therefore on static approximations to *GW*, that will render the calculation theoretically simpler and practically tractable. In this context, we intend to provide answers to the following questions: *Can we and should we calculate band structure using self-consistent QP methods? Are pseudopotential suited to perform this kind of calculations?* As a by product, we will make explicit "when" and

TABLE I. Summary of the theoretical schemes applied in the present work.

LDA	Local density approximation of DFT
HF	Self-consistent Hartree-Fock approximation to the self-energy
SEX	Self-consistent screened exchange approximation to the self-energy
COHSEX	Self-consistent Coulomb hole plus screened exchange approximation to the self-energy
LDA+G_0W_0	Standard perturbative "one-shot" GW calculation based on LDA inputs
COHSEX+G_0W_0	Perturbative "one-shot" GW calculation based on COHSEX inputs
QPscGW	Self-consistent static approximation to the GW self-energy of Ref. 16

"why" the usual "one-shot" procedure is justified by analyzing the quality of LDA inputs.

The static approximations considered here are historical approximations:[1,18] Hartree-Fock (HF), screened exchange (SEX), screened exchange plus Coulomb hole (COHSEX); and a recently proposed static version of GW by Faleev *et al.* in Ref. 16, called QPscGW by those authors and in the following.

Our calculated QPscGW and LDA wave functions have sometimes a much lower overlap than the usually claimed 99.9%, even for bulk silicon. This finding demands a careful study of the effect of wave functions on the QPscGW band structure.

We have evaluated the quality of SEX and COHSEX self-consistent eigenvalues and wave functions. Based on the result, we suggest to use COHSEX inputs (eigenfunctions and eigenvalues) as the starting point for a "one-shot" GW calculation (method called COHSEX+G_0W_0 in the following). In fact, because the COHSEX approximation is a straightforward approximation to the GW self-energy, the result should be closer to GW than LDA, which justifies a first-order perturbation treatment. Moreover, the study of the quality of SEX and COHSEX approximations is a very interesting point, since a modeled version of SEX is used in practice within the generalized KS framework.[19–21] To our knowledge, the COHSEX approximation has never been evaluated self-consistently.

To test the accuracy of the different methods, the present work focuses on simple solids. Bulk silicon is the test case for *sp* bonded semiconductors. Aluminum is a simple metal, with an almost free electron band structure. Solid argon is representative for the atomiclike insulators with a huge band gap [14.2 eV (Ref. 22)].

The paper is organized as follows. In Sec. II we recapitulate the basics of Green's functions and the derivation of the approximations used here. In Sec. III we provide the parameters used in the practical calculations. In Sec. IV we assess the technical approximations by performing single shot calculations. We provide the results for band structures of simple solids, silicon, aluminum, and argon in Sec. V. The analysis of the QP wave functions and their role in the GW calculation is the topic of Sec. VI. The conclusions are drawn in Sec. VII. The different theoretical schemes used in this paper are gathered in Table I to help the reader.

II. QUASIPARTICLE FRAMEWORK

A. Green's functions and quasiparticles

The usual one-particle picture with single particle wave functions and eigenvalues is not sufficient to define the band structures of the many interacting electrons of solids. Instead, the one-particle Green's function G is the central object in the determination of band structures within the MBPT. It describes the propagation of an extra particle (electron or hole) in the system and hence it is often referred to as a propagator.

Due to the many-body interaction, one cannot consider simply single bare particles. The QP concept should be used instead.[6,7] The propagation of the extra particle consists of QP contributions that resemble much well defined single particle excitations, plus other (usually weaker) many-body contributions, e.g., the satellites. The QP wave functions $\phi_i^{\mathrm{QP}}(\mathbf{r})$ may look similar to the usual single particle wave functions, but they intrinsically carry many-body features, and hence they do not correspond strictly speaking to a unique particle. They satisfy a Schrödinger-like equation of motion (spin degrees of freedom are omitted throughout the present work)

$$\left[-\frac{\nabla^2}{2}+\int d\mathbf{r}'\rho(\mathbf{r}')v_{\mathrm{ext}}(\mathbf{r})+\int d\mathbf{r}'\rho(\mathbf{r}')v(\mathbf{r}-\mathbf{r}')\right]\phi_i^{\mathrm{QP}}(\mathbf{r})$$
$$+\int d\mathbf{r}'\Sigma(\mathbf{r},\mathbf{r}',\epsilon_i^{\mathrm{QP}})\phi_i^{\mathrm{QP}}(\mathbf{r}')=\epsilon_i^{\mathrm{QP}}\phi_i^{\mathrm{QP}}(\mathbf{r}), \quad (1)$$

where index i runs over states and $\mathbf{k}$ points in a solid. The usual kinetic and external potential can be recognized. The electron-electron interaction is accounted for via two terms: the classical Hartree term $v_H=\int\rho v$ and the exchange-correlation term, the self-energy Σ. The whole complexity of the problem is contained in the self-energy, which is nonlocal, non-Hermitian, and energy dependent. Consequently, the QP wave functions are not orthogonal and the QP energies ϵ_i^{QP} have an imaginary component that accounts for the damping of the QP excitations, i.e., finite lifetimes. The real

part of ϵ_i^{QP} defines the electron addition or removal energy, or in other words, the band structure of the system.

To calculate the key observable ϵ_i^{QP}, efficient and reliable approximations to the self-energy have to be found. This can be done with MBPT.

B. Hartree-Fock approximation

Original MBPT proposes a solution of the many-electron problem in terms of powers of the electron-electron interaction, i.e., the Coulomb potential v.[7] Beyond Hartree approximation ($\Sigma=0$), in which electronic interaction is just considered as the interaction of classical charges, the only first-order contribution in v is the Fock self-energy. This self-energy Σ_X is also named the exchange operator, since it accounts for the exchange of the indistinguishable electrons. It reads

$$\Sigma_X(1,2) = iG(1,2)v(1^+,2), \tag{2}$$

where 1 is a shorthand notation for $(\mathbf{r}_1,t_1)$. Here, 1^+ means that the limit of times $t_1+\eta$ with vanishing positive η has to be considered and $v(1^+,2)$ stands for $v(\mathbf{r}_1-\mathbf{r}_2)\delta(t_1+\eta-t_2)$. The δ function ensures instantaneity and will produce a static self-energy.

The definition of G corresponding to a static self-energy is simply of the independent-particle form

$$G(\mathbf{r}_1t_1,\mathbf{r}_2t_2) = -i\sum_i \phi_i(\mathbf{r}_1)\phi_i^*(\mathbf{r}_2)e^{-i\epsilon_i(t_1-t_2)}[\theta(t_1-t_2)\theta(\epsilon_i-\mu) - \theta(t_2-t_1)\theta(\mu-\epsilon_i)], \tag{3}$$

with purely real energies ϵ_i. μ stands for the chemical potential. If this expression is introduced into Eq. (2), it is easily Fourier transformed to frequency space, thanks to the δ function in $v(1^+,2)$

$$\Sigma_X(\mathbf{r}_1,\mathbf{r}_2) = -\sum_i \theta(\mu-\epsilon_i)\phi_i(\mathbf{r}_1)\phi_i^*(\mathbf{r}_2)v(\mathbf{r}_1-\mathbf{r}_2), \tag{4}$$

where θ is a step function (or a smeared distribution for a finite temperature). The time t_1^+ was essential here, as it retains only terms corresponding to occupied states in the Green's function, i.e., below the chemical potential.

Unfortunately, the perturbative approach with respect to the bare Coulomb interaction v breaks down: already some of the second order terms diverge.[7] To circumvent this, the need to change the coupling constant of the perturbative approach has been identified.

C. *W*-based approximations

In this context, Hedin established in 1965[1] a set of equations that solves in principle the many-body problem. The key quantity of these equations in no longer the bare Coulomb interaction v, but the screened Coulomb interaction W

$$W(1,2) = \int d3\varepsilon^{-1}(1,3)v(3,2), \tag{5}$$

where $\varepsilon(1,3)$ is the time-ordered dielectric matrix of the system within the random-phase approximation.[7] The dynamically screened Coulomb interaction is the bare Coulomb interaction renormalized by the screening due to all the electrons. This quantity is expected to be smaller than v in solids and therefore, more suited to be used as a perturbation.

According to Hedin's equations, the self-energy symbolically reads $\Sigma=iGW\Gamma$, where Γ is the vertex function. By retaining only the first order in W, or equivalently "neglecting the vertex correction" in the self-energy, one obtains the so-called GW approximation:[1,18,23]

$$\Sigma_{GW}(1,2) = iG(1,2)W(1^+,2). \tag{6}$$

Beside the static exchange Σ_X of Eq. (4), the GW self-energy has a dynamical part that accounts for correlation effects $\Sigma_C=iGW_p$, with $W_p=W-v$. This correlation part is both non-Hermitian, giving rise to the finite lifetime of QP, and dynamical, producing, e.g., satellite structures in the spectral functions. Furthermore, the conservation laws are satisfied only for self-consistent calculations.[24] In practice, several works[25,26] show that for nonself-consistent calculations the number of electrons indeed fluctuates.

In order to identify the physical contributions in the GW self-energy, Hedin introduced[1,18] approximations to GW. The screened exchange approximation, labeled SEX in the following, can be obtained by assuming straightforwardly that the screening (and hence W) is instantaneous (instead of dynamical). This yields a function $\delta(t_1+\eta-t_2)$ in the self-energy

$$\Sigma_{SEX}(1,2) = iG(1,2)W(1,2)\delta(\tau+\eta), \tag{7}$$

where τ is t_1-t_2. Exactly as for the Fock self-energy, only contributions corresponding to occupied states are retained. The resulting self-energy is simply a statically screened version of the exchange interaction

$$\Sigma_{SEX}(\mathbf{r}_1,\mathbf{r}_2) = -\sum_i \theta(\mu-\epsilon_i)\phi_i(\mathbf{r}_1)\phi_i^*(\mathbf{r}_2)W(\mathbf{r}_1,\mathbf{r}_2,\omega=0), \tag{8}$$

where the Fock term can be recognized except that the bare Coulomb interaction is replaced by the statically screened one. It justifies the approximated screened exchangelike term commonly used in the generalized Kohn-Sham scheme.[19–21]

The assumption of an instantaneous screening is however drastic; the solid line of Fig. 1 shows that W_p is far from being instantaneous. The figure describes the response of the charge density to an external δ perturbation at time $\tau=0$. As the screening is time ordered, $\mathrm{Re}\{W_p\}$ is an even function of time. The dynamical part of the screening W_p behaves like a damped oscillator, its frequency being close to the classical plasma frequency, whereas SEX supposes it to be a δ function.

SEX has another major drawback: it erases the distinction between time ordered and advanced screening. In SEX, the term $\delta(\tau+\eta)$ retains only the value of W_p at time $\tau=0^-$, whereas the value of W_p at time $\tau=0^+$ is disregarded. This obviously breaks the parity in time of W_p and simulates more likely the advanced screening (full circles in Fig. 1), than the time-ordered one (solid line in Fig. 1).

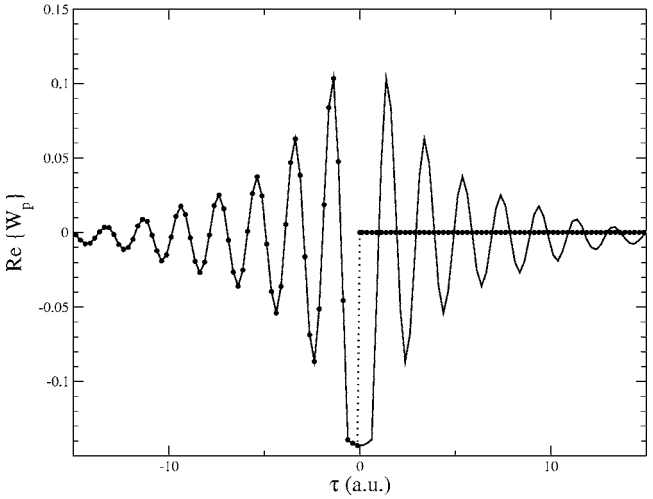

FIG. 1. Aluminum: real part of W_p as a function of time. The solid line represents the time-ordered W_p and the full circles the advanced W_p. The component for $\mathbf{G}=\mathbf{G}'=0$ and $q=(1/4,0,0)$ is drawn. The vertical part of the full circle curve is a guide to the eyes.

A simple way to improve over the SEX approximation and recover the correct symmetry in W_p is to slightly shift the δ function to $\delta(\tau)$ (i.e., neglect the small η). This still static approximation yields

$$\Sigma_{\mathrm{C}}(1,2) = iG(1,2)W_p(1,2)\delta(\tau). \tag{9}$$

Using the algebraic relation $\delta(\tau)\theta(\tau)=1/2\delta(\tau)$ and a Fourier transform, the correlation part Σ_{C} becomes

$$\Sigma_{\mathrm{C}}(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{2}\sum_i \phi_i(\mathbf{r}_1)\phi_i^*(\mathbf{r}_2)W_p(\mathbf{r}_1,\mathbf{r}_2,\omega=0) \times[\theta(\epsilon_i-\mu)-\theta(\mu-\epsilon_i)]. \tag{10}$$

The completeness relation $\Sigma_i\phi_i(\mathbf{r}_1)\phi_i^*(\mathbf{r}_2)=\delta(\mathbf{r}_1-\mathbf{r}_2)$ allows one to recast the self-energy into two terms: a screened exchange term [given by Eq. (8)], and a so-called Coulomb hole term Σ_{COH}

$$\Sigma_{\mathrm{COH}}(\mathbf{r}_1,\mathbf{r}_2) = \tfrac{1}{2}\delta(\mathbf{r}_1-\mathbf{r}_2)W_p(\mathbf{r}_1,\mathbf{r}_2,\omega=0). \tag{11}$$

This is the COHSEX approximation[1,18] of the *GW* self-energy. This approximation has an additional term with respect to the SEX term, which accounts for the classical attraction of the electrons of the system upon removal of a point charge electron or, reversely, for the classical repulsion of the electrons of the system upon addition of a point charge electron.

Both SEX and COHSEX, though approximate, already contain much physics. Their evaluation is much simpler than the *GW* self-energy, since they do not involve explicitly the empty states, as opposed to *GW*. The corresponding self-energies are static and Hermitian. These are great advantages when performing intensive self-consistent calculations.

D. Standard *GW* calculations

The evaluation of the *GW* self-energy, and hence of the *GW* band structure, is complex. From the first jellium[1] and empirical linear combination of atomic orbitals[27] calculations to the accurate *ab initio* calculations,[8,9] numerous additional technical approximation have been used.

First of all, self-consistency topic has been disregarded. Whereas the work on jellium was based on the free-electron Green's function and the Lindhard polarizability in *W*, most state-of-the-art *GW* calculations employ the more realistic KS Green's function G_0 and the corresponding screening for W_0. This is the LDA+G_0W_0. A few works chose to start from Hartree-Fock[28] or from the exact-exchange KS scheme.[29]

Second, the KS wave functions may be thought as good approximations to the QP ones. Indeed, the KS equations

$$\left[-\frac{\nabla^2}{2}+\int d\mathbf{r}'\rho(\mathbf{r}')v_{\mathrm{ext}}(\mathbf{r})+\int d\mathbf{r}'\rho(\mathbf{r}')v(\mathbf{r}-\mathbf{r}') + v_{xc}[\rho](\mathbf{r})\right]\phi_i^{\mathrm{KS}}(\mathbf{r}) = \epsilon_i^{\mathrm{KS}}\phi_i^{\mathrm{KS}}(\mathbf{r}), \tag{12}$$

are identical to the equations of motion of the QP wave functions [Eq. (1)], except that the exchange-correlation potential of DFT replaces the self-energy and except that hence orthogonal KS wave functions are involved (instead of the in general nonorthogonal QP wave functions). For system of lower dimension, it is known that KS and QP wave functions are not always close.[30,31] For bulk systems, their overlap has been checked by the earliest *ab initio* works.[10]

Considering the similarities between Eqs. (1) and (12), a perturbative approach is reasonable. The quantity $(\Sigma-v_{xc})$ is to be regarded as small with respect to the entire Hamiltonian. The first-order perturbative approach is equivalent to the assumption that KS and QP wave functions are equal. Then, the difference between Eqs. (1) and (12), projected on KS wave functions ϕ_i^{KS}, is merely

$$\epsilon_i^{GW}-\epsilon_i^{\mathrm{KS}} = \langle\phi_i^{\mathrm{KS}}|\Sigma^{GW}(\epsilon_i^{GW})-v_{xc}[\rho]|\phi_i^{\mathrm{KS}}\rangle. \tag{13}$$

This is an evaluation of the *GW* energies from KS ones, via the calculation of diagonal expectation values only. Equation (13) for the *GW* QP energy is the main equation of most perturbative implementations.

Following the original work of Hybertsen and Louie,[10] a single plasmon-pole model can be further used to model the frequency behavior of *W*, leading to an analytic expression for the frequency behavior of the *GW* self-energy. In time domain, a single plasmon pole leads to a sinusoidal screening with no damping. According to Fig. 1, one can see that this approximation is rather realistic.

Moreover, Eq. (13) is generally linearized in the vicinity of ϵ_i^{KS}

$$\epsilon_i^{GW}-\epsilon_i^{\mathrm{KS}} = Z\langle\phi_i^{\mathrm{KS}}|\Sigma^{GW}(\epsilon_i^{\mathrm{KS}})-v_{xc}[\rho]|\phi_i^{\mathrm{KS}}\rangle, \tag{14}$$

where the renormalization factor $Z=1/(1-\partial\Sigma/\partial\epsilon)$ accounts for the dynamical behavior of Σ.

All these rules and assumptions form the most common implementation of the *GW* self-energy calculations (referred to as G_0W_0).

TABLE II. Parameters used in practice for the calculation of LDA ground state, of the screened Coulomb interaction W, and of the self-energy Σ. PW stands for the number of plane waves and E_{cut} for the plane-wave cutoff energy. The unitary transform matrices $c_{\mathbf{k}ij}$ are defined in the text.

	Si	Ar	Al
Lattice parameter (a.u.)	10.263	9.932	7.652
E_{cut} for LDA (Ha)	12	20	8
Electronic Temp. (Ha)	—	—	0.05
Bands used for W	35	30	50
PW for $\phi(\mathbf{G})$ for W	169	169	51
PW for $W_{\mathbf{GG}'}$	169	169	51
corresponding to E_{cut} (Ha)	5.1	5.4	3.7
Bands used for G in Σ	100	100	50
PW for $\phi(\mathbf{G})$ for Σ	169	459	89
PW for $\Sigma_{X\mathbf{GG}'}$	169	459	89
corresponding to E_{cut} (Ha)	5.1	11.2	6.4
Frequencies along imaginary axis	4	8	6
Frequencies along real axis	20	20	20
Size of matrices $c_{\mathbf{k}ij}$	30	30	30

E. Self-consistent self-energies

There are two main paths that might improve the standard LDA+G_0W_0 results: either include some vertex corrections or perform calculations self-consistently. The effect of first-order vertex for the band gap of simple materials being small,[32,33] this work turns to self-consistent calculations. The complexity of the GW self-energy and corresponding Green's function is however such that it is hardly feasible nowadays to implement full self-consistency in practice. Furthermore, fully self-consistent GW has been ruled out for the spectral properties of jellium,[11] because the QP part in the Green's functions is then underestimated. Instead, the spatial behavior of the Green's function needs to be improved.

Therefore we choose to turn to static approximations of the GW self-energy that avoid the dynamical trap, but still have the correct spatial behavior. The Green's function is then fully represented by a sum of single QP excitations [Eq. (3)]. This approximation to the full Green's function is both simpler and safer.

SEX and COHSEX approximations seem appropriate in this context. They are static and further yield orthogonal QP wave functions, since their self-energies are Hermitian. We perform self-consistent SEX and COHSEX and, if meaningful, use these as an input for "one-shot" GW calculations.

Trying to find static approximations that better stick to the GW self-energy leads us to the interesting attempt of Faleev and co-workers.[16] These authors constrained the dynamical GW self-energy, such that it becomes static and Hermitian, but still remains as close as possible to the original self-energy. Their model QPscGW self-energy reads

$$\langle \phi_i|\Sigma|\phi_j\rangle = \tfrac{1}{2}\Re[\langle \phi_i|\Sigma(\epsilon_i)|\phi_j\rangle + \langle \phi_i|\Sigma(\epsilon_j)|\phi_j\rangle], \quad (15)$$

where $\Re$ means that one retains only the Hermitian part of the matrix $\Re[\Sigma]=\frac{1}{2}(\Sigma+{}^t\Sigma^*)$. Along with self-consistency, the diagonal elements of the self-energy are better and better approximations to the true GW diagonal terms, as each of them is finally evaluated for the correct GW energy. Only the off-diagonal elements are modeled. This approximation to the self-energy has the same technical advantages as the previously proposed SEX, COHSEX self-energies, but has of course a much higher computational cost.

In practice, the implementation is in any case much simplified within the framework of static and Hermitian self-energies, since the number of particles is straightforwardly conserved, in contrast to the original GW self-energy.[25] Moreover, the wave functions are all eigenvectors of the *same* matrix, instead of a matrix for each energy in the real GW case.

III. TECHNICAL DETAILS AND NOTATIONS

Calculations are based on the plane wave plus pseudopotential method.[34] The valence configuration is $3s^23p^2$ for silicon, $3s^23p^6$ for argon, and $3s^23p^1$ for aluminum. The influence of the LDA interaction contained implicitly in the pseudopotential is expected to be weak, since the spatial overlap between the core and the valence is very small for these materials. This has been explicitly demonstrated for some semiconductors in Ref. 35. These three solids crystallize in a face-centered cubic Bravais lattice. Calculations are performed at the experimental lattice parameters, specified in Table II. For the three solids, the **k** point grid is the same: a regular grid centered in Γ containing 256 **k** points in the full Brillouin zone that corresponds to 19 **k** points in the irreducible wedge of the Brillouin zone. The other parameters are summarized in Table II.

In the implementation of self-consistent QP approaches, we recalculated with the updated wave functions the new kinetic, pseudopotential, and Hartree contributions to the QP

Hamiltonian matrix for each iteration. The lowest eigenvectors and eigenvalues are updated, whereas the highest are kept fixed to the original LDA ones. We have checked that the influence of the non-updated eigenvalues on the band gap region is extremely weak. It makes therefore no difference to keep them constant or to shift them up using a scissor operator. The limit between the updated and fixed wave functions and energies is governed by the size of the unitary transform matrices displayed in the last row of Table II. The role of this parameter and its convergence will be made clearer in the following section. Self-consistency is stopped, when the eigenvalues are stable within 0.01 eV. Generally, from 6 to 10 iterations are needed in practice.

For the *GW* self-energy, it has been necessary to calculate explicitly the frequency convolution of G and W, without the use of a plasmon-pole model. We have used the convolution following the contour integral method devised in Ref. 17 and implemented with slight improvements in Ref. 36. The plasmon-pole model has been found unsatisfactory for quasiparticle energies and wave functions far from the Fermi level.

When plotting wave functions, we choose to present wave functions that maximize the changes between the different approximations. We found the maximum differences for points that have a large weight: these are points of the irreducible Brillouin zone that yield the greatest number of equivalent points in the full Brillouin zone. According to this statement, we will focus our comparisons for the wave functions on the **k** point $\mathbf{k}=(-1/8,-3/8,1/4)$, which is one of the most representative points available in our **k** point set.

IV. ASSESSMENT OF THE METHODS RETAINED AFTER ONE SINGLE ITERATION

A. Limited number of LDA states as a basis set

The self-consistency requires the calculation and the storage of the updated wave functions, different from the LDA ones. In order to keep calculations tractable, we have to find an efficient and reliable way to represent the QP wave functions. We choose to expand them in the basis set of LDA wave functions

$$|\phi_{\mathbf{k}i}^{\mathrm{QP}}\rangle = \sum_j c_{\mathbf{k}ij} |\phi_{\mathbf{k}j}^{\mathrm{LDA}}\rangle. \tag{16}$$

The expansion coefficients $c_{\mathbf{k}ij}$ form unitary matrices, ${}^t c_{\mathbf{k}}^* \cdot c_{\mathbf{k}} = 1$, since they transform the orthogonal LDA wave functions into the orthogonal QP ones. As LDA eigenvectors are furthermore a complete set, this procedure is exact as long as *all* the LDA states are employed. The practical interest of our method is precisely that the number of basis functions can be reduced drastically, while the description of QP wave functions remains very good.

To check the accuracy of a limited LDA wave functions basis set, we examine the representation of a HF conduction wave function of solid argon after a single iteration, because this wave function widely differs from its LDA counterpart. We perform the calculation either in our limited LDA basis set or in the plane wave one. Of course, the use of plane waves is not efficient from the computational point of view, because we had to evaluate the quantity

$$\langle \mathbf{k}+\mathbf{G}|\Sigma_{\mathrm{X}}|\mathbf{k}+\mathbf{G}'\rangle = -\frac{4\pi}{V}\sum_{\mathbf{k}_1\mathbf{G}_1}\sum_{i\ \mathrm{occ}} \frac{\phi_{\mathbf{k}_1 i}(\mathbf{G}_1-\mathbf{G})\phi_{\mathbf{k}_1 i}^*(\mathbf{G}_1-\mathbf{G}')}{|\mathbf{k}_1-\mathbf{k}+\mathbf{G}_1|^2}, \tag{17}$$

where $\phi_{\mathbf{k}i}(\mathbf{G})$ stands for the Fourier transform of $\phi_{\mathbf{k}i}(\mathbf{r})$. In the present section, HF results are single-shot ones using LDA wave functions to construct the Fock operator. This is harmless for our purposes here, since the first iteration is always by far the most significant.

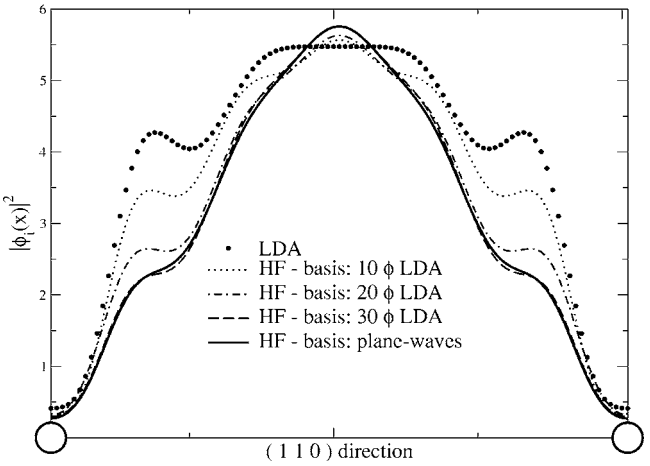

FIG. 2. Second conduction band of solid argon: squared modulus of wave functions along the direction (1 1 0) at $\mathbf{k}=(-1/8, -3/8, 1/4)$. The white circles represent the location of the argon atoms.

Figure 2 represents a cut of the squared modulus of the wave functions corresponding to the second conduction band of argon, within LDA (full circles) and HF (lines), for the **k** point chosen in Sec. III. The HF wave function is represented either in the reference plane-wave basis set (solid line) or in a LDA basis set limited to the 10 first states (dotted line), 20 first states (dot-dashed line), or 30 first states (dashed line). The LDA and the "exact" HF wave function differ noticeably (dot product equal to 0.992). The dotted, dot dashed, and dashed curves of Fig. 2 show how the flexibility of the LDA basis set is improved by increasing the number of basis functions. With as few as 30 basis function, the wave function looks already almost indistinguishable from the unbiased plane-wave result.

In order to confirm these qualitative statements, it is necessary to quantify the impact of the basis set onto the interesting observable, here the band structure. In Table III, we provide the energy deviation with respect to the "exact" plane-wave result for the limited basis sets for the same **k** point as previously. The magnitude of the effect of changes in the wave functions is relatively small. For instance, the error made by using one single LDA basis function (first column) is equivalent to the error of a first-order perturbation approach, where the bra and ket of expectation values are simply taken as LDA wave functions. The maximum effect ($\sim$100 meV) occurs for conduction bands. (It increases for

TABLE III. HF band structure of solid argon: deviation for the eigenvalues (meV) at $\mathbf{k}=(-1/8,-3/8,1/4)$ calculated in the limited basis set from the "exact" HF energies obtained in the plane-wave basis set.

Size of the basis set	1	10	20	30	100
Top VB	6	10	0.7	0.5	0.4
1st CB	106	119	24	12	0.6
2nd CB	112	116	50	14	0.8

higher conduction bands, but this has weak effects for the present work that concentrates on the band gap region.) The convergence as a function of the number of basis functions is as fast for the energies as for the wave functions (Fig. 2). 100 basis functions give an almost perfect agreement and 30 functions allow one to obtain energy within a 10 meV accuracy. For all applications in the present work (silicon, argon, aluminum), a basis set of 30 LDA wave functions will be employed (see last line of Table II). For silicon and aluminum, the convergence is even faster than for argon.

Let us finally illustrate the gain of computational time for solid argon. For this material, 459 plane waves are necessary to describe the pseudowave functions. Therefore, the nonlocal self-energies require 459×459 matrix elements given by Eq. (17). By using 30 LDA wave functions as a basis set, only 30×30 matrix elements

$$\langle\phi_{\mathbf{k}j}|\Sigma_{\mathrm{X}}|\phi_{\mathbf{k}l}\rangle=-\frac{4\pi}{V}\sum_{\mathbf{k}_1\mathbf{G}_1}\sum_{i\ \mathrm{occ.}}\frac{\tilde{\rho}^{*}_{\mathbf{k}_1 i\mathbf{k}j}(\mathbf{G}_1)\tilde{\rho}_{\mathbf{k}_1 i\mathbf{k}l}(\mathbf{G}_1)}{|\mathbf{k}_1-\mathbf{k}+\mathbf{G}_1|^2} \quad (18)$$

will be computed, with a rather similar accuracy. The terms $\tilde{\rho}_{\mathbf{k}_1 i\mathbf{k}j}(\mathbf{G})$ are defined and obtained as (fast) Fourier transforms of products $\phi_{\mathbf{k}_1 i}(\mathbf{r})\phi_{\mathbf{k}j}(\mathbf{r})$. This procedure allows us to perform from now on, fully self-consistent calculations using nonlocal, but still static self-energies.

B. Quality of the model self-energy

The model in Eq. (15) for the QPscGW self-energy introduced in Ref. 16 will be considered as the most achieved calculation in the following. Though reasonable, this model remains an approximation. We propose to establish its validity by considering the QP wave functions obtained from the model and from the original GW self-energy evaluated at the QP energy, after one single iteration. This comparison is already meaningful because, again, the first iteration carries the main part of the changes in the wave functions. We postpone the self-consistent results to the next sections.

Figure 3 shows the dependence of the QP wave functions on the energy for which Σ_{GW} is evaluated, for (a) the top valence band and (b) the second conduction band of solid argon at the $\mathbf{k}$ point specified in Sec. III. The original GW self-energy is evaluated for the energy of the first valence band (energy ϵ_1, dotted line), for the energy of a very high conduction band (energy ϵ_{12}, dashed line), and for the chemical potential (energy μ, gray line). When evaluated for the QP energy [ϵ_4 in (a) or ϵ_6 in (b)], the corresponding GW wave function is represented with the open circles. The wave

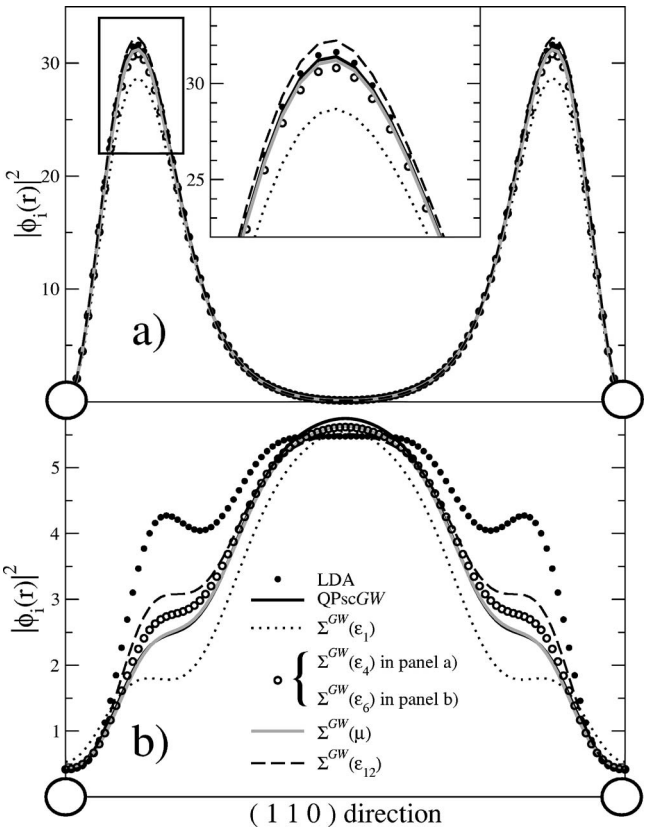

FIG. 3. Solid argon: squared modulus of the wave functions along the direction (1 1 0) at $\mathbf{k}=(-1/8,-3/8,1/4)$ for (a) the top valence band, (b) the second conduction band. The white circles represent the location of the argon atom. The inset is a close-up of the region around the maximum.

function corresponding to the model QPscGW self-energy of Eq. (15) is given by the solid line and LDA wave functions by the full circles.

First, the QP wave functions differ from the LDA ones, slightly for the valence band, largely for the conduction band. For both conduction and valence, the dependence on the energy is large: the wave functions arising from self-energies evaluated for ϵ_1 or for ϵ_{12} are drastically different. According to Fig. 3, the model QPscGW self-energy gives results very close to the self-energy evaluated for the chemical potential. Both $\Sigma(\mu)$ and the model yield wave functions that are rather good approximations to the original GW QP wave functions (given by the open circles). The approximation to evaluate the self-energy at the chemical potential only and the model QPscGW have both been suggested in Ref. 16. The difference between the model QPscGW and the original GW wave functions remains visible, especially for the conduction band, but it is much smaller than the deviation of LDA with respect to QPscGW. The task of finding wave functions, which are all eigenvectors of one single matrix and which are equal to the true GW wave functions, is of course impossible. However, the model of Eq. (15) is a good attempt towards this goal. Let us exemplify the dilemma with one matrix element involved in the calculation of both wave functions of Fig. 3: $\langle\phi_4^{\mathrm{LDA}}|\Sigma(\epsilon)|\phi_6^{\mathrm{LDA}}\rangle$. When $\epsilon=\epsilon_4$ or $\epsilon=\epsilon_6$, its value varies from -1.24 to -1.38 eV. The first

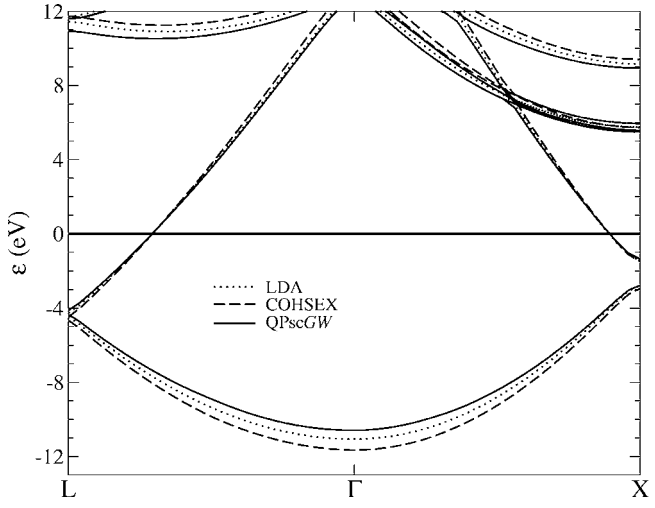

FIG. 4. Aluminum: band structure within LDA (dotted line), self-consistent COHSEX (dashed line), and QPscGW (solid line).

value is used to calculate ϕ_4, whereas the second one is employed for ϕ_6. On the contrary, the model QPscGW self-energy uses the same off-diagonal element (the mean value) in both cases. That is why the model QPscGW self-energy cannot give the exact GW answer, as proved by Fig. 3, but sticks to it rather efficiently.

In the following, the model QPscGW self-energy plays the role of the reference result in order to evaluate the quality of the other simpler approximations. The QPscGW being an approximation, the difference between the wave functions stemming from the original GW and from the model QPscGW self-energies is to be considered as the error bar of the present study of QP wave functions. From now on, all the QP wave functions under consideration are orthogonal, thanks to the approximations made (HF, SEX, COH, or QPscGW).

V. SELF-CONSISTENT BAND STRUCTURES OF SIMPLE MATERIALS

A. Metal: aluminum

It is first interesting to test the approximations for the case of a simple metal, which has a rather homogeneous electronic density. This is a typical case where LDA is supposed to perform best, since the system is as close as possible to the jellium model. The LDA occupied band width is 11.06 eV, whereas photoemission experiments[37] measure 10.6 eV. The LDA result lies very close to the experimental result, even though rigorously KS band structures are not meant to be interpreted.

Quasiparticle techniques instead are dedicated to provide meaningful band structures. However, the HF band structure is unrealistic for aluminum, because for metals, the role of the screening is prominent. On the contrary, the self-consistent SEX and COHSEX approaches provide interesting results. In Fig. 4, the COHSEX band structure (dashed line) is compared to the LDA one (dotted line). SEX results are not represented here, since they are almost identical to COHSEX. The present results are important, as the common knowledge about COHSEX approximation is that it overestimates the band gap between occupied and empty states of semiconductors.[10] In the case of a metal, one could have feared a unrealistic behavior of the COHSEX self-energy in the vicinity of the chemical potential. In fact, SEX and COHSEX band structures of aluminum are meaningful.

The resulting occupied band width is 11.69 eV within SEX, and 11.72 eV within COHSEX. The similarity between SEX and COHSEX comes from the fact that Σ_{COH} is essentially $\mathbf{k}$ point independent for homogeneous systems. Though large $\langle\phi_i|\Sigma_{\mathrm{COH}}|\phi_i\rangle \sim -9.4$ eV in absolute, this part of the self-energy accounts merely for a rigid shift of the whole band structure.

The QPscGW band structure, the solid line in Fig. 4, ends up with a band width of 10.58 eV, which lies impressively close to experimental value. For this material, a calculation with a plasmon-pole model would be rather crude when considering the bottom of the valence bands, of which the energy is close to the classical plasmon energy ($\sim$15.8 eV). As a comparison, the calculation with plasmon-pole model gives a band width of 9.75 eV, which is compatible with the plasmon-pole calculations of Ref. 38.

The self-consistency is not crucial in the calculation without plasmon-pole model: the LDA$+G_0W_0$ band width is 10.54 eV. Following Ref. 38, it is important to calculate self-consistently the occupations of the states used for the Green's function G for metals, when the chemical potential gets displaced along with iterations. However, the QPscGW modified the absolute value of the chemical potential by only 0.10 eV with respect to LDA. Furthermore, self-consistency in W yields very weak contributions: using the screening of LDA or of QPscGW in W changes the band width by 0.04 eV. This is due to the quality of the LDA initial guess: the W_0 calculated from LDA is impressively close to the QPscGW screening.

In conclusion, LDA, SEX, and COHSEX slightly overestimate the band width of aluminum. A COHSEX$+G_0W_0$ approach would yield results really close to the QPscGW ones, if the GW energies are updated in G, because, as explained in the next sections, all the wave functions (QP or KS) are almost the same. The impressive success of the GW approximation is not surprising: it is consistent with the very good results that have been obtained in the first application of GW to the homogeneous electron gas.[1] This system is in fact a good model for aluminum. This is also in agreement with the QPscGW results obtained for sodium, another simple metal.[39]

B. Semiconductor: silicon

Bulk silicon is the test case of all band structure calculations. Except a deviation of 0.2 eV on the band gap between all-electron[17] and pseudopotential[8] based methods, the band structure seems to be accurately described by the standard perturbative method LDA$+G_0W_0$.

The primary purpose of the present paragraph is to prove that self-consistency in a pseudopotential based model GW calculation does not destroy the former agreement between

TABLE IV. Band structure of bulk silicon at high-symmetry points (eV).

	This work				LMTO (Ref. 16)	
	LDA	LDA+G_0W_0	QPscGW	COHSEX+G_0W_0	QPscGW	Expt. (Ref. 40)
Γ_{1v}	−11.9	−11.4	−11.9	−11.6	−12.3	−12.5±0.6
Γ'_{25v}	0.00	0.00	0.00	0.00	0.00	
Γ_{15c}	2.57	3.20	3.54	3.69	3.40	3.05 (Ref. 41), 3.40
X_{1c}	0.65	1.29	1.60	1.68	1.28	1.32
L_{1c}	1.46	2.08	2.41	2.56	2.24	2.04
E_g	0.51	1.14	1.47	1.56	1.14	1.17

LDA+G_0W_0 and experiment, as opposed to the alarmist results of Refs. 13 and 42. According to Table IV, our LDA+G_0W_0 band structure is in very good agreement with experiment. When self-consistency is achieved, the band gap gets overestimated by 0.3 eV and the direct band gap at Γ by 0.13 eV. Such an overestimation is not calamitous and much better than the results announced in Refs. 13 and 42. In fact, going from LDA+G_0W_0 to QPscGW opens the band gap by 0.33 eV. A similar value (~0.24 eV) is found in all-electron linear muffin-tin orbital (LMTO) based method.[16,43] The statement is that in general (for perturbative LDA+G_0W_0 or for self-consistent QPscGW), there is a constant discrepancy between all-electron methods and pseudopotential based approaches. There is no pseudopotential catastrophe when achieving self-consistency. The poor results of Ref. 13 might be due to the fact that the satellites of the Green's function are included in the self-consistency. As pointed out in Ref. 11, the self-consistent QPscGW approximation is not reliable to determine the satellites of the spectral function. Additional vertex corrections have been shown to be necessary for a consistent treatment of the latter.[44] Therefore, it makes sense to avoid the calculation of the satellites within GW for example by using the static model for the self-energy of Eq. (15).

The second purpose is to check the quality of the COHSEX self-energy to determine inputs (wave functions, energies) for a subsequent perturbative GW calculation. The corresponding results are labeled "COHSEX+G_0W_0" in Table IV. The band gap gets overestimated by 0.39 eV and the direct one by 0.29 eV. This method provides hence reasonable results compared to self-consistent GW ones. For a computational price that is one order of magnitude lower. However, for silicon, starting the GW step from LDA turns out to yield the better result.

C. Insulator: argon

Solid argon is a rare-gas solid with a huge band gap [14.2 eV (Ref. 22)], made of very weakly interacting atoms. This is an interesting test case, since it is one of the few simple materials for which the standard LDA+G_0W_0 fails significantly. In fact, it is known that the performance of the GW approximation for atoms is poor and may require additional vertex contributions.[46] Originally the GW approximation was designed for rather homogeneous and polarizable media.

According to Table V, the LDA+G_0W_0 band gap underestimates by 1.2 eV the experimental result. The discrepancy is larger than the one calculated in Ref. 47, but that author used a partially self-consistent approach (on energies in G).[48] However, looking at the LDA band gap (8.20 eV), it is not surprising that, for example, the dielectric constant of solid argon is strongly overestimated in W_0. Therefore, solid argon is a typical case for which self-consistency is expected to yield large effects. Using the self-consistent GW approach, the band structure ends in good agreement with respect to the available experimental data (a deviation lower than 4%). Furthermore, the small occupied band width (Γ_{15v}-L'_{2v}) is improved in the self-consistent GW calculation with respect to the standard LDA+G_0W_0 one.

Table V also proves the very good performance of the cheap COHSEX+G_0W_0 method. It agrees almost perfectly with the self-consistent GW calculation for a much lower computational time.

As far as band structures are concerned, we can now draw general conclusions for the reliability of the approximations used. For the metal, the best results have been obtained from self-consistent GW and LDA+G_0W_0; for the semiconductor, standard LDA+G_0W_0 has performed the best; for the insulator, self-consistent GW and COHSEX+G_0W_0 produced the

TABLE V. Band structure of solid argon at high-symmetry points (eV).

	LDA	LDA+G_0W_0	QPscGW	COHSEX+G_0W_0	Expt.
Γ_{15v}	0.00	0.00	0.00	0.00	
Γ_{1c}	8.20	12.98	14.84	14.72	14.2 (Ref. 22)
X'_{5v}	−0.46	−0.52	−0.53	−0.54	
L'_{3v}	−0.15	−0.18	−0.18	−0.15	
L'_{2v}	−1.41	−1.47	−1.57	−1.59	−1.70 (Ref. 45)

nicest results. The poorer the LDA starting point, the stronger the need for self-consistent QP schemes. Furthermore, we stress that the self-consistent QP schemes have yielded rather good results for all materials, whereas no other method has been found as consistently reliable.

VI. QUASIPARTICLE WAVE FUNCTIONS

This section is dedicated to the detailed study of the impact of changing the LDA wave functions to the QP ones. This part of the analysis is to our knowledge missing in the literature.

A. Are LDA wave functions similar to *GW* wave functions?

The standard LDA+G_0W_0 calculations are based on the assumption that LDA and *GW* wave functions are similar. In the 1980's, the similarity between LDA and *GW* wave functions was checked.[10,49] In Ref. 10, Hybertsen and Louie found an overlap of 99.9% between LDA and *GW* wave functions. In our opinion, it is worthwhile to open this issue again in the light of our self-consistent *GW* results.

Citing Ref. 10, the mentioned value for the overlap between LDA and *GW* wave functions was obtained "for several states at symmetry points in the Brillouin zone." This methodology is fully justified when calculating LDA $+G_0W_0$ band structures. Let us illustrate this fact with an example: to calculate the LDA+G_0W_0 band structure at Γ point by means of a first-order perturbative method, only the value of $|\langle \phi_{\Gamma i}^{\mathrm{LDA}} | \phi_{\Gamma i}^{\mathrm{QPsc}GW} \rangle|$ matters. However, when turning to self-consistent *GW*, the overlaps for all the other **k** points $|\langle \phi_{\mathbf{k}i}^{\mathrm{LDA}} | \phi_{\mathbf{k}i}^{\mathrm{QPsc}GW} \rangle|$ have an influence on the band structure at Γ via the Hartree potential (through the density) and the self-energy (through the wave functions). Therefore, we have to extend the study of the overlaps to all **k** points, not only to high-symmetry points.

In complete agreement with the early *GW* results,[10] Fig. 5 shows that the overlap of the LDA and QPsc*GW* wave functions at Γ point (full circles) is very large for both silicon [panel (a)] and argon [panel (b)]. For all the valence bands and the first conduction states, the overlap at Γ is indeed greater than 99.9%. On the contrary, for the other **k** points used in the calculations that may not be high-symmetry points, Fig. 5 proves that the overlap may be much smaller. Our study adds an important contribution with respect to the original *GW* works: even for simple materials, the LDA and *GW* wave functions may differ noticeably. The similarity of the different wave functions for high-symmetry points is forced by the symmetry constrains. For instance, let us consider the worst case by comparing HF and the LDA wave functions of silicon at Γ. One may expect that HF and LDA wave functions differ strongly. However, the overlap between LDA and HF is higher than 99.97% for the valence states and higher than 99.89% for the first conduction states. In fact, if one considers the example of the top valence bands with symmetry Γ'_{25}, the next states having the same symmetry (i.e., that can yield nonvanishing off-diagonal expectation values) are the eighth to the tenth conduction bands located at about 11 eV above the band gap. Looking solely at the overlaps at the Γ point is therefore misleading when performing self-consistent calculations.

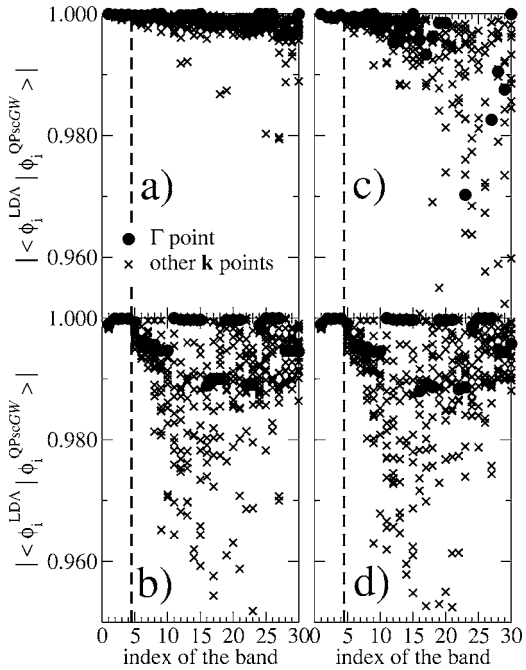

FIG. 5. Overlap between LDA and QPsc*GW* wave functions for (a) silicon and (b) argon as a function of the index of the band without the plasmon-pole model approximation. Panels (c) and (d) display the corresponding quantities using the approximation of the plasmon-pole model. The dashed line separates occupied and empty states.

Here, we want to insist on the role of the calculations without plasmon-pole model for the overlap between LDA and QPsc*GW* wave functions. Using a plasmon-pole model in silicon in Fig. 5 panel (c) yields overlaps in the conduction bands much smaller than the ones depicted in Fig. 5 panel (a). For states in the vicinity of the Fermi level plus the classical plasmon energy (states number 10 to 30), the wave functions are very sensitive to the description of the correlation through the self-energy with or without plasmon-pole model. On the contrary, concerning solid argon in Fig. 5 panel (d), the use of the plasmon-pole model is harmless, since the main effect in the wave functions is carried by the bare exchange operator in this atomiclike system.

We showed above that the LDA and QPsc*GW* wave functions are indeed different. It is now interesting to understand how influential are the changes in the wave functions for the evaluation of the band structures. Aluminum results are not shown here and in the following, because all the wave functions (LDA, SEX, COHSEX, and QPsc*GW*) agree almost perfectly.

B. Effect of the quasiparticle wave functions onto band structures

The self-consistent QP schemes examined in this work update both the wave functions and the energies used in the Green's functions from Eq. (3). The results for the band structures in Sec. V showed the significance of self-

TABLE VI. Argon: Differences in eV in the diagonal expectation values of operators at the Γ point when achieving self-consistency on energies and wave functions minus self-consistency on energies only for the top valence band (Γ_{15v}), for the bottom conduction band (Γ_{1c}), and for the second conduction at the point $\mathbf{k}=(-1/8,-3/8,1/4)$.

Band Approx.	Γ_{15v}		Γ_{1c}		Second conduction band	
	COHSEX	QPsc*GW*	COHSEX	QPsc*GW*	COHSEX	QPsc*GW*
ΔH_H	0.42	0.03	−0.19	−0.13	−0.38	−0.41
$\Delta\Sigma_X$	−0.16	−0.01	0.31	0.21	0.42	0.40
$\Delta\Sigma_C$	−0.01	−0.07	0.07	0.05	0.11	0.06
$\Delta\epsilon^{\mathrm{QP}}$	0.25	−0.05	0.19	0.13	0.15	0.05

consistency. Are the changes in the band structures due to the difference in the wave functions (exhibited in the previous section) or do they come simply from the difference in the energies?

Table VI allows us to appreciate the contribution of the wave functions in the band structure changes of argon at the Γ point and for the **k** point specified in Sec. III. The results for silicon are qualitatively similar, but quantitatively smaller. We therefore concentrate on solid argon in order to make the discussion more significant. The Table VI provides the differences between full self-consistency minus self-consistency on energies, i.e., conserving LDA wave functions, for all the parts of the QP Hamiltonian. The QP Hamiltonian consists of the Hartree Hamiltonian H_H, the Fock operator Σ_X, and of the correlation part of the self-energy Σ_C. The sum of the changes in these parts (ΔH_H, $\Delta\Sigma_X$, and $\Delta\Sigma_C$) gives the change in the QP energy $\Delta\epsilon^{\mathrm{QP}}$.

For the high-symmetry point Γ, the effects of the change in the wave functions can be evidenced, but it is small compared to the band gap of argon (14.2 eV). However, for the nonsymmetric point, the effects are made more obvious. Going from LDA to QP wave functions has a weak effect on the correlation part of the self-energy. On the contrary, the wave functions influence noticeably the Hartree Hamiltonian and the bare exchange, but in an opposite manner. By changing the wave functions, the expectation of the Hartree Hamiltonian and the bare exchange are modified in such a way that the sum of the changes remains small. This can be understood since H_H and Σ_X have terms that appear with opposite signs, in particular the self-interaction term that exactly cancels.

We now turn to the comparison between COHSEX and QPsc*GW* wave functions. As shown in Table VI, the changes induced by the COHSEX and QPsc*GW* wave functions are always in the same direction. In this sense, COHSEX wave functions can be considered as a realistic approximation to the QPsc*GW* wave functions. COHSEX wave functions are particularly suited to describe the change in the conduction bands. The major drawback of COHSEX wave functions is that they overestimate the changes with respect to LDA wave functions for the valence bands.

In conclusion, the main changes in the band structure that occur along with self-consistency come indeed from the update of the energies. The effect of the changes in the wave functions remains rather limited and is subject to cancellations. However, a change for a **k** point has consequences on other **k** points, since all wave functions contribute to the Hartree potential, exchange, and correlations operators.

C. Quality of the electronic densities and density matrices

In standard LDA+G_0W_0, the ground-state density is approximated by the LDA one and consequently, the Hartree potential remains the same within QPsc*GW* and LDA. The self-consistent QP methods studied here give access to updated ground-state densities. This provides fundamental information on the quality of the densities within the different approximations.

The statements drawn from Fig. 6 for silicon and from Fig. 7 for argon lead to the same conclusions. (Once more aluminum results are not displayed, since all densities are superimposed on one another for this material.) The LDA density (full circles) and the QPsc*GW* one (solid line) are almost indistinguishable for both materials. One can notice that QPsc*GW* slightly increases the electronic density along the Si-Si bond (inset of Fig. 6). This confirms the statement of Ref. 25, although the mentioned work was not fully self-consistent. The self-energy part of the Hamiltonian of these authors was calculated only once, but the remainder was

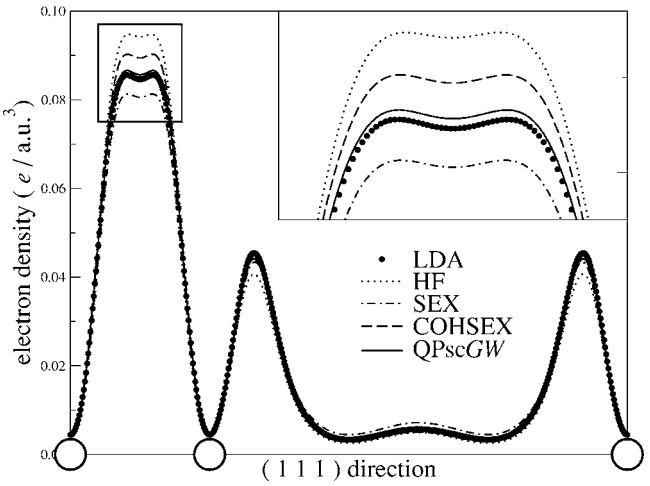

FIG. 6. Silicon: density along the direction (1 1 1) within different approximations. The inset is a close-up of the region around the maximum (Si-Si bond).

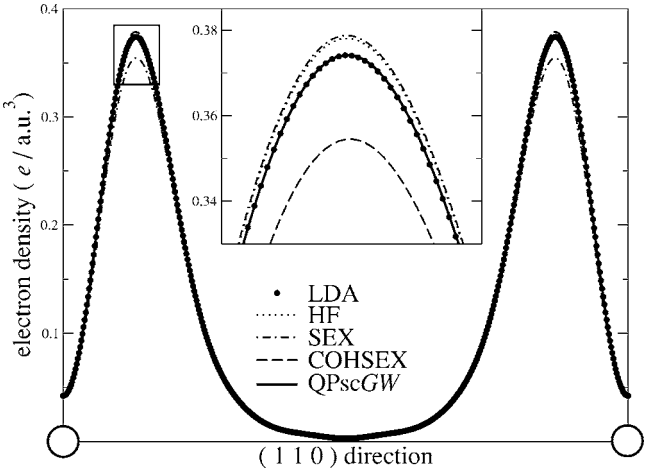

FIG. 7. Argon: density along the direction (1 1 0) within different approximations. The inset is a close-up of the region around the maximum.

evaluated self-consistently. For argon (Fig. 7), the density is slightly higher around the atoms within QPsc*GW* than within LDA. In close agreement with the usual results,[50] HF approximation (dotted line) slightly localizes the electronic density. When adding the screening of the exchange (SEX, dot-dashed line), the density has a clear tendency to differ from all the other results. This singular behavior makes us skeptical concerning the reliability of SEX-only approximation. This qualitative statement will be confirmed in the following. On the contrary, when the COH part of the self-energy is added to SEX, the corresponding COHSEX density (dashed line) lies in between HF and LDA, which is rather convincing. Here we are guided to conclude that SEX only is not satisfying, but once the COH contribution is correctly included the result seems meaningful.

These important conclusions ask for quantitative confirmation. The Hartree energy E_H

$$E_H = \frac{1}{2}\int d\mathbf{r}d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{19}$$

can give a measure of the localization of the electronic density $\rho(\mathbf{r})$. The more localized the density, the higher E_H. Table VII gives the value of the Hartree energy for silicon and argon within the different approximations. As noticed previously, HF approximation induces a localization of the electronic density with respect to LDA. SEX values deviate once again from all the other values. SEX strongly delocalizes the electronic density with respect to LDA. However, when Σ_{COH} is included in the calculation, the COHSEX result is generally in between LDA and HF. Therefore, COHSEX is meaningful, whereas SEX produces poor results. As a consequence, we argue that the screening of the exchange should not be added without the compensating COH term. This result may have resonance in the Generalized Kohn-Sham framework,[19–21] which often uses an approximated SEX interaction together with a local LDA like correction. It might be worthwhile to replace the latter by a COH. Finally, QPsc*GW* Hartree energy lies in the vicinity of the LDA one. This confirms the impressive quality of the LDA electronic density.

One can go a bit further and evaluate the quality the density matrix $\gamma(\mathbf{r},\mathbf{r}')=\Sigma_{i\ \mathrm{occ}}.\ \phi_i(\mathbf{r})\phi_i^*(\mathbf{r}')$ within all these approximations. Whereas LDA ground-state density was theoretically meaningful, the LDA density matrix is not an exact quantity of DFT. A basic way of comparing density matrices is the evaluation of the exchange energy E_{X}

$$E_{\mathrm{X}} = -\frac{1}{2}\int d\mathbf{r}d\mathbf{r}' \frac{\gamma(\mathbf{r},\mathbf{r}')\gamma(\mathbf{r}',\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}. \tag{20}$$

The conclusion drawn for the Hartree energy hold for the exchange energy according to Table VII, except that the signs are opposite. The changes in E_H and E_{X} are almost opposite, which is once again reasonable.

Furthermore, if QPsc*GW* can be considered as a reference calculation, the LDA density matrix appears as very good, since it provides an accurate exchange energy. In fact, all occupied wave functions of the materials studied here show an almost perfect agreement between LDA and QPsc*GW*. This optimistic statement cannot be extended to conduction wave functions.

It is worthwhile to investigate the stability of the QPsc*GW* procedure: we do not know whether there is a unique self-consistent attractor. We have explicitly checked for silicon that the converging point of the QPsc*GW* method is the same starting either from LDA or from HF: one obtains the same band structure and also the same Hartree energy. This is of course not a general proof, however it is rather reassuring.

TABLE VII. Hartree and exchange energies of silicon and argon. The long-range component of the Hartree term is excluded from the values presented here.

	Si		Ar	
Approx.	E_H (Ha)	E_X (Ha)	E_H (Ha)	E_X (Ha).
LDA	0.547	−2.075	4.981	−3.334
HF	0.613	−2.136	5.099	−3.370
SEX	0.467	−2.043	4.594	−3.236
COHSEX	0.593	−2.108	5.109	−3.260
QPsc*GW*	0.557	−2.086	5.011	−3.346

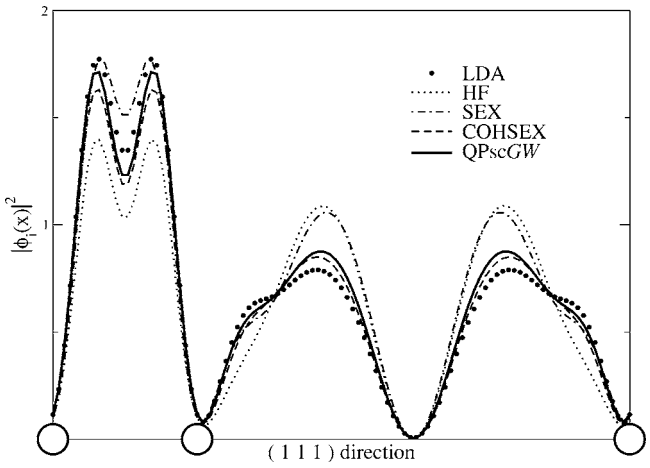

FIG. 8. First conduction band of silicon: squared modulus of the wave function along the direction (1 1 1) at $\mathbf{k}=(-1/8,-3/8,1/4)$. The white circles represent the location of the silicon atoms.

D. Quality of the conduction wave functions

Let us now compare the wave functions for the conduction states. The wave functions of the first conduction band for silicon is plotted in Fig. 8, and the wave functions of the second conduction band for argon are displayed in Fig. 9. The LDA (dotted line) and QPsc*GW* (solid line) wave functions differ noticeably for silicon and strongly for argon. Once again, SEX results (dot-dashed line) are not reliable. On the contrary, the COHSEX wave functions (dashed line) approximate closely the QPsc*GW* ones. In particular, the COHSEX result is indistinguishable from the QPsc*GW* one along the Si-Si bond in Fig. 8. One should note also that the difference between HF, COHSEX, and QPsc*GW* wave functions for the shoulder in the curves of Fig. 9 is of the same order of magnitude as the difference between the QPsc*GW* and the original *GW* in the panel (b) of Fig. 3.

The poor quality of the LDA wave functions is probably the consequence of the negative properties of any KS potential: it treats the occupied and the empty states with the same local potential. Referring to finite systems, it is well known that only the KS eigenvalue of the highest occupied molecular orbital (HOMO) is correct.[51] As a consequence, the energy of the lowest unoccupied molecular orbital (LUMO) of a N electron system is given exactly by the HOMO of the KS potential for the $N+1$ electron system. One should therefore use two different KS potentials for N or $N+1$ electrons to describe the HOMO and the LUMO.

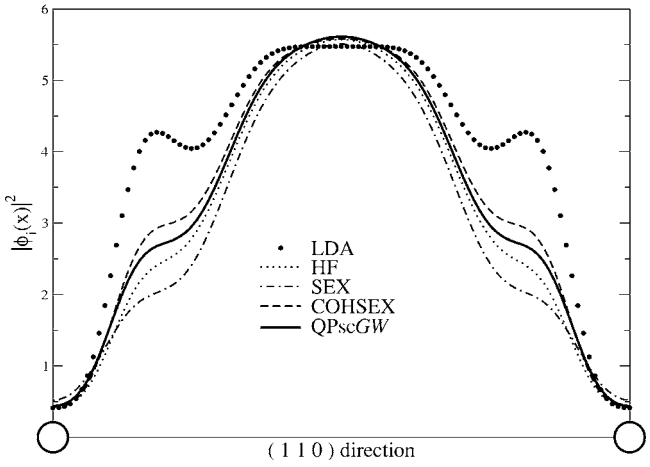

FIG. 9. Second conduction band of argon: squared modulus of the wave function along the direction (1 1 0) at $\mathbf{k}=(-1/8,-3/8,1/4)$. The white circles represent the location of the argon atoms.

In conclusion, as long as only valence states are concerned, LDA is a very reliable approximation. The situation is reversed for conduction bands: LDA is rather poor in this case. COHSEX is good at approximating *GW* conduction wave functions.

VII. CONCLUSION

The present study allows us to draw conclusions of general interest. The achievement of self-consistency in QP calculations (QPsc*GW* or COHSEX+G_0W_0) is a cumbersome procedure even for the simple materials considered here. Furthermore, for twenty years, perturbative QP calculations have been performed with great success. Fortunately, the conclusion of this work is *not* that for practical applications, self-consistency is needed every time, everywhere. Quite logically, the poorer the LDA band structure, the stronger the need for self-consistent QP calculations. Indeed, when the LDA starting point is impressively good (as in the case of aluminum), there is no need for full self-consistency. On the contrary, for argon, the LDA band gap underestimates crudely the experimental value. In this case, self-consistency is unavoidable.

In all cases studied here, self-consistent QP calculations yield good band structures, with a slight tendency to overestimate band gaps. The major changes appear to be due to the update of the QP energies, and not to the changes in the wave functions. In fact, the QP wave functions are sometimes noticeably different from their LDA counterparts, but they induce changes in the different terms of the Hamiltonian that mainly compensate, and their total effect is small.

The evaluation of the QPsc*GW* wave functions allows us to judge the quality of the LDA wave functions. We found that LDA valence wave functions and ground-state densities are in impressive agreement with the QPsc*GW* ones. This assesses *a posteriori* the use of the LDA density (for the Hartree potential) and of the LDA density matrix (for the exchange operator) in the perturbative G_0W_0 approaches. Concerning conduction wave functions, the quality of the LDA wave functions is poorer than the QPsc*GW* ones. KS scheme properly describes valence QP wave functions, but not the conduction ones.

Considering the deficiencies of LDA conduction wave functions, we propose to use the COHSEX approximation as a starting point to initiate a perturbative G_0W_0 calculation (COHSEX+G_0W_0 method). COHSEX conduction wave functions are indeed good approximations to the *GW* ones. The final COHSEX+G_0W_0 band structures lie close to the QPsc*GW* ones, at a much lower computational cost. The procedure may be even improved by combining self-consistent

COHSEX wave functions and self-consistent *GW* energies.

Along the way to COHSEX, we studied the accuracy of the SEX approximation with full RPA screening and we found it not reliable. The use of SEX without the COH term is to be avoided. Since the computational cost of the calculation of the COH term is insignificant, there is no reason not to consider both COH and SEX.

The self-consistent QP methods are theoretically more well grounded than the usual G_0W_0 recipe and their results are in rather good agreement with experiments. This work evaluates carefully the accuracy of the *GW* approximation and can be thought of as a firm basis for the further study of the missing pieces, in particular the vertex corrections.

ACKNOWLEDGMENTS

DFT-LDA calculations were performed using ABINIT package.[34] Self-consistent SEX, COHSEX, and *GW* methods are developments based on the standard *GW* routines of ABINIT package, and are now merged into the publicly available version of ABINIT project.[52] We are grateful for discussions with P. Rinke, T. Kotani, E. K. Shirley, and V. A. Popa, for support by the EU's 6th Framework Programme through the NANOQUANTA Network of Excellence (NMP4-CT-2004-500198), and for computer time from IDRIS (project 544) and from CEA/DSM (project p93).

*Present address: Chemistry and Applied Biosciences, ETH-Zurich, USI-Campus, Via Giuseppe Buffi 13, CH-6900 Lugano, Switzerland.

[1] L. Hedin, Phys. Rev. **139**, A796 (1965).
[2] W. G. Aulbur, L. Jönsson, and J. W. Wilkins, Solid State Phys. **54**, 1 (1999), and references therein.
[3] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[4] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[5] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory* (Springer-Verlag, Berlin, 1990).
[6] L. D. Landau and E. M. Lifschitz, *Statistical Physics Part II* (Pergamon, Oxford, 1980).
[7] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particule Systems* (Dover, Mineola, New York, 1971, 2003).
[8] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985).
[9] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).
[10] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[11] B. Holm and U. von Barth, Phys. Rev. B **57**, 2108 (1998).
[12] F. Aryasetiawan and O. Gunnarsson, Phys. Rev. Lett. **74**, 3221 (1995).
[13] W.-D. Schöne and A. G. Eguiluz, Phys. Rev. Lett. **81**, 1662 (1998).
[14] W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **89**, 126401 (2002).
[15] K. Delaney, P. García-González, A. Rubio, P. Rinke, and R. W. Godby, Phys. Rev. Lett. **93**, 249701 (2004).
[16] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004).
[17] S. Lebègue, B. Arnaud, M. Alouani, and P. E. Bloechl, Phys. Rev. B **67**, 155208 (2003).
[18] L. Hedin and S. Lundqvist, Solid State Phys. **23**, 1 (1969).
[19] D. M. Bylander and L. Kleinman, Phys. Rev. B **41**, 7868 (1990).
[20] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy, Phys. Rev. B **53**, 3764 (1996).
[21] S. Picozzi, A. Continenza, R. Asahi, W. Mannstadt, A. J. Freeman, W. Wolf, E. Wimmer, and C. B. Geller, Phys. Rev. B **61**, 4677 (2000).
[22] R. Haensel, Phys. Rev. Lett. **23**, 1160 (1969).
[23] G. Strinati, Riv. Nuovo Cimento **11**, 1 (1988).
[24] G. Baym and L. P. Kadanoff, Phys. Rev. **124**, 287 (1961).
[25] M. M. Rieger and R. W. Godby, Phys. Rev. B **58**, 1343 (1998).
[26] A. Schindlmayr, P. García-González, and R. W. Godby, Phys. Rev. B **64**, 235106 (2001).
[27] G. Strinati, H. J. Mattausch, and W. Hanke, Phys. Rev. B **25**, 2867 (1982).
[28] S. Galamić-Mulaomerović and C. H. Patterson, Phys. Rev. B **71**, 195103 (2005).
[29] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler, New J. Phys. **7**, 126 (2005).
[30] O. Pulci, L. Reining, G. Onida, R. Del Sole, and F. Bechstedt, Comput. Mater. Sci. **20**, 300 (2001).
[31] I. D. White, R. W. Godby, M. M. Rieger, and R. J. Needs, Phys. Rev. Lett. **80**, 4265 (1998).
[32] P. A. Bobbert and W. van Haeringen, Phys. Rev. B **49**, 10326 (1994).
[33] F. Bruneval, F. Sottile, V. Olevano, R. Del Sole, and L. Reining, Phys. Rev. Lett. **94**, 186402 (2005).
[34] X. Gonze, G.-M. Rignanese, M. Verstraete, J.-M. Beuken, Y. Pouillon, R. Caracas, F. Jollet, M. Torrent, G. Zerah, M. Mikami *et al.*, Z. Kristallogr. **220**, 558 (2005).
[35] M. L. Tiago, S. Ismail-Beigi, and S. G. Louie, Phys. Rev. B **69**, 125212 (2004).
[36] F. Bruneval, Ph.D. thesis, Ecole Polytechnique, Palaiseau, France (2005), URL http://theory.lsi.polytechnique.fr/people/bruneval/bruneval_these.pdf.
[37] P. Livins and S. E. Schnatterly, Phys. Rev. B **37**, 6731 (1988).
[38] J. E. Northrup, M. S. Hybertsen, and S. G. Louie, Phys. Rev. B **39**, 8198 (1989).
[39] M. van Schilfgaarde, Takao Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006).
[40] Landolt-Börnstein, *Numerical Data and Functional Relationships in Science and Technology* (Springer, New-York, 1982), New Series Group III, Vol. 17, Pt A.
[41] J. E. Ortega and F. J. Himpsel, Phys. Rev. B **47**, 2130 (1993).
[42] V. Popa, G. Brocks, and P. Kelly, cond-mat/0507013 (unpublished).
[43] T. Kotani and M. van Schilfgaarde, Solid State Commun. **121**, 461 (2002).
[44] F. Aryasetiawan, L. Hedin, and K. Karlsson, Phys. Rev. Lett. **77**, 2268 (1996).
[45] N. Schwentner, F.-J. Himpsel, V. Savle, M. Skibowski, W. Steinmann, and E. E. Koch, Phys. Rev. Lett. **34**, 528 (1975).
[46] E. L. Shirley and R. M. Martin, Phys. Rev. B **47**, 15404 (1993).

[47] E. L. Shirley, Phys. Rev. B **58**, 9579 (1998).
[48] E. L. Shirley (private communication).
[49] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10159 (1988).
[50] S. Massidda, M. Posternak, and A. Baldereschi, Phys. Rev. B **48**, 5058 (1993).
[51] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).
[52] URL http://www.abinit.org