Efficient scheme for *GW* quasiparticle band-structure calculations with applications to bulk Si and to the Si(001)-(2×1) surface

Michael Rohlfing, Peter Krüger, and Johannes Pollmann
Institut für Theoretische Physik II — Festkörperphysik, Universität Münster, D-48149 Münster, Germany
(Received 10 February 1995)

We report an efficient scheme for evaluating the quasiparticle corrections to local-density-approximation (LDA) band structures within the *GW* approximation. In this scheme, the *GW* self-energy corrections are evaluated in a sufficiently flexible Gaussian orbital basis set instead of using plane-wave Fourier representations of the relevant two-point functions. It turns out that this set has to include orbitals up to *f*-type symmetry, when in the LDA calculations Gaussian orbitals up to *d*-type symmetry are needed for convergence. For bulk Si, both schemes yield virtually identical quasiparticle band structures and the demand on computer time is roughly the same. For the Si(001)-(2×1) surface, the *GW* Gaussian orbital scheme is a factor of 5 faster. In our calculations for Si(001)-(2×1) the dynamic dielectric matrix is obtained by applying a plasmon-pole approximation. The static dielectric matrix of the Si(001) surface is fully calculated within the random-phase approximation (RPA). In addition, we have performed quasiparticle surface band-structure calculations employing two model dielectric matrices. Our respective results are compared with those obtained employing the full RPA dielectric matrix as well as with results of previous calculations by other authors which were based on model dielectric matrices.

I. INTRODUCTION

It has been shown within the last decade that well-known shortcomings of the local-density approximation (LDA) for calculating band structures of semiconductors and their surfaces can be surmounted by first-principles quasiparticle calculations employing the many-body formalism presented by Hedin,[1] and Hedin and Lundqvist.[2] The basic quantity of this theory is the nonlocal, non-Hermitian, and energy-dependent self-energy operator $\Sigma(\mathbf{r},\mathbf{r}',E)$, a two-point function with respect to its spatial degrees of freedom. In lowest approximation, Σ is given as a product of the Green's function G times the screened Coulomb interaction W. This approximation is usually referred to as the *GW* approximation (*GW*A). In their landmark contribution to the field, Hybertsen and Louie[3] developed practicable schemes for evaluating the many-body corrections within the *GW*A and they arrived at theoretical results that showed excellent agreement with a whole body of experimental data. One important feature for the success of those calculations was the inclusion of local fields and dynamic effects in the screened Coulomb interaction W. The required dielectric matrices are calculated within the random-phase approximation[4,5] (RPA) for the static case and are extended to finite frequency using a generalized plasmon-pole model based on sum rules. In the calculations, both the one-point wave functions of the LDA and the two-point functions of the *GW*A have to be represented by appropriate basis sets. For example, Hybertsen and Louie,[3] as well as, other authors[6–8] employed plane-wave basis sets both for the LDA and the *GW*A calculations.

A basis set of localized Gaussian orbitals can equally well be used to represent the LDA wave functions. Small LDA Gaussian orbital basis sets yield results that are in excellent agreement with the results of respective plane-wave calculations, as we have shown, e.g., in a previous paper.[9] In all previous *GW* calculations reported so far,[3,6–9] the matrix elements of the self-energy operator have been represented using plane waves, i.e., a Fourier representation was used. Since the operators of the *GW*A are two-point functions with respect to their spatial degrees of freedom, their Fourier representation in the plane-wave approach is extremely demanding, in particular when surfaces or bulk systems with occupied *d* orbitals are considered.

In this paper, we present an efficient scheme for representing the operators entering the *GW* quasiparticle calculations in a Gaussian orbital basis set. To classify the various possible approaches more clearly, we use the following abbreviations throughout this paper. They refer to the LDA and to the *GW*A, respectively. Each method can be evaluated using either a basis set of plane waves (PW-LDA, PW-*GW*A) or a basis set of Gaussian orbitals (GO-LDA, GO-*GW*A). Hybertsen and Louie,[3] e.g., combined the PW-LDA with the PW-*GW*A in their calculations, while we combined the GO-LDA with the PW-*GW*A in our previous work.[9] In this paper, we present and employ a combination of the GO-LDA with the GO-*GW*A. The size of the problem in this GO-*GW*A approach with respect to the necessary CPU size and computation time is reduced drastically, as compared to the PW-*GW*A, in particular for surfaces and for bulk solids with occupied *d* orbitals like II-VI semiconductors. Using the GO-*GW*A approach, we have achieved the calculation of the quasiparticle band structure of the Si(001)-

0163-1829/95/52(3)/1905(13)/$06.00

(2×1) surface incorporating the full RPA static dielectric matrix.

The paper is organized as follows. In Sec. II, we briefly address the *GW*A in general and summarize the basic equations needed for the following discussions. In Sec. III, we describe in detail the representation of the relevant two-point functions of the *GW*A in the Gaussian orbital representation and their evaluation. In Sec. IV, we apply our scheme to bulk Si and to the Si(001)-(2×1) surface. Our results are presented and discussed in comparison with pertinent literature data. Both the full static RPA matrix and two different model functions for the static dielectric matrix are employed. The usefulness of these model functions is assessed by comparisons with our full RPA results and results from the literature that are based on model functions. A short summary concludes the paper in Sec. V.

II. THE *GW* APPROXIMATION

For the calculation of the electronic band structure of a semiconductor system from first principles, the exchange and correlation effects of the interacting electrons play a very important role. The major difficulty stems from an adequate treatment of the dynamical correlations of the electrons in the solid with an energy gap and a strongly inhomogeneous charge density. Within the widely used density functional theory (DFT), these effects are described approximately by a local potential $V_{\rm xc}(\mathbf{r})$. This approximation enables a very reliable determination of the ground-state properties of the system. However, such an exchange-correlation potential is not well suited to compute single-particle excitation energies. For this task $V_{\rm xc}(\mathbf{r})$ has to be replaced by a nonlocal, energy-dependent self-energy operator $\Sigma(\mathbf{r},\mathbf{r}',E)$, which enters the quasiparticle equation[1,2]

$$\left\{-\frac{\hbar^2}{2m}\nabla^2+V_{ps}(\mathbf{r})+V_H(\mathbf{r})\right\}\psi_{n\mathbf{k}}(\mathbf{r}) + \int\Sigma(\mathbf{r},\mathbf{r}',E_{n\mathbf{k}})\psi_{n\mathbf{k}}(\mathbf{r}')d^3r' = E_{n\mathbf{k}}\psi_{n\mathbf{k}}(\mathbf{r}) \ . \quad (1)$$

As was pointed out by Hedin,[1] and Hedin and Lundqvist,[2] the self-energy operator can be expanded in a series containing the Green's function G and the screened interaction W of the system. The first term of this series constitutes the *GW* approximation:

$$\Sigma(\mathbf{r},\mathbf{r}',E)=\frac{i}{2\pi}\int e^{-i\omega 0^+}G(\mathbf{r},\mathbf{r}',E-\omega)W(\mathbf{r},\mathbf{r}',\omega)d\omega \ . \quad (2)$$

For the computation of the self-energy operator and the evaluation of Eq. (1), we employ the scheme of Hybertsen and Louie.[3] The Green's function $G(\mathbf{r},\mathbf{r}',E)$ of the system is calculated as

$$G(\mathbf{r},\mathbf{r}',E)=\sum_{n\mathbf{k}\sigma}\frac{\psi_{n\mathbf{k}\sigma}(\mathbf{r})\psi^*_{n\mathbf{k}\sigma}(\mathbf{r}')}{E-E_{n\mathbf{k}\sigma}+i0^+\mathrm{sgn}(E_{n\mathbf{k}\sigma}-\mu)} \ . \quad (3)$$

The dynamically screened Coulomb interaction $W=\epsilon^{-1}v$ is evaluated as follows. First, the polarization P is calculated within the random-phase approximation.[4,5] The result is convoluted with the Coulomb interaction v yielding the dielectric function $\epsilon=1-vP$. Then the inverse dielectric function ϵ^{-1} is calculated and the result is again convoluted with the Coulomb interaction. This finally yields the screened interaction W.

Since the evaluation of the RPA expression of the polarization requires very much computation time, we perform it only for the *static* polarization, i.e., at frequency ω=0. The resulting static dielectric function is then extended to the dynamic one by use of a plasmon-pole model,[6] taking Johnson's sum rule[10] into account.

After defining this *GW* self-energy operator, one has to solve Eq. (1) self-consistently. We take as starting values the results of a self-consistent DFT-LDA calculation. It turns out that for many systems the LDA wave functions $\psi^{\rm LDA}_{n\mathbf{k}}(\mathbf{r})$ agree remarkably well with the final *GW* wave functions.[3,7] Therefore, the self-energy operator can be constructed using the results of the LDA calculation (wave functions, energy spectrum, and electronic density) and an iterative evaluation of Eq. (1) is not necessary. The quasiparticle energies are then simply given by

$$E_{m\mathbf{k}}=E^{\rm LDA}_{m\mathbf{k}}+Z_{m\mathbf{k}}\langle\psi^{\rm LDA}_{m\mathbf{k}}|\Sigma(E^{\rm LDA}_{m\mathbf{k}})-V^{\rm LDA}_{\rm xc}|\psi^{\rm LDA}_{m\mathbf{k}}\rangle \ . \quad (4)$$

The renormalization constant $Z_{m\mathbf{k}}$ has been introduced[3] to take into account the energy dependence of the self-energies $\Sigma_{m\mathbf{k}}(E)$ to some extent. Equation (4) can be understood as the result of a perturbation approach that is evaluated to first order. The second-order contributions to the energy can also be calculated. We have tested this for a number of systems like Si, GaAs, or diamond and have found that these contributions nearly vanish. Since the second-order contribution to the energy arises from the first-order contributions to the wave functions, this supports the assumption that $\psi^{\rm LDA}_{n\mathbf{k}}\approx\psi_{n\mathbf{k}}$.

As an alternative approach, the static dielectric function may be approximated by model functions (see Ref. 9 and references therein). By application of such model functions in the *GW*A calculation of the quasiparticle band structures of bulk semiconductors like diamond, Si, Ge, GaAs, and SiC, we have obtained results in good agreement with our full RPA calculations.[9]

III. *GW*A USING LOCALIZED BASIS FUNCTIONS

In this section we present our scheme for calculating quasiparticle band structures within the *GW*A using localized Gaussian orbitals both in the underlying LDA calculations as well as in the *GW*A calculations. Usually the computer demand for the evaluation of the *GW*A scales as $O(N^2)$ or $O(N^3)$ (depending on the respective plasmon-pole model used), where N is the number of electrons in the unit cell. Surface calculations are very

demanding for this reason. The numerical size of the problem also scales as $O(M^2)$ where M is the necessary number of GW basis functions per atom. Therefore systems with strongly localized electronic states like II-VI semiconductors with occupied d orbitals are very demanding when plane waves are used as GW basis functions. Our approach of using Gaussian orbitals instead of plane waves as GW basis functions does not overcome the bad scaling of the GWA with respect to the size N of the problem. But the absolute demand for computer size and computation time is very much reduced due to the small number M of GW basis functions per atom required both for surfaces and for systems with strongly localized electronic states.

A. The GO-LDA basis set

To perform the LDA calculations, we use norm-conserving *ab initio* pseudopotentials. Only the valence electrons are considered. For the representation of the respective wave functions, we use Gaussian orbitals that are centered at the atomic positions in the unit cell. To obtain sufficient flexibility of the basis set, we employ several shells of Gaussians with different decay constants for each atom. Each shell contains s, p, d, and s^* orbitals. The following linear combination of atomic Gaussians yields Bloch functions that can be used as basis functions in the crystal:

$$\chi_\alpha(\mathbf{k},\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot(\mathbf{R}+\boldsymbol{\tau}_\alpha)} \phi_\alpha(\mathbf{r}-\mathbf{R}-\boldsymbol{\tau}_\alpha) \quad . \quad (5)$$

The index α specifies the orbital character (s, p_x,..., d_{xy},...,) of the Gaussian functions $\phi_\alpha(\mathbf{r})$ and labels Gaussians with different decay constants. The atomic position in the unit cell is $\boldsymbol{\tau}_\alpha$ and $\mathbf{R}$ is a Bravais lattice vector.

It turns out that these basis sets yield results for the structural and electronic properties of semiconductors agreeing very well with those of converged plane-wave calculations. (See Ref. 9 for that matter.) The calculated structural properties also agree well with experimental results. Of course, the band structures suffer from the well-known gap problem that is typical of all DFT-LDA calculations.

B. The GO-*GW*A basis set

So far, the space dependence of two-point functions $f(\mathbf{r},\mathbf{r}')$ appearing in the GW approximation ($\epsilon, v, W,...$) has been expressed in terms of their Fourier transforms, i.e., by a matrix representation $f_{\mathbf{G},\mathbf{G}'}(\mathbf{q})$ with respect to plane waves (see, e.g., Refs. 3, 8, and 9). The number of plane waves that is necessary to obtain sufficient accuracy depends very much on the system under consideration. Typical values for standard bulk semiconductors range from about 100 to 200 plane waves.[9] If strongly localized wave functions are to be described, e.g., the d electrons of transition metal elements and their compounds, many more plane waves are required. The description of such systems is much more efficient when localized basis functions are used, since the required number of respective functions remains modest even for strongly localized states. This has also been shown by Aryasetiawan and Gunnarson for the calculation of the dielectric matrix of Ni,[11] as well as for the calculation of the electronic structure of NiO within the GWA.[12]

A two-point function $f(\mathbf{r},\mathbf{r}')$ obtaining the translational invariance of the crystal can be represented by any complete set of functions $\{\chi_\beta(\mathbf{q},\mathbf{r})\}$. Since a finite set is never rigorously complete, at least a sufficiently flexible basis set should be used. In this work we use a basis set of Gaussian orbitals of the *same type* as in the GO-LDA calculations [see Eq. (5)]:

$$f(\mathbf{r},\mathbf{r}') = \sum_{\mathbf{q}} \sum_{\beta\beta'} \chi_\beta(\mathbf{q},\mathbf{r}) \langle f \rangle_{\beta\beta'}(\mathbf{q}) \chi^*_{\beta'}(\mathbf{q},\mathbf{r}') \quad . \quad (6)$$

Similar to the common Fourier transform, f is represented in Eq. (6) by a matrix with elements $\langle f \rangle_{\beta\beta'}(\mathbf{q})$, but with respect to the *localized* basis. It should be noted that the orbital types and decay constants of the GO's in Eq. (6), in principle, are entirely independent from those of the GO-LDA basis set in Eq. (5). In order to obtain the representation matrix $\langle f \rangle$, the elements of the following matrix $[f]$ have to be calculated first. They are defined as

$$[f]_{\beta\beta'}(\mathbf{q}) = \int \chi^*_\beta(\mathbf{q},\mathbf{r}) f(\mathbf{r},\mathbf{r}') \chi_{\beta'}(\mathbf{q},\mathbf{r}') d^3r d^3r' \quad . \quad (7)$$

The two matrices $[f]$ and $\langle f \rangle$ are related to each other in the following way:

$$\langle f \rangle(\mathbf{q}) = S^{-1}(\mathbf{q})[f](\mathbf{q})S^{-1}(\mathbf{q}) \quad , \quad (8)$$

where S is the overlap matrix of the basis functions:

$$S_{\beta\beta'}(\mathbf{q}) = \int \chi^*_\beta(\mathbf{q},\mathbf{r}) \chi_{\beta'}(\mathbf{q},\mathbf{r}) d^3r \quad . \quad (9)$$

For a plane-wave basis set, the overlap matrix would be the unit matrix and the two matrices in Eqs. (7) and (8) would be identical. For a Gaussian orbital basis set, however, the two matrices are different since the Gaussians are nonorthogonal to each other. This has to be taken into account throughout the calculations.

The matrix of the static polarization reads in RPA:

$$[P]_{\beta\beta'}(\mathbf{q}) = 4 \sum_{\mathbf{k}} \sum_{m\in \mathrm{Val}} \sum_{n\in \mathrm{Con}} \frac{M^{mn}_\beta(\mathbf{k},\mathbf{q}) \left[M^{mn}_{\beta'}(\mathbf{k},\mathbf{q})\right]^*}{E_{m\mathbf{k}} - E_{n,\mathbf{k}+\mathbf{q}}}, \quad (10)$$

with

$$M^{mn}_\beta(\mathbf{k},\mathbf{q}) = \int \psi^*_{m\mathbf{k}}(\mathbf{r}) \chi^*_\beta(\mathbf{q},\mathbf{r}) \psi_{n,\mathbf{k}+\mathbf{q}}(\mathbf{r}) d^3r \quad . \quad (11)$$

The evaluation of the integrals $M^{mn}_\beta(\mathbf{k},\mathbf{q})$ is described below in more detail (see Sec. III C). By a convolution of P with the Coulomb interaction v, one obtains the

dielectric function $\epsilon = 1 - vP$. In order to prevent divergencies, we use the symmetrized form of the dielectric function $\tilde{\epsilon} = 1 - \tilde{v}P\tilde{v}$. The double convolution of P with the auxiliary function $\tilde{v}(\mathbf{r},\mathbf{r}') = e\pi^{-3/2}/|\mathbf{r}-\mathbf{r}'|^2$ replaces its single convolution with the Coulomb interaction v (see the Appendix). Using Gaussian orbitals, this convolution is performed by matrix multiplications:

$$[\tilde{\epsilon}](\mathbf{q}) = S(\mathbf{q}) - [\tilde{v}](\mathbf{q}) \cdot S^{-1}(\mathbf{q}) \cdot [P](\mathbf{q}) \cdot S^{-1}(\mathbf{q}) \cdot [\tilde{v}](\mathbf{q}) \ . \tag{12}$$

In order to extend the static dielectric function to nonzero frequencies, we introduce a plasmon-pole model equivalent to that used in Refs. 6, 8, and 9. First we perform a Choleski decomposition of the overlap matrix $S=LL^\dagger$. After multiplication of $[\tilde{\epsilon}]$ with the inverse of this Choleski matrix, we calculate eigenvectors and eigenvalues

$$\left[L^{-1}[\tilde{\epsilon}](L^{-1})^\dagger\right]_{\beta\beta'}(\mathbf{q}) \equiv \sum_l \phi^l_\beta(\mathbf{q})\lambda_{\mathbf{q}l}\left[\phi^l_{\beta'}(\mathbf{q})\right]^* \ . \tag{13}$$

In order to obtain a dielectric function for nonzero frequencies, one assumes[6] that only the eigenvalues $\lambda_{\mathbf{q}l}$ in Eq. (13) depend on frequency ω while the eigenvectors are taken to be frequency independent. Thus, the dynamic dielectric matrix becomes

$$[\tilde{\epsilon}]_{\beta\beta'}(\mathbf{q},\omega) = \sum_l \Phi^l_\beta(\mathbf{q})\lambda_{\mathbf{q}l}(\omega)\left[\Phi^l_{\beta'}(\mathbf{q})\right]^* , \tag{14}$$

with

$$\Phi^l_\beta(\mathbf{q}) = \sum_{\beta'} L_{\beta\beta'}\phi^l_{\beta'}(\mathbf{q}) \ . \tag{15}$$

The eigenvalues $\lambda_{\mathbf{q}l}(\omega)$ are assumed to have the following frequency dependence:[6]

$$\lambda^{-1}_{\mathbf{q}l}(\omega) = 1 + \frac{z_{\mathbf{q}l}\omega_{\mathbf{q}l}}{2}\left(\frac{1}{\omega - (\omega_{\mathbf{q}l} - i0^+)} - \frac{1}{\omega + (\omega_{\mathbf{q}l} - i0^+)}\right) \ . \tag{16}$$

The parameters $z_{\mathbf{q}l}$ are chosen such that Eq. (14) reproduces the static dielectric function for ω=0. The parameters $\omega_{\mathbf{q}l}$ are determined by evaluating Johnson's sum rule.[10]

It should be mentioned that the frequency-independent part of the eigenvalues (which equals one) would result in $[\tilde{\epsilon}^{-1}](\mathbf{q})$=[1], i.e., this part describes the Hartree-Fock-like exchange contribution to the self-energy. In contrast, the frequency-dependent part of $\lambda^{-1}_{\mathbf{q}l}(\omega)$ yields the correlation contribution to the self-energy.

The inverse dielectric function is obtained easily from Eq. (14). Another convolution with $\tilde{v}$ results in the dynamically screened interaction $W=\tilde{v}\tilde{\epsilon}^{-1}\tilde{v}$. Finally the self-energy of an electronic state $\psi_{m\mathbf{k}}$ is given as

$$\langle\psi_{m\mathbf{k}}|\Sigma(E)|\psi_{m\mathbf{k}}\rangle = \sum_{\mathbf{q}}\sum_{n,l}\left|\sum_\beta M^{mn}_\beta(\mathbf{k},\mathbf{q})B_{\beta l}(-\mathbf{q})\right|^2 \times \begin{cases} \left(-1 + \dfrac{z_{-\mathbf{q}l}\omega_{-\mathbf{q}l}}{2}\dfrac{1}{E - E^{\mathrm{LDA}}_{n,\mathbf{k}+\mathbf{q}} + \omega_{-\mathbf{q}l}}\right) & \text{for } n \in \mathrm{Val} \\ \dfrac{z_{-\mathbf{q}l}\omega_{-\mathbf{q}l}}{2}\dfrac{1}{E - E^{\mathrm{LDA}}_{n,\mathbf{k}+\mathbf{q}} - \omega_{-\mathbf{q}l}} & \text{for } n \in \mathrm{Con}, \end{cases} \tag{17}$$

with

$$B_{\beta l}(\mathbf{q}) = \sum_{\beta'\beta''} S^{-1}_{\beta\beta'}(\mathbf{q})[\tilde{v}]_{\beta'\beta''}(\mathbf{q})\Phi^l_{\beta''}(\mathbf{q}) \ . \tag{18}$$

C. Details of the method

The LDA wave functions $\psi_{n\mathbf{k}}$ enter the *GW* formalism in the integrals $M^{mn}_\beta(\mathbf{k},\mathbf{q})$. These occur in the evaluation of the RPA polarization [see Eq. (10)] as well as in the final formula for the self-energy of a state $\psi_{m\mathbf{k}}$ [Eq. (17)]. Their evaluation requires most of the computation time of our method. Since we express the wave functions in terms of Gaussian orbitals $\chi_\alpha(\mathbf{k},\mathbf{r})$, the calculation of $M^{mn}_\beta(\mathbf{k},\mathbf{q})$ involves the evaluation of three-center overlap integrals

$$\int \phi_\alpha(\mathbf{r}-\mathbf{R}-\boldsymbol{\tau}_\alpha)\phi_\beta(\mathbf{r}-\tilde{\mathbf{R}}-\boldsymbol{\tau}_\beta)\phi_{\alpha'}(\mathbf{r}-\mathbf{R}'-\boldsymbol{\tau}_{\alpha'})d^3r \ . \tag{19}$$

We compute these integrals for all orbital types in an iterative way following the prescription of Obara and Saiki.[13]

The three-center overlap integrals in Eq. (19) tend to zero if the atomic sites $\mathbf{R}+\boldsymbol{\tau}_\alpha$, $\tilde{\mathbf{R}}+\boldsymbol{\tau}_\beta$, and $\mathbf{R}'+\boldsymbol{\tau}_{\alpha'}$ are far away from each other. In slab calculations for surfaces, many of the occurring integrals need not be calculated for that reason.

As can be seen in Eq. (11), the *GW* basis functions $\chi_\beta(\mathbf{q},\mathbf{r})$ must be suitable to represent *products* of wave functions $\psi^*_{m\mathbf{k}}(\mathbf{r}) \cdot \psi_{n,\mathbf{k}+\mathbf{q}}(\mathbf{r})$. Since the wave functions themselves contain contributions from *s*-, *p*-, and *d*-like LDA basis functions, their products will consist of terms with *s*, *p*, *d*, *f*, and *g* character (and even higher angu-

lar momenta, since an orbital that has a defined angular momentum with respect to its localization center has no longer a single angular momentum when it is viewed from another lattice site). Therefore, the *GW* basis set must contain functions of higher angular momenta than the LDA basis set. We find that for Si the use of *GW* basis functions with angular momenta l=0, 1, 2, and 3 yields sufficient accuracy (see below). Other systems may require even higher angular momentum functions in the basis for the *GW* calculations.

All two-point functions occurring in the *GW*A are represented using Gaussian orbitals. Some quantities, however, are set up in terms of their common Fourier representation first, before they are transformed into a matrix representation with respect to the Gaussian basis (see below). First, most model dielectric functions $\epsilon_{\mathbf{G},\mathbf{G}'}(\mathbf{q})$ are constructed in terms of plane waves.[9] Second, the auxiliary function $\tilde{v}$ that is used to describe the Coulomb interaction v is represented by a diagonal matrix in Fourier space (see the Appendix). Its representation matrix with respect to Gaussians is obtained very fast as described below. [The direct calculation of $[\tilde{v}]_{\beta\beta'}(\mathbf{q})$ according to Eq. (7) would converge only very slowly with respect to the number of neighboring cells $\mathbf{R}$ because of the long-range behavior of $\tilde{v}$.] Third, the evaluation of Johnson's sum rule[10] usually is carried out in Fourier space, as well.

If the common Fourier transform $f_{\mathbf{G},\mathbf{G}'}(\mathbf{q})$ of a two-point function is known, it is easily transformed into the representation matrix $[f]_{\beta\beta'}(\mathbf{q})$ with respect to the Gaussian orbital basis set by

$$[f]_{\beta\beta'}(\mathbf{q}) = \sum_{\mathbf{G},\mathbf{G}'} \left[K_{\beta}^{\mathbf{G}}(\mathbf{q})\right]^{*} f_{\mathbf{G},\mathbf{G}'}(\mathbf{q}), K_{\beta'}^{\mathbf{G}'}(\mathbf{q}), \qquad (20)$$

where $K_{\beta}^{\mathbf{G}}(\mathbf{q})$ is the Fourier transform of a Gaussian orbital, which is given analytically. With the help of Eq. (20) the Fourier-represented quantities described above can be transformed into the Gaussian representation in a very efficient way.

The long-range behavior of the dielectric function is difficult to compute within RPA. In the Fourier representation, the "head" of the dielectric matrix $\epsilon_{\mathbf{00}}(\mathbf{q})$ (for $\mathbf{q} \to \mathbf{0}$) contains most of the information about long-range effects. Its convergent calculation by a Brillouin-zone integration requires many more $\mathbf{k}$ points than that of the other matrix elements. As can be seen in Eq. (20), the head of the Fourier representation of the dielectric function contributes to all elements of the Gaussian orbital representation matrix $[\epsilon]_{\beta\beta'}(\mathbf{q})$. In order to account for the exact long-range behavior of the screening, we exclude the head of the dielectric function from the Gaussian orbital matrix and calculate it in Fourier representation using a sufficiently large number of $\mathbf{k}$ points. Thereafter, the Fourier representation of the head is transformed to the Gaussian orbital representation according to Eq. (20) and is added to the dielectric matrix.

IV. RESULTS

We have applied our approach first to bulk Si in order to test its accuracy in direct comparison with a conventional *GW* calculation within the Fourier representation. Next, we have studied the Si(001)-(2×1) surface which, to our knowledge, has not yet been treated by a full RPA *GW* approach, so far.

A. Bulk Si

To calculate the wave functions and the band-structure energies entering the *GW* scheme, we perform a GO-LDA calculation first. In this calculation, we use 20 basis functions for each atom with decay constants of 0.15 and 0.5 (in atomic units). The separable pseudopotential of Stumpf, Gonze, and Scheffler[14] is employed. The exchange-correlation data are taken from Ceperley and Alder[15] as parametrized by Perdew and Zunger.[16] Ten special points are used in the Brillouin-zone summations. We have shown in Ref. 9 that our LDA basis set consisting of 40 Gaussian orbitals per unit cell leads to an LDA band structure for bulk Si in excellent agreement with the results of highly converged plane-wave calculations performed with 450 plane waves. These LDA calculations together with the *GW* computations within the Fourier representation, as described in Ref. 9, were carried out for the experimental lattice constant of 5.43 Å. The theoretical lattice constant of 5.37 Å, which we obtain by total energy minimization using the Ceperley-Alder exchange-correlation energy, is roughly 1% smaller than the experimental value. This is in agreement with previous results obtained with basis sets of localized orbitals,[17] as well as of plane waves.[18] Here we present our results for bulk Si calculated at the *theoretical* lattice constant for reasons of consistency with the structure optimization in the case of the Si(001)-(2×1) surface reported in the next section. In consequence, the fundamental LDA gap of 0.46 eV obtained in this work is a little bit smaller than that of 0.56 eV obtained for the experimental lattice constant, as reported in Ref. 9.

Based on this GO-LDA calculation, two *GW*A calculations have been performed. First, we use 60 Gaussian orbitals in a GO-*GW*A calculation according to the method described in Sec. III B. The dielectric matrix is calculated within the RPA. We obtain a bulk dielectric constant ϵ_{∞}=13.0, which is slightly larger than the value of 12.7 which we obtained for the experimental lattice constant, previously.[9]

As the most important effect of the *GW* calculations, the LDA valence band energies are shifted to lower energies by −0.40 to −0.65 eV according to Eq. (4), depending on the particular band and wave vector considered. The quasiparticle shifts are displayed in Fig. 1 with respect to the LDA energy of the respective state. The shape and dispersion of the valence bands remain more or less unchanged. The *GW* valence band width amounts to 12.17 eV as compared to the LDA value of 12.11 eV. On the other hand, the LDA conduction band energies are shifted to higher energies by about +0.25 . . . + 1.0 eV with the largest shifts for the high-energy conduction band states. The fundamental gap is increased by +0.77 eV and amounts to 1.23 eV in very good agreement with

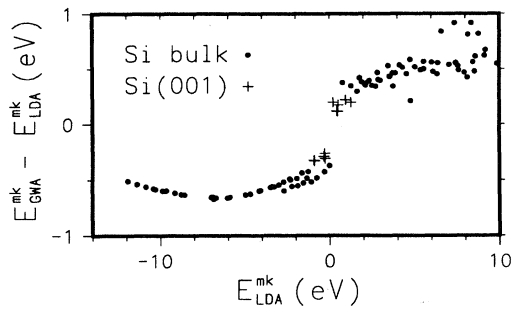

FIG. 1. Calculated quasiparticle shifts $E^{mk}_{GWA} - E^{mk}_{LDA}$ for the Si bulk states (dots) and for the dangling-bond states of the Si(001)-(2×1) surface (crosses).

the experimental value of 1.17 eV.[19] This gap energy is a little bit smaller than the value of 1.31 eV, which we obtained in Ref. 9 for the experimental lattice constant.

Next we have performed a second *GW* calculation for the same system using the common PW-*GW*A method with Fourier transforms instead of the Gaussian orbital representation. In that calculation we employ 113 plane waves as *GW* basis functions (cf. Ref. 9). All other aspects of this PW-*GW*A calculation are identical to the GO-*GW*A method described above. We again evaluate the RPA expression for the static dielectric function employing the plasmon-pole model. Also for the plane-wave basis, we obtain a bulk dielectric constant of ϵ_∞=13.0.

The quasiparticle shifts obtained by the PW-*GW*A and the GO-*GW*A are the same within 0.03 eV for the valence band states and 0.05 eV for the conduction band states with few exceptions at high energies (e.g., 0.08 eV for L'_{2c}). The gap energy is 1.20 eV for the PW-*GW*A and 1.23 eV for the GO-*GW*A, respectively. The valence band width amounts to 12.18 eV (PW's) and 12.17 eV (GO's), respectively. The two band structures are displayed in Fig. 2. Solid lines show the GO-*GW*A result and dashed lines show the PW-*GW*A result, respectively. The results are virtually identical. The figure thus proves the equivalence of both approaches for calculating the self-energy corrections to the Si bulk band structure.

As Gaussian *GW* basis functions we use *s*, *p*, *d*, and *s** orbitals with decay constants of 0.15 and 0.5 for each atom in the unit cell as in our GO-LDA basis set. But, in addition, *f* orbitals with a decay constant of 0.5 had to be employed to achieve good agreement with the PW-*GW*A calculations mentioned above. Without *f* orbitals in the *GW* basis set, the band-structure energies change by up to 0.5 eV and disagree with the PW-*GW*A results. The gap energy, e.g., is reduced to 0.97 eV while the valence band width is increased to 12.28 eV. Therefore, we have included *f* orbitals in the *GW* basis set in all calculations for the Si(001)-(2×1) surface, to be reported below, in order to obtain accurate quasiparticle band structures. It should be noted at this point that higher angular momenta are included in plane-wave basis sets automatically.

In Table I, we give some information on basis-set size and necessary CPU time for the calculations in the different schemes. Our GO-LDA basis set (containing 40 functions) is much smaller in size than an equivalent PW-LDA basis set that requires some 400 functions (corresponding to an energy cutoff of 15–20 Ry). At the same time, a PW-*GW*A basis set requires much less functions (about 113) than the PW-LDA basis, because the self-energy operator has a smoother real-space structure than the LDA wave functions. On the contrary, the number of Gaussian functions required by the GO-*GW*A method (about 60) is larger than that needed for the GO-LDA because of the necessary inclusion of *f* orbitals in the former. Yet, in the case of bulk Si the number of required plane waves for the *GW* approximation is about twice as large as the number of Gaussian orbitals required for the same accuracy. Therefore, all matrix operations are faster to perform when using Gaussians. But the eval-

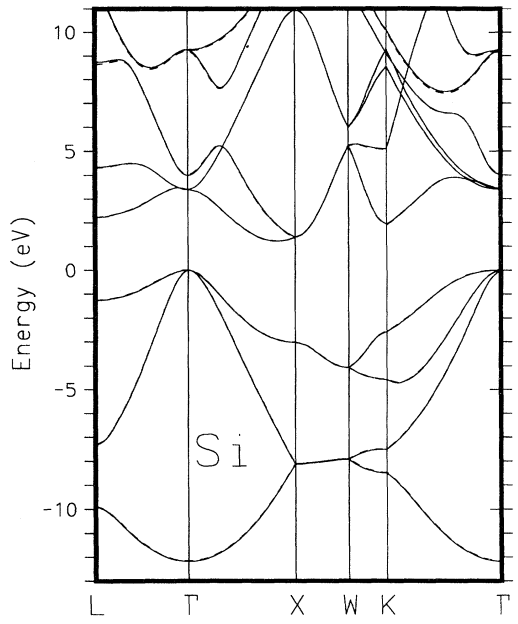

FIG. 2. Calculated quasiparticle band structure for bulk Si, obtained with Gaussian orbitals as *GW* basis functions (GO-*GW*A, solid lines) and with plane waves as *GW* basis functions (PW-*GW*A, dashed lines), respectively.

TABLE I. Numbers of basis functions ν, as required by calculations using plane waves or Gaussian orbitals, respectively. The CPU time τ (given in minutes) refers to the calculation of the self-energy of ten states ($m\mathbf{k}$) at a given $\mathbf{k}$ point for bulk Si or of one state ($m\mathbf{k}_\parallel$) at a given $\mathbf{k}_\parallel$ point for the Si surface. The calculations are carried out on an IBM RS/6000-580.

	LDA PW	LDA GO	*GW*A PW	*GW*A GO
Bulk Si				
ν	~ 400	40	113	60
τ[min]			8	8
Si(001)-(2×1)				
ν	~ 3000	400	~ 1000	520
τ[min]			~ 500	110

uation of the RPA expression of the static polarization [Eq. (10)], as well as the calculation of the self-energies according to Eq. (17) involve the evaluation of the necessary three-center overlap integrals $M_\beta^{mn}(\mathbf{k},\mathbf{q})$. This requires considerable computation time. Nevertheless, we find that both ways of evaluating the *GW* approximation for *bulk* Si require about the same computation time, namely about eight CPU minutes on an IBM RS/6000-580 for the self-energies of ten states at one **k** point (not counting for the calculation of the dielectric matrices). If the studied system contains, however, more localized wave functions the PW-*GW*A requires much more basis functions whereas the number of required Gaussian orbitals in the GO-*GW*A will be roughly the same as in bulk Si. In this case our GO-*GW*A method is much faster than the PW-*GW*A approach. This holds for systems like diamond or SiC as well as for transition metal elements containing strongly localized *d* electrons. We note in passing, that related work on II-VI semiconductors including strongly localized, occupied *d* orbitals is in progress. In that case the GO-*GW*A performs more than an order of magnitude faster than the PW-*GW*A. Here we present applications of the GO-*GW*A method to the Si(001) surface. Also in this case, the GO-*GW*A performs much faster than the PW-*GW*A, as discussed below.

B. The Si(001)-(2×1) surface

Since the (2×1)-reconstructed Si(001) surface has been investigated very carefully both experimentally and theoretically, we consider it as a proper test system for our new approach to the *GW*A. In particular, two *GW*A calculations for the Si(001) surface have been reported recently.[20,21] However, they are both based on model dielectric functions. Kress *et al.*[20] investigated the (2×1)-reconstructed surface while Northrup[21] studied the low-temperature $c(4\times2)$ surface.

It is well known from low-energy-electron-diffraction experiments[22–24] as well as from scanning tunneling microscopy (STM) studies[25] that at room temperature (RT) a nominal (2×1) reconstruction occurs. Formation of dimers by pairing of neighboring surface atoms has been identified as the main building block of the reconstructed surface geometry.[26] Empirical tight-binding calculations by Chadi,[27] as well as DFT calculations[17,28–30] carried out for a (2×1) unit cell show that the total energy of the surface is lowered if the dimers are asymmetric, i.e., if one atom is moved down (down atom) while the other one is moved up (up atom) with respect to the ideal surface plane. As a consequence of this buckling, there are two different types of dangling bonds. The occupied one ($D_{\rm up}$) is mainly localized at the up atom while the empty one ($D_{\rm down}$) is attributed to the down atom. Both dangling-bond states are detected in photoemission and inverse photoemission experiments, respectively.[22–24] The existence of asymmetric dimers is also supported by the results of high-resolution core-level spectroscopy studies.[31] Detailed STM studies[25] have shown that the Si(001)-(2×1) surface is not well ordered at RT. A high degree of defects and a coexistence of asymmetric and symmetric dimers have been observed in these experiments. This is not contradictory to the theoretical result that asymmetric dimers are the building blocks of the reconstruction. If a buckled dimer switches between its left- and right-tilted configuration, as suggested in Ref. 29, within a time period that is much shorter than the time resolution of the experiment, the experimental result will necessarily be a symmetric image. This can be understood as a mean value of two asymmetric dimers with opposite buckling directions.

At low temperatures or if dimer flipping is suppressed by neighboring defects, the interaction among the occupied dangling bonds becomes important. It is energetically favorable when the occupied $D_{\rm up}$ states are as far away from each other as possible. This can be achieved by an alternating orientation of buckling along a row of dimers and leads to a $p(2\times2)$ or to a $c(4\times2)$ reconstruction. Both patterns have been detected in LEED experiments. This can be understood as an indirect proof of the asymmetric buckling of the dimers. Recent STM measurements[32–34] carried out at temperatures below 200 K show that the major part of the Si(001) surface consists of asymmetric dimers.

All our calculations are performed using a supercell geometry containing eight layers of Si and six layers of vacuum repeated periodically. Because of the finite thickness of the Si slab in each unit cell, the dangling bonds of its two surfaces will interact with each other and split. (In our LDA calculations we get a splitting of up to 1 eV at the J' point of the surface Brillouin zone.) This problem is easily avoided by hydrogen saturation of one surface of the slab. We keep the atoms of one-half of the slab in their ideal positions. The dangling bonds of this ideal surface are saturated with hydrogen atoms.

At the opposite, nonideal (2×1)-reconstructed surface three layers of Si are relaxed. The geometry optimization[35] yields an asymmetric dimer configuration with a dimer-bond length of 2.29 Å and a buckling angle of 20.2°. Our LDA calculations are carried out using 20 Gaussian basis functions at each atom (decay constants of 0.15 and 0.5 for Si, 0.25 and 1.5 for H). We employ the separable Si pseudopotential of Stumpf, Gonze, and Scheffler[14] and the H pseudopotential reported by Gygi.[36] Two special points in the irreducible part of the surface Brillouin zone are used.

Concerning the surface band structure, we concentrate on the dangling-bond states. They are displayed in Fig. 3 (LDA, dashed lines) together with the quasiparticle energies (full lines) as obtained by our *GW* correction scheme (see below). All band-structure energies refer to the valence band maximum of the bulk crystal. For this purpose we have aligned the potential in the middle of the slab with the bulk potential.

In LDA we obtain an occupied state $D_{\rm up}$ with a bandwidth of 0.95 eV. Its dispersion along the ΓJ line and along KJ' is very weak because the dangling bonds are separated by 7.60 Å in the direction orthogonal to the dimer rows. The dispersion along JK and $J'\Gamma$ is stronger due to the smaller dangling-bond distance of 3.80 Å along the dimer rows. The unoccupied state $D_{\rm down}$ has a larger

bandwidth of 1.25 eV with its minimum value at J'. We find a weak dispersion orthogonal to the dimer rows and a strong dispersion along the dimer rows for the same reasons as discussed above.

In LDA we obtain an indirect fundamental surface band gap of 0.20 eV. The direct energy gap between the dangling-bond bands ranges from 0.70 eV at Γ to 1.55 eV along the $\Gamma J_2'$ line. It should be noted that the dispersion of the D_{down} state depends very sensitively on the surface geometry, especially on the buckling angle. A reduction of the buckling angle reduces the gap as has also been found by other authors.[29]

Starting from this LDA calculation, the *GW* approximation is evaluated following the method described in Sec. III. We calculate the full static dielectric function of the slab in RPA. In the *GW* basis set we employ the same Gaussian orbitals for the Si atoms as used for the Si bulk system, i.e., s, p, d, and s^* orbitals with decay constants of 0.15 and 0.5 and additional f orbitals with the decay constant 0.5. For the H atoms we use s, p, d, and s^* orbitals with the decay constant 0.3.

For the calculation of the dielectric matrix as well as for the evaluation of the self-energies, we employ two special points in the irreducible part of the surface Brillouin zone. In order to account for the long-range behavior of the head of the dielectric function, we employ eight special $\mathbf{k}_{\|}$ points for the calculation of this quantity in the case of $\mathbf{q} \to \mathbf{0}$ (see Sec. III C). The dielectric constant calculated for the slab geometry amounts to $\epsilon_{\infty}^{\text{slab}}=12.3$. This value is smaller than $\epsilon_{\infty}^{\text{bulk}}=13.0$ calculated for the bulk (see Sec. IV A).

The correction of the LDA energies according to Eq. (4) results originally in a shift of D_{up} to lower energies by –0.35 to –0.20 eV (displayed by crosses in Fig. 1). This energy shift is smaller than that of the valence band maximum (VBM) in bulk Si of –0.40 eV (displayed in Fig. 1). It should be noted that we have defined the valence band maximum as $E_{\text{VBM}} = 0$ eV for both the LDA and the *GW*A results in Fig. 3. This way, relative to E_{VBM} the D_{up} band is eventually shifted to *higher* energies by +0.05–+0.20 eV. The bandwidth of D_{up} is increased a little bit by the *GW* correction and amounts to 1.00 eV (cf. 0.95 eV in LDA). The shape and dispersion of D_{up} is hardly affected by the quasiparticle corrections. The quasiparticle energies of the dangling bonds are displayed in Fig. 3 (solid lines).

The dangling-bond band D_{up} compares well with experimental data from angle-resolved photoemission spectroscopy,[22,23] especially with those of Ref. 23 (filled circles in Fig. 3). This holds for both the LDA and the *GW*A results (dashed and solid lines, respectively). At Γ and J' and between Γ and J we find very good agreement of our *GW* quasiparticle band-structure energies with the experimental values. Between J' and Γ and along the [010] direction, the quasiparticle energies come out a little bit higher than found experimentally. Both the LDA and the *GW*A calculations yield distinct maxima of D_{up} along JK, $J'\Gamma$, and $\Gamma J_2'$, which have been observed in LDA results of other authors as well,[17,29] but they are not observed in photoemission. The two experimentally determined band structures of Refs. 22 and 23 differ from

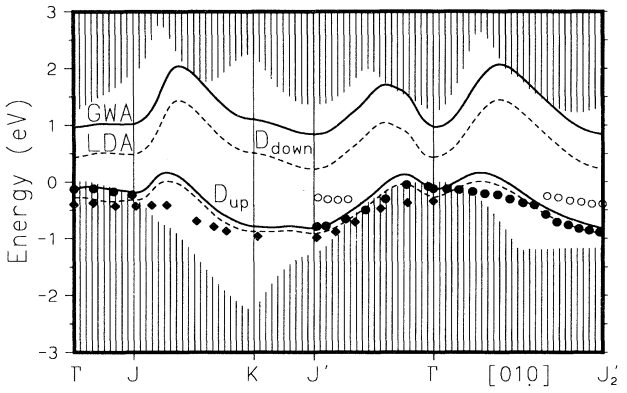

FIG. 3. Calculated dangling-bond bands of Si(001)-(2×1). Solid lines, *GW*A energies (full RPA calculation). Dashed lines, LDA energies. The symbols refer to photoemission experiments by Uhrberg *et al.* (diamonds) (Ref. 22) and by Johansson *et al.* (filled and open circles) (Ref. 23).

each other by up to 0.3 eV. This may be due to difficulties in determining the position of the valence band maximum E_{VBM} with respect to the Fermi level E_F. At J' and J_2' some states (open circles in Fig. 3) are detected at about –0.3 eV.[23] They do not fit to the calculated dispersion of the D_{up} state. These measured features may be explained as a second dangling-bond band resulting from local $c(4\times2)$ regions existing already at RT on the nominal (2×1)-reconstructed surface. In a $c(4\times2)$ reconstruction, for example, $J_{2\times1}$ and $J'_{2\times1}$ are backfolded onto $J_{4\times2}$ and become equivalent. Thus the dangling-bond states D_{up} at $J_{2\times1}$ and $J'_{2\times1}$ may both be detected at the same wave vector for the $J'_{2\times1}$ high-symmetry point even on the nominal (2×1)-reconstructed surface. The experimental bandwidth of the D_{up} state is somewhat smaller than our results from the LDA as well as from the *GW* calculation. This feature may be caused by local $c(4\times2)$ regions existing already at RT, as well. In such regions, the distance between the occupied dangling bonds is increased due to the alternating orientation of the buckled dimers. This leads to a reduction of the bandwidth of the D_{up} state, as has been shown by Northrup's detailed calculation for the $c(4\times2)$ reconstructed surface.[21] The resulting bandwidth is in good agreement with experimental data.

The unoccupied state D_{down} is shifted by the *GW* quasiparticle corrections to higher energies by +0.10–+0.25 eV, which is a smaller shift than found for the lowest bulk conduction band states (+0.25–+0.40 eV; see Fig. 1). If we refer the quasiparticle D_{down} band again to E_{VBM}, the D_{down} state is eventually shifted to higher energies by +0.50–+0.65 eV. Its bandwidth amounts to 1.20 eV, which is a little bit smaller than the LDA value of 1.25 eV. The indirect fundamental surface band gap, which results as 0.20 eV in LDA, is increased by 0.50 eV and amounts to 0.70 eV. The direct energy gap between D_{up} and D_{down} ranges from 1.10 eV at Γ to 2.00 eV along the JK line. There are a few inverse photoemission data on D_{down}.[24] We do not include them in Fig. 3 since their ab-

solute energetic position with respect to E_{VBM} is difficult to determine from the experimental data. From optical spectroscopy experiments on the Si(001)-(2×1) surface, the energy of the empty D_{down} state at Γ can be estimated to be 1.1 eV above E_{VBM}.[37] This agrees well with our theoretical result of 0.95 eV. The fundamental surface gap extracted from optical absorption spectroscopy[37] is 0.44 eV, which is smaller than our calculated value of 0.70 eV. STM measurements,[38] on the contrary, show a larger fundamental surface gap of 0.9 eV. Surface photovoltage measurements[39] have yielded a value of 0.64 eV.

It is quite revealing to analyze the exchange and correlation contributions to the total self-energy at the surface in comparison to the bulk. To this end, we present in Fig. 4 self-energy values of the dangling-bond states and compare them with the self-energies obtained for Si bulk states. For the occupied D_{up} state, the exchange contribution Σ_x to the self-energy is negative and its absolute value is smaller than that of occupied bulk states. The correlation contribution Σ_c is positive and is smaller than that of respective bulk states. In total, D_{up} shows a smaller self-energy (–11.4 to –10.2 eV) than the upper bulk valence band states (–11.65 eV for the VBM). For the unoccupied D_{down} state, the absolute value of the exchange contribution to the self-energy is larger than for the low conduction band bulk states. The absolute value of the correlation part, on the other hand, is a little bit smaller. The total self-energy amounts to –9.9 to –9.2 eV

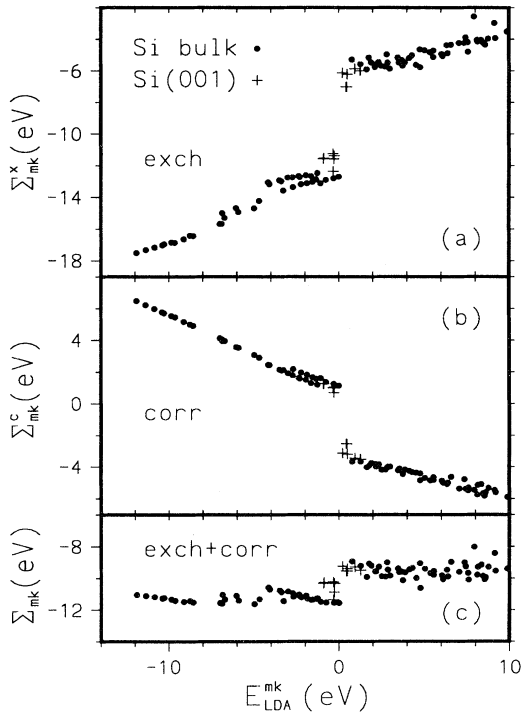

FIG. 4. Calculated self-energies for bulk Si (dots) and for the dangling bonds D_{up} and D_{down} of the Si(001)-(2×1) surface (crosses). We present the exchange contribution (a) and the correlation contribution (b) to the self-energies $\langle m\mathbf{k}|\Sigma(E_{m\mathbf{k}})|m\mathbf{k}\rangle$ as well as the total self-energy (c) which is given by the sum of both.

for D_{down} and is larger than for the low conduction band bulk states [–9.05 eV for the conduction band maximum (CBM)]. In the bulk calculation, the self-energies of the VBM and of the CBM are separated by 2.60 eV. The self-energy separation between occupied and unoccupied states corresponds to the existence of a fundamental gap in the band structure. The large difference between the self-energies of occupied and unoccupied bulk states is strongly reduced at the surface [see Fig. 4(c)]. In consequence, the fundamental gap between the dangling-bond bands at the surface is smaller than the bulk gap. It is interesting to note in Fig. 4 that the total self-energies for the respective states at energy $E_{m\mathbf{k}}$ show only a very weak energy dependence while the separate exchange and correlation contributions show significant but opposite energy dependencies, as has been found for the homogeneous electron gas, as well.[40] For the values of the total self-energy, we observe the following trend:

$$\Sigma(\mathrm{VBM}) < \Sigma(D_{\mathrm{up}}) < \Sigma(D_{\mathrm{down}}) < \Sigma(\mathrm{CBM}) \qquad (21)$$

In Fig. 1, we compare the quasiparticle shifts of the dangling-bond states (crosses) at the Si(001)-(2×1) surface to the quasiparticle shifts of the Si bulk states (dots). As mentioned above, the absolute shifts of the dangling-bond states are smaller than those of the bulk states, resulting in a reduction of the step within this figure. Nevertheless the correction values of the surface states fit well to the bulk values. The distinct step function between occupied and unoccupied bulk states with a correction step of 0.7 eV is smoothed by the corrections for the surface states whose energies lie in the fundamental energy gap.

The calculated renormalization constants $Z_{m\mathbf{k}}$ for the surface and bulk states do not differ very much. The values for both the occupied and empty dangling-bond bands range from 0.76 to 0.78. For the bulk states near the fundamental gap we find somewhat larger values of 0.79–0.81. The dangling bonds are highly localized at the surface. One cannot expect the screening properties at the surface to be the same as in the bulk. Therefore, the perturbation operator $\Sigma(\mathbf{r},\mathbf{r}',E) - V_{\mathrm{xc}}(\mathbf{r})\cdot\delta(\mathbf{r}-\mathbf{r}')$ at the surface will differ from that in the bulk. This real-space difference as well as the energetic differences between dangling-bond states at the surface and bulk states explain the difference between the self-energies $\Sigma_{m\mathbf{k}}$ of respective states and their different quasiparticle corrections, as discussed above. The good fit of the surface state corrections to the bulk quasiparticle corrections (according to Figs. 1 and 4) is a quite revealing result.

Finally we address the efficiency of our GO-*GW*A calculations in comparison with PW-*GW*A calculations (see Table I). A highly converged surface LDA calculation using plane waves would require about 3000 basis functions and the evaluation of the *GW* approximation would require more than 1000 plane waves. Again, the *GW* basis set requires less basis functions than the LDA because of the smooth real-space structure of the *GW* self-energy operator as compared to the LDA wave functions. If Gaussian orbitals are employed as basis functions, only 400 GO's are needed to perform the LDA. In this case, the *GW* calculation again demands more Gaussians due

to the need for including f orbitals but 520 GO's turn out to be sufficient for good convergence (see Sec. IV A).

Contrary to the bulk situation, for the slab system the GO-GWA method performs much faster than the PW-GWA. The Gaussian method requires only about 110 CPU minutes for the self-energy of one state ($m\mathbf{k}_{\|}$) whereas the PW-GWA method would require more than 500 CPU minutes. The same relation holds for the calculation of the dielectric matrix in the RPA for one $\mathbf{q}$ point. It would require more than 100 CPU hours for plane waves as opposed to only 20 CPU hours for Gaussians. In conclusion, we find that the evaluation of the GW approximation for surface systems using Gaussian orbital basis functions is a relatively fast and very efficient method.

C. Model dielectric functions for Si(001)-(2×1)

Finally, we address the question of the usefulness of model dielectric functions for quasiparticle band-structure calculations of surfaces. For more complex systems the evaluation of the RPA expression for the dielectric function may quickly become too demanding, even if localized basis functions are employed. In such cases appropriate dielectric model functions could be particularly useful. In Ref. 9 we have investigated both diagonal and nondiagonal model functions for bulk semiconductors. We found that some nondiagonal functions that account for local fields [labeled (b2) and (c2) in Ref. 9] yield quasiparticle band-structure energies in good agreement with GWA calculations employing the full RPA dielectric function. In addition to the RPA dielectric matrix, we have employed these two model functions in our GO-GWA calculations for the Si surface, as well. In Table II we compare the respective quasiparticle band-structure energies with those resulting from the full RPA GW calculation discussed in Sec. IV B. The nondiagonal model function (b2) has been proposed by Hybertsen and Louie,[41] based on a diagonal model function by Levine and Louie.[42] Using this model function, we have obtained quasiparticle bulk band structures in good agreement with those of our RPA GW calculation in Ref. 9. However, we found that the conduction bands come out somewhat lower in energy than in the RPA GW method. For bulk Si we obtain a fundamental energy gap of 0.95 eV for the theoretical lattice constant using ϵ_∞=13.0, as compared to 1.23 eV calculated with the RPA expression (see above).

For the application to Si(001)-(2×1), we employ the dielectric constant ϵ_∞=12.3 as calculated in the RPA (see above). Contrary to the results presented in Sec. IV B, the D_{up} state is now shifted to lower energies (especially at K and J') with a mean correction of −0.05 eV with respect to the LDA. The correction of the D_{down} state amounts to about +0.35 eV with respect to the LDA. This value is smaller than the one that results when the RPA dielectric function is used, as could be expected from the bulk results. Nevertheless, the fundamental surface gap amounts to 0.65 eV in close agreement to the RPA result of 0.70 eV. This is due to the lower energy of D_{up} resulting for the (b2) model dielectric function.

This same dielectric model function has been employed by Northrup[21] to calculate the quasiparticle energies of the Si(001)-c(4×2) surface. To account for the vacuum in the slab supercell, Northrup used a dielectric constant of ϵ_∞=10.0, which is considerably smaller than the ex-

TABLE II. Calculated quasiparticle energies (in eV) for the dangling-bond states of Si(001)-(2×1) of a full GW calculation including the RPA dielectric function (second column), as well as GW calculations using model dielectric functions (columns 3–5; see text) using dielectric constants of ϵ_∞=12.3 and ϵ_∞=10.0, respectively. The model function (b2) has been proposed by Hybertsen and Louie[41] while the model function (c2) is a combination of the models by Bechstedt *et al.*[43] and by Falter *et al.*[44,45] We display the respective LDA results in the first column and the last column contains experimental data.

Si(001)		LDA	GWA RPA	GWA (b2)	GWA (b2)	GWA (c2)	Expt.
	ϵ_∞			12.3	10.0	12.3	
D_{up}	Γ	−0.30	−0.15	−0.25	−0.30	−0.15	−0.4[a],−0.1[b]
	J	−0.30	−0.20	−0.30	−0.35	−0.15	−0.4[a],−0.2[b]
	K	−0.85	−0.80	−1.00	−1.05	−0.85	−1.0[a]
	J'	−0.90	−0.85	−1.05	−1.05	−0.90	−1.0[a],−0.8[b]
	bandwidth	0.95	1.00	1.05	1.05	1.05	0.6[a],0.8[b]
	mean shift		+0.15	−0.05	−0.10	+0.10	
D_{down}	Γ	0.45	0.95	0.75	0.80	0.85	1.1[c]
	J	0.50	1.00	0.80	0.85	0.90	
	K	0.50	1.10	0.90	0.95	1.00	
	J'	0.20	0.85	0.65	0.70	0.70	
	bandwidth	1.25	1.20	1.10	1.10	1.10	
	mean shift		+0.50	+0.35	+0.40	+0.40	
$E_{\mathrm{gap}}^{\mathrm{ind}}$		0.20	0.70	0.65	0.70	0.55	0.44[c],0.9[d]

[a]See Ref. 22. [c]See Ref. 23.
[b]See Ref. 37. [d]See Ref. 38.

perimental value of 11.7 for bulk Si. If we use $\epsilon_\infty=10.0$ instead of our RPA result $\epsilon_\infty^{\mathrm{RPA}}=12.3$ in our model calculation, an additional downshift of D_{up} by −0.05 eV to lower energies results, whereas the D_{down} state is shifted to higher energies by some 0.05 eV, in addition. Our final quasiparticle corrections of −0.10 eV (D_{up}) and +0.40 eV (D_{down}) relative to E_{VBM} for the (2×1) surface agree well with Northrup's results of −0.15 eV and +0.33 eV for the $c(4\times2)$ surface, respectively. This close agreement indicates that the evaluation of the *GW* approximation using localized basis functions yields very reliable results agreeing well with those of PW-*GW*A calculations. Furthermore, it shows that the quasiparticle shifts of the *GW* approach are very similar for both the Si(001)-(2×1) and the Si(001)-$c(4\times2)$ surface. The agreement is even more obvious, if we consider the opening of the surface gap, which results from the downward shift of D_{up} and the upward shift of D_{down}. It amounts to $\Delta E_{\mathrm{gap}}=0.48$ eV in Northrup's and to $\Delta E_{\mathrm{gap}}=0.50$ eV in our results. It is thus only the absolute energy position of the two bands D_{up} and D_{down} with respect to E_{VBM} that differs for the (2×1) and the $c(4\times2)$ surface. Their relative energy separation is virtually the same for the two configurations.

In addition, we have investigated the usefulness of a second model dielectric function [labeled (c2) in Ref. 9] that uses a model function proposed by Bechstedt *et al.*[43] for the diagonal matrix elements while the nondiagonal matrix elements are taken into account in the form suggested by Falter *et al.*[44,45] When applied to bulk Si with $\epsilon_\infty=13.0$, this model leads to a fundamental energy gap of 1.09 eV [larger than for (b2), but smaller than for the RPA calculation]. For the Si(001)-(2×1) surface, we obtain (with $\epsilon_\infty=12.3$) an upward shift of +0.40 eV for D_{down} that is again larger than for (b2) and smaller than for the RPA. Referred to E_{VBM}, the D_{up} state is shifted to higher energies by about +0.10 eV. This agrees roughly with our RPA result. In total, the fundamental surface gap amounts to 0.55 eV in this case, which is smaller than for the RPA dielectric function and for the (b2) model calculation. Kress *et al.*[20] calculated quasiparticle corrections for the dangling bonds of the Si(001)-(2×1) surface, employing a simplified *GW* approach using a model dielectric function similar to (c2). The authors obtained a correction of the D_{up} state at Γ relative to E_{VBM} of −0.03 eV (+0.15 eV in our calculation) while the D_{down} state at Γ shifts to higher energies by +0.85 eV (cf. +0.40 eV in our calculation). The fundamental surface gap calculated in Ref. 20 thus amounts to 0.80 eV as compared to our result of 0.55 eV for this model. These deviations may arise from systematic differences between the approach of Ref. 20 and our method.

From our results we conclude, that both nondiagonal model dielectric functions (b2) and (c2) give quasiparticle band-structure energies for Si(001)-(2×1), which agree well with those of our calculations based on the full RPA dielectric matrix. Differences in the approach for the evaluation of the *GW*A, as encountered between our method and that of Ref. 20, give rise to deviations of the order of 0.25 eV for the surface gap. The close agreement of the calculated quasiparticle corrections using the model (b2) between our results for the (2×1) surface and Northrup's results for the $c(4\times2)$ surface is quite rewarding. We conclude from these comparisons of our calculations based on model dielectric functions with the results of Kress *et al.*[20] and Northrup,[21] as well as, with our results based on the full RPA dielectric matrix that appropriate model functions can certainly be very useful to treat more complex reconstructions with larger unit cells.

V. SUMMARY

The *GW* approximation of Hedin and Lundqvist has been applied in the literature with great success to real, inhomogeneous systems, especially to bulk semiconductors. However, more complex systems like surfaces are very difficult to treat by this method since it scales very badly with the size of the system ($\sim N^3$). The investigation of such systems by the *GW* scheme requires more efficient algorithms than the common evaluation of Fourier transforms enforcing large numbers of plane waves. In this paper we have represented all two-point functions of the *GW*A using a relatively small basis set of Gaussian orbitals. This way large Fourier transform matrices are avoided. We have tested this GO-*GW*A method in the case of bulk Si and found excellent agreement of all results with those of an equivalent PW-*GW*A calculation. This proves the high flexibility and the usefulness of our GO basis set.

With the use of such basis functions we have calculated the quasiparticle band structure for the Si(001)-(2×1) surface including the static dielectric function fully computed within the random-phase approximation and extending it to finite frequencies by a generalized plasmon-pole model. We find that the dangling-bond states D_{up} and D_{down} shift nearly rigidly in energy with respect to the LDA results so that their dispersion remains almost unchanged. The calculated band structure of the occupied state D_{up} agrees well with results from photoemission experiments. The fundamental surface gap is opened by 0.50 eV with respect to the LDA gap of 0.20 eV by the quasiparticle corrections and amounts to 0.70 eV in agreement with the scarce experimental data. We have tested two model dielectric functions in our surface calculations, in addition, and found results that are similar to those obtained employing the full RPA dielectric matrix. However, the final quasiparticle shifts resulting for the model functions show some systematic deviations from the results obtained with the RPA dielectric matrix. The energy shifts resulting from our calculations for one of the model dielectric functions (b2) applied to Si(001)-(2×1) agree very well with those of a previous PW-*GW*A study by Northrup for Si(001)-$c(4\times2)$ based on the same model dielectric function. This additionally confirms the appropriateness of our basis set for surface systems. In summary, we have presented a method for evaluating the *GW* approximation that is considerably more efficient than the PW-*GW*A approach because it performs much faster than the conventional plane wave

approach based on Fourier representations of the occurring two-point functions. This method will turn out to be particularly useful if more complex surface reconstructions or systems including highly localized orbitals are to be investigated.

ACKNOWLEDGMENT

We would like to thank M. Sabisch for calculating the optimized surface geometry of the Si(001)-(2×1) surface within the LDA and for fruitful discussions.

APPENDIX: THE SYMMETRIZED DIELECTRIC FUNCTION

Starting from the polarization P, the dielectric function ϵ is calculated by a convolution with the Coulomb interaction, $\epsilon=1-vP$. If evaluated, e.g., in Fourier space, this matrix is not Hermitian. Even worse, its $(\mathbf{G}=0,\mathbf{G}'\neq 0)$ elements $\epsilon_{\mathbf{0},\mathbf{G}'}(\mathbf{q})$ diverge as $1/q$ if $\mathbf{q}\to\mathbf{0}$ because of the divergency of the Coulomb interaction, $v_{\mathbf{G},\mathbf{G}'}(\mathbf{q})=4\pi e^2/|\mathbf{q}+\mathbf{G}|^2\delta_{\mathbf{G},\mathbf{G}'}$. The localized basis set of Gaussian orbitals used in this work leads to the same complications as can be seen from the close relationship between the Gaussian representation and the Fourier representation [see Eq. (20)]. To overcome this problem, an auxiliary dielectric function $\tilde{\epsilon}$ can be defined in Fourier space as

$$\tilde{\epsilon}_{\mathbf{G},\mathbf{G}'}(\mathbf{q}) := \frac{|\mathbf{q}+\mathbf{G}|}{|\mathbf{q}+\mathbf{G}'|}\epsilon_{\mathbf{G},\mathbf{G}'}(\mathbf{q}) \tag{A1}$$

$$= \sum_{\mathbf{G}'',\mathbf{G}'''} |\mathbf{q}+\mathbf{G}|\delta_{\mathbf{G},\mathbf{G}''}\cdot\epsilon_{\mathbf{G}'',\mathbf{G}'''}(\mathbf{q})\cdot\frac{1}{|\mathbf{q}+\mathbf{G}'|}\delta_{\mathbf{G}''',\mathbf{G}'} \tag{A2}$$

$$= \delta_{\mathbf{G},\mathbf{G}'} - 4\pi e^2 \sum_{\mathbf{G}'',\mathbf{G}'''} \frac{1}{|\mathbf{q}+\mathbf{G}|}\delta_{\mathbf{G},\mathbf{G}''}\cdot P_{\mathbf{G}'',\mathbf{G}'''}(\mathbf{q})\cdot\frac{1}{|\mathbf{q}+\mathbf{G}'|}\delta_{\mathbf{G}''',\mathbf{G}'} \quad . \tag{A3}$$

In a basis-independent notation, these equations read

$$\tilde{\epsilon} := \tilde{v}^{-1}\epsilon\tilde{v} = 1-\tilde{v}P\tilde{v} \tag{A4}$$

with an auxiliary function $\tilde{v}$. Its Fourier representation $\tilde{v}_{\mathbf{G},\mathbf{G}'}(\mathbf{q})$ is given by

$$\tilde{v}_{\mathbf{G},\mathbf{G}'}(\mathbf{q}) = \frac{2\pi^{1/2}e}{|\mathbf{q}+\mathbf{G}|}\delta_{\mathbf{G},\mathbf{G}'} \tag{A5}$$

as can be seen from comparing Eqs. (A3) and (A4). Its transformation into real space results in

$$\tilde{v}(\mathbf{r},\mathbf{r}') = \frac{e}{\pi^{3/2}\cdot|\mathbf{r}-\mathbf{r}'|^2} \quad . \tag{A6}$$

The convolution of the auxiliary function $\tilde{v}$ with itself yields the Coulomb interaction:

$$\int \tilde{v}(\mathbf{r},\mathbf{r}'')\tilde{v}(\mathbf{r}'',\mathbf{r}')d^3r'' = v(\mathbf{r},\mathbf{r}'), \tag{A7}$$

or

$$\sum_{\mathbf{G}''}\tilde{v}_{\mathbf{G},\mathbf{G}''}(\mathbf{q})\tilde{v}_{\mathbf{G}'',\mathbf{G}'}(\mathbf{q}) = v_{\mathbf{G},\mathbf{G}'}(\mathbf{q}) \quad . \tag{A8}$$

We also note the following relation that results from Eq. (A7):

$$\tilde{v}^{-1}v = \tilde{v} \quad . \tag{A9}$$

If represented with respect to basis functions, the symmetric dielectric function of Eq. (A4) is given by a Hermitian matrix with elements that remain finite if $\mathbf{q}\to\mathbf{0}$. After inversion of $\tilde{\epsilon}$, we calculate the screened interaction:

$$W = \epsilon^{-1}v = \left(\tilde{v}\tilde{\epsilon}^{-1}\tilde{v}^{-1}\right)v = \tilde{v}\tilde{\epsilon}^{-1}\tilde{v} \quad . \tag{A10}$$

We have written Eqs. (A4) and (A10) in a general, basis-independent notation. With respect to a set of basis functions, the convolutions must be replaced by the respective matrix multiplications [see, e.g., Eq. (12)]. We use the respective symmetric dielectric matrices throughout our work.

[1] L. Hedin, Phys. Rev. **139**, A796 (1965).
[2] L. Hedin and S. Lundqvist, in *Solid State Physics: Advances in Research and Application*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969), Vol. 23, p. 1.
[3] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[4] S. L. Adler, Phys. Rev. **126**, 413 (1962).
[5] N. Wiser, Phys. Rev. **129**, 62 (1963).
[6] W. von der Linden and P. Horsch, Phys. Rev. B **37**, 8351 (1988).
[7] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10 159 (1988).
[8] R. Hott, Phys. Rev. B **44**, 1057 (1991).
[9] M. Rohlfing, P. Krüger, and J. Pollmann, Phys. Rev. B **48**, 17 791 (1993).
[10] D. L. Johnson, Phys. Rev. B **9**, 4475 (1974).
[11] F. Aryasetiawan and O. Gunnarson, Phys. Rev. B **49**,

16 214 (1994).
[12] F. Aryasetiawan and O. Gunnarson, Phys. Rev. Lett. **74**, 3221 (1995).
[13] S. Obara and A. Saika, J. Chem. Phys. **84**, 3963 (1986).
[14] R. Stumpf, X. Gonze, and M. Scheffler (unpublished); X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).
[15] D. M. Ceperley and B. I. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[16] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[17] P. Krüger and J. Pollmann, Phys. Rev. B **47**, 1898 (1993).
[18] V. Fiorentini, Phys. Rev. B **46**, 2086 (1992).
[19] *Semiconductors. Physics of Group IV Elements and III-V Compounds*, edited by K.-H. Hellwege and O. Madelung, Landolt-Börnstein, New Series, Group III, Vol. 17, Pt. a (Springer, Berlin, 1982); *Semiconductors. Intrinsic Properties of Group IV Elements and III-V, II-VI, and I-VII Compounds*, edited by K.-H. Hellwege and O. Madelung, Landolt-Börnstein, New Series, Group III, Vol. 22, Pt. a (Springer, Berlin, 1982).
[20] C. Kress, M. Fiedler, and F. Bechstedt, in *Proceedings of the Fourth International Conference on the Formation of Semiconductor Interfaces*, edited by B. Lengeler, H. Lüth, W. Mönch, and J. Pollmann (World Scientific, Singapore, 1994), p. 19.
[21] J. E. Northrup, Phys. Rev. B **47**, 10 032 (1993).
[22] R. I. G. Uhrberg, G. V. Hansson, J. M. Nicholls, and S. A. Flodström, Phys. Rev. B **24**, 4684 (1981).
[23] L. S. O. Johansson, R. I. G. Uhrberg, P. Mårtensson, and G. V. Hansson, Phys. Rev. B **42**, 1305 (1990).
[24] L. S. O. Johansson and B. Reihl, Surf. Sci. **269/270**, 810 (1992).
[25] R. J. Hamers, R. M. Tromp, and J. E. Demuth, Phys. Rev. B **34**, 5343 (1986).
[26] See, e.g., D. Haneman, Rep. Prog. Phys. **50**, 1045 (1983).
[27] D. J. Chadi, Phys. Rev. Lett. **43**, 43 (1979).
[28] M. T. Yin and M. C. Cohen, Phys. Rev. B **24**, 2303 (1981).
[29] J. Dabrowski and M. Scheffler, Appl. Surf. Sci. **56**, 15 (1992).
[30] P. Krüger and J. Pollmann, Phys. Rev. Lett. **74**, 1155 (1995).
[31] E. Landemark, C. J. Karlsson, Y.-C. Chao, and R. I. G. Uhrberg, Phys. Rev. Lett. **69**, 1588 (1992).
[32] R. A. Wolkow, Phys. Rev. Lett. **68**, 2636 (1992).
[33] H. Tochihara, T. Amakusa, and M. Iwatsuki, Phys. Rev. B **50**, 12 262 (1994).
[34] D. Badt, H. Wengelnik, and H. Neddermeyer, J. Vac. Sci. Technol. B **12**, 2015 (1994).
[35] M. Sabisch (private communication).
[36] F. Gygi, Phys. Rev. B **47**, 11 692 (1993).
[37] Y. J. Chabal, S. B. Christman, E. E. Chaban, and M. T. Yin, J. Vac. Sci. Technol. A **1**, 1241 (1983).
[38] R. J. Hamers and U. K. Köhler, J. Vac. Sci. Technol. A **7**, 2854 (1989).
[39] W. Mönck, P. Koke, and S. Krueger, J. Vac. Sci. Technol. **19**, 313 (1981).
[40] See, e.g., Sec. 25 of Ref. 2.
[41] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **37**, 2733 (1988).
[42] Z. H. Levine and S. G. Louie, Phys. Rev. B **25**, 6310 (1982).
[43] F. Bechstedt, R. Enderlein, and R. Wischnewski, Phys. Status Solidi B **107**, 637 (1981).
[44] C. Falter (private communication).
[45] C. Falter, M. Klenner, and W. Ludwig, Phys. Status Solidi B **167**, 85 (1991).