PHYSICAL REVIEW B VOLUME 53, NUMBER 7 15 FEBRUARY 1996-I

Generalized Kohn-Sham schemes and the band-gap problem

A. Seidl, A. Görling, P. Vogl, and J. A. Majewski
Walter Schottky Institut and Physik Department, Technische Universität München, 85748 Garching, Germany

M. Levy
Department of Chemistry and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118

(Received 24 August 1995)

As an alternative to the standard Kohn-Sham procedure, other exact realizations of density-functional theory (generalized Kohn-Sham methods) are presented. The corresponding generalized Kohn-Sham eigenvalue gaps are shown to incorporate part of the discontinuity Δ_{xc} of the exchange-correlation potential of standard Kohn-Sham theory. As an example, a generalized Kohn-Sham procedure splitting the exchange contribution to the total energy into a screened, nonlocal and a local density component is considered. This method leads to band gaps far better than those of local-density approximation *and* to good structural properties for the materials Si, Ge, GaAs, InP, and InSb.

I. INTRODUCTION

First-principles calculations based on the Kohn-Sham[1] scheme of density-functional theory[2] (DFT) have successfully predicted and explained a wide range of solid-state properties.[3] Strictly speaking, however, this is true only for cohesive and structural properties, whereas band gaps are typically underestimated by a factor of 2. One approach to solve this gap problem is to consider the energies of quasiparticles and to calculate the electron self-energy in terms of perturbation theory. This approach has been followed by several authors[4–6] invoking Hedin's GW approximation.[7] While this procedure has been quite successful, it does not allow one to calculate structural properties together with energy gaps in a self-consistent way. Since, however, principal band gaps E_g of semiconductors are differences of ground-state energies of N and $N\pm1$ particle systems, $E_g=E(N+1)+E(N-1)-2E(N)$, they are in principle accessible by DFT.

So far, DFT has been applied almost exclusively within the Kohn-Sham formalism where E_g is the difference of single-particle eigenvalues plus a contribution that originates in the discontinuity Δ_{xc} of the exchange-correlation potential at integer particle numbers. In practice, the Kohn-Sham scheme is usually carried out within the local-density approximation (LDA), whose exchange-correlation potential exhibits no discontinuity ($\Delta_{\mathrm{xc}}=0$). Indeed, the energy gaps within LDA show a large discrepancy with experiment that is caused both by the lack of the discontinuity and errors in the single-particle eigenvalues resulting from the approximative nature of the functional.

In this paper we show that there are numerous exact realizations of DFT besides the standard Kohn-Sham procedure that yield the correct total energy of the system and lead to self-consistently determined single-particle eigenvalues of the N-particle system whose differences already incorporate part of the discontinuity Δ_{xc}. In particular, these schemes lend themselves to approximations that allow the self-consistent calculation of both structural properties and energy gaps of semiconductors in good agreement with experiment, in contrast to LDA. We shall call these alternative realizations of DFT *generalized Kohn-Sham (GKS) schemes*.

We show that the GKS framework that we develop in this paper constitutes a rigorous basis for employing nonlocal potentials such as Hartree-Fock-like exchange potentials or screened nonlocal exchange potentials within Kohn-Sham-type approaches. An example of such an approach with screened nonlocal exchange and LDA correlation potentials was first proposed by Bylander and Kleinman on empirical grounds,[8] and was shown to give significantly better energy gaps in Si. This method was also applied to atoms in Refs. 9 and 10. In this work, we show that this method is firmly rooted within DFT and we present results for structural properties and band gaps in several semiconductors based on this screened-exchange GKS scheme. We show that this approach, which we shall call the screened-exchange LDA method (sX-LDA), compares very favorably with experiment.

The paper is organized as follows. In Sec. II, the GKS formalism is developed, generalizing earlier work of Ref. 11. Several examples are given in Sec. III. In particular, a scheme based on screened nonlocal exchange functionals is discussed. The analog of the local-density approximation of this scheme is shown to lead to the sX-LDA method. In Sec. IV, we analyze the band gaps emerging from the various GKS schemes and compare them with the standard Kohn-Sham eigenvalue gaps. Based on nonlocal pseudopotentials and a plane-wave representation, we present energy gaps, valence-band widths, lattice constants, and bulk moduli calculated within the sX-LDA method for Si, Ge, GaAs, InP, and InSb in Sec. V. In addition, we develop a perturbative version of the sX-LDA scheme that is computationally much less demanding but still gives results in fair agreement with the self-consistent version. In addition, several screening models for the nonlocal exchange potential are discussed. A summary is given in Sec. VI. In Appendix A, we prove the Hohenberg-Kohn theorem[2] for the GKS formalism. Finally, some variational properties of the GKS solutions are proven in Appendix B. Atomic units are used throughout the paper except where noted.

0163-1829/96/53(7)/3764(11)/$06.00

II. DERIVATION OF GENERALIZED KOHN-SHAM SCHEMES

Virtually all concrete applications of density-functional theory to the ground state of many-electron systems are based on the well-known Kohn-Sham[1] scheme. In this work, we will develop generalized Kohn-Sham schemes that include the standard Kohn-Sham scheme as a special case. In deriving these GKS schemes, we will invoke the constrained-search formulation of DFT.[3,12–15]

We start by considering the Schrödinger equation of N electrons,

$$(\hat{T}+\hat{V}_{ee}+\hat{v})\Psi_0[v]=E_0[v]\Psi_0[v]. \tag{2.1}$$

In Eq. (2.1), $\hat{T}$ and $\hat{V}_{ee}$ are the operators of the kinetic energy and the electron-electron interaction, respectively. The notation emphasizes the fact that the ground-state energy E_0 and wave function Ψ_0 are functionals of the external potential $v(\mathbf{r})$. Equation (2.1) is equivalent to the minimization[3,12–14]

$$E_0[v]=\min_{\rho(\mathbf{r})\to N}\left\{F[\rho]+\int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r})\right\}, \tag{2.2a}$$

$$F[\rho]=\min_{\Psi\to\rho(\mathbf{r})}\langle\Psi|\hat{T}+\hat{V}_{ee}|\Psi\rangle. \tag{2.2b}$$

The minimizations in Eqs. (2.2) are performed within the space of all densities yielding N electrons and all antisymmetric wave functions yielding the density $\rho(\mathbf{r})$, respectively. Equation (2.2b) defines the Hohenberg-Kohn functional $F[\rho]$.[2]

The basic idea of the standard Kohn-Sham scheme is to replace the calculation of $\Psi_0[v]$ by that of a single Slater determinant that represents a noninteracting model system and yields the same ground-state density as $\Psi_0[v]$. However, the expectation value of the Hamiltonian with this determinant only gives part of the total energy whereas the remaining exchange-correlation contribution to the total energy is not directly accessible by the determinant.

Alternatively, one may try to introduce (interacting) model systems that take into account the electron-electron interaction to some extent and incorporate at least part of the exchange and correlation contribution to the total energy, but can still be represented by a single Slater determinant. The latter is important to obtain tractable single-particle equations. In this way one may hope to obtain single-particle equations with eigenvalues that more faithfully reflect the physical excitation energies and energy gaps.

Let us define an energy functional $S[\Phi]$ of N-electron Slater determinants Φ. Throughout this work, Slater determinants will be denoted by Φ to distinguish them from general many-electron wave functions Ψ. $S[\Phi]$ defines a functional $S[\{\phi_i\}]$ of the N unitary (spinor) orbitals that generate Φ. According to its definition, the functional $S[\{\phi_i\}]$ is invariant with respect to unitary transformations of the orbitals. We give three examples of such functionals $S[\Phi]$ that will be seen to be relevant for the further discussion, namely

$$S[\Phi]=\langle\Phi|\hat{T}|\Phi\rangle, \tag{2.3}$$

$$S[\Phi]=\langle\Phi|\hat{T}+\hat{V}_{ee}|\Phi\rangle=\langle\Phi|\hat{T}|\Phi\rangle+U_H[\{\phi_i\}]+E_x[\{\phi_i\}], \tag{2.4}$$

$$S[\Phi]=\langle\Phi|\hat{T}|\Phi\rangle+U_H[\{\phi_i\}]+E_x^{sx}[\{\phi_i\}]. \tag{2.5}$$

In Eqs. (2.3)–(2.5), U_H is the Hartree energy and E_x is the exchange energy that arises from the determinant Φ. Note that E_x differs in general from the Hartree-Fock exchange energy. E_x^{sx} denotes a statically screened exchange interaction with a Thomas-Fermi screening constant k_{TF},

$$E_x^{sx}[\Phi]=-\sum_{i<j}^{N}\int d\mathbf{r}d\mathbf{r}' \times\frac{\phi_i^*(\mathbf{r})\phi_j^*(\mathbf{r}')e^{-k_{\mathrm{TF}}|\mathbf{r}-\mathbf{r}'|}\phi_j(\mathbf{r})\phi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{2.6}$$

In contrast to the first example, the functionals $S[\Phi]$ in Eqs. (2.4) and (2.5) contain part of the electron-electron interaction besides the kinetic energy. Based on these functionals $S[\Phi]$, we now define functionals $F^S[\rho]$ of the density $\rho(\mathbf{r})=\Sigma_i|\phi_i(\mathbf{r})|^2$,

$$F^S[\rho]=\min_{\Phi\to\rho(\mathbf{r})}S[\Phi]=\min_{\{\phi_i\}\to\rho(\mathbf{r})}S[\{\phi_i\}]. \tag{2.7}$$

The minimization process in Eq. (2.7) searches all Slater determinants or unitary orbitals $\{\phi_i\}$ that yield the density $\rho(\mathbf{r})$. The minimizing determinant is denoted by $\Phi^S[\rho]$. As will become clear below, the functional $F^S[\rho]$ plays an analogous role as the noninteracting kinetic energy in the standard Kohn-Sham formalism but, in addition, contains parts of the electron-electron interaction energy. It depends not only on the density but also on the chosen $S[\Phi]$, and is defined by a single Slater determinant rather than with the N-particle wave function as is the case of the Hohenberg-Kohn functional, Eq. (2.2b). Such functionals F^S will be shown to lead to an exact realization of DFT, provided the corresponding functionals $S[\Phi]$ obey in all cases of interest the following conditions that guarantee the existence of self-consistent single particle equations.

(C1) The minimum $F^S[\rho]$ defined by Eq. (2.7) and its functional derivative with respect to $\rho(\mathbf{r})$ exist.

(C2) Define the energy

$$E^S[\{\phi_i\};v_{\mathrm{eff}}]=S[\{\phi_i\}]+\int d\mathbf{r}v_{\mathrm{eff}}(\mathbf{r})\rho(\mathbf{r}), \tag{2.8}$$

where v_{eff} is an arbitrary local multiplicative potential. The minimization of this energy by the usual Lagrange procedure is required to lead to a set of canonical single-particle equations that can be cast into the form

$$\hat{O}^S[\{\phi_i\}]\phi_j+\hat{v}_{\mathrm{eff}}\phi_j=\varepsilon_j\phi_j \quad \text{with} \quad j=1,\ldots,N, \tag{2.9}$$

where the operator $\hat{O}^S$ may depend on the orbitals and the functional S, but *not explicitly on the potential* $\hat{v}_{\mathrm{eff}}$, and is invariant with respect to unitary transformations of the orbitals. Furthermore we require that the orbitals ϕ_j that minimize $E^S[\{\phi_i\};v_{\mathrm{eff}}]$ are the N energetically lowest eigenstates of the Hamiltonian $\hat{O}^S+\hat{v}_{\mathrm{eff}}$ in Eq. (2.9). Note that the operator $\hat{O}^S$ is nonlocal in general. We denote the density that results from the minimization of the energy E^S in Eq. (2.8) by $\rho_0^S([v_{\mathrm{eff}}];\mathbf{r})]$.

(C3) The involved densities are v representable, i.e., every physically realized ground-state density $\rho_0([v];\mathbf{r})$ of Eq.

(2.1) equals the density $\rho_0^S([v_{\rm eff}];\mathbf{r})$ for some potential $v_{\rm eff}(\mathbf{r})$ and, conversely, every density $\rho_0^S([v_{\rm eff}];\mathbf{r})$ is the ground-state density $\rho_0([v];\mathbf{r})$ for some external potential $v(\mathbf{r})$ in Eq. (2.1).

We follow the general practice in density-functional theory to merely assume the first and third condition rather than attempting to prove them. We note, however, that the third condition is less stringent than the condition that any given arbitrary density must be the ground-state density $\rho_0([v];\mathbf{r})$ of a Schrödinger equation with a suitable external potential (which has been shown to be invalid[16,17]).

In any practical application of the GKS scheme, one will choose functionals $S[\Phi]$ that lead to single-particle equations of a form that is required by the second condition. In the example Eq. (2.3), one has $\hat{O}^S=\hat{T}$, whereas Eq. (2.4) leads to $\hat{O}^S=\hat{T}+\hat{v}_x^{\rm NL}+\hat{u}$, where $\hat{v}_x^{\rm NL}$ has the form of the nonlocal Hartree-Fock exchange potential and $\hat{u}$ is the classical Coulomb potential of the density given by the orbitals $\{\phi_i\}$.

Next, the total energy $E_0[v]$ of the interacting system in Eq. (2.1) is divided into the total energy of the model system and the remainder. In order to do this, we denote the difference between $F^S[\rho]$ of Eq. (2.7) and the Hohenberg-Kohn functional $F[\rho]$ of Eq. (2.2b) by the functional $R^S[\rho]$,

$$F[\rho]=F^S[\rho]+R^S[\rho]. \tag{2.10}$$

Then, we can write

$$\begin{aligned}E_0[v]&=\min_{\rho(\mathbf{r})\to N}\left\{F^S[\rho]+R^S[\rho]+\int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r})\right\}=\min_{\rho(\mathbf{r})\to N}\left\{\min_{\Phi\to\rho(\mathbf{r})}S[\Phi]+R^S[\rho]+\int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r})\right\}\\&=\min_{\Phi\to N}\left\{S[\Phi]+R^S[\rho[\Phi]]+\int d\mathbf{r}v(\mathbf{r})\rho([\Phi];\mathbf{r})\right\}=\min_{\{\phi_i\}\to N}\left\{S[\{\phi_i\}]+R^S[\rho[\{\phi_i\}]]+\int d\mathbf{r}v(\mathbf{r})\rho([\{\phi_i\}];\mathbf{r})\right\}.\end{aligned} \tag{2.11}$$

Note that the functional $R^S[\rho]$ and the term $\int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r})$ in Eq. (2.11) depend on the determinant Φ or on the orbitals $\{\phi_i\}$ only indirectly through the density. The latter is written as functional $\rho([\Phi];\mathbf{r})$ or $\rho([\{\phi_i\}];\mathbf{r})$. The ground-state density $\rho_0([v];\mathbf{r})$ is determined by minimizing $E_0[v]$ within the space of unitary orbitals. Because of condition (C2), a Lagrange procedure leads to the GKS equations

$$\hat{O}^S[\{\phi_i\}]\phi_j+\hat{v}_R\phi_j+\hat{v}\,\phi_j=\varepsilon_j\phi_j \quad \text{with } j=1,\ldots,N, \tag{2.12}$$

where

$$v_R(\mathbf{r})=\frac{\delta R^S[\rho]}{\delta\rho(\mathbf{r})}. \tag{2.13}$$

While the exact form of the functional $R^S[\rho]$ and of its functional derivative $v_R(\mathbf{r})$ are not known, suitable approximations can be found as will be discussed in Sec. III.

Importantly, the orbitals resulting from these GKS equations yield the exact ground-state density $\rho_0([v];\mathbf{r})$ of the Schrödinger equation (2.1) because they are obtained through the minimization of the true ground-state energy $E_0[v]$ as given in Eq. (2.11). Note that the GKS equations have exactly the form of Eq. (2.9) with

$$v_{\rm eff}(\mathbf{r})=v(\mathbf{r})+v_R(\mathbf{r}). \tag{2.14}$$

After the GKS equations have been solved self-consistently, the ground-state energy $E_0[v]$ of the Schrödinger equation (2.1) can be evaluated according to the equation

$$E_0[v]=F^S[\rho_0^S[v_{\rm eff}]]+R^S[\rho_0^S[v_{\rm eff}]]+\int d\mathbf{r}v(\mathbf{r})\rho_0^S([v_{\rm eff}];\mathbf{r}), \tag{2.15}$$

which follows by inserting the expression Eq. (2.10) for $\rho_0([v];\mathbf{r})=\rho_0^S([v_{\rm eff}];\mathbf{r})$ into Eq. (2.11). The first and the third contribution on the right-hand side of Eq. (2.15) can be calculated exactly. $F^S[\rho_0^S[v_{\rm eff}]]$ follows from evaluating $S[\{\phi_i\}]$ with the self-consistent GKS orbitals. The contribution $R^S[\rho_0^S([v_{\rm eff}];\mathbf{r})]$, as well as the potential $v_R(\mathbf{r})$ in the GKS equations, needs to be approximated. This procedure is analogous to the treatment of the exchange-correlation functionals in the standard Kohn-Sham scheme. In contrast to the latter scheme, however, important portions of the exchange correlation energy are already contained in F_S that is treated exactly.

Note that the self-consistently determined density $\rho_0^S([v_{\rm eff}];\mathbf{r})=\rho_0([v];\mathbf{r})$ is the minimizing density not only of the minimization Eq. (2.11) but also of

$$E_0^S[v_{\rm eff}]=\min_{\rho(\mathbf{r})\to N}\left\{F^S[\rho]+\int d\mathbf{r}v_{\rm eff}(\mathbf{r})\rho(\mathbf{r})\right\}, \tag{2.16}$$

with $v_{\rm eff}(\mathbf{r})$ being determined by Eq. (2.14). This follows from requirement (C2) because the preceding Eq. (2.16) results from the minimization of the energy $E^S[\{\phi_i\};v_{\rm eff}]$.

We will now clarify the role of conditions (C1) and (C3). The first requirement (C1) guarantees that the functional $F^S[\rho]$ and its functional derivative $\delta F^S[\rho]/\delta\rho(\mathbf{r})$ exist. Consequently, we can conclude that $R^S[\rho]$ and $\delta R^S[\rho]/\delta\rho(\mathbf{r})$ exist for all densities $\rho(\mathbf{r})$ provided the Hohenberg-Kohn functional $F[\rho]$ and its functional derivative exist. Indeed, the existence of $F[\rho]$ has already been proven in Refs. 16 and 18, whereas that of $\delta F[\rho]/\delta\rho(\mathbf{r})$ is merely assumed in the Kohn-Sham formalism.

The first part of requirement (C3) guarantees the existence of an effective potential $v_{\rm eff}(\mathbf{r})$ leading to a density $\rho_0^S([v_{\rm eff}];\mathbf{r})$ that equals the ground-state density $\rho_0([v];\mathbf{r})$.

Indeed, this existence follows *a posteriori* whenever the GKS equations lead to a self-consistent solution.

We show in Appendix B that the second part of requirement (C3) guarantees the calculated density $\rho_0^S([v_{\text{eff}}];\mathbf{r})$ to represent the minimum of $F[\rho]+\int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r})$ instead of merely some stationary solution. In a similar way, one may also show that the solution of the GKS equations is unique. This is proven in Appendix A. Thus, the condition (C3) guarantees that $\rho_0^S([v_{\text{eff}}];\mathbf{r})$ equals the ground-state density $\rho_0([v];\mathbf{r})$ of Eq. (2.1).

In summary, the GKS scheme replaces the original Schrödinger equation (2.1) by a set of one-particle equations, Eqs. (2.9), that are much easier to handle. In this respect, this scheme is equivalent to the standard Kohn-Sham formalism. The key difference is that not only the noninteracting kinetic energy but also the part of the total energy given by $F^S[\rho]$ is treated exactly. $F^S[\rho]$ is determined by the choice of the functional $S[\{\phi_i\}]$. Different realizations of the GKS scheme are therefore characterized by the choice of $S[\{\phi_i\}]$ and the approximations for $R^S[\rho]$ [Eq. (2.10)]. The major advantage of this GKS procedure is that suitable choices of F_S can result in R^S to be small compared to the total energy $E_0[v]$. Thus, errors in approximating R^S have only a small effect on the energy.

The derivation given above does not make use of a Hohenberg-Kohn type theorem.[2] For an interacting many-electron system, the Hohenberg-Kohn theorem states that there are not two external potentials differing by more than a constant that give the same ground-state density. For the GKS equations (2.9), the equivalent theorem states that there are not two effective potentials $v_{\text{eff}}(\mathbf{r})$ differing by more than a constant that give the same density $\rho_0^S([v_{\text{eff}}];\mathbf{r})$ as a result of minimizing $E^S[\{\phi_i\};v_{\text{eff}}]$ of Eq. (2.8). We present a proof of this latter theorem in Appendix A.

III. EXAMPLES OF GENERALIZED KOHN-SHAM SCHEMES

A. Standard Kohn-Sham method

By choosing $S[\{\Phi\}]$ to be equal to the kinetic energy of a Slater determinant, Eq. (2.3), one obtains the standard Kohn-Sham scheme.[1,3,12–15] Strictly speaking, this is true only for systems where the wave function of the model system can be represented by a single Slater determinant. This is the case, for example, in nondegenerate systems. More general situations have been discussed in Ref. 19.

With $S[\{\Phi\}]$ from Eq. (2.3), the functional $R^S[\rho]$ is given by

$$R^S[\rho]=U[\rho]+E_x[\rho]+E_c[\rho], \tag{3.1}$$

where $U[\rho]$ is the classical Coulomb energy, $E_x[\rho]$ is the exchange energy,

$$E_x[\rho]=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\int d\mathbf{r}\int d\mathbf{r}'\times\frac{\phi_i^*([\rho];\mathbf{r})\phi_j^*([\rho];\mathbf{r}')\phi_j([\rho];\mathbf{r})\phi_i([\rho];\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{3.2}$$

and $E_c[\rho]$ is the correlation energy. The corresponding functional derivatives with respect to the density are denoted by $u([\rho];\mathbf{r})$, $v_x([\rho];\mathbf{r})$, and $v_c([\rho];\mathbf{r})$ for the Coulomb, exchange, and correlation potential, respectively. The standard Kohn-Sham equations read

$$[\hat{T}+\hat{v}+\hat{u}[\rho]+\hat{v}_x[\rho]+\hat{v}_c[\rho]]\phi_i=\varepsilon_i\phi_i \quad \text{with } j=1,\ldots,N. \tag{3.3}$$

Since Eq. (3.2) does not allow one to directly calculate the functional derivative $v_x([\rho];\mathbf{r})$ and the correlation energy is not known exactly, both exchange and correlation energies and potentials are determined from approximate density functionals (such as LDA).

Henceforth, $T[\rho]$ refers to the kinetic energy evaluated with the self-consistent orbitals of Eq. (3.3).

B. Hartree-Fock–Kohn-Sham scheme

If the functional $S[\Phi]$ is chosen as the sum of the kinetic and the electron-electron energy as in Eq. (2.4), one obtains a GKS scheme that resembles the Hartree-Fock method. The resulting procedure is known as the Hartree-Fock–Kohn-Sham (HF-KS) scheme.[11,20,21] To refer to this case, we shall denote the (unknown) potential $v_R(\mathbf{r})$ as $v_c^{\text{HF}}([\rho];\mathbf{r})$ to indicate that it contains the correlation effects. The corresponding generalized Kohn-Sham equations are given by

$$-\tfrac{1}{2}\nabla^2\phi_i(\mathbf{r})+v(\mathbf{r})\phi_i(\mathbf{r})+u([\rho];\mathbf{r})\phi_i(\mathbf{r})-\int d\mathbf{r}'v_x^{\text{NL}}(\mathbf{r},\mathbf{r}')\phi_i(\mathbf{r}')+v_c^{\text{HF}}([\rho];\mathbf{r})\phi_i(\mathbf{r})=\varepsilon_i\phi_i, \tag{3.4}$$

$$v_x^{\text{NL}}(\mathbf{r},\mathbf{r}')=-\sum_{j=1}^{N}\frac{\phi_j(\mathbf{r})\phi_j^*(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.5}$$

Here, the correlation potential (as well as the energy) differs from the standard Kohn-Sham case since the orbitals $\{\phi_i\}$ obey different one-particle equations but it is believed that this difference is small.[22]

Apart from the (unknown) potential term $v_c^{\text{HF}}([\rho];\mathbf{r})$, Eq. (3.4) corresponds to the Hartree-Fock equations. In contrast to the Hartree-Fock method, however, this procedure is formally exact.

C. Screened, nonlocal exchange — sX-LDA scheme

Motivated by the observation that Hartree-Fock band gaps in solids exceed the observed ones by a large amount, it appears plausible to invoke a screened exchange potential in the single-particle equations. Indeed, this approach has been used on the basis of the quasiparticle scheme[4–6] to derive perturbative methods to solve quasiparticle equations. The latter approach, however, does not easily allow one to compute total energies.

The generalized Kohn-Sham method outlined in this paper offers the possibility to split up the total energy in such a way that the variational single-particle equations contain a screened exchange potential. As will be demonstrated below, this procedure yields single-particle eigenvalues leading to more accurate band gaps than both Hartee-Fock and LDA, while the quality of the total energy remains comparable to LDA.

To put these ideas into a rigorous framework, we choose $S[\rho]$ as given in Eq. (2.5). The Thomas-Fermi screening constitutes the simplest version of screening the exchange energy, particularly if one evaluates the Thomas-Fermi constant with the average total density.

This choice of $S[\rho]$ implies a functional $R^S[\rho]$ that is given by

$$R^S[\rho]=E_{\mathrm{xc}}^{sx}[\rho]=\{E_x[\rho]-E_x^{sx}[\rho]+T[\rho]-T^{sx}[\rho]\}+E_c[\rho]. \tag{3.6}$$

As indicated, we shall refer to this functional as $E_{\mathrm{xc}}^{sx}[\rho]$. The term in the curly brackets takes into account the energy difference between the exchange energy $E_x[\rho]$ in the standard Kohn-Sham scheme and the screened exchange energy $E_x^{sx}[\rho]$ as given in Eq. (2.6) and the corresponding difference $T[\rho]-T^{sx}[\rho]$ in the kinetic energies. Note that the correlation energy $E_c[\rho]$ in Eq. (3.6) is the same as in the standard Kohn-Sham procedure.

The resulting generalized Kohn-Sham equations (*sX-LDA equations*) are

$$\begin{aligned}&-\tfrac{1}{2}\nabla^2\phi_i(\mathbf{r})+v(\mathbf{r})\phi_i(\mathbf{r})+u([\rho];\mathbf{r})\phi_i(\mathbf{r})\\&\quad-\int d\mathbf{r}'v_x^{sx,\mathrm{NL}}(\mathbf{r},\mathbf{r}')\phi_i(\mathbf{r}')+v_{\mathrm{xc}}^{sx}([\rho];\mathbf{r})\phi_i(\mathbf{r})=\varepsilon_i\phi_i,\end{aligned} \tag{3.7}$$

$$v_x^{sx,\mathrm{NL}}(\mathbf{r},\mathbf{r}')=-\sum_{j=1}^{N}\frac{\phi_j(\mathbf{r})e^{-k_{\mathrm{TF}}|\mathbf{r}-\mathbf{r}'|}\phi_j^*(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{3.8}$$

where $v_{\mathrm{xc}}^{sx}([\rho];\mathbf{r})$ is the functional derivative of $E_{\mathrm{xc}}^{sx}[\rho]$ with respect to the density and $\hat{v}_x^{sx,\mathrm{NL}}$ is the nonlocal screened exchange operator. If k_{TF} is set equal to zero, the sX-LDA scheme is identical to the HF-KS scheme of the previous section III.B.. So far the formalism has been exact. For concrete applications, the functionals contributing to R^S in Eq. (3.6) are approximated by

$$E_x[\rho]\simeq E_x^{\mathrm{LDA}}[\rho]=\int d\mathbf{r}\rho(\mathbf{r})\epsilon_x[\rho], \tag{3.9}$$

$$E_x^{sx}[\rho]\simeq E_x^{sx,\mathrm{LDA}}[\rho]=\int d\mathbf{r}\rho(\mathbf{r})\epsilon_{sx}[\rho], \tag{3.10}$$

$$E_c[\rho]\simeq E_c^{\mathrm{LDA}}[\rho], \tag{3.11}$$

$$T[\rho]\simeq T^{sx}[\rho], \tag{3.12}$$

with[23]

$$\epsilon_x[\rho]=-\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3}, \tag{3.13}$$

$$\epsilon_{sx}[\rho]=-\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3}F(z), \tag{3.14}$$

$$F(z)=1-\frac{4}{3}z\ \arctan\frac{2}{z}-\frac{z^2}{6}\left[1-\left(\frac{z^2}{4}+3\right)\ln\left(1+\frac{4}{z^2}\right)\right]. \tag{3.15}$$

Here, $z=k_{\mathrm{TF}}/\bar{k}_F$ and $\bar{k}_F$ is the Fermi wave vector corresponding to the average density. Thus, these energy contributions are essentially approximated by functionals corresponding to the homogeneous electron gas, invoking the spirit of the local-density approximation. Note that the usage of an average Fermi wave vector is an additional approximation going slightly beyond LDA.[8] The difference $T[\rho]-T^{sx}[\rho]$ is simply set to zero (see Ref. 22 for reasons to justify this).

In effect, the same type of procedure has been employed by Bylander and Kleinman.[8] The derivations in this section put this procedure on a solid basis within the constrained search formulation of DFT.

IV. BAND GAPS IN GENERALIZED KOHN-SHAM SCHEMES

The fundamental energy gap E_g in a semiconductor can be expressed as a total energy difference. Alternatively, it can be expressed entirely in terms of KS eigenvalues:[24,25]

$$\begin{aligned}E_g&=\varepsilon_{N+1}^{\mathrm{KS}}(N)-\varepsilon_N^{\mathrm{KS}}(N)+\varepsilon_{N+1}^{\mathrm{KS}}(N+1)-\varepsilon_{N+1}^{\mathrm{KS}}(N)\\&=\varepsilon_g^{\mathrm{KS}}+\Delta_{\mathrm{xc}}^{\mathrm{KS}}=\varepsilon_g^{\mathrm{KS}}+\Delta_x^{\mathrm{KS}}+\Delta_c^{\mathrm{KS}}\end{aligned} \tag{4.1}$$

with

$$\begin{aligned}\varepsilon_g^{\mathrm{KS}}&=\varepsilon_{N+1}^{\mathrm{KS}}(N)-\varepsilon_N^{\mathrm{KS}}(N)\\\Delta_{\mathrm{xc}}^{\mathrm{KS}}&=\varepsilon_{N+1}^{\mathrm{KS}}(N+1)-\varepsilon_{N+1}^{\mathrm{KS}}(N).\end{aligned} \tag{4.2}$$

$\varepsilon_i^{\mathrm{KS}}(M)$ denotes the energetically ith lowest Kohn-Sham eigenvalue of the M particle system. Thus, $\varepsilon_M^{\mathrm{KS}}(M)$ is the highest occupied and $\varepsilon_{M+1}^{\mathrm{KS}}(M)$ the lowest unoccupied orbital energy. The contribution $\varepsilon_g^{\mathrm{KS}}$ is the eigenvalue gap, given by the difference in energy of lowest unoccupied and highest occupied level. The contribution $\Delta_{\mathrm{xc}}^{\mathrm{KS}}$ results from the discontinuity of the exchange-correlation potential with respect to particle number[24,25] and involves the eigenvalues of the $(N+1)$-particle system. The magnitude of $\Delta_{\mathrm{xc}}^{\mathrm{KS}}$ in the exact Kohn-Sham formalism is not known quantitatively[26–28] but it is generally believed that a fair fraction of the discrepancy between the LDA eigenvalue gaps and the experimental gaps originates in this discontinuity.[29]

The crucial point in the presently introduced GKS scheme is that it incorporates part of the discontinuity $\Delta_{\mathrm{xc}}^{\mathrm{KS}}$ already into the eigenvalue gap of the N-particle GKS system, thereby substantially reducing the discrepancy between theory and experiment.

In order to demonstrate this, we write the exchange part Δ_x^{KS} of the discontinuity in the form[30,31]

$$\begin{aligned}\Delta_x^{\mathrm{KS}}&=\langle N+1|\Delta\hat{v}_x|N+1\rangle-\langle N|\Delta\hat{v}_x|N\rangle,\\\Delta\hat{v}_x&=\hat{v}_x^{\mathrm{NL}}[\{\phi_i^{\mathrm{KS}}\}]-\hat{v}_x[\rho(N)].\end{aligned} \tag{4.3}$$

Here, $|N+1\rangle$ and $|N\rangle$ denote the Nth and $(N+1)$th KS orbital of an N-particle system and $\hat{v}_x^{\mathrm{NL}}[\{\phi_i^{\mathrm{KS}}\}]$ is the N-particle nonlocal exchange operator specified in Eq. (3.5) but evaluated with exact Kohn-Sham orbitals. In addition,

$\hat{v}_x[\rho]$ is the exact, local Kohn-Sham exchange potential of the N-particle density $\rho(N)$ entering the standard Kohn-Sham equations, Eq. (3.3).

The GKS equations (3.4) for the HF-KS scheme can be written in the form of standard Kohn-Sham equations plus correction terms,

$$\begin{aligned}[\hat{T}+\hat{v}+\hat{u}[\rho]+\hat{v}_x[\rho]+\hat{v}_c[\rho]]\phi_i+\Delta\hat{v}_x[\{\phi_i\}]\phi_i+\Delta\hat{v}_c\phi_i \\ =\varepsilon_i\phi_i \qquad (4.4)\end{aligned}$$

and we see indeed that the operator $\Delta\hat{v}_x$ from Eq. (4.3) enters as a potential term in the HF-KS equations. Within a perturbative approach, this operator simply yields an additive term to the KS eigenvalues,

$$\varepsilon_i \simeq \varepsilon_i^{\mathrm{KS}}+\langle i|\Delta\hat{v}_x|i\rangle. \qquad (4.5)$$

We note that there is a subtle difference between $\Delta\hat{v}_x$ in Eq. (4.4) and Eq. (4.3). The former operator contains HF-KS orbitals instead of standard Kohn-Sham orbitals. In addition, $\Delta\hat{v}_c$ is the operator that emerges from the difference between the standard Kohn-Sham and the HF-KS correlation potential. However $\Delta\hat{v}_c$ and the differences in the operators $\Delta\hat{v}_x$ are likely to be very small[22] and are neglected here.

Finally, the eigenvalue gap ε_g of the HF-KS equations is given by

$$\begin{aligned}\varepsilon_g&=\varepsilon_{N+1}(N)-\varepsilon_N(N)\\ &\simeq\varepsilon_{N+1}^{\mathrm{KS}}(N)-\varepsilon_N^{\mathrm{KS}}(N)+\langle N+1|\Delta\hat{v}_x|N+1\rangle-\langle N|\Delta\hat{v}_x|N\rangle,\\ &=\varepsilon_g^{\mathrm{KS}}+\Delta_x^{\mathrm{KS}}. \qquad (4.6)\end{aligned}$$

Thus, the exchange contribution Δ_x^{KS} to the discontinuity $\Delta_{\mathrm{xc}}^{\mathrm{KS}}$ of Eq. (4.1) is already contained in the eigenvalue gap in the HF-KS method. Analogously, the eigenvalue gap in the sX-LDA formalism of Sec. III C can be shown to contain parts of the discontinuity $\Delta_{\mathrm{xc}}^{\mathrm{KS}}$. Invoking perturbation theory once more, one may write

$$\begin{aligned}\varepsilon_i^{sx}&\simeq\varepsilon_i^{\mathrm{KS}}+\langle i|\Delta\hat{v}_x^{sx}|i\rangle,\\ \Delta\hat{v}_x^{sx}&=\hat{v}_x^{sx,\mathrm{NL}}-\hat{v}_x^{sx}, \qquad (4.7)\end{aligned}$$

where $\hat{v}_x^{sx}$ is the operator to the functional derivative of the screened exchange energy $E_x^{sx}[\rho]$[Eqs. (2.6) and (3.10)].

The GKS single-particle equations resemble the *GW* quasiparticle equations.[4–6] However, the GKS approach is a ground-state formalism and yields the correct ground-state density and energy. One may use the perturbative expression Eq. (4.7) to calculate the sX-LDA eigenvalues with the standard LDA Bloch orbitals. This grossly reduces the numerical effort since the nonlocal screened Fock exchange operator is not needed in the self-consistency procedure. We show in Sec. V D that even this perturbative approach yields very favorable results for the energy gaps.

V. IMPLEMENTATION AND RESULTS FOR BULK SEMICONDUCTORS

In order to test the proposed GKS scheme, we have implemented it in the form of the sX-LDA equations, Eq. (3.7), and applied it to energy gaps and structural properties of several semiconductors.

A. Numerical procedure

The basis of our computations is the plane-wave pseudopotential method.[32,33] Technically, the sX-LDA scheme resembles a Hartree-Fock calculation with a screened exchange potential. Since we are going to consider bulk solids, the orbitals in Eq. (3.7) are Bloch states and will be denoted by $\psi_{n,\mathbf{k}}$. In a plane-wave basis, the sX-LDA Hamiltonian matrix in Eq. (3.7) reads

$$\begin{aligned}\langle\mathbf{k}+\mathbf{G}|\hat{H}|\mathbf{k}+\mathbf{G}'\rangle&=\tfrac{1}{2}|\mathbf{k}+\mathbf{G}|^2\delta_{\mathbf{G},\mathbf{G}'}+\langle\mathbf{k}+\mathbf{G}|\hat{V}_{\mathrm{ps}}|\mathbf{k}+\mathbf{G}'\rangle\\ &\quad+4\pi\frac{\rho(\mathbf{G}-\mathbf{G}')}{|\mathbf{G}-\mathbf{G}'|^2}+v_{\mathrm{xc}}^{sx,\mathrm{LDA}}(\mathbf{G}-\mathbf{G}')\\ &\quad-\frac{4\pi}{\Omega}\sum_{n,\mathbf{q}}\sum_{\mathbf{G}_1}\\ &\quad\times\frac{\langle n,\mathbf{q}|\mathbf{q}+\mathbf{G}+\mathbf{G}_1\rangle\langle\mathbf{q}+\mathbf{G}'+\mathbf{G}_1|n,\mathbf{q}\rangle}{|\mathbf{q}-\mathbf{k}+\mathbf{G}_1|^2+k_{\mathrm{TF}}^2}. \qquad (5.1)\end{aligned}$$

Here, the vectors $\mathbf{k}$ and $\mathbf{q}$ lie in the first Brillouin zone and $\mathbf{G},\mathbf{G}',\mathbf{G}_1$ denote reciprocal-lattice vectors. We have employed the standard *ab initio* norm-conserving semilocal pseudopotentials $\hat{V}_{\mathrm{ps}}$ of Bachelet, Hamann, and Schlüter (BHS).[34] The numerically most demanding part is the computation of the plane-wave matrix elements of the nonlocal exchange operator, given by the last term in Eq. (5.1). For all investigated semiconductors, we have evaluated the $\mathbf{k}$ sums in the electronic density and in Eq. (5.1) with six special $\mathbf{k}$ points in the irreducible wedge of the Brillouin zone.[35] The plane-wave Hamiltonian matrix has been cut off at a kinetic energy of 20 Ry for Si, Ge, and GaAs, 30 Ry for InSb and 45 Ry for InP, respectively. It turned out that in the case of InP and InSb a smaller cutoff of 20–30 Ry for the $\mathbf{G}_1$ sum in Eq. (5.1) changes the calculated band gaps by less than 1 meV. We have used the analytic Perdew-Zunger parametrization[36] of the electron gas correlation energy[37] in Eq. (3.11). For a finite screening wave vector k_{TF}, the integrand in the last term of Eq. (5.1) is smooth, justifying the use of only six $\mathbf{k}$ points. In the limiting case of $k_{\mathrm{TF}}=0$, however, it contains a singularity. It can be handled efficiently by adding and subtracting another term in the wave-vector integral that compensates this singularity and can be integrated exactly.[38] In order to be able to investigate the sX-LDA method for any value of k_{TF}, we have used this approach throughout.

B. Band gaps and structural properties in sX-LDA

We have performed total-energy calculations for several semiconductors employing the sX-LDA method of Sec. III C. All results in this section have been obtained with the Thomas-Fermi (TF) dielectric function, using the average valence electron density in the screening constant k_{TF}. In the next section we will discuss alternative screening models.

Figure 1 shows the significant increase in the principal band gaps of Si, Ge, GaAs, InP, and InSb obtained by the sX-LDA method compared to LDA results that have been

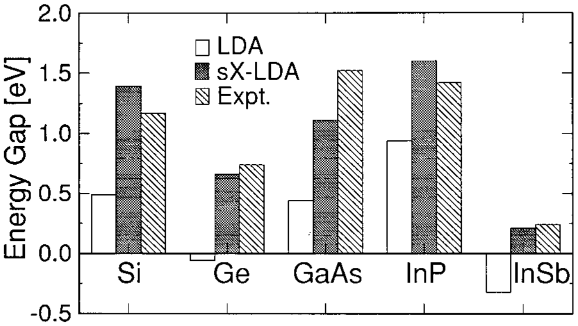

FIG. 1. Fundamental energy gaps in Si, Ge, GaAs, InP, and InSb calculated in sX-LDA method compared to the LDA and experimental values.

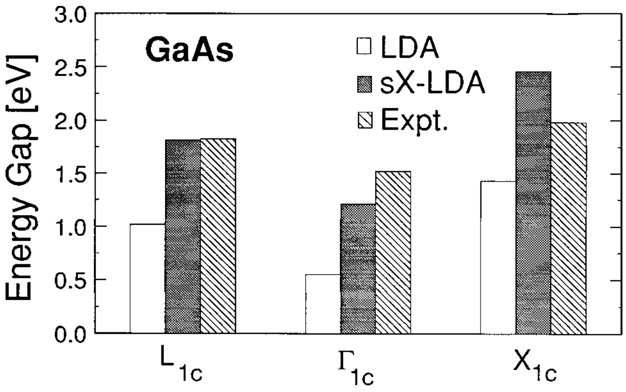

FIG. 2. Direct (Γ_{1c}) and indirect (L_{1c} and X_{1c}) energy gaps, relative to the valence-band top, in GaAs calculated in sX-LDA method compared to the LDA and experimental values.

obtained with the same BHS pseudopotentials and with the same experimental lattice constants. For materials that contain heavier anions (Ge, GaAs, and especially InSb), the spin-orbit interactions cannot be neglected. It is well established that the spin-orbit effects can be added perturbatively *a posteriori* in the LDA-pseudopotential framework. This procedure reproduces the experimental spin-orbit splitting energies in semiconductors very well.[39,40] Since we are interested only in the principal energy gaps, we have sidestepped this calculation and simply decreased the calculated nonrelativistic gaps by one-third of the valence-band spin-orbit splitting at Γ ($\frac{1}{3}\Delta_0$). The energy gaps between the lowest conduction-band states in Γ, X, and L points and the top of the valence band and the valence-band widths are shown in Table I. As a first remark, we notice that Ge and InSb have metallic properties within the LDA. In InSb, even nonrelativistic LDA calculations that use pseudopotentials[39,41] or the linearized augmented plane-wave method[41] predict a *negative* gap at Γ, whereas the present nonrelativistic sX-LDA procedure gives a *positive* gap of 0.495 eV. Taking the experimental value for the spin-orbit splitting in InSb, one obtains an sX-LDA band gap of 0.23 eV that perfectly matches the experimental band gap of 0.235 eV. In Ge, the calculated energy gap at Γ is still smaller than the one at the L point by an amount of 0.373 eV. Thus, this theory is still not able to reproduce the indirect energy gap in Ge. Table I and Fig. 2 illustrate the significant overall improvement of the predicted band gaps throughout the Brillouin zone by the sX-LDA method.

As a further check, we list the sX-LDA electronic energy eigenvalues at high-symmetry points for the occupied and lowest-lying unoccupied states of GaAs in Table II and compare them with experimental results and GW calculations of

TABLE I. Energy gaps between the lowest conduction states and the valence-band edge and valence-band widths (VBW) of Si, Ge, GaAs, InP, and InSb from present sX-LDA with Thomas-Fermi screening and LDA calculations compared to available experimental results taken from Ref. 42 except where noted. Energies are in eV.

	Si	Ge	GaAs	InP	InSb
$E(\Gamma)$	3.37	0.28	1.11	1.60	0.21
$E^{\mathrm{LDA}}(\Gamma)$	2.54	−0.06	0.44	0.94	−0.32
$E^{\mathrm{expt}}(\Gamma)$	3.05[a], 3.4	0.89	1.52	1.42	0.24
$E(X)$	1.55	1.45	2.35	2.75	1.82
$E^{\mathrm{LDA}}(X)$	0.61	0.61	1.32	1.64	1.06
$E^{\mathrm{expt}}(X)$	1.25[a]	1.3±0.2	1.98	2.38	1.79
$E(L)$	2.18	0.66	1.70	2.45	1.02
$E^{\mathrm{LDA}}(L)$	1.44	0.07	0.91	1.56	0.35
$E^{\mathrm{expt}}(L)$	1.65±0.01 2.1[b], 2.4±0.15[c]	0.74	1.81	2.03	—
VBW	12.47	13.41	13.40	11.91	11.02
$\mathrm{VBW}^{\mathrm{LDA}}$	11.94	12.80	12.40	11.20	10.48
$\mathrm{VBW}^{\mathrm{expt}}$	12.5±0.6	12.6,12.9±0.3[d]	13.1	11.0	11.7, 11.2[e]

[a] Reference 43.
[b] Reference 44.
[c] Reference 45.
[d] Reference 46.
[e] Reference 47.

TABLE II. Electronic energy levels without spin-orbit corrections (in eV) for GaAs calculated at the experimental lattice constants. Experimental values are from Ref. 48 except where noted. *GW* stands for quasiparticle band-structure calculations of Ref. 6 in the *GW* approximation.

	LDA	sX-LDA	Expt.	*GW*
Γ_{1v}	−12.56	−13.29	−13.1	−13.03
Γ_{15v}	0.0	0.0	0.0	0.0
Γ_{1c}	0.55	1.21	1.52	1.22
Γ_{15c}	3.54	4.67	4.72	4.48
X_{1v}	−10.25	−11.02	−10.75	−10.69
X_{3v}	−6.70	−6.72	−6.70	−7.19
X_{5v}	−2.58	−2.47	−2.80	−2.87
X_{1c}	1.43	2.46	2.01	2.01
X_{3c}	1.64	2.55	2.58	2.24
L_{1v}	−10.95	−11.71	−11.24	−11.41
L_{1v}	−6.52	−6.46	−6.70	−6.97
L_{3v}	−1.09	−1.05	−1.30	−1.28
L_{1c}	1.02	1.81	1.85	1.64
L_{3c}	4.52	5.63	5.45[a]	5.40

[a] Reference 43.

Ref. 6. One can see that the sX-LDA method predicts energy gaps in similar good agreement with experiment as the *GW* method does.

Table III displays the calculated structural properties of various bulk semiconductors as obtained by the sX-LDA method. The relative errors in the lattice constants a are below 0.5%, when compared to experiment. Since these errors are so small, we have calculated all energy gaps at the experimental lattice constants. The effect of zero point motion of the ions on a is one order of magnitude smaller than the remaining discrepancies between theory and experiment and has been neglected. Table III also shows that the bulk moduli predicted by sX-LDA underestimate the experimental values by typically 20%, and agree less well with experiment than the standard LDA results.

We point out that the present sX-LDA calculations contain, strictly speaking, an inconsistency or at least an ambiguity in treating the core and valence electrons. Since we employ BHS pseudopotentials, the core electrons are treated within LDA whereas the valence electrons are subject to the sX-LDA. As a consequence, we have determined only relative structural energies but not the total cohesive energy. It would clearly be desirable to employ the same sX-LDA Hamiltonian for both core and valence electrons. On the other hand, it is not clear why the same constant k_{TF} should work well for all electrons. To be consistent, one could use a screening parameter that depends on the *local* density rather than the average valence density but this procedure is computationally impractical and we have not pursued it.

TABLE III. Theoretical lattice constants (in Å) and bulk moduli (in GPa), both obtained from fit to Murnaghan's equation of state (Ref. 49), of Si, Ge, GaAs, and InP in sX-LDA with Thomas-Fermi screening. Experimental values are from Refs. 42 and 50.

	Si	Ge	GaAs	InP
a_0	5.421	5.635	5.627	5.776
a_0^{LDA}	5.37[a]	5.567[a]	5.51[a]	5.74[b]
a_0^{expt}	5.43	5.657	5.652	5.869
B_0	89.3	62.7	63.1	65.5
B_0^{LDA}	96.8[a]	76.2[a]	77.8[a]	76.0[b]
B_0^{expt}	87.6–97.9	73.4–75.8	75.3–76.9	72.5

[a] Reference 51.

[b] Reference 41.

There are several known methods that model the valence-core interactions and core polarization effects more accurately than simple LDA, such as partial core corrections,[52] mixed Hartree-Fock LDA calculations,[53] model core polarization potentials,[54] or self-interaction corrected pseudopotentials.[55] Partial core corrections eliminate a large part of the discrepancy between the experimental lattice constants and the values calculated within LDA.[56–58] Another recently proposed scheme that leads to similar improvements is based on self-interaction-corrected (SIC) pseudopotentials.[59]

In order to study the effect of our choice of BHS pseudopotentials on the lattice constants and the energy gaps, we have constructed a SIC pseudopotential for the case of Ge, using a recently developed rigorous procedure[59] that incorporates the nondiagonal Lagrange multipliers of the SIC method. The most pronounced difference in the band structure is obtained for the Γ point, where the direct energy gap $(\Gamma'_{2c}-\Gamma'_{25v})$ opens by an additional 260 meV in comparison to the value obtained with the standard BHS pseudopotential. Changes of other gaps are minor, e.g., the gaps $L_{1c}-\Gamma'_{25v}$ and $X_{1c}-\Gamma'_{25v}$ change by 50 and −90 meV, respectively. The lattice constant increases by 2%. This is consistent with the increase found earlier within an LDA calculation with SIC pseudopotentials.[59] Overall, the results of the present sX-LDA calculations do not seem to get significantly altered by using more sophisticated pseudopotentials.

C. Perturbative sX-LDA eigenvalues: Results

As pointed out in Sec. IV, one may obtain the eigenvalues of the sX-LDA equations in a much simpler albeit approximate way from Eq. (4.7) with LDA wave functions by using first-order perturbation theory. For all semiconductors investigated in this paper, these approximate energy gaps differ by less than 10% from the full self-consistent sX-LDA eigenvalues. For example, in the case of Ge, the differences in perturbative and "exact" sX-LDA energy gaps between the lowest conduction states in high-symmetry points L, X, and Γ and the valence-band edge are equal to −0.02, −0.08, and 0.09 eV, respectively. This good agreement between perturbation theory and the full sX-LDA calculation indicates that the orbitals emerging from the sX-LDA and LDA procedures, respectively, are quite similar. This is consistent with previous results by von der Linden *et al.*, who pointed out the similarity between LDA and Hartree-Fock wave functions for semiconductors.[60]

D. Alternative screening models: From HF-KS to sX-LDA

In the formulation of the sX-LDA equations [(3.7) and (5.1)], we have employed a Thomas-Fermi type of screening.

TABLE IV. Theoretical lattice constant a_0 (in Å), bulk modulus B_0 (in GPa), and band gaps (in eV) of Ge from the present sX-LDA calculations with various model dielectric functions. TF, HUBB, and SEM refer to the Thomas-Fermi, Hubbard (Ref. 61), and semiconductor (Ref. 62) model dielectric functions, respectively. Experimental values are from Ref. 42.

	TF	HUBB	SEM	Expt.
a_0	5.63	5.66	5.38	5.657
B_0	63	60	104	77
$\Gamma'_{25v}\rightarrow L_{1c}$	0.76	0.63	3.76	0.74
$\Gamma'_{25v}\rightarrow \Gamma'_{2c}$	0.38	0.23	3.39	0.89
$\Gamma'_{25v}\rightarrow X_{1c}$	1.55	1.43	4.59	1.3

More generally, we can write the screened nonlocal exchange contribution in Eq. (5.1) in the form

$$\langle \mathbf{k}+\mathbf{G}|H_x^{sx,\mathrm{NL}}|\mathbf{k}+\mathbf{G}'\rangle = -\frac{4\pi}{\Omega}\sum_{n,\mathbf{q}}\sum_{\mathbf{G}_1}\frac{\langle n,\mathbf{q}|\mathbf{q}+\mathbf{G}+\mathbf{G}_1\rangle\langle \mathbf{q}+\mathbf{G}'+\mathbf{G}_1|n,\mathbf{q}\rangle}{Q^2\epsilon(Q)} \tag{5.2}$$

with $Q=|\mathbf{q}-\mathbf{k}+\mathbf{G}_1|$ and with a general dielectric function $\epsilon(Q)$.

For germanium, we have investigated two screening models besides the Thomas-Fermi case, namely a metallic Hubbard model dielectric function[61] and a semiconductor model dielectric function.[62] The latter dielectric function contains two parameters, $\alpha=1.563$ and $\epsilon_0=12$, where $\epsilon(Q)\rightarrow\epsilon_0$ for $\mathbf{q}\rightarrow\mathbf{0}$. For each of the different models, the functional derivative $v_x^{sx,\mathrm{LDA}}(\mathbf{r})$ of $E_x^{sx,\mathrm{LDA}}[\rho]$ [Eq. (3.10)] has been determined by numerical integration.

The results of these calculations are summarized in Table IV. We find that the energy gaps depend mostly on the long-wavelength behavior of the model dielectric functions. Both the Thomas-Fermi and the Hubbard model yield complete screening for $Q\rightarrow 0$ and give very similar results. By contrast, the present model of the semiconductor dielectric function does not guarantee complete screening of a long-wavelength external potential in a semiconductor. This is due to the neglect of all off-diagonal elements $\mathbf{G}\neq\mathbf{G}'$ in $\epsilon^{-1}(\mathbf{Q}+\mathbf{G},\mathbf{Q}+\mathbf{G}')$ which play an essential role in a self-consistent calculation. Consequently, the present diagonal semiconductor dielectric function screens the exchange interaction less efficiently than the Thomas-Fermi model and yields energy gaps that are much closer to the unscreened Hartree-Fock–Kohn-Sham scheme. All of these models do not appreciably change the **k** dependence of the matrix element in Eq. (5.2), but give more or less rigid shifts of the band structure.

In the calculations with the Thomas-Fermi screening, the screening constant k_{TF} has been determined according to the average valence electron density. If one considers k_{TF} an adjustable parameter, one obtains the standard LDA method in the limit $k_{\mathrm{TF}}=\infty$. The opposite limit $k_{\mathrm{TF}}=0$, on the other hand, corresponds to the HF-KS scheme of Sec. III B with the approximation $E_c^{\mathrm{HF}}[\rho]=E_c^{\mathrm{LDA}}[\rho]$. We find that the band gaps change linearly as a function of k_{TF} between these two opposite limits. In particular, the calculated energy gaps in the HF-KS scheme, which incorporates an LDA correlation potential, are only marginally different from those obtained from standard Hartree-Fock calculations for the valence electrons (also employing BHS pseudopotentials). In Ge, for example, the energy gap at Γ is 7.02 eV in HF-KS and 6.97 eV in Hartree Fock. Similar results hold for other **k** points and materials. Thus, the improvement of the band gaps calculated with sX-LDA method is due to the fact that the screening of the exact exchange potential already takes into account essential correlation effects.

VI. CONCLUSIONS

Employing a constrained-search formulation of DFT, we have developed a framework to derive various exact realizations of DFT in addition to the standard Kohn-Sham scheme. Such generalized Kohn-Sham (GKS) schemes minimize not only the kinetic energy for fixed density such as the standard Kohn-Sham method but also part of the electron-electron energy. As a result, one gains more flexibility in constructing generalized exchange-correlation functionals. Generally, the GKS scheme leads to single-particle equations that contain nonlocal potentials such as the Fock exchange operator. A key point of the GKS method is that the single-particle eigenvalues reflect much more faithfully the lowest excitation energies (such as band gaps) of the many-electron system than the Kohn-Sham method. The reason is that the discontinuity of the exchange-correlation potential with respect to the particle number is, to a large extent, already incorporated in the GKS single-particle eigenvalues.

We have shown that a previously suggested method based on a total-energy functional with a nonlocal screened exchange and LDA correlation energy[8] is firmly rooted within the presented GKS framework. In particular, we have applied this approach, that we call sX-LDA method, to Si, Ge, GaAs, InSb, and InP and obtained eigenvalue gaps in excellent overall agreement with true band gaps in these materials. The improvements are comparable to the perturbative *GW* approach but the sX-LDA method allows a self-consistent determination of ground-state properties as well.

The lattice constants, as calculated by sX-LDA, agree well with experiment, whereas the bulk moduli are underestimated by typically 20%. In part, we believe this to be caused by using pseudopotentials that do not model core-valence polarization effects with sufficient accuracy.

We have developed a perturbative scheme for the sX-LDA single-particle eigenvalues that basically requires only a standard LDA calculation and yields semiconductor band gaps that agree well with the self-consistent sX-LDA eigenvalue gaps.

ACKNOWLEDGMENTS

Helpful comments by Professor Del Sole and financial support by the Deutsche Forschungsgemeinschaft (SFB 348), Bayerische Forschungsverbund (FOROPTO), and Volkswagenstiftung are gratefully acknowledged. One of us (A.G.) thanks the Deutsche Forschungsgemeinschaft for a Habilitationsstipendium.

APPENDIX A: PROOF OF HOHENBERG-KOHN THEOREM FOR THE GKS SCHEME

The Hohenberg-Kohn theorem corresponding to the GKS equations is given by the statement that two different potentials $v_{\text{eff}}(\mathbf{r})$ cannot yield the same density $\rho_0^S(\mathbf{r})$ as a result of minimizing $E^S[\{\phi_i\};v_{\text{eff}}]$ of Eq. (2.8). Its proof consists of two steps. First, it is shown that two different local potentials $v_{\text{eff}}(\mathbf{r})$ that give the same density $\rho_0^S(\mathbf{r})$ also yield the same Slater determinant Φ_0^S. Potentials that differ by a constant are considered equivalent. Second, we show that two different potentials $v_{\text{eff}}(\mathbf{r})$ cannot give the same Slater determinant Φ_0^S, in contradiction to the first statement. Altogether, we establish in this way the one-to-one mapping between $v_{\text{eff}}(\mathbf{r})$ and $\rho_0^S(\mathbf{r})$.

First, it follows from the minimization of $E_0^S[v_{\text{eff}}]$ given in Eq. (2.16) that once the ground-state density $\rho_0^S(\mathbf{r})$ is known, the ground-state wave function Φ_0^S is completely determined as a functional of that density by Eq. (2.7), where $\rho(\mathbf{r})$ is set equal to $\rho_0^S(\mathbf{r})$. This implies that two different effective potentials that yield the same ground-state density $\rho_0^S(\mathbf{r})$ also give the same ground-state wave function Φ_0^S.

For the second step of the proof, we note that Φ_0^S is built of the orbitals obeying the single particle Eqs. (2.9). Therefore, the determinant Φ_0^S is the ground-state wave function of the N-electron Schrödinger equation

$$\{\hat{O}^S[\{\phi_i\}]+\hat{v}_{\text{eff}}\}\Phi_0^S[v_{\text{eff}}]=E_{\text{sum}}^S[v_{\text{eff}}]\Phi_0^S[v_{\text{eff}}]. \tag{A1}$$

Here, $\hat{O}^S[\{\phi_i\}]$ is the N-electron equivalent of the one-particle operator that appears in the GKS equations (2.9) or (2.12) but is denoted by the same symbol. Note that the energy $E_{\text{sum}}^S[v_{\text{eff}}]$ is nothing but the sum of the eigenvalues ε_j of the GKS equations (2.9) or (2.12).

By dividing the preceding equation by Φ_0^S, one obtains

$$-\frac{[\hat{O}[\{\phi_i\}]\Phi_0^S(\mathbf{r}_1\ldots\mathbf{r}_i\ldots\mathbf{r}_N)]}{\Phi_0^S(\mathbf{r}_1\ldots\mathbf{r}_i\ldots\mathbf{r}_N)}=\left[\sum_{i=1}^{N}v_{\text{eff}}(\mathbf{r}_i)\right]-E_0^S. \tag{A2}$$

We note that one has to form a scalar product of Eq. (A1) with the spin part of Φ_0^S if the latter function contains spinors, in order to make the division well defined. Equation (A2) determines the effective potential uniquely and therefore establishes the Hohenberg-Kohn theorem for the GKS equations.

The given proof of the Hohenberg-Kohn theorems is a direct consequence of the requirements (C1) and (C2). One may therefore consider (C1) and (C2) as an equivalent formulation of the Hohenberg-Kohn theorem in the constrained-search formulation of the GKS formalism.

APPENDIX B: VARIATIONAL PROPERTIES OF GKS SOLUTIONS

Here, we show that the second part of requirement (C3) in Sec. II guarantees the calculated density $\rho_0^S([v_{\text{eff}}];\mathbf{r})$ to represent the minimum of $F[\rho]+\int d\mathbf{r}v(\mathbf{r})\rho(\mathbf{r})$ instead of merely some stationary solution. According to (C3), there exists some external potential $\tilde{v}(\mathbf{r})$ for the interacting Schrödinger equation (2.1) that yields a ground-state density equal to $\rho_0^S([v_{\text{eff}}];\mathbf{r})$. We now show that, up to a constant, $\tilde{v}(\mathbf{r})=v(\mathbf{r})$. From the Euler equation that corresponds to Eq. (2.2a), it follows that this external potential $\tilde{v}(\mathbf{r})$ is given by the functional derivative $-\delta F[\rho]/\delta\rho(\mathbf{r})$ evaluated at the density $\rho_0^S([v_{\text{eff}}];\mathbf{r})$. The potential $v_{\text{eff}}(\mathbf{r})$ is equal to the functional derivative $-\delta F^S[\rho]/\delta\rho(\mathbf{r})$ evaluated at the same density because of the Euler equation corresponding to Eq. (2.16). This leads, together with Eq. (2.10), to the identity

$$\tilde{v}(\mathbf{r})=v_{\text{eff}}(\mathbf{r})-v_R(\mathbf{r}),$$

$$v_R(\mathbf{r})=\delta R^S[\rho]/\delta\rho[\mathbf{r}]\big|_{\rho(\mathbf{r})=\rho_0^S([v_{\text{eff}}];\mathbf{r})}. \tag{B1}$$

In the course of solving the GKS equations self-consistently, Eq. (2.14) connects $v_{\text{eff}}(\mathbf{r})$ to the given real external potential $v(\mathbf{r})$. Thus, once convergence of the GKS equations has been achieved, the identity $v(\mathbf{r})=v_{\text{eff}}(\mathbf{r})-v_R(\mathbf{r})$ holds and comparison with Eq. (2.1) leads to $\tilde{v}(\mathbf{r})=v(\mathbf{r})$.

---

[1] W. Kohn and l. J. Sham, Phys. Rev. **140**, A1133 (1965).

[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[3] For review see, e.g., R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).

[4] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. B **37**, 10 159 (1988).

[5] M. S. Hybertsen and S. G. Louie, Phys. Rev. Lett. **55**, 1418 (1985); Phys. Rev. B **34**, 5390 (1986).

[6] X. Zhu and S. G. Louie, Phys. Rev. B **43**, 14 142 (1991).

[7] L. Hedin, Phys. Rev. **139**, A796 (1965).

[8] B. M. Bylander and L. Kleinman, Phys. Rev. B **41**, 7868 (1990).

[9] F. R. Vukajlovic, E. L. Shirley, and R. M. Martin, Phys. Rev. B **43**, 3994 (1991).

[10] H. Yamagami, Y. Takada, H. Yasuhara, and A. Hasegawa, Phys. Rev. A **49**, 2354 (1994).

[11] A. Görling and M. Levy, Phys. Rev. B **47**, 13 105 (1993).

[12] M. Levy and J. P. Perdew, in *Density Functional Methods in Physics*, Vol. 123 of *NATO Advanced Science Institutes Series*, edited by R. M. Dreizler and J. da Providencia (Plenum Press, New York, 1985), Chap. 2, pp. 11–30.

[13] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, International Series of Monographs on Chemistry, No. 16 (Oxford University Press, New York, 1989).

[14] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory* (Springer, Berlin, 1990).

[15] M. Levy, in *Advances in Quantum Chemistry*, edited by S. B. Trickey (Academic Press, San Diego, 1990), Vol. 21, p. 69.

[16] E. Lieb, in *Density Functionals for Coulomb Systems*, edited by H. Feshbach and A. Shimony (MIT, Cambridge, 1982).

[17] M. Levy, Phys. Rev. A **26**, 1200 (1982).

[18] E. Lieb, Int. J. Quantum Chem. **23**, 243 (1983).

[19] A. Görling, Phys. Rev. A **47**, 2783 (1993).

[20] S. Baroni and E. Tuncel, J. Chem. Phys. **79**, 6140 (1983).

[21] H. Stoll and A. Savin, in *Density Functional Methods in Physics* (Ref. 12), Chap. 7, pp. 177–207.

[22] A. Görling and M. Ernzerhof, Phys. Rev. A **51**, 4501 (1995).

[23] J. E. Robinson, F. Bassani, R. S. Knox, and J. R. Schrieffer, Phys. Rev. Lett. **9**, 215 (1962).
[24] J. P. Perdew and M. Levy, Phys. Rev. Lett. **51**, 1884 (1983).
[25] L. J. Sham and M. Schlüter, Phys. Rev. Lett. **51**, 1888 (1983).
[26] R. W. Godby, M. Schlüter, and L. J. Sham, Phys. Rev. Lett. **56**, 2415 (1986).
[27] O. Gunnarsson and K. Schönhammer, Phys. Rev. Lett. **56**, 1968 (1986).
[28] K. Schönhammer and O. Gunnarsson, J. Phys. C **20**, 3657 (1987).
[29] J. P. Perdew and M. Levy, in *Many-Body Phenomena at Surfaces*, edited by H. S. D. Langreth (Academic, Orlando, 1984), Chap. 1, pp. 71–89.
[30] J. P. Perdew, in *Density Functional Methods in Physics* (Ref. 12), Chap. 10, pp. 265–308.
[31] A. Görling and M. Levy, Phys. Rev. A **52**, 4493 (1995).
[32] J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C **12**, 4409 (1979).
[33] W. E. Pickett, Comp. Phys. Rep. **9**, 117 (1989).
[34] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).
[35] H. J. Monkhost and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).
[36] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[37] D. M. Ceperly and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[38] F. Gygi and A. Baldereschi, Phys. Rev. B **34**, 4405 (1986).
[39] M. P. Surh, Ming-Fu Li, and S. G. Louie, Phys. Rev. B **43**, 4286 (1991).
[40] L. A. Hemstreet, C. Y. Fong, and J. S. Nelson, Phys. Rev. B **47**, 4238 (1993).
[41] S. Massidda, A. Continenza, A. J. Freeman, T. M. de Pascale, F. Meloni, and M. Serra, Phys. Rev. B **41**, 12 079 (1990).
[42] *Numerical Data and Functional Relationships in Science and Technology*, edited by K.-H. Hellwege, Landolt-Börnstein, New Series, Group III, Vols. 17 and 22, edited by O. Madelung, M. Schulz, and H. Weiss (Springer, Berlin, 1982), Vol. 23a edited by A. Goldman and E.-E. Koch (Springer, Berlin, 1989).
[43] J. E. Ortega and F. J. Himpsel, Phys. Rev. B **47**, 2130 (1993).
[44] R. Hulthen and N. G. Nilsson, Solid State Commun. **18**, 1341 (1976).
[45] D. Straub, L. Ley, and F. J. Himpsel, Phys. Rev. Lett. **54**, 142 (1985).
[46] A. L. Wachs, T. Miller, T. C. Hsieh, A. P. Shapiro, and T.-C. Chiang, Phys. Rev. B **32**, 2326 (1985).
[47] D. E. Eastman, W. D. Grobman, J. L. Freeouf, and M. Erbudak, Phys. Rev. B **9**, 600 (1974).
[48] E. P. O'Reilly, in *Properties of Gallium Arsenide*, 2nd ed., EMIS Datareviews Series No. 2 (INSPEC, London, 1990), p. 113.
[49] F. D. Murnaghan, Proc. Natl. Acad. Sci. U.S.A. **30**, 244 (1944).
[50] J. D. Wiley, in *Semiconductors and Semimetals*, edited by R. K. Willardson and A. C. Beer (Academic, New York, 1975), Vol. 10, p. 134.
[51] V. Fiorentini, Phys. Rev. B **46**, 2086 (1992).
[52] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
[53] B. M. Bylander and L. Kleinman, Phys. Rev. B **43**, 12 070 (1991).
[54] E. L. Shirley and R. M. Martin, Phys. Rev. B **47**, 15 413 (1993).
[55] M. M. Rieger and P. Vogl, Phys. Rev. A **52**, 282 (1995).
[56] A. Qteish and R. J. Needs, Phys. Rev. B **43**, 4229 (1991).
[57] A. Garcia and M. L. Cohen, Phys. Rev. B **47**, 6751 (1993).
[58] A. Dal Corso, S. Baroni, R. Resta, and S. de Gironcoli, Phys. Rev. B **47**, 3588 (1993).
[59] M. M. Rieger and P. Vogl, Phys. Rev. B **52**, 6 567 (1995).
[60] W. von der Linden, P. Fulde, and K.-P. Bohnen, Phys. Rev. B **34**, 1063 (1986).
[61] J. Hubbard, Proc. R. Soc. London Ser. A **243**, 336 (1957).
[62] G. Cappellini, R. DelSole, L. Reining, and F. Bechstedt, Phys. Rev. B **47**, 9892 (1993).