PSEUDOPOTENTIAL METHODS IN CONDENSED MATTER APPLICATIONS

Warren E. PICKETT

Complex Systems Theory Branch, Naval Research Laboratory, Washington, DC 20375-5000, USA

1989

NORTH-HOLLAND – AMSTERDAM

Contents

Computer Physics Reports 9 (1989) 115–198
North-Holland, Amsterdam

PSEUDOPOTENTIAL METHODS IN CONDENSED MATTER APPLICATIONS

Warren E. PICKETT

Complex Systems Theory Branch, Naval Research Laboratory, Washington, DC 20375-5000, USA

Received 22 July 1988

The generalization from empirically determined screened pseudopotentials to self-consistently screened *ab inito* pseudopotentials has led to widespread use of the method in solid state applications. The method is reviewed here, beginning with the formal basis in density functional theory. Algorithms for solving the one-particle equations and evaluating the density functional expression for the energy are presented. Special attention is given to the developments in iteration to self-consistency, and to the implications of the variational nature of the energy. An extensive list is given of the broad range of applications of the self-consistent pseudopotential method for describing properties of condensed matter systems.

1. Introduction

In solving the Schrödinger equation for condensed aggregates of atoms, space can be divided into two regions with quite different properties. The region near the nuclei, the "core regions", are composed primarily of tightly bound core electrons which respond very little to the presence of neighboring atoms, while the remaining volume contains the valence electron density which is involved in the bonding together of the atoms. Although the potential in the core is strongly attractive for valence electrons, the requirement that valence wave functions be orthogonal to those of the core produces a large kinetic energy which contributes an effective repulsive potential for valence states. The pseudopotential formalism grew out of the Orthogonalized Plane Wave (OPW) method [1], in which valence wave functions were expanded in a set of plane waves (PW) which are orthogonalized to all of the core wavefunctions $|\psi_c\rangle$:

$$|\mathrm{OPW},\ \boldsymbol{K}\rangle = |\mathrm{PW},\ \boldsymbol{K}\rangle - \sum_{c}^{\mathrm{core}} |\psi_c\rangle\langle\psi_c|\mathrm{PW},\ \boldsymbol{K}\rangle, \tag{1.1}$$

where the wavevector $\boldsymbol{K}$ labels the PW or OPW.

The construction of a pseudopotential can be demonstrated in terms of the exact core and valence states $|\psi_c\rangle\,|\psi_v\rangle$, which satisfy

$$H|\psi_i\rangle = E_i|\psi_i\rangle, \quad i = \mathrm{c},\ \mathrm{v}. \tag{1.2}$$

The valence states are smoothed in the core region by subtracting out the core orthogonality wiggles, leading to pseudostates $|\phi_v\rangle$ given by

$$|\phi_v\rangle = |\psi_v\rangle + \sum_c |\psi_c\rangle\alpha_{cv}, \tag{1.3}$$

with $\alpha_{cv} = \langle\psi_c|\phi_v\rangle$. Applying H to $|\phi_v\rangle$ gives

$$\begin{aligned} H|\phi_v\rangle &= E_v|\psi_v\rangle + \sum_c E_c|\psi_c\rangle\alpha_{cv} \\ &= E_v|\phi_v\rangle + \sum_c (E_c - E_v)|\psi_c\rangle\alpha_{cv} \end{aligned} \tag{1.4}$$

or

$$\left\{H + \sum_c (E_v - E_c)|\psi_c\rangle\langle\psi_c|\right\}|\phi_v\rangle = E_v|\phi_v\rangle. \tag{1.5}$$

Thus the valence pseudostates $|\phi_v\rangle$ satisfy a Schrödinger equation with an energy-dependent pseudo-Hamiltonian

$$H^{\mathrm{PS}}(E) = H + \sum_c (E - E_c)|\psi_c\rangle\langle\psi_c|, \tag{1.6}$$

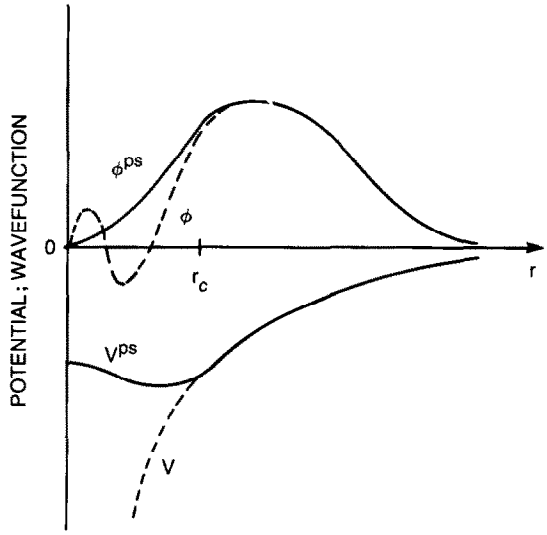

Fig. 1. Schematic representation of the pseudopotential method. The all-electron potential V and orbital ϕ are altered to the pseudopotential V^{ps} and pseudo-orbital ϕ^{ps} inside the core radius r_c.

and one has transformed the core orthogonality wiggles out of the valence electronic structure problem at the expense of introducing an energy-dependent, non-local repulsive potential. The pseudopotential concept is sketched in fig. 1.

The Phillips–Kleinman [2] pseudopotential, given by the sum of the true potential V and the repulsive potential of eq. (1.6),

$$V^{PK} = V + \sum_c (E - E_c) \, |\psi_c\rangle\langle\psi_c| , \tag{1.7}$$

would not be worth the additional complexity were it not for the feature that for many atoms V^{PK} is so weak that it can be handled more easily than the full potential. In the weak pseudopotential limit a solid resembles an electron gas weakly perturbed by pseudopotentials, and this picture applies reasonably accurately to the alkali metals but not to many other materials. A more useful feature is that the pseudopotential often is rapidly convergent in a PW basis set, and this feature can be utilized for describing most semiconductors as well as metals with no d or f bands (or 2p bands) in the valence region. Applications using empirical pseudopotentials have been reviewed extensively elsewhere [3].

The purpose of the present paper is not to review extensively the development of pseudopotentials, but rather to describe the current implementation of electronic pseudopotentials in modern electronic structure calculations. The formal basis for the great majority of recent electronic structure calculations in condensed matter is Density Functional Theory (DFT), which will be reviewed briefly in section 2. DFT provides an exact (in principle) all-electron formalism for determining the density and the energetics of electronic systems, and for pseudopotentials to be useful they must reproduce more than simply the valence eigenvalues as discussed above. They must produce valence wavefunctions which faithfully describe the bonding among aggregates of

atoms and produce accurate densities, and in this section we investigate the validity of their use in DFT. There are important questions, firstly, about the applicability of non-local potentials in DFT and, secondly, about how accurate a pseudopotential (i.e., frozen core) approach can be in describing bonding in crystals. Discussion of these questions is also included in section 2. Section 3 will be devoted to the increasingly sophisticated methods of constructing pseudopotentials in current use.

The purpose of this paper is to provide a fairly complete and detailed presentation of the formal and numerical methods necessary to carry out pseudopotential-based electronic structure calculations. In sections 4 and 5 the formalism, equations and algorithms for solving the Kohn–Sham equations are collected from the literature. The focus will be on PW expansions of the wavefunctions, since this implementation of pseudopotentials has been the most prevalent. There have, however, been several cases of localized basis or mixed basis (localized + PWs) implementations, so the discussion will point out basis independent procedures as well as considering aspects related to localized basis functions.

The goal of electronic structure studies is to be able to calculate physical properties for comparison with experimental data. When agreement is obtained, one is justified in placing confidence in the calculational methods themselves. This confidence allows an identification and detailed interpretation of the microscopic bases of observed phenomena, followed by the prediction of new and unexpected properties and materials. Section 6 is devoted to a brief summary of the wide variety of applications for which the pseudopotential method has proven successful.

2. Density functional theory

2.1. The Kohn–Sham equations

There are several extensive discussions in the literature [4–8] of the foundations of DFT and the approximations which are used in its implementation. Here we will be content to present and describe the formalism and to investigate the use of pseudopotentials in DFT. First it is necessary to establish some notation. We consider a periodic system with atoms at positions $\boldsymbol{R} + \boldsymbol{\tau}$, where $\boldsymbol{R}$ is a Bravais lattice vector locating a unit cell of the crystal and $\boldsymbol{\tau}$ is a basis vector giving the positions within the unit cell. The Hamiltonian of the system is given by

$$H = -\sum_{i=1}^{N} \frac{1}{2m} \nabla_i^2 + \sum_{i=1}^{N} V_{\text{ion}}(\boldsymbol{r}_i) + \tfrac{1}{2} \sum_{i \neq j}^{N} v(\boldsymbol{r}_i - \boldsymbol{r}_j) + V_{\text{I-I}}, \tag{2.1}$$

which describes N electrons at positions $\boldsymbol{r}_i$ interacting via the Coulomb potential $v(\boldsymbol{r}) = e^2/|\boldsymbol{r}|$ and moving in the potential of the static ions given by

$$V_{\text{ion}}(\boldsymbol{r}) = \sum_{m,s} V_{\text{ion}}^{(s)}(\boldsymbol{r} - \boldsymbol{R}_m - \boldsymbol{\tau}_s) \tag{2.2}$$

in terms of ionic potentials $V_{\text{ion}}^{(s)}$. The ion–ion repulsion is given by

$$V_{\text{I-I}} = \frac{1}{2} \sum_{\substack{mm' \\ ss'}}{}' \frac{Z_s Z_{s'} e^2}{|\boldsymbol{R}_m + \boldsymbol{\tau}_s - \boldsymbol{R}_{m'} - \boldsymbol{\tau}_{s'}|}, \tag{2.3}$$

where the prime indicates the $m = m'$, $s' = s$ term is to be omitted from the summation. We will use throughout this paper Rydberg atomic units $\hbar = 1$, $2m = 1$, $e^2 = 2$, wherein distances are measured in bohr and energies in rydbergs. The e^2 and $2m$ factors will often be retained in explicit expressions for energies and potentials for clarity.

The first three terms in eq. (2.1) are often lumped together as the electron energy operator H^{el}. Obviously all electronic properties of this system are functionals of the ionic potential V_{ion}, since only fundamental constants appear elsewhere in H. A central result of DFT is that the ground state electronic energy E^{el} of the system is given by

$$E^{\text{el}} = T_0[n] + \int V_{\text{ion}}(\boldsymbol{r}) n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} + E_{\text{h}}[n] + E_{\text{xc}}[n]. \tag{2.4}$$

Here $T_0[n]$ is the kinetic energy of a system of non-interacting electrons with density $n(\boldsymbol{r})$, E_{h} is the classical interaction energy

$$E_{\text{h}}[n] = \tfrac{1}{2} \int\int n(\boldsymbol{r}) v(\boldsymbol{r} - \boldsymbol{r}') n(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}\, \mathrm{d}\boldsymbol{r}' \tag{2.5}$$

and E_{xc} is the remaining energy, termed the exchange-correlation energy, which is a functional of n alone.

A second result of DFT is the E^{el} is minimized, with respect to number-conserving variations of the density, by the true ground state density $n_0(\boldsymbol{r})$:

$$\delta \left\{ E^{\text{el}}[n] - \mu \int n(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \right\} = 0, \tag{2.6a}$$

$$\left. \frac{\delta E^{\text{el}}[n]}{\delta n(\boldsymbol{r})} \right|_{n = n_0} = \mu, \tag{2.6b}$$

where μ is the (constant) chemical potential of the electronic system.

This equation is satisfied if the following set of Kohn–Sham equations [4–8] is solved self-consistently:

$$h\psi_i(\boldsymbol{r}) = \left\{ -\frac{1}{2m} \nabla^2 + V_{\text{eff}}(\boldsymbol{r};\, n) \right\} \psi_i(\boldsymbol{r}) = \epsilon_i \psi_i(\boldsymbol{r}), \tag{2.7}$$

$$V_{\text{eff}}(\boldsymbol{r};\, n) = V_{\text{ion}}(\boldsymbol{r}) + V_{\text{h}}(\boldsymbol{r};\, n) + V_{\text{xc}}(\boldsymbol{r};\, n), \tag{2.8}$$

$$n(\boldsymbol{r}) = \sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2. \tag{2.9}$$

Here $\{\psi_i\}$ are orthonormal eigenfunctions and $\{\epsilon_i\}$ are the associated eigenvalues of the one-body Hamiltonian h in eq. (2.7), and the sum in eq. (2.9) extends over the N lowest eigenvalues. The Hartree and exchange-correlation (XC) potentials are given by

$$V_{\rm h}(\boldsymbol{r};\, n) = \int v(\boldsymbol{r}-\boldsymbol{r}')n(\boldsymbol{r}')\,{\rm d}\boldsymbol{r}', \tag{2.10}$$

$$V_{\rm xc}(\boldsymbol{r};\, n) = \frac{\delta E_{\rm xc}[n]}{\delta n(\boldsymbol{r})}. \tag{2.11}$$

This Kohn–Sham procedure introduces a one-body Hamiltonian which has the natural interpretation, but without formal justification, as describing an "electron" in the effective mean field $V_{\rm eff}$. In fact, the non-interacting kinetic energy $T_0[n]$ was introduced into the expression (2.4) so a single particle kinetic energy term $(-(1/2m)\nabla^2)$ would appear in eq. (2.7). In spite of the fact that DFT itself assigns no formal interpretation to ψ_i and ϵ_i, it has been recognized for decades that the eigensolutions of the related Hartree–Fock–Slater equation closely resemble single particle excitation energies and wavefunctions in many systems. This interpretation can be shown [8] to follow from Green's function theory, which provides a formalism for single particle excitations. Excitations are described by an equation analogous to eq. (2.7), with the complication that $V_{\rm eff}$ is replaced by a non-Hermitian energy-dependent self-energy (see Pickett [8] and references therein). This self-energy, however, is often similar in its effect to $V_{\rm eff}$, hence the close resemblance of ψ_i and ϵ_i to excited state properties of the many body system. Although this interpretation will not be relied on in this review, h, ψ and ϵ will be referred to as the single electron Hamiltonian, wavefunction and eigenvalue, respectively.

One point which is of concern here is that the potential imposed on the electrons in DFT, denoted $V_{\rm ion}$ in eq. (2.1), is a *local function* (simple real function of $\boldsymbol{r}$). This is a requirement for the formal validity of DFT – the external potential must be local. The nuclear potential,

$$V_{\rm ion}^{(s)}(\boldsymbol{r}) = -\frac{Z_s e^2}{|\boldsymbol{r}-\boldsymbol{R}_s|}, \tag{2.12}$$

satisfies this requirement trivially. As we will discuss below, there have been extensive, very successful applications of DFT in which $V_{\rm ion}$ *is taken to be a non-local pseudopotential.* Somewhat surprisingly, there has been very little justification of this procedure, except in terms of its results.

2.2. Validity of non-local potentials in density functional theory

Gilbert [9] considered non-local potentials in the context of DFT and concluded that there is *no* DFT in such a case. He established rather that for a general non-local external potential of the form $V_{\rm ext}(\boldsymbol{r},\, \boldsymbol{r}')$, the *density matrix* $\mu(\boldsymbol{r},\, \boldsymbol{r}')$ is analogous to the density in DFT. There is a *density matrix* functional theory, and a Kohn–Sham-like Schrödinger equation for the orbitals and eigenvalues, but in addition to the non-local external potential the exchange-correlation

potential is also non-local. Berrondo and Goscinski [10] and Levy [11] have also investigated the introduction of non-local potentials in DFT and come to similar conclusions.

In principle, then, DFT is not formally valid with non-local potentials. However, Gilbert's conclusion allows it to be generalized to an analogous *density matrix functional theory* with a *non-local exchange-correlation potential* $V_{xc}(\boldsymbol{r}, \boldsymbol{r}')$,

$$V_{xc}(\boldsymbol{r}, \boldsymbol{r}'; \mu) \equiv \frac{\delta E_{xc}[\mu]}{\delta \mu(\boldsymbol{r}, \boldsymbol{r}')}, \tag{2.13}$$

occurring in the Kohn–Sham equations. (Gilbert notes that there are formal difficulties in proving the existence of this functional derivative, but the successful applications to be surveyed in section 6 support his suspicion that this is not a fundamental problem.) Since a practitioner of DFT is faced with making approximations for $E_{xc}[n]$ in order to proceed, the approximate functional $\tilde{E}_{xc}[n]$ may just as well be considered as an approximation $\tilde{E}_{xc}[\mu]$ to $E_{xc}[\mu]$ which depends only on $n(\boldsymbol{r}) = \mathrm{Tr}\,\mu$. The result then is that the use of non-local potentials is equally as valid as local potentials within current applications of DFT.

2.3. Validity of frozen core approximation in total energy studies

Another point of concern here is the validity of the pseudopotential approximation – that is, the frozen core ansatz – in carrying out density functional studies of energetics. As described in section 3, the pseudopotential is defined by the condition that it reproduce one electron properties (eigenvalues and wavefunctions) accurately, but no condition involving energetics is imposed. It is less obvious then that pseudopotentials will be successful in such applications, even if one restricts the use to situations where core overlap is negligible.

This concern was noted by Janak [12] in the course of comparing the energies of competing crystal structure using all-electron methods. According to the virial theorem, the total energy is equal to the negative of the kinetic energy K. Janak used the Xa approximation for the XC potential, where $V_{xc}(\boldsymbol{r})$ is proportional to $n^{1/3}(\boldsymbol{r})$, for which he reasoned that the XC contribution to K vanishes. Then K can be partitioned into core and valence parts K_c and K_v given by

$$K_c = \sum_c \int \psi_c(\boldsymbol{r}) \left(-\frac{1}{2m} \nabla^2 \right) \psi_c(\boldsymbol{r})\, \mathrm{d}^3 r, \tag{2.14}$$

$$K_v = \sum_v \int \psi_v^*(\boldsymbol{r}) \left(-\frac{1}{2m} \nabla^2 \right) \psi_v(\boldsymbol{r})\, \mathrm{d}^3 r. \tag{2.15}$$

Calculations for Be in the fcc and bcc structures revealed the difference in $-K_c$ (that is, the core contribution to the total energy) between structures to be three times as large as the difference in total energies, and of different sign. Janak questioned therefore the ability of pseudopotential calculations, which neglect core energies entirely, to reproduce structural energy differences.

Several detailed studies since that time have confirmed, however, that the pseudopotential method does predict structural energies properly, except in certain cases where very polarizable semicore states contribute to bonding in the crystal. Von Barth and Gelatt [13] resolved this

paradox by an unorthodox application of density functional theory. They first dramatized the paradox by noting that the core kinetic energy of Mo changes by 5 eV in forming the crystal, and also that the change in K_c between bcc and fcc Mo is 2.7 eV compared to the total energy difference of 0.5 eV per atom. These facts appear to invalidate the pseudopotential approach and the "frozen core" assumption which underlies it. The core pseudopotential is considered to be a fixed quantity, unchanging when an atom moves to a different environment. This "transferability" precept appears to rely on the assumption that different environments will not alter the core and its properties, and Janak's all-electron calculations show this assumption to be inadequate.

Von Barth and Gelatt [13] approached this problem by defining an energy functional of the core and valence densities $n_c(\boldsymbol{r})$ and $n_v(\boldsymbol{r})$ separately. Although the last three terms of eq. (2.4) can be generalized trivially by the replacement $n(\boldsymbol{r}) \to n_c(\boldsymbol{r}) + n_v(\boldsymbol{r})$, the kinetic energy term is less straightforward. Nevertheless, they established the existence of a functional $T_0[n_c,\ n_v]$ (which is *not* equal to $T_0[n]$) which is sufficient to describe the system to second order in $n_v - n_v^0$ and $n_c - n_c^0$ around the physical density $n^0 = n_c^0 + n_v^0$. (The original work should be consulted for more details.) Using this functional they derived an expression for the error in total energy made by using the frozen core approximation,

$$\delta E_{fc} = \int \left[\rho_c^{fc}(\boldsymbol{r}) - \rho_c(\boldsymbol{r})\right]\left[V_{eff}^{*}(\boldsymbol{r}) - V_{eff}^{fc}(\boldsymbol{r})\right] \mathrm{d}^3 r. \tag{2.16}$$

Here ρ_c^{fc} is the frozen core density, V_{eff}^{fc} is an effective potential (see eq. (2.8)) which will produce ρ_c^{fc}, and

$$V_{eff}^{*}(\boldsymbol{r}) \equiv V_{eff}\left(\boldsymbol{r};\ n_c^{fc} + n_v^{*}\right) \tag{2.17}$$

is the effective one-body potential in the self-consistent frozen core calculation, for which the valence density is n_v^*. This energy correction is *second order* in the difference between frozen core density and true core density, the result which is essential to the validity of the frozen core approximation.

Calculations performed by von Barth and Gelatt confirmed both the second order nature of δE and the accuracy of eq. (2.16). They found that a simple way to understand how the first order terms vanish is to consider the relaxation of the frozen core approximation in two stages: first let the core relax in fixed valence density n_v^*, then let the valence electrons relax with this new (relaxed) core held fixed. The stationarity of their functional with respect to core and valence densities separately guarantee that the energy change is second order is each stage. They concluded that total energy errors in the frozen core approximation should typically be no more than ~ 0.1 eV; subsequent studies suggest it is an order of magnitude better than this figure, as it must be to be useful in predicting pressure-driven structural transformations.

Finally one is led to Janak's argument – what was overlooked in his analysis? In his analysis he was relating energy differences to kinetic energy differences using the virial theorem. There are two weaknesses in applying this approach to *approximate* densities. First, the error in the kinetic energy is *first order* in the error in density, which makes this an unacceptable method for estimating errors. Second, and specific to the frozen core situation, is that the virial theorem is not satisfied by the frozen core approximation. As a result, energy differences which are actually of second order appear to be of first order.

3. Construction of pseudopotentials

3.1. Background

The methods for constructing pseudopotentials have evolved considerably over the years. The initial (empirical) pseudopotentials were constrained only to reproduce one-electron eigenvalues and to be as weak as possible, and these conditions led to non-Hermitian model potentials [14]. The Phillips–Kleinman [2] procedure was to form valence pseudoorbitals by adding small amounts of core orbitals of the same symmetry to the valence orbital. The more modern methods, which we now discuss, place the emphasis on making the valence pseudoorbital resemble as much as possible the valence orbital beyond some fiducial core radius r_l (which depends on the principal angular momentum quantum number l) and imposes other conditions (which usually includes "smoothness" in some form) to complete the assignment.

The earliest introduction of ionic pseudopotentials into self-consistent electronic structure calculations occurred in the context of Hartree–Fock and related methods. Since the interest here is primarily in extended condensed matter systems, no attempt will be made to review in any comprehensive way the use of pseudopotentials in molecular studies. Much of the work up to 1970 has been discussed by Weeks, Hazi and Rice [15] and by Kahn and Goddard [16]. The interest in going beyond the fitting of pseudopotentials to experimental data led to the development of "ab initio" potentials which were parameter-free. The approach can be illustrated with the Hartree–Fock method for a single valence electron atom, with the one-electron equation for the valence orbital given by

$$h_{\mathrm{v}}^{\mathrm{HF}}\phi_{\mathrm{v}} = \left[-\frac{1}{2m}\nabla^2 - \frac{Ze^2}{r} + V_{\mathrm{core}}^{\mathrm{HF}} \right]\phi_{\mathrm{v}} = \epsilon_{\mathrm{v}}\phi_{\mathrm{v}}. \tag{3.1}$$

Here $V_{\mathrm{core}}^{\mathrm{HF}}$ is the Coulomb and non-local exchange potentials arising from the core orbitals, and can be considered as an effective potential seen by the valence orbitals. (The off-diagonal core–valence terms, which are neglected in eq. (3.1), can be included as shown by Weeks and Rice [17].)

In going to the molecule, one may assume the core orbitals are unchanged and thus take $V_{\mathrm{core}}^{\mathrm{HF}}$ as individual ionic potentials seen by the valence orbitals. Because of the non-locality of $V_{\mathrm{core}}^{\mathrm{HF}}$, however, the solution of the valence problem is not simplified in any essential way. One would like to replace $V_{\mathrm{core}}^{\mathrm{HF}}$ by an equivalent local potential. The "local potential" defined by

$$u_{\mathrm{c}}^{\mathrm{HF}}(\boldsymbol{r}) = V_{\mathrm{core}}^{\mathrm{HF}}\phi_{\mathrm{v}}(\boldsymbol{r})/\phi_{\mathrm{v}}(\boldsymbol{r}) \tag{3.2}$$

leads to a simple problem

$$\left\{ -\frac{1}{2m}\nabla^2 - \frac{Ze^2}{r} + u_{\mathrm{c}}^{\mathrm{HF}}(\boldsymbol{r}) \right\}\phi_{\mathrm{v}}(r) = \epsilon_{\mathrm{v}}\phi_{\mathrm{v}} \tag{3.3}$$

with the same solution ϕ_{v} as eq. (3.1) for the atomic valence orbital. For many purposes this potential might be useful, but it carries some undesirable features. Since the valence orbital ϕ_{v}

must be orthogonal to core orbitals of the same symmetry, it will contain one (or more) node, say, at r_0. The non-locality of $V_{\text{core}}^{\text{HF}}$ ensures that the numerator of eq. (3.2) will not generally vanish at r_0, with the result that u_c^{HF} contains singularities of the form $|r - r_0|^{-1}$ which will force nodes of *all* valence orbitals at $|\boldsymbol{r}| = r_0$. This feature illustrates the kind of anomalous behavior which can occur when a direct inversion of the Schrödinger equation is carried out to construct a pseudopotential.

This approach was carried over to the local density formalism by Topiol, Zunger and Ratner [18]. Noting that the one-electron eq. (2.7) heuristically can be written

$$\epsilon_{\text{nl}} = \frac{h\psi_{\text{nl}}(\boldsymbol{r})}{\psi_{\text{nl}}(\boldsymbol{r})}, \tag{3.4}$$

one can define a pseudo-eigenproblem by

$$\epsilon_{\text{nl}}^{\text{ps}} = \frac{h_l^{\text{ps}}\psi_{\text{nl}}^{\text{ps}}(\boldsymbol{r})}{\psi_{\text{nl}}^{\text{ps}}(\boldsymbol{r})}. \tag{3.5}$$

With h given by

$$h = -\frac{1}{2m}\nabla^2 + V_{\text{ion}}(\boldsymbol{r}) + V_{\text{hxc}}(\boldsymbol{r};\, n), \tag{3.6}$$

and h_l^{ps} is given by (V_{hxc} is the sum of Hartree and XC potentials of eqs. (2.10) and (2.11))

$$h_l^{\text{ps}} = -\frac{1}{2m}\nabla^2 + V_{l,\text{ion}}^{\text{ps}}(\boldsymbol{r}) + V_{\text{hxc}}(\boldsymbol{r};\, n^{\text{ps}}), \tag{3.7}$$

the requirement $\epsilon_{\text{nl}}^{\text{ps}} = \epsilon_{\text{nl}}$ that the pseudopotential equation lead to the all-electron eigenvalues allows eqs. (3.4) and (3.5) to be solved for the l-dependent pseudopotential $V_{l,\text{ion}}^{\text{ps}}$, if $\psi_{\text{nl}}^{\text{ps}}$ is specified. The pseudodensity n^{ps} which goes into (3.7) is given by the sum of the valence pseudo-orbital densities.

The original papers can be consulted for the formal solution for $V_{l,\text{ion}}^{\text{ps}}$. Topiol, Zunger and Ratner [18] defined their pseudo-orbital as the valence orbital plus a linear combination of core orbitals of the same l, with coefficients adjusted to make the pseudo-orbital nodeless. This choice of pseudo-orbital leads to a non-singular pseudopotential (except at $r \to 0$) and also allows the pseudowavefunctions in the application (molecule or solid) to be orthogonalized to the core if that is desired.

This local density pseudopotential was applied by Zunger and Cohen [19] to solid state problems, and the pseudopotentials for the first three rows of the periodic table were fit to analytic forms by Lam, Cohen and Zunger [20]. Although they appear to give reliable results, these pseudopotentials have not gained wide use subsequently. This is due primarily to the "hard-core" nature of the potentials, that is, they diverge as $r \to 0$. This divergence leads to very long range tails in their Fourier transforms, and therefore they are not useful for plane-wave based calculational methods.

Hamann, Schlüter and Chiang (HSC) [21] utilized the fact that there is actually much more flexibility in defining the atomic pseudo-orbital (and therefore the pseudopotential) than had been fully taken advantage of, and that this flexibility can be used to generate "soft-core" pseudopotentials, which are regular at $r \to 0$ and short-range in reciprocal space. Such soft-core pseudopotentials, usually in local (l-independent) form, had been used by the Bell Labs [22] and Berkeley [23] groups in early self-consistent local density calculations. These pseudopotentials were determined by fitting functional forms with a few parameters to give satisfactory fits to atomic eigenvalues and orbital shapes, but suffered from some shortcomings. Primary among these shortcomings were (1) the lack of an objective procedure for generating the pseudopotential, and (2) the recognized feature that the pseudo-orbitals did not reproduce the true orbitals well outside the core; rather they differed typically by a few percent although displaying the correct shape. In terminology to be developed later by HSC, these pseudopotentials did not conserve the normalization of the density outside the core region (not *norm-conserving*), and therefore they will not describe the valence charge density accurately in molecular or solid state applications.

3.2. Norm-conserving pseudopotentials

The norm-conserving concept, which was first used by Topp and Hopfield [24] in the context of empirical pseudopotentials and was incorporated into ionic potentials by Starkloff and Joannopoulos [25], was made the central feature by HSC. In a procedure which was refined somewhat by Bachelet, Hamann and Schlüter [26], they freed the construction of the atomic pseudo-orbital from reliance on core states. The construction proceeds as follows.

1. Choosing an appropriate atomic reference configuration, which often differs somewhat from the ground state, the self-consistent field Dirac equations are solved in the spherical approximation for the local density potential and the radial wavefunctions F_κ and G_κ and the corresponding eigenvalue ϵ_κ. F_κ and G_κ are the small and large component for quantum number κ, where $\kappa = l$ for $j = l - \frac{1}{2}$, and $\kappa = -(l+1)$ for $j = l + \frac{1}{2}$. The non-relativistic approximation can be obtained in the following equations by the substitution $\kappa \to l$. Dirac's equations are replaced by a Schrödinger-like equation for valence electrons outside the core region given by

$$\left\{ -\frac{1}{2m}\frac{\mathrm{d}^2}{\mathrm{d}r^2} - \frac{\kappa(\kappa+1)}{2mr^2} + V_\kappa(r) \right\} G_\kappa(r) = \epsilon_\kappa G_\kappa(r). \tag{3.8}$$

   This replacement is valid [27,28] to order α^2, where α is the fine structure constant.

2. A first-step pseudopotential $V_\kappa^{(1)}$ which accomplishes the pseudizing is constructed by cutting off the singularity at the nucleus:

$$V_\kappa^{(1)}(r) = V(r)\left[1 - f(r/r_{c\kappa})\right] + c_\kappa f(r/r_{c\kappa}), \tag{3.9}$$

   where $r_{c\kappa}$ is the fiducial core radius for quantum number κ. The restrictions are that $f(x)$ is a smooth "cutoff function" which is unity at the origin, cuts off around $x \sim 1$ and vanishes rapidly as $x \to \infty$. The constant c_κ is determined by requiring that the lowest nodeless

solution $W_\kappa^{(1)}$ of the radial equation with potential $V_\kappa^{(1)}$ has an energy eigenvalue $\epsilon_\kappa^{(1)} = \epsilon_\kappa$, that is, identical to the all-electron eigenvalue. When $W_\kappa^{(1)}$ is normalized, we can write

$$\gamma_\kappa W_\kappa^{(1)}(r) \to G_\kappa(r), \quad r > r_{c\kappa}, \tag{3.10}$$

since they are solutions for the same potential in this region. The form of cutoff function

$$f(x) = \exp(-x^\lambda), \quad \lambda = 3.5, \tag{3.11}$$

was found to produce optimum results. The core radius $r_{c\kappa}$ must lie outside the outer node of the full-core valence wavefunction; the larger $r_{c\kappa}$ is, the smoother the resulting pseudopotential, while the smaller $r_{c\kappa}$ is, the more accurately it reproduces the valence wavefunction. This parameter typically is chosen between the aforementioned node and the peak position $r_{\max,\kappa}$ of the valence wavefunction,

$$r_{c\kappa} = r_{\max,\kappa}/C, \tag{3.12}$$

with $C = 1.5$–2.

3. Now the norm conservation is built in. The pseudoorbital $W_\kappa^{(1)}$ is modified to $W_\kappa^{(2)}$ by adding a correction in the core region:

$$W_\kappa^{(2)}(r) = \gamma_\kappa\left[W_\kappa^{(1)}(r) + \delta_\kappa g_\kappa(r)\right], \tag{3.13}$$

where $g_\kappa(r)$ must vanish at least as fast as r^{l+1} for small r to give a regular pseudopotential at the origin, and it must vanish rapidly for $r > r_{c\kappa}$ because $W_\kappa^{(1)}$ is the desired solution in that region. BHS chose the form

$$g_\kappa(r) = r^{l+1} f(r/r_{c\kappa}). \tag{3.14}$$

Normalization of $W_\kappa^{(2)}$ requires that δ_κ be the smaller solution of the quadratic equation

$$\gamma_\kappa^2 \int \left|W_\kappa^{(1)}(r) + \delta_\kappa g_\kappa(r)\right|^2 \mathrm{d}r = 1. \tag{3.15}$$

4. The final *screened* pseudopotential $V_\kappa^{(2)}$, defined as that potential which produces the nodeless pseudoorbital $W_\kappa^{(2)}$ with eigenvalue $\epsilon_\kappa^{(2)} = \epsilon_\kappa$, is found by inverting the radial Schrödinger equation. The result is given by

$$V_\kappa^{(2)}(r) = V_\kappa^{(1)}(r) - \left[g_\kappa^{-1}(r) D g_\kappa(r) + V_\kappa^{(1)}(r) - \epsilon_\kappa\right] / \left\{1 + W_\kappa^{(1)}(r) / \left[\delta_\kappa g_\kappa(r)\right]\right\}, \tag{3.16}$$

where

$$D \equiv -\frac{1}{2m}\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{\kappa(\kappa+1)}{2mr^2},$$

in terms of general cutoff functions $f(x)$ and $g_\kappa(r)$.

5. The final step is to obtain the ionic pseudopotential $V_\kappa^{\rm ion}$ by *unscreening* $V_\kappa^{(2)}$ using the nodeless pseudo-orbitals $W_\kappa^{(2)}$. With the atomic pseudocharge density $n_{\rm v}^*$ given by

$$n_{\rm v}^*(r) = \sum_i^{\rm occ\,val} \left| W_\kappa^{(2)}(r)/r \right|^2, \tag{3.17}$$

the unscreening procedure gives

$$V_\kappa^{\rm ion}(r) = V_\kappa^{(2)}(r) - \{ V_{\rm h}(r;\, n_{\rm v}^*) + V_{\rm xc}(r;\, n_{\rm v}^*) \} \tag{3.18}$$

in terms of the notation of eqs. (2.10), (2.11). Defined in this way, $V_\kappa^{\rm ion}$ will give rise, when applied to the self-consistent valence electron problem, to screened pseudopotential $V_\kappa^{(2)}$ and pseudo-orbitals $W_\kappa^{(2)}$ which are self-consistent by construction.

The paper of BHS should be read for details, such as for the feature that different configurations can be used for different values of κ, and for which applications this procedure may be desirable. It should also be noted that the unscreening procedure in eq. (3.18) implies a linearization of the exchange-correlation potential in $n = n_{\rm c} + n_{\rm v}$, which however should be of small consequence for most applications.

It is useful to define l-dependent average $V_l^{\rm ion}$ and spin–orbit $V_l^{\rm so}$ potentials by [26]

$$V_l^{\rm ion} = \frac{1}{2l+1}\left[l V_{l-1/2}^{\rm ion} + (l+1) V_{l+1/2}^{\rm ion} \right], \tag{3.19}$$

$$V_l^{\rm so} = \frac{2}{2l+1}\left[V_{l+1/2}^{\rm ion} - V_{l-1/2}^{\rm ion} \right], \tag{3.20}$$

for which the total pseudopotential can be represented

$$V^{\rm ion}(r) = \sum_l P_l \left[V_l^{\rm ion}(r) + V_l^{\rm so}(r) \boldsymbol{L}\cdot\boldsymbol{S} \right] P_l. \tag{3.21}$$

This treatment is analogous to the "scalar relativistic plus spin–orbit" representations used in all-electron calculations [29]. For many atoms, up to $Z \sim 50$, spin–orbit corrections are negligible for many properties. Even for heavier systems it can be useful to iterate to self-consistency initially for the spin–orbit-less case, for which the secular equation is a factor of two smaller and, for systems with inversion symmetry, real rather than complex.

The process of obtaining the norm-conserving pseudopotential is illustrated in fig. 2 for the valence 4d, 5s and 5p states of Mo. On the left side of fig. 2 the all-electron potential (V), the initial pseudized potential ($V^{(1)}$), the final ionic pseudopotential ($V^{\rm ps}$) and the limiting form $-Z_{\rm v}e^2/r$ for large r, are all shown. On the right side the all-electron wavefunction (ϕ) and the pseudowavefunction ($\phi^{\rm ps}$) are shown. The vertical arrows denote the chosen core radii 0.85, 1.70, 2.18 a.u., using the criteria $r_{\rm c}/r_{\rm max} = 0.625$, 0.588, 0.555 for 4d, 5s and 5p respectively. With this choice of core radii the norm-conservation step (3) (eqs. (3.13)–(3.15)) leads to changes $W^{(1)} \to W^{(2)} \equiv \phi^{\rm ps}$ and $V^{(1)} \to V^{(2)}$ which are hardly visible on the scale of fig. 2. The core correction and renormalization factors which appear in eq. (3.13) are $\delta = 0.075$, 0.005, -0.013

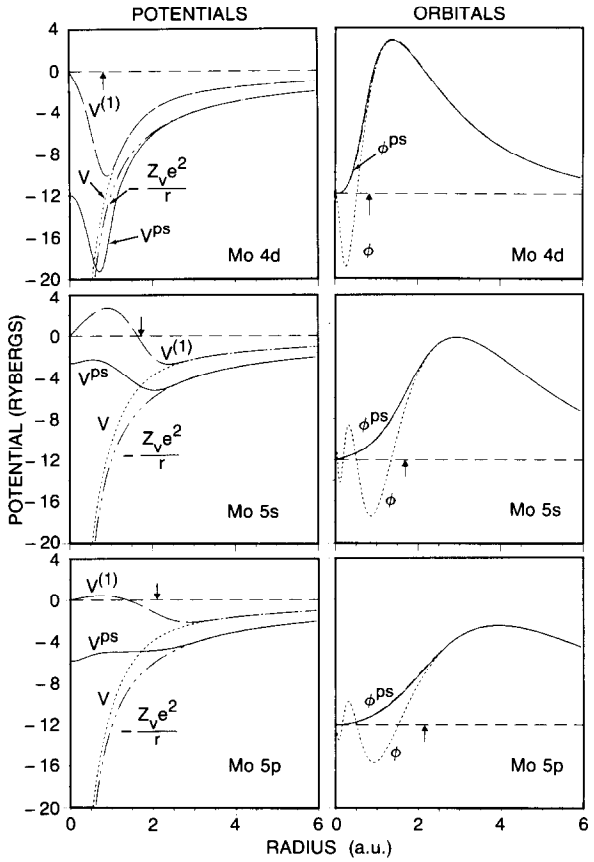

Fig. 2. Illustration of the steps in the calculation of the norm-conserving pseudopotential, using the Mo 4d, 5s and 5p states as an example. The atomic all-electron potential V is first "pseudized" inside the core radius (vertical arrow) to form $V^{(1)}$. For the present choice of core radii the difference between $V^{(1)}$ and $V^{(2)}$ is very small and is not shown (see text). Unscreening by the valence pseudo-orbitals ϕ^{ps} produces the ionic pseudopotential V^{ps}.

and $\alpha = 0.993$, 0.999, 1.002, respectively. More often, δ is not so small and therefore α differs more from unity. The result, for Mo, is a very strong 4d pseudopotential (almost -20 Ry) and a strongly peaked but rather smooth radial pseudofunction, and weak and smooth 5s and 5p pseudopotentials.

These norm-conserving pseudopotentials have been calculated and published (in a form to be discussed below) for the atoms H through Pu ($Z = 94$) by BHS. It must be emphasized that, even within the procedure outlined here, these pseudopotentials are not unique. They depend, first of all, on the choice of what is to be incorporated into the core. For most atoms there is no ambiguity. For a few, however, such as the divalent post-transition elements Zn, Cd and Hg, there may be some applications for which it is useful to assign the highest d states (which lie at ~ -10 eV in the atoms) to the core. To study compound formation, on the other hand, the d states are known to participate in bonding and therefore must be treated as valence electrons. BHS have presented pseudopotentials to cover both cases for these atoms. Greenside and Schlüter [30] have provided an alternative set of pseudopotentials for the 3d transition atoms. An

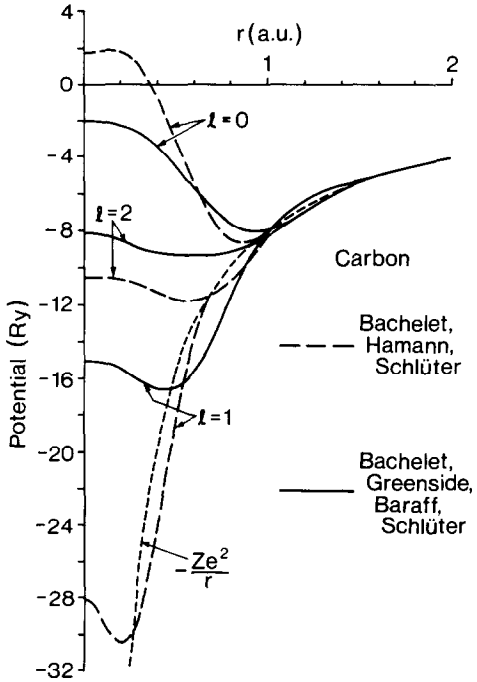

Fig. 3. The s, p and d ($l = 0$, 1 and 2) pseudopotentials for the carbon atom using two different choices of the core radii. The dashed potentials of Bachelet, Hamann and Schlüter [26] use smaller core radii and are more transferable, but require many more plane waves for their representation. The potentials shown as solid lines were taken from Bachelet, Greenside, Baraff and Schlüter [110].

example of the change in potentials when the core radius is changed is shown in fig. 3 for carbon.

The lack of uniqueness extends even further. The pseudopotential depends on the atomic configuration which is chosen, and this choice is not always obvious. There is also flexibility in choosing the pseudizing functions f in eq. (3.11) and g in eq. (3.13), to which, we return below. Finally, there is the crucial matter of the choice of the core radii $r_{c\kappa}$. BHS have emphasized that the core radii are *not* to be considered as free parameters, since varying them results in pseudopotentials of different quality. The smaller the core radius is chosen, the nearer the core the pseudowavefunctions reproduce the true wavefunctions, and the more "transferable" the pseudopotential is, which means the larger the energy region is for which reasonable pseudowavefunctions are obtained. The choice of BHS is only a particular choice and, it turns out, one which usually involves a conservative choice of core radius. This means, of course, that the pseudopotential displays sharper structure than may be necessary for some applications and therefore becomes less suitable for plane-wave-based methods.

3.3. Smoothness considerations

It may be noted that the "smoothness" of the pseudopotential has not been explicitly invoked in the construction outlined above. Kerker [31] proposed an alternative scheme which focuses on

the smoothness of the pseudowavefunction and norm conservation. We will not present the procedure here, but in spite of its relatively rare use in recent calculations it is somewhat simpler to calculate and may have some advantages over the BHS scheme for certain applications.

Vanderbilt [32] has confronted the smoothness criterion by modifying the particulars of the BHS procedure. He noted that the two occurrences of f in eq. (3.9) can be replaced by separate functions f_1 and f_2 since they serve different purposes. More generally, however, the potential $V_\kappa^{(1)}$ can be replaced by a polynomial inside the cutoff radius and fixed by continuity of the function and a few of its derivatives. He also suggested replacing the additive correction in eq. (3.13) with a multiplicative correction to $W^{(1)}$ given by

$$W_\kappa^{(2)}(r) = \gamma_\kappa[1 + \delta_\kappa g_\kappa(r)]\, W_\kappa^{(1)}(r), \tag{3.22}$$

which may be less likely to produce structured pseudopotentials.

The point of mentioning here the methods of Kerker and of Vanderbilt is simply to emphasize that the published BHS potentials can be improved on in some cases for applications using plane wave basis sets. Vanderbilt has illustrated [32] using the case of Si the additional smoothness which can be obtained using his procedure rather than the BHS procedure. He finds that $q^2 V_l^{\text{ion}}(q)$ is negligible beyond $q = 7$ a.u. using his method while the BHS scheme vanishes similarly only beyond $q = 10$ a.u. As a result the number of plane waves necessary to achieve absolute convergence could be as much as a factor of $(10/7)^3 = 2.9$ larger with the BHS method. The corresponding plot for oxygen, which has strongly localized 2p orbitals, is shown in fig. 4. Vanderbilt's scheme results in an $l = 1$ ionic pseudopotential which is negligible beyond $q = 22$ a.u., while the BHS pseudopotential has not vanished yet at $q = 30$ a.u.

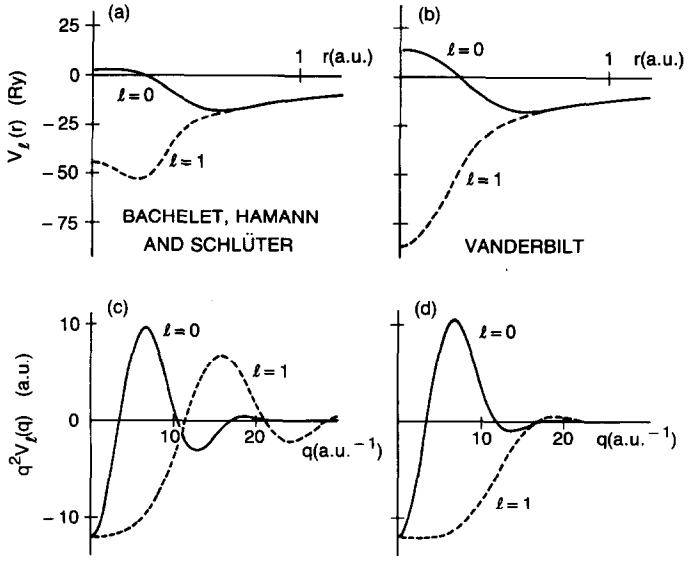

Fig. 4. Comparison of the oxygen 2s (solid) and 2p (dashed) ionic pseudopotentials using the Bachelet–Hamann–Schlüter [26] scheme ((a) and (c)) and the Vanderbilt [32] method ((b) and (d)). This figure illustrates the greater smoothness (faster falloff of $q^2 V_l^{\text{ion,ps}}(q)$) using the Vanderbilt scheme. Note that the very differently-appearing 2p pseudopotentials are essentially equivalent. Courtesy of D. Vanderbilt.

3.4. Numerical considerations

Since the BHS pseudopotentials have become rather widely used, some unusual features related to their *representation for publication* must be pointed out. BHS chose to write

$$V_l^{\text{ion}} = V_{\text{core}} + \Delta V_l^{\text{ion}}, \tag{3.23}$$

where the l-independent part V_{core} heuristically can be considered to originate from Gaussian-distributed effective core charges, giving the form of potential

$$V_{\text{core}}(r) = -\frac{Z_v e^2}{r} \sum_{i=1}^{2} c_i^{\text{core}} \operatorname{erf}\left[(\alpha_i^{\text{core}})^{1/2} r\right], \tag{3.24}$$

where Z_v is the valence charge. The l-dependent parts ΔV_l were represented as

$$\Delta V_l^{\text{ion}}(r) = e^2 \sum_{i=1}^{3} \left(A_i + r^2 A_{i+3}\right) \mathrm{e}^{-\alpha_i r^2}, \tag{3.25}$$

since with this form matrix elements in (i) plane waves alone, (ii) Gaussian linear combination of atomic orbitals, and (iii) mixed basis planewave-Gaussians, can be expressed analytically.

When the c_i^{core}, α_i^{core}, A_i and α_i parameters are fit to the pseudopotentials, a very good fit can be obtained. Two unfortunate features appear, however, due to the *extreme non-linearity* of the fitting functions. The first is that equally good fits can be obtained with quite different sets of parameters, which is harmless numerically but results in parameters which have no physical significance and therefore have no regular behavior through the periodic table. A potentially more serious problem is that the six terms in eq. (3.25) usually display enormous cancellations, i.e., the individual terms have differing signs and may be much larger (orders of magnitude!) than the net value of ΔV_l. Again, this is harmless in a numerical sense *as long as it is handled carefully*.

The result of this latter anomaly is that the A_i parameters may (and do) become so large in absolute value that it is impractical to tabulate the necessary number of significant figures (around ten). To deal with this problem BHS performed a transformation on the representation functions

$$\Phi_i(r) = \begin{cases} \mathrm{e}^{-\alpha_i r^2} & i = 1, 2, 3 \\ r^2\, \mathrm{e}^{-\alpha_i r^2} & i = 4, 5, 6, \end{cases} \tag{3.26}$$

to an orthonormal set. In terms of the A_i parameters, the expansion parameters C_i in the orthonormal set are given by

$$C_i = -\sum_{j=1}^{6} Q_{ij} A_j, \tag{3.27}$$

where the orthogonality matrix Q is given by

$$Q_{ij} = \begin{cases} 0, & i > j \\ \dfrac{\left[S_{ij} - \sum_{k=1}^{i-1} Q_{ki}^2 \right]^{1/2}}{2}, & i = j \\ \left[S_{ij} - \sum_{k=1}^{i-1} Q_{ki} Q_{kj} \right] / Q_{ii}, & i < j \end{cases} \tag{3.28}$$

and S is the overlap matrix,

$$S_{ij} = \int_0^\infty r^2 \Phi_i(r) \Phi_j(r)\, \mathrm{d}r. \tag{3.29}$$

The published parameters are the C_j, therefore one must reconstruct the expressions of eq. (3.25) by back-transforming

$$A_i = -\sum_{j=1}^{6} Q_{ij}^{-1} C_j. \tag{3.30}$$

Although eq. (3.30) looks harmless enough, it must be recalled that the A_i parameters must be computed to up to about 9–10 significant figures (for the largest $|A_i|$) to reconstruct the potential accurately. In addition to using double precision (REAL*8) arithmetic, it is useful to note that formal expressions for the matrix elements of Q^{-1} can be derived by making use of the lower triangular form of Q and repeatedly solving for successive matrix elements of Q^{-1}. Expressions for Q^{-1} have been given by Pattnaik, Fletcher and Fry [33]. An improved form used by the author, which may give marginally better performance, is given in appendix A, in the form of FORTRAN statements to eliminate the possibility of transcription errors.

Alas! When the procedure implied by eq. (3.30) is implemented and debugged, it is dismaying to find that the same code run on two different computers, or indeed with the same machine and same compiler but with different optimization levels, can produce very different sets of coefficients $\{A_i\}$! The differences arise due to the different orderings of arithmetic operations by different compilers and suggest an unstable numerical procedure. It appears, however, to be self-correcting in the sense that it always produces a *set of parameters* $\{A_i\}$ which reproduces ΔV_l^{ion} in eq. (3.25) very well, that is, to millirydberg accuracy.

As was mentioned above, the highly non-linear fit required in the representation of eq. (3.25) leads to non-unique parameters, and only a *set of parameters* is meaningful. BHS realized this anomaly and published a table of Si pseudopotentials in real space to provide a check of coding of the procedure (a set $\{A_i\}$ would be useless for this purpose). The use of the expressions given in appendix A rather than matrix inversion routines seems to minimize the occurrence of differing sets $\{A_i\}$ for a given set $\{C_i\}$.

The use of these pseudopotentials often requires their momentum space representation. A potential of the form $\mathscr{P}_l V_l(|\boldsymbol{r}|)\mathscr{P}_l$, where $\mathscr{P}_l$ is the angular momentum projection operator, has a Fourier transform given by

$$V_l(\boldsymbol{K}, \boldsymbol{K}') = \frac{1}{\Omega}\int \mathrm{d}\boldsymbol{r}\, \mathrm{e}^{\mathrm{i}\boldsymbol{K}\cdot\boldsymbol{r}}\mathscr{P}_l V_l(|\boldsymbol{r}|)\mathscr{P}_l\, \mathrm{e}^{-\mathrm{i}\boldsymbol{K}'\cdot\boldsymbol{r}}$$

$$= \frac{4\pi}{\Omega}(2l+1)P_l(\hat{\boldsymbol{K}}'\cdot\hat{\boldsymbol{K}}')\int_0^\infty r^2 j_l(|\boldsymbol{K}|r)V_l(r)j_l(|\boldsymbol{K}'|r)\,\mathrm{d}r$$

$$\equiv \frac{4\pi}{\Omega}(2l+1)P_l(\hat{\boldsymbol{K}}\cdot\hat{\boldsymbol{K}}')U_l(|\boldsymbol{K}|, |\boldsymbol{K}'|), \tag{3.31}$$

where P_l denotes the Legendre polynomial of order l and j_l is the spherical Bessel function of order l. U_l is defined by the last equality.

3.5. *Kleinman–Bylander form of pseudopotential*

Kleinman and Bylander (KB) [34] pointed out a curious fact about pseudopotentials: by complicating their expression in real space, it is possible to speed considerably their manipulation in reciprocal space. Coining the term "semilocal" (sl) to denote the usual form eq. (3.31) in which the pseudopotential depends on angular momentum l but not on $|\boldsymbol{r}|$ and $|\boldsymbol{r}'|$ separately, KB defined a "truly non-local" (tnl, our notation) atomic pseudopotential as follows.

One uses a standard procedure, such as those of BHS [26], Kerker [31], or Vanderbilt [32], to arrive at the semilocal potentials $V_l^{\mathrm{ion}}(r)$. These are partitioned in the form

$$V_l^{\mathrm{ion}} = V_{\mathrm{local}} + \delta V_{\mathrm{sl},l}^{\mathrm{ion}}, \tag{3.32}$$

where the local potential V_{local} (which is independent of l) is chosen to make the residual semilocal parts $V_{\mathrm{sl}}^{\mathrm{ion}}$ as small as possible. They define the "truly non-local" potential by

$$\delta V_{\mathrm{tnl}}(\boldsymbol{r}, \boldsymbol{r}') = \sum_{lm} \delta V_{\mathrm{sl},l}^{\mathrm{ion}}(r)R_l^{\mathrm{ps}}(r)Y_{lm}(\hat{\boldsymbol{r}})\langle \delta V_{\mathrm{sl},l}\rangle^{-1}Y_{lm}^*(\hat{\boldsymbol{r}}')R_l^{\mathrm{ps}}(r')\delta V_{\mathrm{sl},l}^{\mathrm{ion}}(r'), \tag{3.33}$$

where the normalizing constant is given by

$$\langle \delta V_{\mathrm{sl},l}\rangle = \int \mathrm{d}^3 r\, R_l^{\mathrm{ps}}(r)Y_{lm}^*(\hat{\boldsymbol{r}})\delta V_{\mathrm{sl},l}^{\mathrm{ion}}(r)Y_{lm}(\hat{\boldsymbol{r}})R_l^{\mathrm{ps}}(r). \tag{3.34}$$

Here R_l^{ps} is the radial part of the pseudo-orbital which is used in defining V_l^{ion}, e.g. the non-relativistic limit of $W^{(2)}$ in section 3.2. (The relativistic generalization is straightforward.) It is readily verified that

$$\int \mathrm{d}^3 r'\, \delta V_{\mathrm{tnl}}(\boldsymbol{r}, \boldsymbol{r}')R_l^{\mathrm{ps}}(r')Y_{lm}(\hat{\boldsymbol{r}}') = \delta V_{\mathrm{sl},l}^{\mathrm{ion}}(r)R_l^{\mathrm{ps}}(r)Y_{lm}(\hat{\boldsymbol{r}}), \tag{3.35}$$

that is, the effect of this "truly non-local" potential on the pseudo-orbital is identical to that of the defining semilocal potential. Thus the two forms are equivalent atomic pseudopotentials insofar as they produce identical pseudoorbitals. Since they are however quite distinct potentials, this KB construction illuminates another dimension in the non-uniqueness of pseudopotentials. While the two forms *will not produce identical results* when applied to molecular or solid state situations, the results must be regarded as equally valid, and therefore must be very similar when the pseudopotential approximation is valid.

The utility of the KB form becomes clear when written in reciprocal space operator notation. While the standard form leads to an operator of the form

$$\sum_{GG'} |\boldsymbol{k}+\boldsymbol{G}\rangle \left[\sum_{l} \delta V^{\text{ion}}_{\text{sl},l}(\boldsymbol{k}+\boldsymbol{G},\, \boldsymbol{k}+\boldsymbol{G}')\right] \langle \boldsymbol{k}+\boldsymbol{G}'|, \tag{3.36}$$

the KB form leads to

$$\sum_{GG'} |\boldsymbol{k}+\boldsymbol{G}\rangle \left[\sum_{l} T_l(\boldsymbol{k}+\boldsymbol{G}) T_l^*(\boldsymbol{k}+\boldsymbol{G}')\right] \langle \boldsymbol{k}+\boldsymbol{G}'|$$

$$= \sum_{l} \left[\sum_{G} |\boldsymbol{k}+\boldsymbol{G}\rangle T_l(\boldsymbol{k}+\boldsymbol{G})\right] \left[\sum_{G'} T_l^*(\boldsymbol{k}+\boldsymbol{G}') \langle \boldsymbol{k}+\boldsymbol{G}'|\right], \tag{3.37}$$

where T can be identified from the Fourier transform of eq. (3.33). The process of taking matrix elements of (3.36) in a basis of N plane waves scales as N^2, while the *separable* form of (3.37) reduces this to a process which scales as N. This distinction becomes exceedingly useful in calculations involving many atoms in the unit cell (see the applications discussed in section 6).

3.6. Further developments

The pseudopotentials discussed so far are l-dependent. Starkloff and Joannopoulos [34] proposed that a local pseudopotential for transition metal atoms can be derived if the semicore s and p states are treated like the d states; that is, the pseudopotential describes the complete outer "valence" shell of electrons. They constructed the pseudopotential by first unscreening the atomic potential by subtracting the valence screening potential. This resulting intermediate potential is then "pseudized" by multiplying it by a smooth cutoff function

$$f(r;\, \lambda,\, r_c) = \frac{1 - e^{-\lambda r}}{1 + e^{-\lambda(r - r_c)}}, \tag{3.38}$$

which makes the resulting pseudopotential go to zero inside the core radius r_c.

The reciprocal length λ and core radius r_c are then varied to maximize the agreement between the all-electron eigenvalues and orbitals and their pseudized counterparts, for all angular momentum values simultaneously. They found that accurate results could be obtained with only these two parameters, suggesting the method provides a relatively "natural" pseudopotential. However, since the pseudopotential describes the semicore states, it requires a *considerably larger*

basis set for convergence if plane waves are used. Bylander and Kleinman [35] extended the application to semicore states of norm-conserving pseudopotentials by reverting to l-dependent potentials.

Another extension of the pseudopotential concept is necessary for the description of magnetic systems. Attempts by several workers have culminated in a method by Louie, Froyen and Cohen [36] which seems to be best. The problem in trying to formulate a spin-polarized pseudopotential method (within the local spin density approximation) stems from the strong non-linearity of the exchange-correlation term and the fact that it depends on the *sum of core and valence densities* for each spin. They resolve the difficulty by redefining the (norm-conserving, say) pseudopotential by unscreening with $V_{xc}(\boldsymbol{r};\, n_c + n_v)$ (core plus valence density) rather than with $V_{xc}(\boldsymbol{r};\, n_v)$ as in the BHS scheme. This makes it necessary to "carry around" the core density n_c in order to carry out a spin-polarized study, but this is a small price to pay for obtaining a useful, spin-independent pseudopotential for application to magnetic systems.

A final extension which we mention is the inversion scheme of Gardner and Holzwarth [37]. It is often useful to regain the information about the core region which is given up by choosing the pseudopotential method. Gardner and Holzwarth reasoned that, with the level of accuracy which is given by norm-conserving pseudopotentials, it should be possible to invert the pseudization process and recover information similar to what would be obtained in an all-electron frozen core calculation. They introduced a procedure, involving the use of the (frozen core) density, in which the all-electron valence wavefunctions are required to match the corresponding pseudowavefunctions outside the core radius and to become self-consistent (via an iterative process) with the resulting all-electron potential. Applications to several atoms [37] indicated the procedure provides reasonable accuracy, but the technique apparently has not yet been applied in solid state studies.

4. Formalism and algorithms

4.1. The secular equation

In density functional theory one solves iteratively for the density using the Kohn–Sham equations, Poisson's equation and an exchange-correlation contribution to the one-electron potential. The various steps are shown in flow-chart form in fig. 5. The secular equation is

$$\left(-\frac{1}{2m}\nabla^2 + V_{ion} + V_h + V_{xc}\right)\psi_{kn} = \epsilon_{kn}\psi_{kn}. \tag{4.1}$$

Here V_{ion} is the pseudopotential term

$$V_{ion}(\boldsymbol{r},\, \boldsymbol{r}') = \sum_{j,s} V_{ion}^{(s)}\left(\boldsymbol{r} - \boldsymbol{R}_j - \boldsymbol{\tau}_s,\, \boldsymbol{r}' - \boldsymbol{R}_j - \boldsymbol{\tau}_s\right), \tag{4.2}$$

where $V_{ion}^{(s)}$ arises from sth ion at position $\boldsymbol{\tau}_s$ in the unit cell at $\boldsymbol{R}_j$, and V_h, V_{xc} are Hartree and

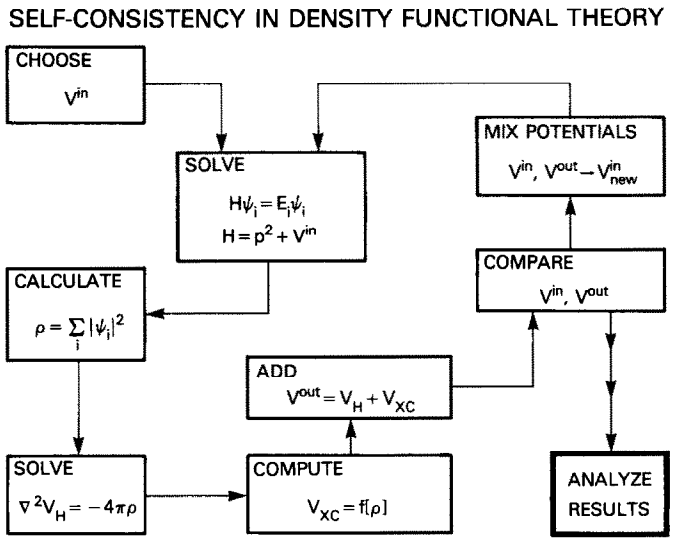

Fig. 5. Schematic diagram showing the steps in attaining self-consistency in Kohn–Sham density functional theory. To begin (upper left) a starting potential V^{in} is chosen, and inserted in the Hamiltonian H. The lowest N eigenfunctions (for an N-electron system) provides the electron density, which lead to the Hartree (V_H) and exchange-correlation (V_{xc}) potentials via Poisson's equation and a simple, usually local, functional, respectively. These potentials are combined to form the output screening potential, which is compared to the input potential. If the output potential is sufficiently close to the input potential, the system is self-consistent and analysis of results can proceed. Otherwise the output potential is mixed with the input potential to form a new input potential, and the procedure is iterated until convergence is attained.

XC potentials. The eigenvalues and eigenfunctions are denoted by ϵ_{kn} and ψ_{kn}, with k and n denoting wavevector and band index in a periodic system. In a plane wave basis ψ is expanded as

$$\psi_{kn}(r) = e^{ik\cdot r}\sum_G c_G(kn)\, e^{iG\cdot r}, \tag{4.3}$$

and the secular equation (4.1) becomes

$$\sum_{G'} H_{GG'}(k)c_{G'}(kn) = \epsilon_{kn}c_G(kn), \tag{4.4}$$

where

$$H_{GG'}(k) = \frac{1}{2m}|k+G|^2\delta_{G,G'} + V_{ion}(k+G,\, k+G') + V_h(G-G') + V_{xc}(G-G'). \tag{4.5}$$

In terms of the Fourier transform $V^{(s)}_{ion}(q, q')$ of $V^{(s)}_{ion}(r, r')$ from (3.31), the external pseudo-

potential can be written

$$V_{\text{ion}}(\boldsymbol{k}+\boldsymbol{G},\ \boldsymbol{k}+\boldsymbol{G}') = \sum_s V_{\text{ion}}^{(s)}(\boldsymbol{k}+\boldsymbol{G},\ \boldsymbol{k}+\boldsymbol{G}')\ \mathrm{e}^{\mathrm{i}(\boldsymbol{G}-\boldsymbol{G}')\cdot\boldsymbol{\tau}_s}. \tag{4.6}$$

If the crystal has a center of inversion which is chosen as the origin, the secular equation (4.4) is real, otherwise it is necessarily complex.

The secular equation (4.4) is infinite but tends to be diagonally dominant, both because the kinetic energy terms are large (except for very small $|\boldsymbol{G}|$) and because the potential terms decrease perpendicular to the diagonal (i.e., increasing $|\boldsymbol{G}-\boldsymbol{G}'|$). In constructing the density from

$$n(\boldsymbol{r}) = \sum_{kn} |\psi_{kn}(\boldsymbol{r})|^2 f(\epsilon_{kn}) \tag{4.7}$$

(where f is the zero temperature limit of the Fermi–Dirac distribution) only the lowest eigensolutions are needed, and these converge as the dimension N of the secular equation is increased. Unfortunately the diagonalization time increases as N^3, which tends to impose severe limits on the size or kind of system which can be treated.

4.2. Löwdin perturbation theory

Löwdin [38] suggested a procedure which can be used to extend the limit on the basis set size with an acceptable loss of accuracy. Consider the matrix $H_{nm} \equiv H_{G_n,G_m}$ partitioned as shown in fig. 6. Cutoffs $E_1 = (1/2m)G_1^2$, $E_2 = (1/2m)G_2^2$ are defined, with N_1 (resp. N_2) being the

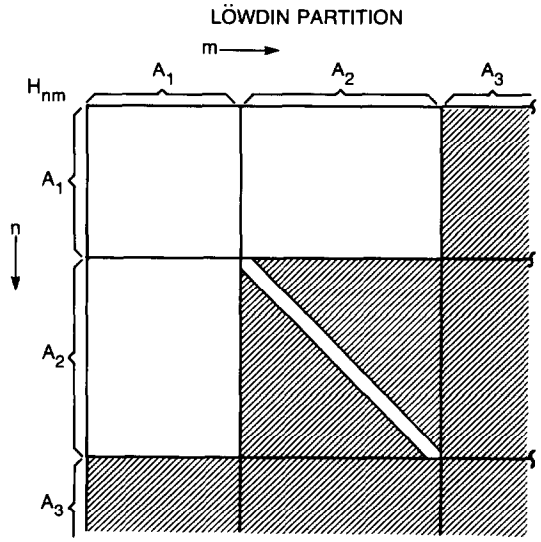

Fig. 6. The Löwdin partition into sets of basis functions A_1, A_2 and A_3. Hatched regions indicate matrix elements of the Hamiltonian H which are never calculated.

number of plane waves with $|\boldsymbol{G}| \leq G_1$ (resp. G_2), which separate the matrix into blocks A_1, A_2 and A_3 as illustrated in fig. 6. The high-lying basis states in A_3, which can be thought of as very short wavelength plane waves, are to be ignored. Löwdin derives the expressions

$$\sum_{n}^{A_1} (U_{mn} - E_i \delta_{mn}) c_n^{(i)} = 0, \quad m \text{ in } A_1 \tag{4.8}$$

$$c_m^{(i)} = \sum_{n}^{A_1} \frac{U_{mn}}{E - H_{mm}} c_n^{(i)}, \quad m \text{ in } A_2 \tag{4.9}$$

where to second order in H_{mn}, U is given by

$$U_{mn} = H_{mn} - \sum_{\alpha}^{A_2} K_{m\alpha} K_{n\alpha}^*, \quad m, n \text{ in } A_1, \tag{4.10}$$

with

$$K_{m\alpha} = K_{m\alpha}(E) = H_{m\alpha}/(H_{\alpha\alpha} - E)^{1/2}. \tag{4.11}$$

The N_1 eigenvalues E_i and eigenvector components $c_n^{(i)}$ for n in A_1 are given by eq. (4.8), and the remaining eigenvector components for m in A_2 are given by eq. (4.9). The cross-hatched region in fig. 6 indicates which matrix elements are never used.

These equations provide a Brillouin–Wigner-like perturbative approximation to the "exact" equation for the subspace $A_1 + A_2$,

$$\sum_{m}^{A_1, A_2} (H_{mn} - E\delta_{mn}) c_n = 0, \quad n \text{ in } A_1 \text{ and } A_2, \tag{4.12}$$

and they work best for the lowest eigenvalues, which however are the eigenvalues of interest. Thus one obtains eigensolutions approximating those of the $N_2 \times N_2$ matrix by diagonalizing only an $N_1 \times N_1$ matrix. The cost is that U becomes dependent on E, that is, the eigenvalue being sought, and if the $N_1 \times N_1$ matrix eq. (4.8) were to be solved exactly for the eigenvalues, such as by iterative diagonalization for each eigenvalue, the cost would be far too great. With a high enough choice of cutoffs, however, it is reasonable to assign E a particular value $\bar{E}$, which is chosen to be where the most accuracy is desired – in the band gap of a semiconductor or at the Fermi energy of a metal. Then a single N_1^3 diagonalization gives good approximations to the desired eigensolutions of an N_2^3 problem. Typical values which occur in practice are $N_2 \sim 2N_1$, which leads to nearly an order of magnitude ($2^3 = 8$) decrease in diagonalization time, or alternatively allows the treatment of twice as many basis functions (i.e., atoms) with fixed computational resources. Note that, when the eigenvector components in (4.9) are appended to those from (4.8), they must be renormalized to unity and reorthogonalized, such as by a Gram–Schmidt process.

4.3. Calculation of the density

Once the secular equation is solved, the density is constructed from eq. (4.7), which can be written, using a normalized summation convention where $\Sigma_k 1 = 1$,

$$n(\boldsymbol{r}) = \sum_{kn} n_{kn}(\boldsymbol{r}) f(\epsilon_{kn}) \tag{4.13}$$

in terms of orbital densities $n_{kn}(\boldsymbol{r})$. Straightforward algebra gives

$$n_{kn}(\boldsymbol{r}) = |\psi_{kn}(\boldsymbol{r})|^2 = \sum_{G''G'} c^*_{G''}(\boldsymbol{kn}) c'_{G}(\boldsymbol{kn})\, e^{i(G'-G'')\cdot r}, \tag{4.14}$$

which, in terms of the Fourier transform of $n_k n_{kn}$, is

$$n_{kn}(\boldsymbol{G}) = \sum_{G'} c^*_{G'-G}(\boldsymbol{kn}) c_{G'}(\boldsymbol{kn}). \tag{4.15}$$

However, there is a more efficient method which takes advantage of complex fast Fourier transform techniques (CFFT) [39]. The wavefunction coefficients $c_G(\boldsymbol{kn})$ are CFFTed into real space giving $u_{kn}(\boldsymbol{r})$, the periodic part of the Block function,

$$u_{kn}(\boldsymbol{r}) = \sum_{G} c_G(\boldsymbol{kn})\, e^{iG\cdot r}, \tag{4.16}$$

on a discrete real space mesh. Then $u_{kn}(\boldsymbol{r})$ is squared and summed over $\boldsymbol{k}$, n indices (methods of summation are discussed below). Finally the inverse CFFT provides the desired reciprocal space representation of the density. This procedure speeds the computation considerably, since eq. (4.15) scales as M^2, where M is the number of Fourier components in the density, while the CFFT technique scales as M log M.

Although the full Brillouin zone (BZ) sum in eq. (4.13) is required to construct the density, the space group symmetry allows the sum to be reduced to the *irreducible* BZ (IBZ). In addition, time reversal symmetry requires (dropping bands indices for simplicity).

$$\psi_{-k}(\boldsymbol{r}) = \psi^*_k(\boldsymbol{r}), \tag{4.17}$$

so ψ_k and ψ_{-k} always lead to the same orbital density. Here the IBZ will be used in a generalized sense in which $\boldsymbol{k}$ and $-\boldsymbol{k}$ never both appear in the IBZ. For example, with this convention the IBZ of a zincblende compounds is 1/48th of the BZ (as for the diamond structure) in spite of only 24 operations in the space group.

The zone sum can be written

$$n(\boldsymbol{G}) = \sum_{k}^{\mathrm{BZ}} n_k(\boldsymbol{G}) = \sum_{k}^{\mathrm{IBZ}} \sum_{\mathbf{S}} n_{\mathbf{S}k}(\boldsymbol{G}), \tag{4.18}$$

where $\mathbf{S}$ is the point group part of the space group operation S of the crystal which operates on a Bloch function as

$$S\psi_k(\boldsymbol{r}) = \psi_k(\mathbf{S}\boldsymbol{r} + \boldsymbol{t}_\mathbf{S})$$
$$= \sum_{\boldsymbol{G}} c_{\boldsymbol{G}}(\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}(\boldsymbol{k}+\boldsymbol{G})\cdot(\mathbf{S}\boldsymbol{r}+\boldsymbol{t}_\mathbf{S})}. \tag{4.19}$$

Here $\boldsymbol{t}_\mathbf{S}$ is the translation vector associated with a (non-symmorphic) symmetry operation S.

Since S is a symmetry operation of the crystal, its action on ψ_k produces another eigenfunction,

$$S\psi_k(\boldsymbol{r}) = \psi_{\mathbf{S}^{-1}k}(\boldsymbol{r}) \tag{4.20}$$

which is degenerate with ψ_k. Using

$$\psi_{\mathbf{S}^{-1}k}(\boldsymbol{r}) = \sum_{\boldsymbol{G}'} c_{\boldsymbol{G}'}(\mathbf{S}^{-1}\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}(\mathbf{S}^{-1}\boldsymbol{k}+\boldsymbol{G}')\cdot\boldsymbol{r}}$$
$$= \sum_{\boldsymbol{G}} c_{\mathbf{S}^{-1}\boldsymbol{G}}(\mathbf{S}^{-1}\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}(\boldsymbol{k}+\boldsymbol{G})\cdot\mathbf{S}\boldsymbol{r}} \tag{4.21}$$

which, together with (4.19) gives

$$c_{\mathbf{S}^{-1}\boldsymbol{G}}(\mathbf{S}^{-1}\boldsymbol{k}) = c_{\boldsymbol{G}}(\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}\boldsymbol{G}\cdot\boldsymbol{t}_\mathbf{S}} \tag{4.22}$$

or

$$c_{\boldsymbol{G}}(\mathbf{S}^{-1}\boldsymbol{k}) = c_{\mathbf{S}\boldsymbol{G}}(\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}\mathbf{S}\boldsymbol{G}\cdot\boldsymbol{t}_\mathbf{S}}. \tag{4.23}$$

From these equations and eq. (4.15) the useful result is

$$n_{\mathbf{S}^{-1}k}(\boldsymbol{G}) = \mathrm{e}^{\mathrm{i}\mathbf{S}\boldsymbol{G}\cdot\boldsymbol{t}_\mathbf{S}} n_k(\mathbf{S}\boldsymbol{G}) \tag{4.24}$$

so it follows that

$$n(\boldsymbol{G}) = \sum_{\mathbf{S}} \mathrm{e}^{\mathrm{i}\mathbf{S}\boldsymbol{G}\cdot\boldsymbol{t}_\mathbf{S}} \sum_{\boldsymbol{k}}^{\mathrm{IBZ}} n_k(\mathbf{S}\boldsymbol{G}) f(\epsilon_k). \tag{4.25}$$

The result is that the Fourier components of the density can be obtained from the state densities calculated within the IBZ, by carrying out the appropriate sums (4.25) over *stars* of $\boldsymbol{G}$. As mentioned above, time reversal requires ψ_k and ψ_{-k} to lead to the same state densities, so the IBZ can be reduced by a factor of two in eq. (4.25) (and an overall factor of two inserted) if the space group does not include the inversion operator.

4.4. Brillouin zone summation

Once the $\boldsymbol{k}$-point densities $n_{\boldsymbol{k}}(\boldsymbol{G})$ have been obtained, it is necessary to evaluate the sum over $\boldsymbol{k}$ in the BZ, or equivalently, the IBZ, in eqs. (4.13) or (4.25). This sum is done by discrete sampling, and several methods are in use. The arguments are elaborations of the "mean value point" of Baldereschi [40], who asked which *single point* in the Brillouin zone would give the best approximation to the zone average of an arbitrary periodic function.

Chadi and Cohen [41] (see also Chadi [42]) generalized the procedure as follows. One wants the average over the BZ of a smoothly varying periodic function g($\boldsymbol{k}$), which can be expanded as

$$g(\boldsymbol{k}) = \sum_{\boldsymbol{R}} g(\boldsymbol{R})\, \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}}. \tag{4.26}$$

Without loss of generality we can assume that g has the full Γ_1 symmetry of the crystal, since any function with lower symmetry has vanishing average over the BZ. Then g can be expanded in symmetrized plane waves ("star functions") $A_m(\boldsymbol{k})$ as

$$g(\boldsymbol{k}) = \sum_{m=0}^{\infty} g_m A_m(\boldsymbol{k}), \tag{4.27}$$

where

$$A_m(\boldsymbol{k}) \equiv \frac{1}{n_s} \sum_{\boldsymbol{R}}^{\text{star } m} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}}. \tag{4.28}$$

A star of lattice vectors is a set $\{\boldsymbol{R}\}$ which is transformed onto itself by the n_s point group operations, and these stars are ordered according to non-decreasing length. Since $A_m(\boldsymbol{k})$, $m \geq 1$, averages to zero over the BZ, the desired average of g is given by

$$g_0 = \sum_{\boldsymbol{k}} g(\boldsymbol{k}). \tag{4.29}$$

(Recall our convention that the symbol "$\Sigma_{\boldsymbol{k}}$" incorporates normalization factors such that $\Sigma_{\boldsymbol{k}} 1 = 1$.)

The strategy for achieving the approximate zone average is to choose a sequence of sets (labelled by M) of points $\{\boldsymbol{k}_j^{(M)}\}$ and associated weights $\{W_j^{(M)}\}$ which satisfy the conditions

$$\sum_{j=1}^{j_M} W_j^{(M)} A_m\left(\boldsymbol{k}_j^{(M)}\right) = 0, \quad m = 1, 2, \ldots, M, \tag{4.30}$$

$$\sum_{j=1}^{j_M} W_j^{(M)} = 1. \tag{4.31}$$

Then the Mth approximation to the average is given by the sum

$$\bar{g}^{(M)} = \sum_{j=1}^{j_M} W_j^{(M)} g\left(\boldsymbol{k}_j^{(M)}\right),$$

$$= \sum_{m=0}^{\infty} g_m \sum_{j=1}^{j_M} W_j^{(M)} A_m\left(\boldsymbol{k}_j^{(M)}\right)$$

$$= g_0 + \sum_{m=M+1}^{\infty} g_m \sum_{j=1}^{j_M} W_j^{(M)} A_m\left(\boldsymbol{k}_j^{(M)}\right), \tag{4.32}$$

which gives an approximation to g_0 that omits contributions only from stars $M+1$ and beyond. Note that condition (4.30) depends only on the crystal lattice; that is, the symmetrized plane waves, so the $\boldsymbol{k}$ point sets are independent of the function to be averaged.

For a smooth function the expansion coefficients g_m decrease rapidly with growing m, and this method provides a rapidly convergent zone sampling. An example of a smooth function $g(\boldsymbol{k})$ is the charge density $n_{\boldsymbol{k}}$ of a non-metal (summed over all occupied bands). For a metal, however, the sharp Fermi surface gives discontinuities in $n_{\boldsymbol{k}}$ which may in principle introduce large expansion coefficients for large m. Nevertheless, the method is also widely used for metals, usually with reasonable results. This success probably reflects the likelihood that the discontinuities introduced by the Fermi surface are very small compared to the density itself.

Baldereschi [40] considered the best approximation that could be obtained from a single point. Since a three dimensional point $\boldsymbol{k}$ has three coordinates with which to satisfy eqs. (4.30) and (4.31), in general an $M=3$ approximation can be obtained. For the simple cubic lattice this "mean value point" is $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})\pi/a$. For the fcc and bcc lattices there is no single point which gives an $M=3$ solution, while there is curve of points which give $M=2$ solutions. Baldereschi defined the mean value point in such cases as the point on this curve which minimizes $|A_3(\boldsymbol{k})|$, obtaining the result $(0.6223, 0.2953, 0.0)2\pi/a$ for fcc and $(1/6, 1/6, 1/2)2\pi/a$ for bcc. He demonstrated that this "mean value point" gave a more accurate approximation to the sum of occupied eigenvalues in Ge, GaAs and ZnSe than an 8 point sampling proposed by Kleinman and Phillips [43]. (N.B. the term "mean value point" is a misnomer, since it does not give the mean value point but only an approximation to it. By the mean value theorem, a true mean value point (and usually many of them) does exist, but depends on the function in question. A more accurate term for this optimal point is the "Baldereschi point.")

Although the Chadi–Cohen condition (4.30) leads to a highly non-linear problem, the periodicity and point group symmetry of the symmetrized plane waves lead to useful simplifications. Chadi and Cohen have given procedures for generating point sets and have presented small sets for cubic and hexagonal lattices. Monkhorst and Pack [44] have developed a more systematic prescription based on equally spaced points. They obtain the result

$$\bar{g}_m = \sum_{j=1}^{J} W_j g(\boldsymbol{k}_j) A_m(\boldsymbol{k}_j) \tag{4.33}$$

where the $\boldsymbol{k}$-point set is given, in terms of the primitive reciprocal lattice vectors $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, $\boldsymbol{G}_3$, by

$$\boldsymbol{k}_{\mathrm{prs}} = \alpha_1(p)\boldsymbol{G}_1 + \alpha_2(r)\boldsymbol{G}_2 + \alpha_3(s)\boldsymbol{G}_3. \tag{4.34}$$

The fractions α_i are given by

$$\alpha_i(n) = (2n - N_i - 1)/2N_i, \quad n = 1, 2, 3, \ldots, N_i, \tag{4.35}$$

and N_1, N_2, N_3 are indices which determine the number of points in the set. MacDonald [45] has presented special point formulas based on product Gauss–Chebyshev integration formulas which may be more efficient for large special point sets. It is not clear that any "best" $\boldsymbol{k}$ point sampling can be specified, but the Monkhorst–Pack scheme appears to be reliable and it is straightforward to code for a general situation.

4.5. Iteration to self-consistency

In any calculation the majority of the computational time is spent in iterating the Kohn–Sham eqs. (2.7)–(2.9) to obtain a self-consistent density. The practice is to use, by methods discussed below, the "output" density from eq. (2.9) and the "input" density which was used to construct V_{eff} in eq. (2.7) to obtain a new "input" density which is nearer to the self-consistent one. Due to the long-range nature of the Coulomb interaction, a small change in the input density can lead to a relatively large change in the output density. As a result, it is usually not possible to use the output density itself as the input density for the next iteration, since large unstable charge oscillations arise; rather it is essential to mix input and output densities in an appropriate manner to obtain a new input density. (It should be mentioned here that, due to the linearity of Poisson's equation, mixing potentials is nearly equivalent to mixing densities. A small difference arises from the exchange-correlation potential, but it is local and leads to no problem in attaining self-consistency. In the discussion to follow we will use the density as the quantity which is mixed. Some investigators (Yaffe and Goddard [46], Harris [47]) have suggested however that there may be some advantage in mixing Coulomb and xc potentials separately, but this practice has not gained widespread use.)

An important factor, and one which is rarely discussed, is the best choice for a starting density or potential. In the earliest self-consistent calculations the common practice was to use overlapping, spherical screened pseudopotentials. Typically these were simple functional forms constrained to fit 3 "form factors" $V_{\mathrm{emp}}(\boldsymbol{G})$ determined empirically from the elemental solid (such as for Si and Al) or a simple compound (such as for Ga and As from GaAs). For simple calculations, such as for an elemental solid or even a high symmetry binary compound, the starting configuration, and even the mixing procedure, is unimportant because the calculation is small and converges easily. One of the earliest application of the self-consistent pseudopotential method was to the study of surfaces and interfaces, which quickly brought out the difficulties which can occur.

The determination of the electronic structure of surfaces and interfaces was studied using a supercell geometry in which one deals with a periodically repeated film which is itself a few atomic layers thick (in the case of surfaces) or a multilayer geometry (for the interface problem).

The unit cell has a very large dimension d perpendicular to the surface or interface, leading to small Fourier components ($G_z = 2\pi/d$) of the Coulomb potential which are very sensitive to shift of charge across the surface or interface. This "sloshing" of charge led to great difficulties in achieving self-consistency and spurred the search for more successful iteration schemes.

The determination of a good starting configuration is still something of a "black art," with each practitioner developing a personal intuition. The use of overlapping potentials is one possibility, but the relative position (in energy) for inequivalent atoms involves guesswork. Problems with this approach have been discussed by Manghi and Molinari [48]. Overlapping atomic densities, usually taken to be spherical, gives a well-defined potential for the system but still requires deciding what atomic (ionic) configuration is most appropriate. Although this question deserves further study, the greatest effort has been expended on developing more sophisticated iteration procedures, for which general systematic procedures can be developed and analyzed.

At each iteration M of the Kohn–Sham procedure one has an input density n_M^{in} which defines the screening potential and the resulting output density n_M^{out}. One can regard the Kohn–Sham procedure as a functional $F[n]$ which satisfies

$$n_M^{\text{out}} = F\left[n_M^{\text{in}}\right]. \tag{4.36}$$

The problem is to find the self-consistent density which satisfies the fixed point condition

$$n = F[n]. \tag{4.37}$$

The most widely used procedure is *linear mixing*, for which

$$n_{M+1}^{\text{in}} = \alpha n_M^{\text{out}} + (1-\alpha) n_M^{\text{in}} = n_M^{\text{in}} + \alpha\left(F\left[n_M^{\text{in}}\right] - n_M^{\text{in}}\right) \tag{4.38}$$

where α, $0 < \alpha \leq 1$, is the mixing parameter that tempers the large charge oscillations that may occur. For "small enough" α this procedure is relatively safe, but in some applications it can take *hundreds* of iterations to give a reasonable level of self-consistency. Occasionally a "stiff" situation is reached for which continued iteration does not improve the "distance" D between input and output potentials. The simplest definition of a distance is

$$D[n_1, n_2] = \left\{ \frac{1}{\Omega_c} \int_{\Omega_c} \mathrm{d}^3 r \left[n_1(\boldsymbol{r}) - n_2(\boldsymbol{r})\right]^2 \right\}^{1/2}, \tag{4.39}$$

although a weighted difference could be used. A common practice when a "stiff" position is reached is to greatly increase α *for one iteration*, often by taking a full output density as the new input density ($\alpha = 1$). This procedure usually moves the density well away from the region of stiffness and leads to further progress toward self-consistency. Reasons for this stiffness will arise in the discussion below.

Before reviewing the most successful iteration schemes, it is useful to consider the problem implied by eqs. (4.36) and (4.37) as generally as possible. Dederichs and Zeller (DZ) [49] have analyzed this problem in some detail, making the (almost universally made) assumption that one

is near self-consistency. At the fixed point the functional derivative $K(\boldsymbol{r}, \boldsymbol{r}')$ of $F[n]$ with respect to $n(\boldsymbol{r}')$, i.e. $\delta n^{\text{out}}(\boldsymbol{r})/\delta n^{\text{in}}(\boldsymbol{r}')$, is directly tied to the static dielectric response of the system. DZ consider the repeated linear mixing procedure applied to an approximate density $n_i = n + \delta n_i$, where n is the fixed point solution. Then the Mth mixing produces

$$\begin{aligned}
\delta n_M &= \alpha\big(F[n + \delta n_{M-1}] - n\big) + (1-\alpha)\delta n_{M-1} \\
&= \alpha K \delta n_{M-1} + (1-\alpha)\delta n_{M-1} \\
&= \alpha(K-1)\delta n_{M-1} + \delta n_{M-1} \\
&= \alpha(K-1)\big[\alpha(K-1)\delta n_{M-2} + \delta n_{M-2}\big] + \big[\alpha(K-1)\delta n_{M-2} + \delta n_{M-2}\big] \\
&= \big[1 - \alpha(1-K)\big]^M \delta n_0 .
\end{aligned} \tag{4.40}$$

where the last equality follows by induction. DZ also show that K is related by

$$1 - K = e^{\mathsf{T}}, \tag{4.41}$$

to the transpose of the static dielectric function. Expanding ϵ^{T} in its (left and right) eigenvectors and its eigenvalues ϵ_m,

$$\epsilon^{\mathsf{T}} = \sum_m \epsilon_m |m_l\rangle\langle m_r|, \quad \langle m_r | m_l' \rangle = \delta_{mm'}, \tag{4.42}$$

one obtains

$$\delta n_M = \sum_m (1 - \alpha\epsilon_m)^M |m_l\rangle\langle m_r| \delta n_0 . \tag{4.43}$$

The iteration procedure is convergent as $M \to \infty$ only if every eigenvalue ϵ_m satisfies

$$0 < \epsilon_m < 2/\alpha, \tag{4.44}$$

or, alternatively, only if

$$\alpha < 2/\max\{\epsilon_m\}, \tag{4.45}$$

Two important general conclusions follow immediately: (1) there is a *maximum mixing parameter* $\alpha^{\max}$ for which linear mixing will lead to convergence, and (2) for any system there is *some finite value* of α which will produce convergence by linear mixing. These conclusions of course assume one is close enough to the fixed point that the functional derivative K describes the response of the system. DZ provide a more detailed analysis of linear mixing and discuss several improved mixing schemes for which their original paper (and a related paper by Ferreira [50]) should be consulted. One of their examples may however be related to the stiffness mentioned

above. After several iterations it may happen that only one component j of δn is dominant, corresponding to

$$\alpha\epsilon_j \ll 1, \quad \text{or } \beta \equiv 1 - \alpha\epsilon_j \approx 1, \tag{4.46}$$

which leads to a slowly decreasing residual density

$$\delta n_{M+1} \cong \beta \delta n_M \cong \beta^2 \delta n_{m-1}. \tag{4.47}$$

DZ show that the optimal choice of β for further iteration is

$$\beta = \frac{n_{M+1}(\boldsymbol{r}) - n_M(\boldsymbol{r})}{n_M(\boldsymbol{r}) - n_{M-1}(\boldsymbol{r})} = \frac{\delta n_{M+1}(\boldsymbol{r}) - \delta n_M(\boldsymbol{r})}{\delta n_M(\boldsymbol{r}) - \delta n_{M-1}(\boldsymbol{r})}, \tag{4.48}$$

rather than $\beta = \delta n_{M+1}/\delta n_M$ as implied by (4.47). In the asymptotic region β will be independent of $\boldsymbol{r}$, and it clearly carries the flavor of a quadratic, rather than linear, extrapolation. The result should be to lead to an increased value of α, similar to the common practice for dealing with stiff regions in the iteration procedure.

Recently several iteration schemes which transcend linear mixing have been developed. A straightforward generalization of simple linear mixing due to Anderson [51] uses a linear combination of input and output densities of *two successive iterations* to provide an optimized input $\bar{n}^{\text{in}}$ and output $\bar{n}^{\text{out}}$:

$$\bar{n}^{\text{in,out}} = \alpha_M n_m^{\text{in,out}} + (1 - \alpha_M) n_{M-1}^{\text{in,out}}. \tag{4.49}$$

Note that input densities are mixed to give a new input, and output densities are mixed to give a new output density ("*predicted*" not calculated!). In addition the mixing parameter changes at each iteration. One would like to require $\bar{n}_M^{\text{out}} = \bar{n}_M^{\text{in}}$, but since each is a function of $\boldsymbol{r}$, this cannot be achieved by a single constant α_M. Rather the procedure is to vary α_M to *minimize* the distance, defined in eq. (4.39), between $\bar{n}_M^{\text{out}}$ and $\bar{n}_M^{\text{in}}$, leading to the value

$$\alpha_M = -\frac{1}{\Omega_c} \int \mathrm{d}^3 r \, \Delta_{M-1}(\boldsymbol{r}) [\Delta_M(\boldsymbol{r}) - \Delta_{M-1}(\boldsymbol{r})] / D^2[\Delta_M, \Delta_{M-1}] \tag{4.50}$$

where D is the metric (4.39) and

$$\Delta_M = n_M^{\text{out}} - n_M^{\text{in}} \tag{4.51}$$

is the difference between output and input densities at the Mth iteration. Then the new input density is obtained by mixing $\bar{n}^{\text{in}}$ and $\bar{n}^{\text{out}}$:

$$n_{M+1}^{\text{in}} = \gamma \bar{n}_M^{\text{out}} + (1 - \gamma) \bar{n}_M^{\text{in}}; \tag{4.52}$$

generally $\gamma = 1$ will not lead to convergence, and the usual practice is to treat γ like the mixing parameter in the simple linear mixing scheme. Unlike simple linear mixing where one makes the

restriction $0 < \alpha \leq 1$, eq. (4.50) can, and often does, lead to values of α_M which are either negative or also significantly larger than unity. The author's experience is that modest deviations of α_M from the interval [0, 1] can be helpful in speeding convergence, while large deviations from this interval are detrimental. A reasonable compromise seemed to be to restrict α_M to the interval $[-1, 3]$, by assigning it the nearest endpoint value when the calculated value is outside. Another practice has been to alternate between simple linear mixing and Anderson mixing, which seems to temper the irregularities which can arise when Anderson mixing alone is used. The Anderson scheme was apparently introduced to condensed matter studies by Hamann [52]; see Mattheiss and Hamann [53] for a discussion. The scheme was also implemented by Ng [54] in plasma studies.

Ho, Ihm and Joannopoulos [55] introduced a mixing scheme based on linear response which is particularly useful for plane-wave-based methods with large unit cells, such as surface calculations using a supercell approach. In such cases the potential contains long wavelength Fourier components which are especially sensitive to fluctuations of the density perpendicular to the plane of the surface, so it is more illustrative to discuss the potential rather than the density. The idea is to determine, at a stage where the input potential V^{in} has led to the output potential V^{out}, an alteration δV^{in} which produces self-consistency:

$$V^{\text{in}} + \delta V^{\text{in}} = V^{\text{out}} + \delta V^{\text{out}}. \tag{4.53}$$

Linear response gives

$$\delta n_G = \sum_{G'} \chi_{GG'} \delta V^{\text{in}}_{G'}, \tag{4.54}$$

with the non-interacting susceptibility χ given by

$$\chi_{GG'} = 2 \sum_{\substack{\boldsymbol{k} \\ nn'}} \frac{f_{kn'} - f_{kn}}{\epsilon_{kn'} - \epsilon_{kn}} \langle \boldsymbol{kn} | e^{-i\boldsymbol{G}\cdot\boldsymbol{r}} | \boldsymbol{kn}' \rangle \langle \boldsymbol{kn}' | e^{i\boldsymbol{G}'\cdot\boldsymbol{r}} | \boldsymbol{kn} \rangle. \tag{4.55}$$

The matrix elements are given by the standard form

$$\begin{aligned} \langle \boldsymbol{kn} | e^{i\boldsymbol{G}\cdot\boldsymbol{r}} | \boldsymbol{kn}' \rangle &= \frac{1}{\Omega} \int_{\Omega} d^3\boldsymbol{r}\, \psi^*_{\boldsymbol{kn}}(\boldsymbol{r})\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}} \psi_{\boldsymbol{kn}'}(\boldsymbol{r}), \\ &= \frac{1}{\Omega_c} \int_{\Omega_c} d^3\boldsymbol{r}\, u^*_{\boldsymbol{kn}}(\boldsymbol{r})\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}} u'_{\boldsymbol{kn}}(\boldsymbol{r}), \end{aligned} \tag{4.56}$$

where $u_{\boldsymbol{kn}}$ is the cell-periodic part of the Bloch wavefunction.

The self-consistent field electron dielectric matrix (DM) is defined by

$$\epsilon_{GG'} = \delta_{GG'} - \sum_{G''} (v_G \delta_{GG''} + v_{\text{xc},GG''}) \chi_{G''G'} \tag{4.57}$$

and of course is the physical DM only at self-consistency. Here v_G is the Fourier transform of the Coulomb interaction $e^2/|\boldsymbol{r}-\boldsymbol{r}'|$ given by

$$v_G = \frac{4\pi e^2}{\Omega_c G^2} \tag{4.58}$$

and v_{xc} is the Fourier transform of the xc *interaction* $v_{xc}(\boldsymbol{r}, \boldsymbol{r}')$, which is the functional derivative of the xc potential

$$v_{xc}(\boldsymbol{r}, \boldsymbol{r}') = \frac{\delta V_{xc}(\boldsymbol{r}; n)}{\delta n(\boldsymbol{r}')} = \frac{\delta^2 E_{xc}[n]}{\delta n(\boldsymbol{r})\delta n(\boldsymbol{r}')}. \tag{4.59}$$

In LDA v_{xc} is a local function

$$v_{xc}(\boldsymbol{r}, \boldsymbol{r}') \to v_{xc}^{LDA}(\boldsymbol{r})\delta(\boldsymbol{r}-\boldsymbol{r}') = \left[n(\boldsymbol{r})\epsilon_{xc}''(n(\boldsymbol{r})) + 2\epsilon_{xc}'(n(\boldsymbol{r}))\right]\delta(\boldsymbol{r}-\boldsymbol{r}'), \tag{4.60}$$

which follow from the LDA expression [4–8]

$$E_{xc}[n] = \int d^3r\, n(\boldsymbol{r})\epsilon_{xc}(n(\boldsymbol{r})). \tag{4.61}$$

In eq. (4.60) primes denote derivatives with respect to $n(\boldsymbol{r})$.

Equation (4.53) then leads to the condition for δV^{in}:

$$\sum_{G'} \epsilon_{GG'}\delta V_{G'}^{in} = V_G^{out} - V_G^{in}, \tag{4.62}$$

which has the interpretation that the input potential must be changed by the lack of self-consistency in the potential (ΔV) screened by the response of the system: $\delta V^{in} = \epsilon^{-1}\,\Delta V$. This scheme is a generalization of a scheme suggested by Kerker [56] which used a model dielectric matrix (e.g. Thomas–Fermi) in the diagonal approximation. Kerker's scheme, which had been applied in similar form earlier [57,58] in jellium-based calculations, provided a $\boldsymbol{G}$-dependent mixing which itself greatly improved the attainment of self-consistency.

Ho et al. applied their DM scheme to large surface calculations and noted the following features:

(1) Only the low Fourier components (small $|\boldsymbol{G}|$) are strongly coupled, so the DM need be calculated for only a small number of $\boldsymbol{G}$'s (~ 50 for their surface calculation). Linear mixing is sufficient for the remaining $\boldsymbol{G}$'s.

(2) Calculation of the resulting small DM is not time-consuming, requiring typically one-tenth the time of an iteration in the case of a plane wave basis.

(3) Elements of the DM typically change by less than 10% in approaching self-consistency, so it is not necessary to re-calculate it at each iteration.

(4) The predicted components of δV^{in} can be much different than any linear mixing would predict, e.g. the value of $V_G^{in,M+1}$ may not lie between $V_G^{in,M}$ and $V_G^{out,M}$. This feature of "overshooting," also allowed in the Anderson [51] scheme discussed above, appears to be a necessary component of improved interation procedures.

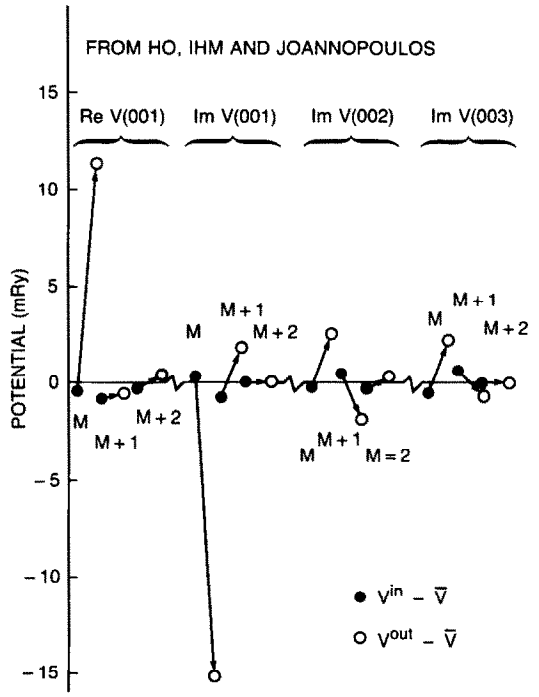

Fig. 7. Behavior of three of the small reciprocal lattice vector components of the potential in a Ag surface calculation by Ho, Ihm and Joannopoulos [55]. M, $M+1$ and $M+2$ denote three successive iterations, and the ordinate measures the difference between input (●), and output (○), potential and the self-consistent value. Note, for example, that for Re $V(001)$, the input for $M+1$ does not lie between the input and output of iterations M, reflecting the strong coupling with other Fourier components.

Feature (4) contains the important result: the DM provides new aspects which are unattainable with linear mixing, or with any mixing which does not couple different Fourier components. Ho et al. presented results for the five longest wavelength components for a nine-layer Ag surface calculation. For the iteration history they presented, 3 of the 5 components "overshot" due to coupling, and the distance between V_G^{out} and V_G^{in} decreased by factors of 3–10 for the three small $|G|$ components, which were the only ones a significant distance from self-consistency. Clearly it would have taken many linear mixes, or even Kerker mixes, to achieve the same improvement. The result is illustrated for $G = (0, 0, n)2\pi/c$ in fig. 7 for $n = 1$, 2 and 3.

Feature (3) reflects an important aspect of improving iteration schemes – linear mixing is such a crude procedure for pursuing self-consistency that it is not necessary to be very sophisticated to obtain drastic improvements to it. It rather is necessary only to introduce coupling in a manner which is qualitatively and semi-quantitatively correct. The *Broyden scheme* [59] is such a procedure which has been achieving great success in this field.

Broyden's method was introduced to solid state calculations by Bendt and Zunger [60] and discussed in more detail by Srivistava [61]. It is useful to introduce some notational conveniences. An unlabeled density indicates an input: $n_M \equiv n_M^{in}$, and we introduce an L-vector notation for densities, $n(r) \rightarrow |n\rangle$, where the components of $|n\rangle$ denote the L coefficients used in the

representation of $n(\boldsymbol{r})$, such a coefficients of plane waves, local orbitals, or the like. Finally, we introduce a notation for difference

$$\delta q_M \equiv q_M - q_{M-1} \tag{4.63}$$

for any quantity q.

One considers the quantity

$$|\Delta(n)\rangle = |F(n)\rangle - |n\rangle, \tag{4.64}$$

with F given by eq. (4.36), so $|\Delta(n)\rangle$ is the difference between output and input densities. Self-consistency is achieved when $|\Delta(n)\rangle = |0\rangle$. Expanding at the Mth iteration gives

$$|\Delta(n)\rangle = |\Delta(n_M)\rangle + J_M(|n\rangle - |n_M\rangle) \tag{4.65}$$

in terms of the Jacobian

$$J_{M,ij} \equiv \left.\frac{\delta|\Delta\rangle_i}{\delta|n\rangle_j}\right|_{n=n_M}. \tag{4.66}$$

The new density is given by the quasi-Newton–Raphson expression (4.65) by requiring $|\Delta(\mathrm{n}_{\mathrm{M}+1})\rangle$ to vanish,

$$|n_{M+1}\rangle = |n_M\rangle - J_M^{-1}|\Delta_M\rangle, \tag{4.67}$$

where the notation has been shortened by defining $|\Delta_M\rangle = |\Delta(n_M)\rangle$. Equation (4.67) can be seen to be equivalent to the DM expression (4.62). The rate of convergence is determined by the quality of the Jacobian, which is necessarily approximate, and a bad Jacobian can lead to divergence.

Since it is J_M^{-1} rather than J_M which is ultimately needed in eq. (4.67), Srivastava [61] suggested sidestepping the matrix inversion by using Broyden's "Method 2." Assuming a reasonable initial value of J^{-1}, which can be a diagonal constant corresponding to linear mixing

$$J_0^{-1} = -\alpha I, \tag{4.68}$$

Broyden's method consists of improving J^{-1} successively at each iteration. At each iteration it is required that the new J_M^{-1} will be one which would have given the result of the iteration just completed,

$$|g\rangle \equiv |\delta n_M\rangle - J_M^{-1}|\delta\Delta_M\rangle = |0\rangle, \tag{4.69}$$

which provides L constraints on the L^2 components of J_M^{-1}. The other $L(L-1)$ components are determined by the requirement that changes predicted by J_M^{-1} in *every* direction perpendicular to $|\delta\Delta_M\rangle$ are identical to that predicted by J_{M-1}^{-1}, which is written as

$$J_M^{-1}|x\rangle = J_{M-1}^{-1}|x\rangle \quad \text{for every } |x\rangle \quad \text{for which } \langle x|\delta\Delta_{M-1}\rangle = 0. \tag{4.70}$$

This condition expresses the restriction that, in the absence of any new information about these perpendicular directions, no change should be allowed. This procedure is equivalent to a "minimum change" criterion on J^{-1} using the Frobenius norm, which can be restated in the form

$$Q \equiv \| J_M^{-1} - J_{M-1}^{-1} \| \equiv \sum_{ij} \left(J_{M,ij}^{-1} - J_{M-1,ij}^{-1} \right)^2 \text{ be minimized,} \tag{4.71}$$

subject to condition (4.69). Using the Lagrange multiplier method and minimizing the functional

$$K = Q + \langle \lambda \mid g \rangle \tag{4.72}$$

with respect to J_M^{-1} yields

$$J_M^{-1} = J_{M-1}^{-1} + \tfrac{1}{2} \mid \lambda \rangle \langle \delta \Delta_M \mid . \tag{4.73}$$

Multiplying by $\mid \delta \Delta_M \rangle$ and using eq. (4.69) gives the Lagrange multiplier vector

$$\mid \lambda \rangle = 2\left(\mid \delta n_M \rangle - J_{M-1}^{-1} \mid \delta \Delta_M \rangle \right) / \langle \delta \Delta_M \mid \delta \Delta_M \rangle . \tag{4.74}$$

Replacement in eq. (4.73) gives the updating procedure

$$J_M^{-1} = J_{M-1}^{-1} + \frac{\left(\mid \delta n_M \rangle - J_{M-1}^{-1} \mid \delta \Delta_M \rangle \right) \langle \delta \Delta_M \mid}{\langle \delta \Delta_M \mid \delta \Delta_M \rangle} . \tag{4.75}$$

This can be rewritten in a form which is somewhat more illuminating,

$$J_M^{-1} = J_{M-1}^{-1} \left(1 - \mid \delta \hat{\Delta}_M \rangle \langle \delta \hat{\Delta}_M \mid \right) + \mid \delta \hat{n}_M \rangle \langle \delta \hat{\Delta}_M \mid \left(\frac{\langle \delta n_M \mid \delta n_M \rangle}{\langle \delta \Delta_M \mid \delta \Delta_M \rangle} \right)^{1/2} \tag{4.76}$$

in terms of normalized "directions"

$$\mid \hat{x} \rangle \equiv \mid x \rangle / \langle x \mid x \rangle^{1/2} . \tag{4.77}$$

Thus, in terms of its operation on a vector, the first term indicates that J_M^{-1} differs from J_{M-1}^{-1} by projecting out anything parallel to $\mid \delta \hat{\Delta}_M \rangle$ while the second term indicates that the part parallel to $\mid \delta \Delta_M \rangle$ will be rotated into the direction of $\mid \delta n_M \rangle$ with altered amplitude. The Broyden method as well as many other quasi-Newton methods have been discussed in detail by Dennis and Moré [62].

Numerous tests by several investigators have indicated that this procedure often leads to dramatic improvements, and selected examples will be summarized below. These successes make it puzzling that, although Broyden insisted the procedure is formally valid, in his numerical tests he found it to be "unsatisfactory" and even "quite useless," and therefore opted for slightly different procedures. In the light of this method's recent successes, it appears the difficulty was only with Broyden's model system or his implementation rather than with the method itself.

Vanderbilt and Louie [63] suggested an extension of the Broyden method which they found to give even better performance. Their approach has been adapted by Johnson [64] to the inverse Jacobian and I will follow this line to provide continuity from the discussion above. The argument is that the updating of J^{-1} to produce exactly the results of the iteration just completed is too severe an alteration, since it will no longer produce the results of any previous iteration. It is suggested instead that information be incorporated from all previous iterations at each update by carrying out a weighted least-squares minimization of the "error function"

$$Q' = \sum_{l=0}^{M-1} w_l \left| \, |\, \delta n_l \rangle - J_M^{-1} \, |\, \delta \Delta_l \rangle \right|^2 + w'^2 \left\| J_M^{-1} - J_0^{-1} \right\|. \tag{4.78}$$

This approach varies J_M^{-1} to minimize the discrepancy between "predicted" and actual output for *all previous iterations* l with preassigned weights w_l. The term in w' eliminates any arbitrary changes in J^{-1} which would not affect the first term.

Vanderbilt and Louie noted some desirable features of this extended Broyden method. The weights w_l and w' introduce the flexibility to allow more recent iterations to have a greater influence than earlier, less accurate (i.e. less relevant) ones. It is usually appropriate to give small weight to the stabilizing term involving w'. Perhaps most importantly, they established for a simple example that with the modified procedure the approximate Jacobian rapidly approaches the true Jacobian while for the straightforward Broyden method this approach is very slow.

Singh, Krakauer and Wang [65] have reported tests of the Broyden method (as outlined by Srivastava) for the $W(001)$ surface. Although their calculations did not pseudopotentials, they used the plane-wave-based Linearized Augmented Plane Wave method and their results should be representative of what can be achieved in pseudopotential studies. Using the distance between input and output densities defined in eq. (4.39), their results for twenty iterations are compared

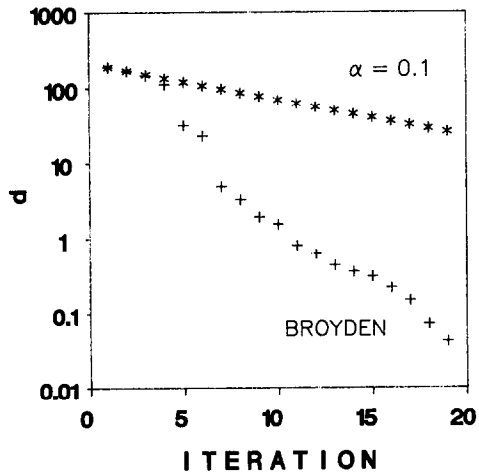

Fig. 8. Distance d between input and output densities for a $W(001)$ surface calculation by Singh, Krakauer and Wang [65]. Crosses denote the results obtained with Broyden mixing, while results of the much slower convergence resulting from 10% linear mixing are denoted by asterisks.

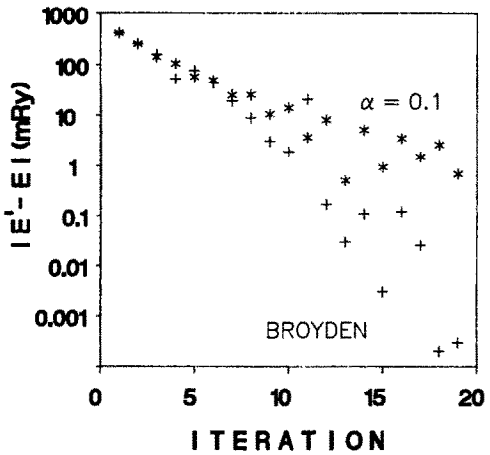

Fig. 9. Same calculation as in fig. 8, but illustrating the convergence of the total energy E' to the converged value E. Note the logarithmic scale for the energy.

in fig. 8 with the linear mixing scheme with 10% mixing. It is apparent that, once a small history is accumulated (four iterations for this example), the Broyden scheme provides a phenomenal improvement over linear mixing. The distance after twenty iterations is three orders of magnitude smaller using Broyden mixing.

The rapid convergence towards self-consistency (vanishing distance between input and output densities) is reflected in the calculated total energy and work function, both of which are important quantities to extract accurately from a surface calculation. Fig. 9 shows the calculated total energy, which is within 0.2 mRy of the converged value after 12 iterations using the Broyden scheme. After 20 iterations of linear mixing the energy is still 1 mRy from convergence, and it is sometimes difficult to achieve a stable, well converged value of the energy (say, to 0.1 mRy) with simple linear mixing. Similar greatly improved convergence of the work function was also obtained with Broyden mixing.

4.6. Localized basis sets

The discussion up to this point has assumed the use of a plane wave (PW) basis set. The PW basis has a number of attractive features. First, and foremost from many points of view, is its simplicity. The basis functions are orthogonal, eliminating any considerations of the overlap matrix. The basis functions are simple; in most cases the expressions are analytic and identical for each function, and the natural reciprocal space formalism is directly applicable. Fast Fourier transformation methods also can be applied readily.

Secondly, the basis set can be systematically improved by including additional plane waves, choosing the longest wavelength members of unused PWs (i.e. higher and higher energy cutoffs $|\boldsymbol{G}^2|/2m$). The reciprocal space expansion of the pseudopotential can even be used to determine when *exact* convergence of the wavefunctions has been obtained.

Thirdly, and very important in geometrically "open" systems such as surfaces and vacancies, is that the basis set is independent of the atom positions and therefore does not bias the orbitals,

and thereby the resulting charge density, in any way. At a surface, for example, the density is highly corrogated near the surface layer of atoms but approaches a nearly uniform value further from the surface. This behavior is difficult to represent accurately with atom-centered orbitals, but is very natural for a PW basis. The same reasoning has been applied to justify the PW basis as an ideal one for describing covalent bounding, in which there is charge build-up midway between neighboring atoms.

This last attribute of the plane wave basis can become, from another point of view, a detriment. Whereas the PW basis is not biased in where it represents charge, it also puts basis set freedom in regions where is may be completely useless. Again an example is provided by the application to surface studies (discussed in the next section), where a supercell geometry has been introduced in order to retain Bloch's theorem. The unit cell consists of a film slab with two surfaces, and enough open space ("vacuum") to separate the slabs from one another. The PW basis set provides the same amount of variational freedom in the vacuum as it does to represent the wavefunctions in the solid, and as a result the PW basis, though accurate, is inefficient. Another example of this inefficiency is provided by materials with atomic states which are strongly peaked near the atom, such as is the case for d states in transition metals, 2p states in the biologically important elements carbon through fluorine, and the f states in lanthanides and actinides. Very few calculations using PW basis sets have been performed in such materials, because short-wavelength PWs are needed to describe the peaks of these "localized" states but are useless elsewhere in the unit cell.

This inefficiency led Louie, Ho and Cohen [66] to construct a mixed-basis method for performing self-consistent pseudopotential calculations. In their approach they chose to augment the PW basis set with Bloch sums of Gaussian orbitals. Matrix elements of the Hamiltonian are set up in this mixed-basis; however, in all other parts of the computation a PW representation was used. That is, the Gaussians were introduced solely to reduce the size of the secular equation, and they demonstrated that for the transition metals Nb and Pd, good convergence could be obtained with 20–40 basis functions/atom. This number was comparable to other (all-electron) methods which treated a general potential (e.g. not simply a muffin-tin form).

(Since the mathematical expressions differ in detail for the various localized bases which will be discussed here, they will not be presented explicity. For the most part the expansions and matrix elements are similar to those used in local-orbital all-electron calculations, such have been presented Lafon, Chaney and Lin [67]. Most of the necessary expressions have been included in the papers which present the methods.)

The crux of the Louie–Ho–Cohen mixed basis representation then is simply to reduce the size of the secular equation while introducing the minimum possible numerical complications. One inescapable complication is that the basis is no longer orthogonal, so the overlap matrix O must be calculated and dealt with. Methods for solving the resulting generalized linear equations

$$\sum_{\alpha'} H_{\alpha\alpha'}(\boldsymbol{k}) d_{\alpha'}(\boldsymbol{k}n) = \epsilon_{\boldsymbol{k}n} \sum_{\alpha'} O_{\alpha\alpha'}(\boldsymbol{k}) d_{\alpha'}(\boldsymbol{k}n) \tag{4.79}$$

are standard. In this equation, which is the non-orthogonal generalization of eq. (4.4), the index α runs over the PWs and the localized basis functions. A complication which can be minimized is that of having to compute the three-center integrals involving localized functions on each of two

atoms overlapping a non-local pseudopotential of a third atom. Fortunately, the non-local part of the pseudopotential is confined to within or near the core region. Louie, Ho and Cohen took advantage of this fact by choosing the localized functions to be rapidly decreasing (e.g. Gaussians) beyond the valence electron peak (the peak in the d function for metal atoms). Then the only term involving the non-local potential is the single-site term, which is $\boldsymbol{k}$-independent.

Louie, Ho and Cohen applied the mixed basis method to Nb and Pd to demonstrate its utility. Subsequently it was applied to some members of the A15 superconductors by Ho, Pickett and Cohen [68–72] and to surface calculations by Louie [73] and others. These and other surface studies will be surveyed in section 6.

The efficiency gained by introducing localized orbitals is not the only reason for their use. Many properties of solids are interpreted in terms of atomic concepts, such as the relative amounts of s and p charge in a tetrahedrally-bonded semiconductor, or the importance of d character in the conduction bands of semiconductors. It is virtually impossible to address such questions solely with PW expansions of wavefunctions. It was primarily to aid interpretation that Chadi [74] introduced a localized orbital description into the empirical (rather than the more recent self-consistent) pseudo-potential method.

With the movement towards describing progressively more complex systems, the emphasis of local orbital methods has been on efficiency. Simultaneously Baraff and Schlüter [75,76] and Bernholc and collaborators [77,78] turned to localized orbitals to describe point defects (vacancies, inpurities, etc.) in semiconductors with the self-consistent pseudopotential approach which has proved so successful for bulk solids and for surfaces. The few self-consistent studies of point defects up to that time had employed the supercell approach, in which a defect is placed at the center of a large cell, and this cell is repeated periodically to recover Bloch's theorem and the well-known numerical techniques which result. The plane wave supercell approach had been applied to the silicon vacancy by Louie, Schlüter, Chelikowsky and Cohen [79], to interstitial hydrogen in germanium by Pickett, Cohen and Kittel [80], to a hydrogen-saturated silicon vacancy model of hydrogenated amorphous silicon by Pickett [81,82] and to the Al vacancy by Chakraborty, Seigel and Pickett [83]. Although some important characteristics of these defects could be extracted from the studies, it was clear that the supercell method was introducing unwanted defect–defect interactions.

The Green's function approach of Baraff and Schlüter [75] and Bernholc, Lipari and Pantelides [78] eliminated these artificial defect–defect interactions by treating an *isolated defect*. Treatment of the point defect dictated the introduction of localized orbital basis sets, both to represent the bulk crystal Green's function as well as (possibly using a different basis set) to calculate the defect electronic structure self-consistently. The success of this approach is due in large part to being able to retain in a useful way the translational symmetry of the host crystal while taking advantage of the short range of the defect potential.

Chelikowsky and Louie [84] developed a pseudopotential-based linear combination of atomic-like orbitals (LCAO) method for use in periodic systems. They used a Bloch sum of Gaussian orbitals, as well as expanding the potential in Gaussians, so the Hamiltonian and overlap matrix elements are given by analytic expressions (which however are not simple for the terms involving the non-local potential). A serious consideration with localized orbital bases is establishing the accuracy that the basis set will afford. On one hand, there is the desirability of using a minimal basis (or as near to it as possible), while on the other hand a known degree of accuracy should be

assured. Chelikowsky and Louie found that s and p orbitals (three each of s, p_x, p_y, p_z with different decay lengths) were sufficient to describe small distortions of diamond around its equilibrium crystal structure, but that the Gaussian decay lengths had to be chosen carefully. They chose these lengths to minimize the energy at a specific volume, and kept them fixed for the structural studies.

Vanderbilt and Louie [85] extended this method to the study of the diamond surface. In the extension of the method to large systems the fitting of the potential to a set of Gaussians becomes an increasingly more tedious problem. For this they implemented a Monte Carlo simulated annealing approach to the functional fitting [86], and varied the coefficients in the representation of the potential in order to minimize the energy. This procedure of varying the potential, which follows a similar approach of Bendt and Zunger [60] and is just as valid within DFT as is varying the density, is a more generally applicable approach which is not directly tied to the use of localized basis functions. The culmination of this local orbital approach is presented in detail by Chan, Vanderbilt, and Louie [86a].

An alternative version of these local-orbital-based approaches to bulk electronic structure has been developed and implemented by Jansen and Sankey [87]. They expand their pseudo-atomic-orbitals in a sum of Slater orbitals, rather than Gaussians, and use the momentum space formalism to evaluate potential matrix elements. That is, the eigenfunctions are expanded in PWs, and then the charge density and screening potential are calculated as in the plane wave formalism. They have demonstrated that this method produces accurate band structures and structural energies for zincblende structure semiconductors, as well as giving good defect energies using a supercell approach. The use of intermediate PW expansions may provide the ultimate limitation in applying this approach to materials with strongly peaked wavefunctions, such as carbon with 2p valence orbitals. The use of localized orbitals does allow the application of the method far beyond that of pure PW methods, however, perhaps to systems containing a factor of ten more atoms.

Feibelman [88] has implemented a more ambitious application of localized orbitals to the study of isolated adsorbate species on surfaces. This application addresses the complexities of both the surface problem and the isolated defect problem. Feibelman's approach was to use the Green's function formalism which (see above) had been used for isolated defects in an otherwise periodic three-dimensional solid. His reference system with Green's function G_0 consists of a semi-infinite solid with a self-consistent clean surface, which coincides with the full problem (absorbate on the surface) far away from the adsorbate, plus the adsorbate atom which is in place but is *decoupled from the surface*. The Green's function formalism generates the interactions between these entities, which finally can be iterated to self-consistency using the Kohn–Sham equations.

In a basis of localized orbitals the Green's function formalism becomes a set of matrix equations whose indices correspond to basis functions. Feibelman noted that it is necessary for G_0 and the full Green's function G to have indices i, j which are in one-to-one correspondence, and that a spatially-compact problem (in which only a finite region of real space is treated explicitly) requires that far from the adsorbate i and j refer to identical orbitals. In the region of the perturbation, however, a given index can refer to different basis functions in G_0 and G, and this freedom can be used to give more accurate wavefunctions, density, energy and forces.

Feibelman found the choice of a good localized orbital basis set for this system to be a

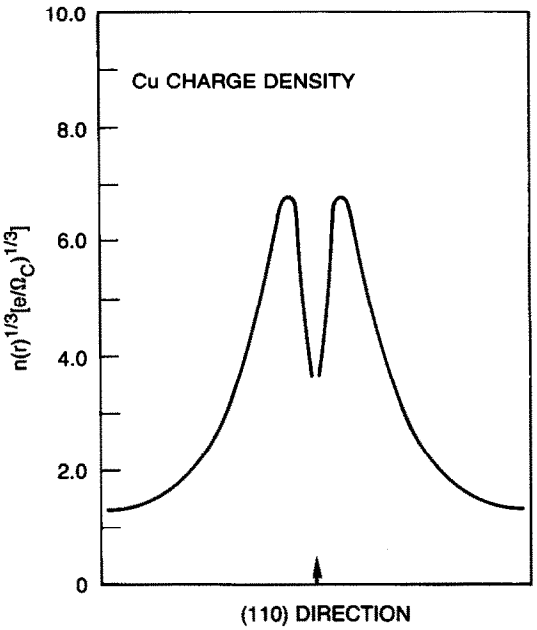

Fig. 10. Valence charge density of copper metal along a (110) direction connecting nearest neighbors. The d electron peaks are rather close to the nucleus and only the tail overlaps with the neighbor.

non-trivial problem. The initial choice for Al was to use atom-centered s- and p-symmetry orbitals which were expanded in two or three Gaussians and were required to fit the Al atom pseudowavefunctions out to at least 3.5 a.u. Since this basis set is not flexible enough to describe the surface charge dipole layer accurately, he supplemented these atomic-centered orbitals with diffuse "floating" (i.e. non-atom-centered) orbitals of p symmetry. The floating orbitals were placed 3 a.u. above the surface layer of Al atoms and resulted in a much improved representation of the work function and surface band structure. Additional atom-centered functions were found to be much less efficient than the floating orbitals. Feibelman reported results for a single Al adatom on an Al(001) surface [88] and for a Si adatom on the same surface [89].

Kang, Tatar, Mele and Soven (KTMS) [90] have used another variation of the mixed basis approach. They add to a modest-sized plane wave set another set of auxiliary functions to represent the more sharply-peaked part of the wave function near the atom, such as occurs in transition metal systems. Their approach is to extend the idea of Louie, Ho and Cohen [66], which was to keep the localized functions so localized that three-center integrals are negligible, one step further. KTMS used localized functions with negligible overlap even between nearest neighbors, so the matrix elements involving this set are $\boldsymbol{k}$-independent and can be calculated by accurate one-dimensional quadrature. They also introduce a "partial fast Fourier transform" procedure which is useful in their method.

The method was applied to copper, for which they investigated the relative stability of the fcc and bcc structures under pressure. In fig. 10 the valence charge density of fcc Cu is shown along the (110) nearest neighbor direction. From this figure it is evident that there should be a useful localized function which represents the strongly peaked density arising from the d states and yet

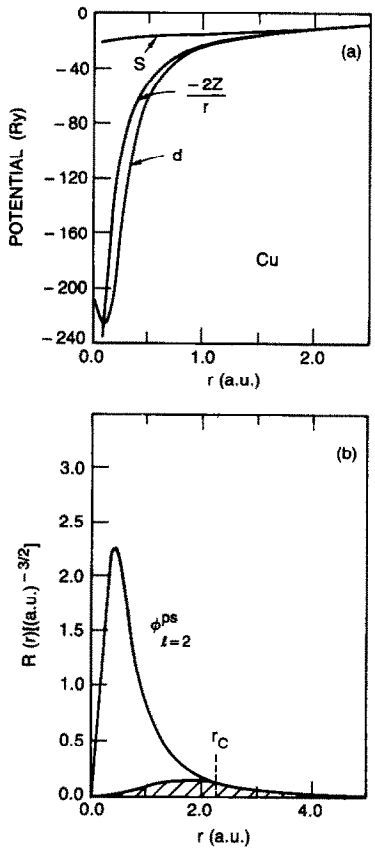

Fig. 11. (a) Pseudopotential (s and d components only) of Cu, from Bachelet, Hamann and Schlüter [25]. (b) The d pseudo-orbital of Cu, illustrating the separation by Kang et al. [90] into local orbital (unshaded) and smooth background (shaded).

will not overlap another such function on the neighboring atom. The method of KTMS for determining the local basis function is illustrated in fig. 11: it is essentially the atomic radial d function, except that a tail which is easily expanded in a few plane waves is subtracted off.

Although there is a clear physical basis for the local functions chosen by KTMS, most LCAO methods (as opposed to mixed basis methods) choose their local orbitals on the basis of computational simplicity rather than on physical grounds. As a result these choices are less likely to produce accurate results with small bases, and this criticism is usually countered by adding addition basis functions, under the guise of "conduction band states" or simply "variational freedom." Itoh and Nakao [91] addressed this question by using numerical local basis functions, a procedure which has been used in several all-electron LCAO codes. The local functions are solutions of an atomic-like problem derived from the cluster or solid being studies, with the use of a Watson sphere [92] if necessary. They have applied the method in the minimal basis et mode for sulfur clusters [91] and for structural defects in sulfur [93], obtaining good agreement with experimentally determined quantities.

5. Total energy and related quantities

5.1. Expressions for the total energy

In DFT the total energy $E_{\rm tot}$ is given by [4–8]

$$E_{\rm tot} = E_{\rm I-I} + T_0[n] + E_{\rm e-i} + E_{\rm Coul} + E_{\rm xc}. \tag{5.1}$$

The ion–ion contribution is given by

$$E_{\rm I-I}(\{R_j\}) = \frac{1}{2}\sum_{\substack{jj'\\ss'}}{}' \frac{Z_s Z_{s'} e^2}{|\boldsymbol{R}_j + \boldsymbol{\tau}_s - \boldsymbol{R}_{j'} - \boldsymbol{\tau}_{s'}|}, \tag{5.2}$$

where the prime indicates that the $(j, s) = (j', s')$ term is to be omitted from the sum. The term T_0 is *defined* [94] as the kinetic energy of a *non-interacting system with the same density*, which is given by

$$T_0[n] = \sum_i \int \psi_i^*(\boldsymbol{r})\left(-\frac{1}{2m}\nabla^2\right)\psi_i(\boldsymbol{r})\, {\rm d}^3r f(\epsilon_i), \tag{5.3}$$

where the density is given by

$$n(\boldsymbol{r}) = \sum_i |\psi_i(\boldsymbol{r})|^2 f(\epsilon_i). \tag{5.4}$$

The interaction of the electrons with the "external" potential, in this case the ion cores, can be written for local potentials as

$$E_{\rm e-i}^{\rm (loc)} = \int V_{\rm ion}(\boldsymbol{r}) n(\boldsymbol{r})\, {\rm d}^3r. \tag{5.5}$$

For the non-local pseudopotentials of more interest here, this is generalized to

$$E_{\rm e-i} = \sum_i \left[\int\int \psi_i^*(\boldsymbol{r}) V_{\rm ion}(\boldsymbol{r}, \boldsymbol{r}')\psi_i(\boldsymbol{r}')\, {\rm d}^3r\, {\rm d}^3r'\right] f(\epsilon_i). \tag{5.6}$$

The classical Coulomb repulsion energy is

$$E_{\rm Coul} = \tfrac{1}{2}\int\int n(\boldsymbol{r})\frac{e^2}{|\boldsymbol{r}-\boldsymbol{r}'|} n(\boldsymbol{r}')\, {\rm d}^3r\, {\rm d}^3r', \tag{5.7}$$

and the exchange-correlation (xc) energy is written as [4–8]

$$E_{\rm xc} = \int \epsilon_{\rm xc}(\boldsymbol{r};\, n) n(\boldsymbol{r})\, {\rm d}^3r. \tag{5.8}$$

Typically a local density form $\epsilon_{xc}(n(\boldsymbol{r}))$ is taken for the xc energy density ϵ_{xc}. It has been shown by Gunnarsson and Lundqvist [95] that eq. (5.8) is a rigorous form for E_{xc}, but the exact functional dependence of ϵ_{xc} on the density is an unsolved problem. Various widely used forms of ϵ_{xc}, and the resulting forms for E_{xc} and V_{xc}, are given in appendix B.

A fundamental problem to be addressed when computing the total energy of a periodic system is the cancellation occurring between the formally divergent terms. Both E_{I-I} and E_{Coul} describe the interaction between an "array" of like charges, while E_{e-i} contains the interaction between unlike charges. The first two are positive and divergent, while the third is negative and divergent. This huge cancellation is handled formally by adding and subtracting terms which leave finite terms to be calculated while the divergent terms can be accounted for by algebraic methods. Specifically, this is done by writing the density in E_{e-i} and E_{Coul} as

$$n(\boldsymbol{r}) = n_0 + \delta n(\boldsymbol{r}), \tag{5.9}$$

where n_0 is the average density, so δn describes no net charge, and by replacing the ionic pseudopotential V_{ion} of an ion of charge Z at the origin by

$$V(\boldsymbol{r}, \boldsymbol{r}') = -\frac{Ze}{r}\delta(\boldsymbol{r}-\boldsymbol{r}') + \delta V(\boldsymbol{r}, \boldsymbol{r}'), \tag{5.10}$$

where of course

$$\delta V(\boldsymbol{r}, \boldsymbol{r}') \equiv V(\boldsymbol{r}, \boldsymbol{r}') + \frac{Ze}{r}\delta(\boldsymbol{r}-\boldsymbol{r}'). \tag{5.11}$$

The important result is that δn and δV lead to *short range interaction* terms. Methods for dealing with the formally divergent terms have been discussed by several workers [96–98].

The resulting expression for the total energy *in reciprocal space* becomes [98,99]

$$E_{tot} = E_{kin} + E'_{e-i} + E'_{Coul} + E_{xc} + E_{Ewald} + E_{rep}, \tag{5.12}$$

where the "kinetic energy" E_{kin} is (treating each spin as a separate band)

$$E_{kin} = \sum_{\boldsymbol{k}n}\sum_{\boldsymbol{G}} |c_{\boldsymbol{G}}(\boldsymbol{k}n)|^2 \frac{|\boldsymbol{k}+\boldsymbol{G}|^2}{2m} f(\epsilon_{\boldsymbol{k}n}), \tag{5.13}$$

and E'_{e-i} is given by

$$E'_{e-i} = \sum_{\boldsymbol{k}n}\sum_{\boldsymbol{G},\boldsymbol{G}'}{}' c^*_{\boldsymbol{G}}(\boldsymbol{k}n) V_{ion}(\boldsymbol{k}+\boldsymbol{G}, \boldsymbol{k}+\boldsymbol{G}') c_{\boldsymbol{G}'}(\boldsymbol{k}n) f(\epsilon_{\boldsymbol{k}n}). \tag{5.14}$$

with V_{ion} given in eqs. (4.2) and 4.6). The prime on the sum indicates that the $\boldsymbol{G} = \boldsymbol{G}' = 0$ terms is to be omitted from the sum for the local part of V_{ion}.

The Coulomb term

$$E'_{Coul}[n] = \frac{\Omega_c}{2}\sum_{\boldsymbol{G}\neq 0}\frac{4\pi e^2}{G^2}|n(\boldsymbol{G})|^2 \tag{5.15}$$

and the xc energy

$$E_{\rm xc}[n] = \Omega_{\rm c} \sum_{\boldsymbol{G}} \epsilon_{\rm xc}(\boldsymbol{G};\, n) n^*(\boldsymbol{G}) \tag{5.16}$$

are straightforward functionals of n, assuming the xc energy density is known (appendix B).

The Ewald energy is given by

$$E_{\rm Ewald} = \frac{e^2}{2} \sum_{s,s'} Z_s \gamma_{ss'} Z_{s'}, \tag{5.17}$$

where one expression for the Ewald constants $\gamma_{s,s'}$ is [100]

$$\gamma_{ss'} = \frac{4\pi}{\Omega_{\rm c}} \sum_{\boldsymbol{G}\neq 0} \frac{1}{G^2} \cos[\boldsymbol{G}\cdot(\boldsymbol{\tau}_s - \boldsymbol{\tau}_{s'})] \exp\left(-\frac{G^2}{4\eta^2}\right) - \frac{\pi}{\eta^2 \Omega_{\rm c}}$$

$$+ \sum_j \frac{\operatorname{erfc}(\eta x_j)}{x_j}\bigg|_{x_j = |\boldsymbol{R}_j + \boldsymbol{\tau}_s - \boldsymbol{\tau}_{s'}|} - \frac{2\eta}{\sqrt{\pi}} \delta_{ss'}. \tag{5.18}$$

Here η is a "convergence parameter" which formally is arbitrary but in practice is adjusted to make both the real space and reciprocal space sums converge rapidly. The last term $E_{\rm rep}$ is the "repulsive" part of the pseudopotential, compared to a Coulomb potential, given by

$$E_{\rm rep} = \left(\sum_s \alpha_s\right)\left(\frac{1}{\Omega_{\rm c}} \sum_s Z_s\right) = \bar{\alpha}\bar{n}, \tag{5.19}$$

where

$$\alpha_s = \frac{1}{\Omega_{\rm c}} \int {\rm d}^3 r \left(V_{\rm ion}^{(s)}(r) + \frac{Z_s e^2}{r}\right). \tag{5.20}$$

$E_{\rm rep}$ is seen to be the product of the average "repulsive" potential $\bar{\alpha}$ and the average valence density $\bar{n}$. Only the $l=0$ (angle independent) term in $V_s^{\rm ion}$ contributes to α_s. The formally divergent terms in $E_{\rm i-i}$, $E_{\rm Coul}$ and $E_{\rm e-i}$ combine to give the finite terms $E_{\rm Ewald}$ and $E_{\rm rep}$. These expressions are formally valid, but very slowly convergent, for an all-electron calculation for which Z_s becomes the nuclear charge and α_s vanishes identically.

The energy expression can be written in another widely used form which makes use of the secular equation. The sum over occupied states of the eigenvalues gives the "band structure" energy

$$E_{\rm bs} \equiv \sum_{\boldsymbol{kn}} \epsilon_{\boldsymbol{kn}} f(\epsilon_{\boldsymbol{kn}}) = \sum_{\boldsymbol{kn}} \langle \boldsymbol{kn} | \left(-\frac{\nabla^2}{2m} + V_{\rm ion}\right) | \boldsymbol{kn} \rangle f(\epsilon_{\boldsymbol{kn}}) + \int {\rm d}^3 r \left[V_{\rm h}(\boldsymbol{r}) + V_{\rm xc}(\boldsymbol{r})\right] n(\boldsymbol{r}). \tag{5.21}$$

The first term on the right hand side gives exactly the terms T_0 and E_{e-i} in the energy functional. The term involving V_h is exactly twice the term E'_{Coul} in eq. (5.15) assuming the usual convention $V_h(\boldsymbol{G}=0)=0$ is used to fix the arbitrary constant, and the xc term can easily be dealt with. Then the energy expression can be written

$$E_{tot}[n] = E_{bs} - E'_{Coul}[n] + \Delta E_{xc}[n] + E_{Ewald} + E_{rep}, \tag{5.22}$$

where the xc correction is

$$\Delta E_{xc}[n] = E_{xc}[n] - \int d^3r\, V_{xc}(\boldsymbol{r};\, n)n(\boldsymbol{r})$$

$$= \int d^3r \left[\epsilon_{xc}(\boldsymbol{r};\, n) - V_{xc}(\boldsymbol{r};\, n)\right] n(\boldsymbol{r}). \tag{5.23}$$

5.2. *Variational nature of the total energy*

A fundamental result of DFT is that the functional $E_{tot}[n]$ is minimized by the true ground state density $n(\boldsymbol{r})$. It follows that if an approximate density $\tilde{n}(\boldsymbol{r})$ is used to evaluate E_{tot}, the error is second order in the error in density

$$E_{tot}[\tilde{n}] - E_{tot}[n] \propto \| \tilde{n}(\boldsymbol{r}) - n(\boldsymbol{r}) \|^2. \tag{5.24}$$

Inaccuracies in the density can arise from an incomplete representation, from approximation in solving the DFT equations (Schrödinger's equation, Poisson's equation, zone integration, etc.) or from lack of self-consistency. The variational nature of the energy functional should allow one to carry out well converged studies of energetics without an unreasonable amount of computational expense.

There is a less obvious reason why a "natural" evaluation of E_{tot}, particularly using expression (5.22), results in a non-quadratic convergence of the calculated energy. The difficulty, perhaps encountered first by Zunger and Cohen [101] and clarified by Chelikowsky and Louie [83], arises because a variational expression is not actually being calculated. If after each iteration, the resulting *wavefunctions* and *density* are used to evaluate expression (5.12), there is no problem. If, on the other hand, the *eigenvalues* and *density* are used to evaluate expression (5.22), quadratic convergence of the energy will not result.

The underlying cause of this situation is that the eigenvalue sum (5.21) is not a functional of either the input density or the output density alone. Rather, it depends on both input and output densities through the expression

$$E_{bs} = T_0[n^{out}] + E_{ext}[n^{out}] + \int d^3r \left[V_h(\boldsymbol{r};\, n^{in}) + V_{xc}(\boldsymbol{r};\, n^{in})\right] n^{out}(\boldsymbol{r})$$

$$= E_{bs}[n^{in},\, n^{out}], \tag{5.25}$$

because the *wavefunctions define the output density*.

The difficulty is easily rectified by explicitly subtracting out the offending terms in eq. (5.25) which depend on both n^{in} and n^{out}, and then adding in the corresponding terms expressed in terms of n^{in} alone. Expression (5.22) can be rendered variational in the output density by the replacement $E_{\text{bs}} \rightarrow E'_{\text{bs}}$, with

$$E'_{\text{bs}}[n^{\text{out}}] = E_{\text{bs}}[n^{\text{in}}, n^{\text{out}}] - \int \mathrm{d}^3 r \big[V_h(\boldsymbol{r}, n^{\text{in}}) + V_{\text{xc}}(\boldsymbol{r}, n^{\text{in}}) - V_h(\boldsymbol{r}, n^{\text{out}}) - V_{\text{xc}}(\boldsymbol{r}, n^{\text{out}}) \big] n^{\text{out}}(\boldsymbol{r}). \tag{5.26}$$

The term under the integral cancels the dependence on n^{in} which is contained in $E_{\text{bs}}[n^{\text{in}}, n^{\text{out}}]$, with the result that E'_{bs} is a functional of the output density alone. Obviously the correction vanishes at self-consistency, where $n^{\text{out}} = n^{\text{in}}$. The author has verified that, in simple calculations involving zincblende semiconductors, the variational expression converges to the level of 0.1 mRy/atom while the non-variational expression (5.23) is still changing by *several tens* of mRy at each iteration. *The variational form is essential for efficient total energy studies.*

Since the energy functional is variational, one can expect that the convergence of the energy during iteration to self-consistency will be more rapid than the convergence of other quantities, such as the density itself, the eigenvalues, or individual contributions to the energy. However, in actual numerical calculations a number of complications may arise that may destroy the quadratic convergence of the energy. For example, one may ask whether, for a given set of calculational parameters (number of PWs in the expansion of the density, energy cutoffs, specified $\boldsymbol{k}$-point sampling, etc.) which limit the accuracy of the density which can be obtained, variational behavior in this restricted problem can result. The calculational fact that variational behavior does survive can be understood by considering that a given numerical calculation restricts one to a specific "subspace" of the full variational freedom implied by the formal theory. If the numerical variational freedom is "reasonable," that is to say, if the self-consistent density in the subspace is "near" the true density, the situation will be as pictured schematically in fig. 12. Within this restricted space the density converges to an approximate value n_{app}, and the energy converges to $E_{\text{app}} = E_{\text{tot}}[n_{\text{app}}]$. If n_{app} lies within the quadratic minimum centered at n, the convergence of the energy to E_{app}, will be quadratic, or nearly so. Then, since n_{app} is near the true density n, E_{app} provides a variational estimate of the energy. The result is that certain types of useful total energy studies can be performed without using extensive basis sets, large cutoffs or sophisticated sampling techniques.

Weinert, Watson and Davenport [102] (WWD) also noted the variational form (5.26), as well as another form which involves only the eigenvalue sum and *input* quantities. They *defined* an expression

$$\tilde{T}_0[n] \equiv E_{\text{bs}}[n^{\text{in}}, n^{\text{out}}] - \int \mathrm{d}^3 r \, n(\boldsymbol{r}) V_{\text{eff}}(\boldsymbol{r}), \tag{5.27}$$

where

$$V_{\text{eff}} \equiv V_{\text{ion}} + V_{\text{h}} + V_{\text{xc}} \tag{5.28}$$

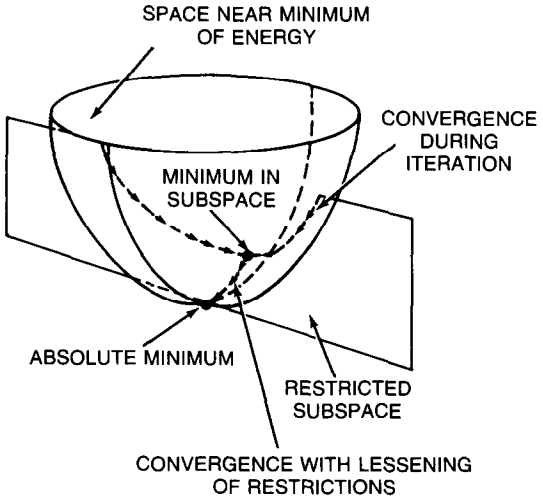

Fig. 12. Schematic representation of an energy minimum, shown as a paraboloid. The plane represents a restricted subspace for the density, with interation to self-consistency taking place along the intersection of these two surfaces (see text).

is the effective one-electron potential. $\tilde{T}_0$ is the kinetic energy of a non-interacting density *only* if $n(\boldsymbol{r})$ is the self-consistent density, i.e. $n = n^{\text{out}} = n^{\text{in}}$. WWD suggested that the energy expression

$$\tilde{E}[n] \equiv \tilde{T}_0[n] + E_{\text{e-i}}[n] + E_{\text{Coul}}[n] + E_{\text{xc}}[n] + E_{\text{i-i}} \tag{5.29}$$

provides a variational estimate of the true total energy. Numerical results (by Weinert and collaborators as well as by the author and collaborators) indeed have established that the convergence of (5.29) during iteration is much more rapid than convergence of (5.22), and at least close to the quadratic convergence

$$\tilde{E}[n + \delta n] = E_{\text{tot}}[n] + \mathcal{O}(\delta n)^2 \tag{5.30}$$

stated by WWD.

The derivation of Harris [103], which leads to the same expression, helps clarify the "variational" nature of (5.29). Harris noted that the difference between $\tilde{E}[n + \delta n]$ and $E_{\text{tot}}[n]$ is indeed of second order but can be of *either sign*. Thus $\tilde{E}$ is not truly variational, since it provides no rigorous bound, but it is rapidly convergent. Moreover, his derivation established that the correction is of second order only if the eigenvalue sum includes only those one-electron states which will be occupied at self-consistency; that is, E_{bs} in (5.27) must be given by

$$\tilde{E}_{\text{bs}} \equiv \sum_{kn} \tilde{\epsilon}_{kn} f(\epsilon_{kn}), \tag{5.31}$$

where $\tilde{\epsilon}_{kn}$ is the eigenvalue corresponding to the (approximate) input density but ϵ_{kn} *is the self-consistent* eigenvalue. Therefore in metallic systems where occupancy changes with iteration

there will be a first order term in (5.30), while in non-metallic systems (semiconductors, clusters) the correction will become quadratic as soon as the correct occupation is achieved.

6. Applications of pseudopotential methods

The previous sections have provided a presentation and discussion of the formalism and many of the algorithms for carrying out pseudopotential-based studies of the electronic structure and total energy of solids. In this final section the various applications which pseudopotential methods have found will be reviewed. Since the applications are many, no attempt will be made to provide a comprehensive list of references. Rather the emphasis will be put on touching on each of the diverse areas of application which have appeared during the past decade. These areas include:

(1) structural studies of solids,

(2) surfaces and interfaces,

(3) point defects,

(4) frozen phonons, forces, and stresses,

(5) dielectric theory and linear response,

(6) dynamical self-energies,

(7) various large unit cell applications,

(8) quantum molecular dynamics and simulated annealing.

These topics will be reviewed separately in this section.

6.1. Structural studies of solids

The pseudopotential method was developed originally to describe the properties of weak-scattering simple metals, but with the development of the sophisticated self-consistent density functional methods there has been very little interest in the alkali and alkaline earth metals. Chelikowsky [104] suggested using a "Thomas-Fermi-pseudopotential" approach in which the kinetic energy functional $T_0[n]$ of eq. (5.3) is approximated by a simple "local" form such as the Thomas–Fermi or von Wiezsacker-like form which depends directly on the density and its gradient. With this approximation the density functional expression for the energy becomes a relatively simple minimization problem $\delta E[n]/\delta n(\boldsymbol{r}) = 0$, for which the density can be varied directly until this equation is satisfied for each $\boldsymbol{r}$. For systems with strong variations of density such approximations are known to be inadequate, but for simple metals there remains some promise. Chelikowsky applied this approach to the alkali metals, Mg, Ca and Al, finding that structural energies are described realistically if a hard-core pseudopotential is used. Non-local effects, from both the kinetic and pseudopotential contributions, are likely to be important in other systems.

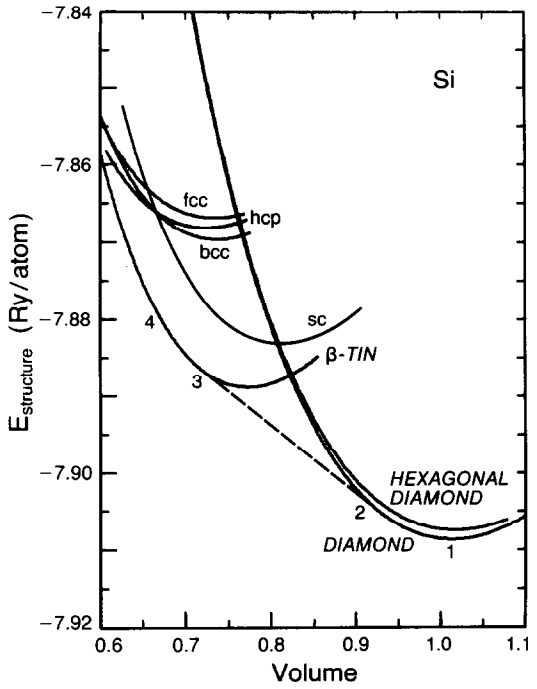

Fig. 13. Equation of state for silicon, calculated by Yin and Cohen [99] for seven crystal structures. The slope of the dashed line gives the critical pressure for transformation from the diamond to the β-tin structure.

The principal area of application has been to elemental semiconductors, especially the ones commonly occurring in the diamond structure. There have been several applications to C (diamond), Si, Ge and α-Sn [83,99,100,105–114]. The predicted lattice constant, cohesive energy, and bulk modulus are of course the first properties to assess, but one of the most impressive results is the theoretical prediction of the critical pressures of pressure-driven structural transformations. An example is shown in fig. 13, where the energy versus volume calculations of Yin and Cohen [99] for both Si and Ge are pictured. Seven crystal structures were considered, and the calculations correctly predicted the diamond-to-β-tin transformation at critical pressures within 20% or better of experiment. Biswas and Hamann [115] also performed a wide variety of total energy studies for several structures, not for their intrinsic interest but to form a database of first-principles results which could be used to determine two-body and three-body potentials. The expectation is to use these parameterized potentials for more complex studies where first-principles studies are not feasible.

Other total energy calculations for semiconductors include those for AlP, AlAs, GaP and GaAs by Froyen and Cohen [116], for GaAs and AlAs by Ihm and Joannopoulos [117] and for GaAs by Kunc and Martin [118]. Needs, Martin and Nielsen [119] applied the method to the semimetal As with good results, while Vanderbilt and Joannopoulos [120] performed studies on the insulator Se. More recent work has seen as emphasis on II–VI semiconducting compounds and alloys, with calculations having been carried out on HgTe, CdTe and $HgCdTe_2$ (i.e., $Hg_{0.5}Cd_{0.5}Te$) [121–124].

These methods have been extended to several transition elements. The self-consistent pseudopotential approach was first applied to calculate the band structure of Nb [125,126], and since has been extended to a number of transition metals, including Mo [126–128], W [127–130], Cu [90], Pd [66], Zr [126] and Ru [131]. The results establish that the structural properties are reproduced just as accurately as had been done for semiconductors.

Beryllium is a divalent metal which is not at all like the heavier alkaline earth metals. Since it has a small 1s core and no core p states, its pseudopotential is strong and Be becomes a low density of states metal, not far from being semiconducting. Chou, Lam and Cohen [132,133] have investigated the electrical and structural properties of Be (lattice constants of the hexagonal structure, cohesive energy, bulk modulus and Poisson's ratio), obtaining good agreement with experiment.

The logical extension of these structural studies is to compounds, for which a number have been reported. A wide variety of zincblende structure and related semiconductor studies (in addition to those listed above) exist [112,124,134], and the method was extended to the ionic insulator NaCl by Andreoni, Maschke and Schlüter [138]. A more complete study of structural stability and transformations in ionic crystals has been carried out by Chelikowsky and coworkers [139,140]. Chan and Louie [141] have studied hydrides of Pd using pseudopotentials. Bhattacharyya, Bylander and Kleinman [142] have provided one of the few fully relativistic studies in their work on $MoSi_2$ and WSi_2.

The implication of this body of work is that the pseudopotential method, applied properly, is valid throughout the periodic table. The papers referenced in this section are certainly not exhaustive; rather, the intention has been to point out work which has in some sense been ground-breaking and provided evidence of the validity of the pseudopotential method in various regions of the periodic table. A large number of total energy studies have been carried out for other than bulk crystals, such as for surfaces, interfaces and defects, and these topics will be discussed separately in following subsections. A recent review of total energy calculations, not limited completely to pseudopotential work, has been given by Ihm [143]. Discussion of several intricacies of total energy studies which are extremely useful to the practitioner have been given by Denteneer and Haeringen [112] and in more detail in the thesis of Denteneer [144].

6.2. Surfaces and interfaces

One of the first uses of the self-consistent pseudopotential method was in the area of electronic structure of surfaces and interfaces. It is important to locate and characterize surface states, because they strongly influence adsorption, catalysis and surface reconstruction as well as having a definite effect on the work function. From the earliest studies it was clear that the position in energy and the character of such states is sensitive to the potential in the surface region, which demanded a self-consistent treatment. Application of the method to intimate interfaces to determine interface bonding and band line-ups also occurred at an early stage.

The application to semiconductor surfaces has been extensive, with several reviews having been published [145–152]. The earliest calculations of Appelbaum, Hamann and Baraff [153–161] utilized a semi-infinite surface system, in which a Green's function method is used to match wavefunctions in the surface region onto bulk solutions a few (2–5) layers under the surface. This method is ideal for extracting detailed spectroscopic information, since it treats a true surface

(contrast to the slab methods discussed below). The method however requires difficult numerical procedures, so it has not been practiced widely.

The alternative calculational approach which has become the method of choice is the slab supercell method. This approach consists of artificially introducing periodicity *perpendicular* to the surface by (1) treating a slab instead of a true isolated surface, and (2) repeating the slabs periodically, in the perpendicular direction. This procedure results in a system with three dimensional periodicity, so the methods for bulk solids become available. To reproduce the electronic structure of real surfaces reasonably, the slabs must be thick enough that the two surfaces do not interact appreciably, and the vacuum region must be wide enough that interaction between slabs through the vacuum is negligible. Nearly all subsequent pseudopotential calculations of surfaces have been done using this approach, and it has naturally been extended to study interfaces using the same "superlattice" geometry.

Chelikowsky and Cohen [162] applied the method to the ideal geometry of the (110) surfaces of GaAs and ZnSe, identifying dangling bond surface states related to each of the cation and anion surface atoms. The method was simultaneously extended to metal/semiconductor Schottky barrier interfaces and semiconductor/semiconductor heterojunction interfaces. Initial application to Schottky barriers treated the metal in the jellium approximation [163,164], but this was later extended to include the pseudopotential of the metal atom (see below).

Since the earliest surface work [162] the self-consistent pseudopotential method has been applied widely to group IV, III-V and II-VI semiconductors (see, for example, refs. [165–167] for Si, refs. [168–170] for Ge, refs. [152,169,171] for C, ref. [172] for Sn, refs. [162,173–179] for GaAs, refs. [180,181] for GaP, refs. [182,183] for InP and refs. [162,184] for ZnSe). The emphasis has been on identifying surface states and comparing with spectroscopic data, and studying relaxations and reconstructions of surface layers to gain an understanding of bonding characteristics at surfaces. A number of surface overlayers have also been studied, including H on Si [185], Al on Si [186], Al on GaAs [187], O on GaP [188], Sb on GaP [189] and H on Ga-based III-V compounds [190].

The method has also been extended to transition metal surfaces such as Nb (001), [191,192], Pd (111) [193], Mo (001)) [194,195], W (001) [196], Ru (0001) [152,197] and Au (110) [198]. With plane wave basis sets alone, such calculations are extremely large, but the introduction of local orbital basis sets have even allowed the study of overlayers on transition metal surfaces. Examples are H on Pd (111) [199], H on Mo (001) [200,201] and subsurface H on Pd (111) [202].

Study of semiconductor interfaces and heterojunctions began with the non-polar (110) interface of the Ge–GaAs and AlAs–GaAs systems [203,204],since a polar interface of a system like Ge–GaAs can be shown [205,206] to be metallic in the absence of a reconstruction which enlarges the surface unit cell. In general, interface states are found [203,204] to occur in semiconductor systems, and often they are sensitive to atomic geometry. A fundamental property which is important in device engineering is the band edge discontinuity, which determines the relative positions of the bandgaps on each side of the interface, and this property has been the main impetus for the growing number of studies of semiconductor interfaces [207,208].

An example of the kind of information which can be obtained from surface and interface calculations is shown in figs. 14–16 for the Ge–GaAs (110) interface [203,204]. Figure 14 shows the calculated interface bands S_1, S_2, B_1, B_2, P_1, P_2 which must lie within mutual gaps of the (110) projected band structures of both components Ge and GaAs. Although there are many interface

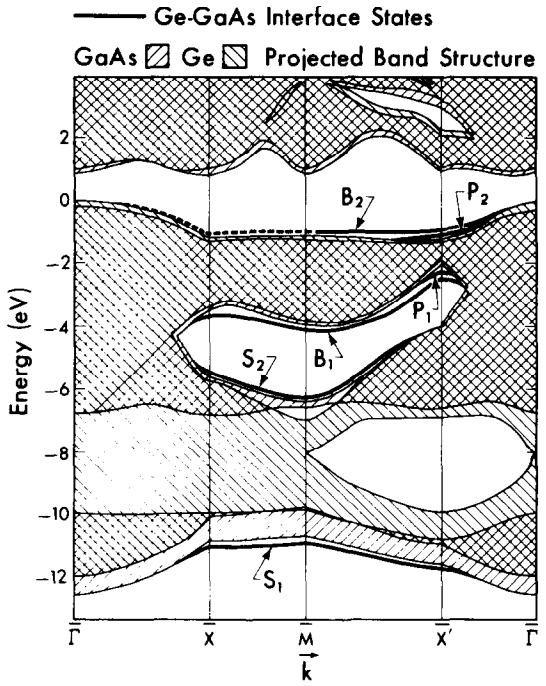

Fig. 14. Interface bands (heavy lines) of (110) Ge–GaAs relative to the projected band structures, shown by cross-hatching. For this system the interface bands remain near the edges of the gaps.

bands, in this system they are located in all cases near the edges of the gaps. The charge densities of these states, shown in figs. 15 and 16, indicate that these interface states can be understood directly in terms of "undersaturated" Ge–Ga bonds and "oversaturated" Ge–As bonds across the interface.

After the initial work and follow-ups on GaAs–AlAs [204,209–213] and Ge–GaAs [206, 214–217], the method was also applied to GaAs–ZnSe [218], InAs–GaSb [219–221], Si–Ge [222,223], Si–GaP [224], InAs–GaAs multilayers [225] and the (111)–(0001) interfaces of ZnS and ZnSe [226]. As for the case of surface calculations mentioned above, this list is intended as representative and is not meant to be complete.

6.3. Point defects

The application of the self-consistent pseudopotential formalism to studies of defects opened up a new era in the study of deep levels in semiconductors. Baraff and Schlüter [75] and Bernholc, Lipari and Pantelides [78] demonstrated how the Green's function scattering theory equations could be put into matrix form and solved self-consistently. The method was applied initially to the Si vacancy, where it was established [227,228] that the unrelaxed vacancy gives rise to a bound state of T_2 symmetry near mid-gap. The method then was applied to deep substitutional sp-bonded impurities (S, Zn, and H) in silicon [229] and to ideal vacancies in GaAs [230].

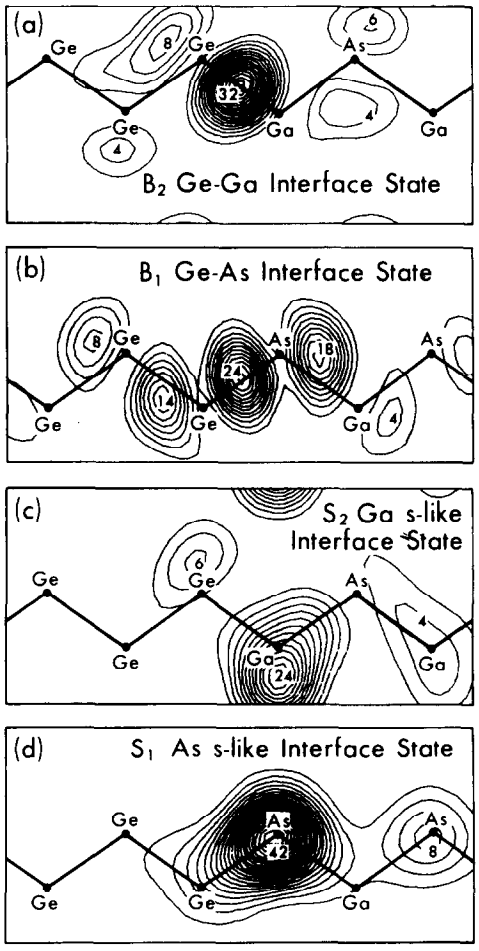

Fig. 15. Contour plots, in a (101) plane parallel to the interface, of the interface states B_2 (a), B_1 (b), S_2 (c), S_1 (d), shown in fig. 14. S_1 and S_2 are s-like states on As and Ga, respectively, whereas B_1 and B_2 are bond-centered Ge–As and Ge–Ga states, respectively.

Lindefelt and Zunger [231] introduced a different approach they called the "quasi-band method". They noted that for impurities which are chemically mismatched to the host crystal wavefunctions, such as for a 3d transition metal in Si, an expansion of the defect wavefunctions may require an unreasonably large (10^2–10^4) number of crystalline eigenstates. The quasi-band wavefunctions they construct include aspects of both host and defect orbitals and thereby reduce the numerical expense.

Kirton and Banks [232] devised a method which exploits symmetry to maximum advantage. Their method explicitly expands the defect wavefunctions in the complete set of host crystal eigenstates, so it is not very appropriate for chemically mismatched impurities. However, the full use of symmetry makes it an efficient method for dealing with vacancies and chemically similar impurities, and the initial application of the method treated vacancies in Si, GaP and ZnSe [233].

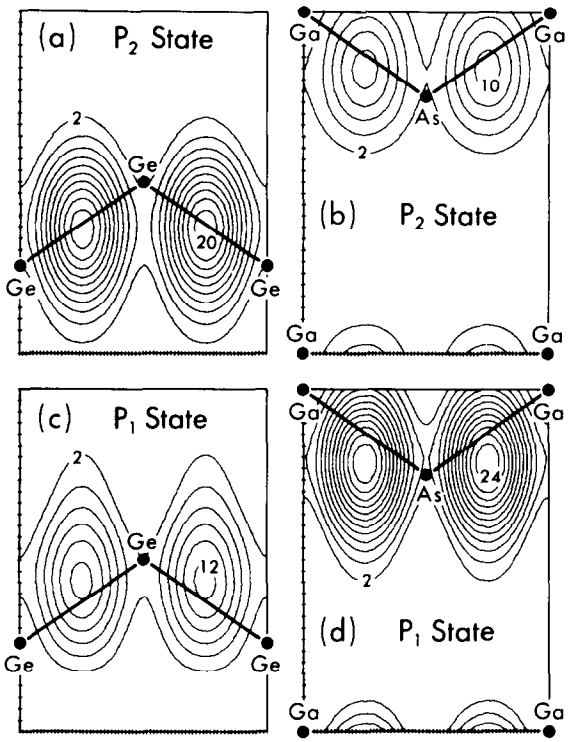

Fig. 16. Contour plots, in a (110) plane parallel to the interface, of the "parallel" interface states P_1 and P_2 shown in fig. 14. These states lie on bonds parallel to the interface rather than on bands perpendicular to the interface as for the B_1 and B_2 states in fig. 15.

Several other methods has been devised for studying defects within the pseudopotential formalism. Itoh and Nakao [234] have used a numerical LCAO basis set approach [91] to investigate structural defects in sulfur. Falck, Stoll and Schwan [235] have performed pseudopotential cluster calculations on hydrogen centers in the florides and chlorides of Li, Na, K and Rb. Ramsey and Smith [236] recast the formalism of Kanamori, Terakura and Yamada [237] and Gautier [238] to make it suitable for studies of transition metal systems. The approach was applied to Pd impurities in Cu, Ag and Au [236].

The original objective of these defect studies was to describe correctly the electronic structure of some idealized defects, such as the unrelaxed vacancy. The field now has progressed well beyond this point. After establishing the procedures necessary to calculate the formation energy of both neutral and charged defects [239,240], Baraff and Schlüter carried out an extensive study of native defects in GaAs. Calculations for the self-interstitials, vacancies, and anti-site defects allowed them to determine which of these defects should be abundant in As-rich or Ga-rich n-type or p-type material.

The LCAO method of Jansen and Sankey [87] has been applied by Jansen, Wolde-Kidane and Sankey [241] to study the deep level positions and energetics of interstitial defects in III–V semiconductors. With a view towards understanding the site preference of interstitial defects and predicting chemical trends in impure semiconductors, these authors carried out calculations of K,

Zn, Ga, Si, As, Se and Te in the two inequivalent tetrahedral interstitial sites in GaAs, AlAs, ZnSe and ZnTe. They reproduced chemical trends in agreement with experiment and made specific predictions of which interstitial site should be favored by these impurities.

The quasi-band method of Lindefelt and Zunger [231] has been applied primarily to transition metal impurities in semiconductors, which are scientifically and technologically important systems which most other calculational methods cannot handle. These applications, their results and implications have been discussed in the very extensive review of the subject by Zunger [242].

6.4. Frozen phonons, forces and stresses

The realization that total energies could be calculated accurately led directly to the computation of selected phonon frequencies by the "frozen phonon" method. This approach relies on the large mass difference between ions and electrons, which implies that electrons remain very close to their ground state configuration (on the "Born–Oppenheimer surface") as the ions move. The total energy is computed for a few displacements along the eigenvector of the phonon, and the frequency of vibration is extracted from the curvature and atomic mass. This approach allows the identification of changes in bonding during the vibration, which is a factor in determining the frequency. This approach, which has been surveyed by Kunc [243] and Louie [151], has been applied to a variety of systems.

Application of the frozen phonon approach to Si [244–247] has shown that the soft transverse acoustic modes at X and Γ are described correctly, and the softness can be related to specific charge rearrangements caused by these phonons. This approach has been extended to other semiconductors [248] as well as to some transition metals (Zr, Nb, and Mo) [249–251] and compounds (NbH) [252].

This direction of study has expanded rapidly. Although forces on atoms can be calculated indirectly, by taking finite differences of total energies as an ion is displaced, it is much more efficient to calculate forces directly. Yin and Cohen [253] have provided the expressions for calculating the analytic derivative of the total energy, i.e. the force on the ion at $\boldsymbol{R}_i$,

$$\boldsymbol{F}_i = -\frac{\mathrm{d}E}{\mathrm{d}\boldsymbol{R}_i}.$$

It is tempting to use the Hellman [254] – Feynman [255] theorem, which states that the first-order change in the wavefunctions (and therefore the density) do not contribute to the force, and the expressions of Yin and Cohen use this feature to arrive at simple analytic expressions. The resulting force is much more sensitive to numerical details than is the total energy itself [256–258]. Its error is first order in the error in density, rather than being second order as the energy is. The error can arise from lack of self-consistency or from inexactness of the representation of the density (such as truncation in the number of plane waves). These errors are straightforward to monitor, however, and since for N atoms in the cell a single self-consistent calculation provides $3N$ Cartesian components of forces (i.e. equivalent to $3N$ total energy calculations), the calculation of forces has been applied widely [259,260].

A novel application of force and frozen phonon calculations leads to an entirely parameter-free evaluation of the strength λ of the electron-phonon interaction (EPI), which is important in

understanding transport behavior as well as the superconducting transition temperature. Lam, Dacorogna and Cohen [261] formulated the method, which uses a set of frozen phonon supercell calculations to (i) evaluate the appropriate electron–phonon matrix elements, (ii) calculate the phonon frequency, and (iii) carry out the (coarse but satisfactory) Brillouin zone sampling. In addition to the information on bonding changes which is obtained from frozen phonon studies, the method gives the electron–phonon contribution $\gamma_{Q\nu}$ to the phonon linewidth and the coupling strength $\lambda_{Q\nu}$ for each wavelength Q and mode ν. This method has been applied to Al [262,263], to the group IV materials Si, Ge and Sn in the β-Sn structure [264], and most particularly to high pressure phases of Si [265–267] and Ge [268], which become superconducting at up to 8 K.

A logical extension of the calculation of forces is the study of pressure–volume and stress–strain relationships. Yin [269] used the virial theorem to derive an expression for the pressure of a system, and applied the method to Ge to obtain the correct value of the bulk modulus and its pressure derivative from a single calculation.

The study of stresses and generalized forces requires a more ambitious program. Nielsen and Martin [270] constructed a quantum-mechanical theory of force and stress, as well as providing an extensive discussion of and comparison with previous work in the field. A number of explicit expressions for carrying out pseudopotential evaluations are given by Nielsen and Martin [271], and results for Si, Ge and GaAs were reported [271,272], including bulk moduli, elastic constants (second and third order), optical phonon frequencies and the internal strain parameter. Nielsen [273] has presented an exhaustive study of the stress-strain relationship in diamond, for very large stresses along the high symmetry ⟨100⟩, ⟨110⟩ and ⟨111⟩ directions.

Needs [274] applied the stress formalism of Nielsen and Martin to surfaces of Al. By evaluating the surface stress tensor for the (111) and (110) surfaces, he established that the stress is tensile, favoring *contraction in the* plane of the surface, and interpreted this tendency as due to the smoothing of the electronic wavefunctions at the surface. Vanderbilt [274a] calculated surface stresses for the Si(111) surface and did not find the anticipated compressive stress, and concluded that surface stress is not driving reconstructions on this surface. Gomes DaCosta, Nielsen and Kunc [275] have provided a useful discussion of practical aspects of stress, force and total energy computations, including guidance in obtaining useful results even when full convergence cannot be achieved.

6.5. Dielectric theory and linear response

In nearly all the applications which are discussed in this section, the pseudopotential approach is a convenience which takes advantage of the inertness of the core with respect to bonding in solids. The application of pseudopotentials to dielectric theory may be an instance where this approach is more nearly a necessity than a convenience. The linear response of a solid to a perturbation – electric field, phonon or so forth – is described by the dielectric matrix (DM) $\epsilon_{GG'}(\boldsymbol{q}, \omega) \equiv \epsilon(\boldsymbol{q}+\boldsymbol{G}, \boldsymbol{q}+\boldsymbol{G}'; \omega)$ where $\boldsymbol{q}$ and ω are the wavevector and frequency of the perturbation and $\boldsymbol{G}$, $\boldsymbol{G}'$ are reciprocal lattice vectors. The evaluation of the DM requires the $\boldsymbol{q}+\boldsymbol{G}$ Fourier coefficient of the one-electron wavefunctions (references given below can be consulted for further details). In the pseudopotential approach the pseudo-orbitals are relatively smooth, their Fourier coefficients fall off systematically with increasing $|\boldsymbol{g}+\boldsymbol{G}|$, and this

provides a natural "cut-off" to the size of the DM. The all-electron wavefunctions, on the other hand, contain core orthogonalization wiggles whose Fourier coefficients extend to *much* larger $|\boldsymbol{q}+\boldsymbol{G}|$. Although there have been no serious attempts to evaluate the full DM from an all-electron viewpoint, there could be difficulty in finding an appropriate cut-off. The pseudopotential avoids these questions and yet appears to be very accurate.

The great majority of calculations have centered on the static ($\omega = 0$) DM in semiconductors. Due to the energy gap, the DM matrix is not only easier to calculate in semiconductors than in metals, but it has much more interesting consequences. The first calculations were made by Walter and Cohen [276], who used the empirical pseudopotential method to evaluate the $\boldsymbol{G} = \boldsymbol{G}' = 0$ element for $\boldsymbol{q}$ along high symmetry directions for Si. Though these results were instructive, they were of limited usefulness. The single most instructive quantity, the static dielectric constant ϵ_0, is determined from the *inverse* DM, which requires an evaluation of the full DM followed by a matrix inversion.

The necessity of understanding static and dielectric screening in semiconductors led initially to the proposal of various model forms of the DM, usually based on physical intuition and constructed to satisfy sum rules [277–281]. The first extensive ab initio calculation was by Louie, Chelikowsky and Cohen [282], who evaluated $\epsilon_{GG'}(\boldsymbol{q}=0,\ \omega)$ in order to obtain the optical absorption spectrum from Im ϵ^{-1}. Their results established that the "local field corrections", i.e. the off-diagonal elements, significantly improved the agreement with experiment.

Baldereschi and Tosatti [283] used the static DM to introduce the idea of the "dielectric band structure," which is defined by the eigenvalues of the DM. The corresponding eigenvectors of the DM describe eigen-perturbations which are merely scaled by the corresponding eigenvalue as a result of the screening process. The dielectric band structure has several uses; it makes it easy to

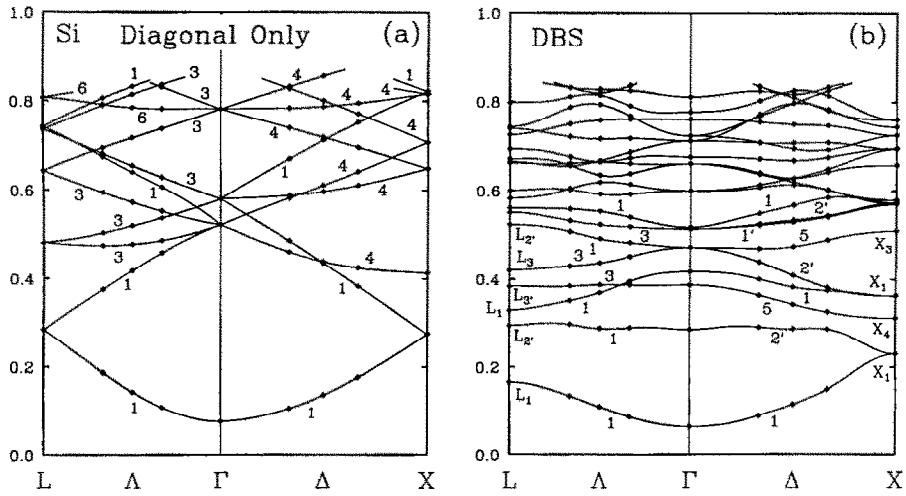

Fig. 17. Dielectric band structure (DBS), that is, the eigenvalues of the inverse dielectric matrix, for Si along the (111) and (100) symmetry directions, as calculated by Hybertsen and Louie [285]. (a) Diagonal terms only, with numbers denoting degeneracies. (b) Full dielectric matrix, including exchange-correlation effects. Numbers denote the symmetry label. These bands are bounded above by unity and the (many) bands above ~ 0.85 have not been plotted. The low bands corresponds to strongly screened eigenpotentials (see text).

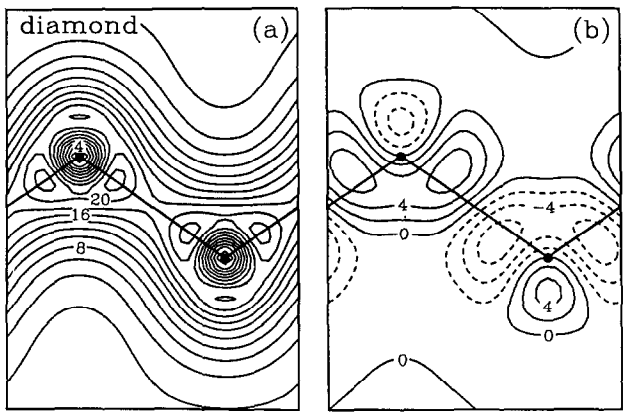

Fig. 18. Effect on the valence pseudochange density of diamond of a constant applied electric field, from Hybertsen and Louie [288]. The field is applied along the y-direction of the figure, and the linear response is calculated from the static dielectric matrix. (a) Density in zero field in a (110) plane, normalized to 8 valence electrons/cell. (b) Change in density, with a magnitude corresponding to an applied field of magnitude $eE = 2\text{Ry/bohr}$, or $E \approx 5 \times 10^9$ volt/cm.

introduce symmetry into the screening problem, it illustrates chemical trends and directly reflects local field effects. Extensive studies of Si [279] and GaAs and ZnSe [284] reflect the trends with increasing ionicity as well as the details of each system separately. Figure 17 shows the dielectric band structure calculated by Hybertsen and Louie [285] for Si.

Dielectric properties have continued to be an active area of development. Baldereschi and Tosatti [286] established that the special point techniques for Brillouin zone summations, discussed in section 2.4, were useful for calculations of the DM. They presented limited results for Si, Ge, α-Sn, MgO and NaCl. Baroni and Resta [287] presented a careful calculation of ϵ_0 for Si, finding that the local density approximation overestimates this quantity by 12%. The numerical aspects of the direct evaluation of static DMs in non-metals has been presented in some detail by Hybertsen and Louie [285], who also present the result of extensive studies of diamond, Si, Ge and LiCl [288].

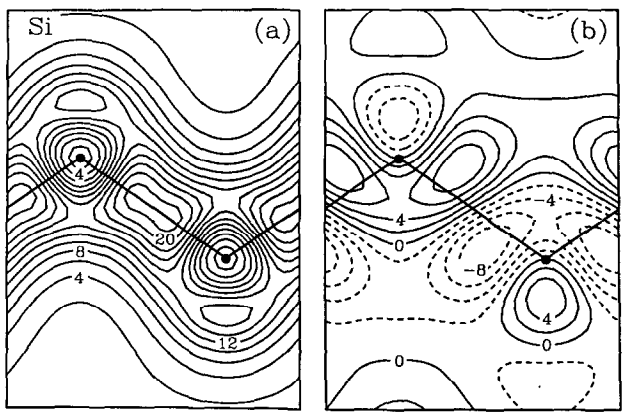

Fig. 19. As in fig. 18, but for silicon. The increase is polarizability compared to diamond (fig. 18) is apparent from the larger rearrangement of change in the bond region.

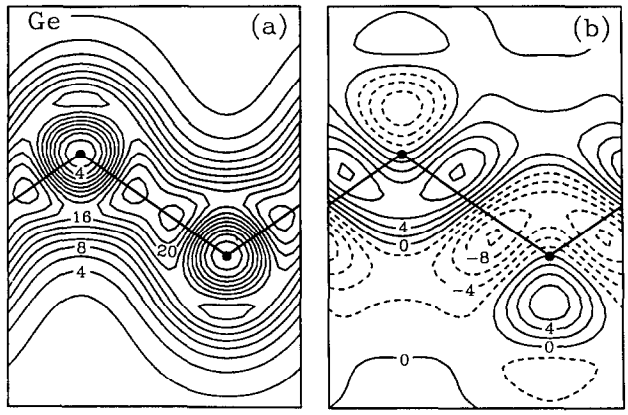

Fig. 20. As in figs. 18 and 19, but for germanium. Note the dipole formed in the bond region, which lies neither along the bond direction nor along the field direction.

One straightforward use of the static DM for investigating microscopic screening in solids is illustrated in figs. 18–20. These figures, taken from Hybertsen and Louie [288], show directly the change in charge density of diamond, Si and Ge due to a constant applied electric field. The field induces a dipolar redistribution of charge in each bond which is not parallel to the field, but the vector sum from all bonds of course will lie along the field direction. The figures also reflect the increasing polarizability in the series diamond → Si → Ge.

Methods have been suggested for obtaining certain screening properties without a full evaluation and inversion of the DM. McKitterick [289] observed that a series of supercell calculations can be used to extract ϵ_0. The basic idea is to perform a discretized $\boldsymbol{q} \to 0$ limit to obtain ϵ_0 (effective charges and the piezoelectric constant can also be obtained in an analogous manner). He used supercells N times larger than the primitive cell in the (1, 1, 1) direction, and added perturbations V_{pert} of wavelength $q = (1/N)$ (1, 1, 1) $2\pi/a$, respectively. The dielectric constant is given by

$$\epsilon^{-1}(q) = 1 - \frac{4\pi n(q)}{q^2 V_{\text{pert}}}.$$

McKitterick used $N = 2$, 4, 6 and 8 and the ansatz

$$\epsilon^{-1}(q) = \epsilon_0^{-1} + \frac{b}{N^2} + \frac{c}{N^4}$$

to arrive at $\epsilon_0 = 10.9$ for GaAs, in excellent agreement with experiment. This method appeared to be somewhat preferable to a related method [290,291] which applied a linear (ramp or sawtooth) perturbing potential to a large supercell and then let self-consistency generate the screening as in the method of McKitterick.

Kunc and Tosatti [292] pointed out the possibility of evaluating the inverse DM directly, rather then by inversion of the DM itself. For any rational $\boldsymbol{q}$, one can find a supercell for which $\boldsymbol{q}$

is a reciprocal lattice vector. Adding the perturbation $V_{\text{pert}} \exp[i(\boldsymbol{q} + \boldsymbol{G}_0) \cdot \boldsymbol{r}]$ to this supercell, one has from linear response theory.

$$\delta n(\boldsymbol{q} + \boldsymbol{G}) = \chi_{GG_0}(\boldsymbol{q}) V_{\text{pert}},$$

so from each self-consistent calculation for a chosen $\boldsymbol{G}_0$, it is possible to obtain an entire column of $\chi(\boldsymbol{q})$. The method is limited of course to high symmetry, highly commensurate values of $\boldsymbol{q}$ by computational considerations. Screening in Ge and GaAs, primarily at $\boldsymbol{q} = 0$, have been studied using this approach by Kunc and Tosatti [292]. Fleszar and Resta [293] applied it to study screening in Si and GaAs at the Γ, X and L points.

Baroni, Giannozzi and Testa [294] used Green's functions to establish an even more powerful method of calculation. Although the alterations required by their formalism are not difficult or extensive, they will not be recounted here. Their method allows the calculation of the response to perturbations of arbitrary wavelength using only the *valence* energy bands and wavefunctions of the unperturbed crystal. As an application they calculated ϵ_0 and the LO and TO phonon frequencies at $\boldsymbol{q} = 0$ for Si. Baroni, Giannozzi and Testa [294a] have also used a linear response approach to calculate elastic constants and the bulk modulus of Si.

6.6. Dynamical self-energies

The applications discussed in the previous subsections reflect the incredible success the local density approximation (LDA) has had in predicting a wide assortment of physical and chemical behavior. One shortcoming of LDA – widely referred to as a failure – is in predicting the bandgaps in semiconductors and insulators. In truth the lack of agreement between experimental gaps and the LDA value is no failure at all, since even the exact density functional does not guarantee the experimental gap [295,296]. It is clear however that LDA sometimes severely underestimates the DFT gap. An illustration of this is Ge, for which LDA predicts a vanishingly small gap, or even a negative gap (semimetallic behavior) when spin-orbit corrections are included [297,298].

Using the self-consistent pseudopotential LDA results as the starting point, Pickett and Wang [299,300] found that including dynamical self-energy corrections so as to describe true single-particle excitation energies brought the gaps of Si and diamond into reasonable agreement with experiment. Their approach, which invoked a local approximation into the real space part of the self-energy which is in principle non-local, was subsequently [301] shown to vastly improve the gaps of SiC and GaP. In fig. 21 GaP is shown as an example of how the quasiparticle bands differ from the LDA bands. This local approach did not produce acceptable gaps in Ge and GaAs, however, which are the two common semiconductors for which LDA seems to give the least satisfactory bandgaps. Because of its computational simplicity this model can be applied to any system for which the DFT equations can be solved, and Sterne and Wang [302] have applied the method to the Si/GaP (110) interface to study band discontinuities at this heterojunction.

Subsequent work has placed more emphasis on the real space non-locality of the self-energy, which requires considerably more effort to evaluate. Hybertsen and Louie [303,304] have suggested efficient, sum-rule-satisfying methods of approximating the wavevector- and frequency-dependent dielectric screening matrix, which is the difficult quantity to evaluate. With this more detailed implementation of the self-energy correction, the gaps and excitation energies

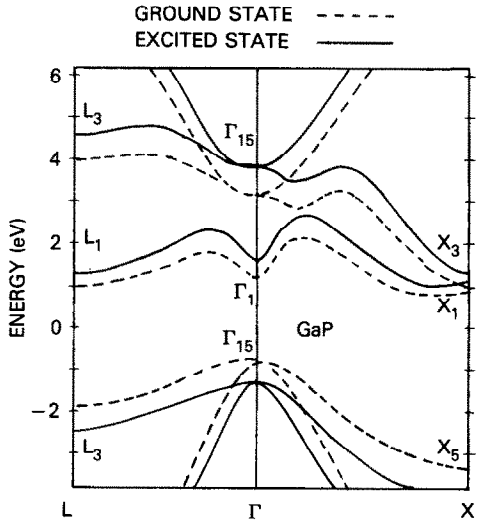

Fig. 21. The local density (dashed) and quasiparticle (solid) bands of GaP along high symmetry directions, from Pickett and Wang [301]. The local density bands were calculated using the Levine–Louie [Phys. Rev. B25 (1982) 6310] exchange-correlation potential. In these calculations the chemical potential is held fixed at midgap, so valence bands move downwards in addition to the upward correction of the conduction bands.

are given accurately for diamond and Si [303,305], for Ge [303] and even for the ionic insulator LiCl [306]. The same approach also works well for alkali metals [307].

The large amount of computational effort necessary to include the full non-locality of the self-energy is a distinct hindrance to the application of this theory to larger, more complex systems which are already computationally demanding. Driven by such considerations, Hybertsen and Louie [308] have introduced a model in which the screening response $\epsilon^{-1}(\boldsymbol{r}, \boldsymbol{r}')$ is approximated by the average of the responses of homogeneous mediums at densities $n(\boldsymbol{r})$ and $n(\boldsymbol{r}')$. This movement back toward the original, computationally simple approach of Pickett and Wang [299,300] leads to a model which is efficient and still accurate enough for most purposes (generally to 0.1–0.2 eV for diamond, Si and Ge).

Godby, Schlüter and Sham [309] have also investigated in detail the full non-locality of the self-energy operator by examining trends in diamond, Si, GaAS and AlAs. Finding good agreement with experiment and insight based on the trends they observed, they also introduced a simplified model which is computationally efficient and yet produces reasonably accurate quasiparticle energies.

6.7. Various large unit cell applications

With the advent of supercomputers and the continuing development of algorithms, it has become possible to attack problems with many atoms per unit cell. Since most real materials have important properties which cannot be understood either from the perfect periodic solid or the isolated point defect point of view, these new situations comprise an important part of the future of condensed matter studies. Here we mention a few of these studies.

Guttman and Fong [310] have used large cells of Si and H to model amorphous silicon hydride [a-(Si, H)]. Empirical pair and bond-bending potentials were used to generate several unit cells of

about 50 Si atoms and 7–13 atomic percent hydrogen. After relaxing the atomic positions with these empirical potentials, the self-consistent electronic structure was calculated. They focussed primarily on the density of states, both in the valence band region and especially near the (pseudo) gap. They concluded that this approach constitutes a promising method of studying gross features of a-(Si, H).

Nelson et al. [311] have followed this up by studying the effects on the vibrational and electronic properties of these models of a-(Si, H) of removing H atoms. The Si-related changes in the vibrational spectra are related to local distortions at Si sites neighboring a H atom which was removed. States introduced into the gap just above the valence band lie 0.5 eV below experimentally detected states thought to be of this origin.

Extended defects such as stacking faults and twin boundaries are crucial in understanding many properties of materials. Chou, Cohen and Louie [312] investigated both intrinsic and extrinsic stacking faults along the [111] direction in Si. They found that, while electron densities near the fault deviated only slightly from the perfect crystal density, the fault introduces defect states in the energy gap. Stacking fault energies were consistent with experimental values. Denteneer and van Haeringen [313] suggested that stacking fault energies can be obtaining without very large supercell calculations by calculating the energies of several polytypes of the crystal, and then parametrizing the energies in terms of interaction constants between layers. The approach was tested for Si and applied to Ge and SiC. They found not only that the parametrization seemed appropriate for these materials, but also that the sign of the layer interaction constants serves as an indicator of the occurrence of polytypism, such as occurs in SiC.

A model of a twin boundary in Si was studied by DiVincenzo et al. [314]. They first used semi-empirical tight-binding methods to relax the structure to the minimum energy configuration. Then they used 36 to 40 atom supercells (with repeated twins) to investigate the charge density, electronic structure and to obtain a twin-boundary energy. This type of application appears be be a fruitful one for future studies.

Qteish and Resta [315] have investigated the atomic structure and stability of Si–Ge solid solutions. They studied supercells involving nine different ordered structures of Si_xGe_{1-x}, finding none of them to be thermodynamically stable. Their results produced the measured x variation of the Ge–Ge and Ge–Si bond lengths and provided a prediction of the x variation of the Si–Si bond distance. The extreme chemical similarity of Si and Ge is responsible for the preference of solid solutions over ordered superstructures.

The pursuit of important materials properties was extended by Goodwin, Needs and Heine [316], who attacked the problem of impurity-induced embrittlement. Choosing aluminium as the host metal and taking Ge and As as impurities, they studied supercells containing up to 18 atoms and a vacuum layer. Within a somewhat idealized model, they found that the [111] interlayer energy of cohesion was increased by substitutional Ge and As impurities. These calculations did not support the decohesion models of impurity-promoted grain boundary embrittlement.

6.8. Quantum molecular dynamics and simulated annealing

Applications of the density functional approach to the properties of condensed matter systems almost universally have followed the prescription that, for example, in following the movement

of atoms as a geometrical relaxation is carried out, the density must be iterated to self-consistency while the ions are held static. Then the energy, forces, etc. can be obtained, the ions moved appropriately, and the process iterated as appropriate. In real systems, of course, the density responds continuously to the motion of ions, and recently there has been more emphasis on simulating the simultaneous dynamical motion of electrons and ions.

It is useful to distinguish two limiting cases for which one might wish to allow ions and electrons to move simultaneously. One case is in "simulated annealing," in which the objective is to approximate the annealing process which encourages the ionic coordinates to find a configuration near, if not at, the minimum of energy. Two simplified forms of this process, which may still be quite complicated, are simple energy minimization (of the type of consecutive electron/ion changes which have been dominant until now) and the process of "simulated quenching," in which one tries to avoid shallow local minima near the starting configuration and find a more representative minimum. For both of these processes it is of no particular use to try to simulate the ionic motion in any realistic way. For the second case, which I will call quantum molecular dynamics, the object is to simulate the actual dynamics of the ions while allowing the electronic density to deform continuously and exert quantum-mechanically-derived forces on the ions. This second case can be expected to be much more demanding than the first one.

Bendt and Zunger [258] presented a formalism for minimizing the energy of a system by varying the ionic configuration and electron density together. They utilized a form of density functional theory in which the potential, rather than the density, is varied until the energy is minimized. Then generalized "density forces" were defined which tend to drive the electronic system toward its minimum. For the ionic coordinates it was shown that approximate forces are sufficient, and much simpler computationally, to drive the ions toward their minimum configuration. Then the condition that the forces on both the ionic and electronic coordinates approach zero are applied together, resulting in a coupled system in which all generalized coordinates are varied together. A guiding principle of this approach is that none of the forces need to be particularly accurate except near the energy minimum. This method then amounts to one approach to simulated annealing (or quenching).

A method which has the potential for doing realistic quantum molecular dynamics simulations was presented by Car and Parrinello (CP) [317]. Part of their approach is, similar to Bendt and Zunger, to regard the density functional part as a complex optimization problem, and apply concepts from simulated annealing theory [318]. The novel part of the CP approach was to introduce a fictitious Lagrangian which includes the "time" derivative of the occupied electronic wavefunctions as one contribution to the kinetic energy, and let the electronic plus ionic energy of the system constitute the potential energy. The derived equations of motion result in desirable properties: (1) at equilibrium (vanishing first and second time derivatives) the electron density is the ground state density, and the forces on the ions vanish; (2) by making the fictitious electronic mass small enough the electronic system can be made to stay arbitrarily close to the Born–Oppenheimer surface (with concomitant numerical expense) so the true dynamics of the system can be followed; and (3) the electronic mass can be increased and/or energy can be extracted from the system to carry out annealing/quenching simulations. After applying this method to model Si systems [317] and suggesting computational gains which may be obtained with real space (rather than reciprocal space) implementations [319], CP found that the description of amorphous Si provided by this approach reproduced its atomic structure, the phonon spectrum and its electronic properties very well.

From one point of view a large part of the content of the CP scheme is in providing a systematic procedure for obtaining a correction to a set of electron wavefunctions resulting from a change in the Hamiltonian, *without* performing an additional diagonalization which becomes very costly for large systems. This part of the CP scheme is not new, but the incorporation into a coupled electron–ion system which simulates continuous (non-zero temperature) behavior is novel. Related methods of updating eigenvectors can be obtained from the work of Davidson [320,321], Wood and Zunger [322], Nex [322a], and Vanderbilt [322b].

The original version of the CP scheme incorporated numerical techniques which no doubt will be improved upon as the method gains use. Payne et al. [323] have reported one such improvement in handling the differential equations, which allows a significantly larger time step to be used. Allan and Teter [324] pointed out that the Kleinman-Bylander [33] form of non-local pseudopotential can be applied to great advantage in the CP scheme, and reported an energy-minimization application of the method to SiO_2 glass. Payne, Bristowe and Joannopoulos [325] studied a twist grain boundary in Ge using the CP scheme. By investigating several rotation-and-translation configurations, they predicted the structure of this grain boundary and found evidence of unusual defects and glasslike tunneling-mode states. Bachelet and De Lorenzi [326] have pursued non-zero temperature applications of the CP method by studying the lattice dynamics (within a quasi-harmonic treatment) of argon and the dynamics around a vacancy in argon. Needels, Payne and Joannopoulos [327] have applied the method to the (100) surface of Ge, predicting the geometry of the $c(4 \times 2)$-symmetry structure and finding an unusual soft energy surface for displacement of dimers on this surface. Payne, Needels and Joannopoulos [328] have discussed some general features related to symmetry and symmetry-breaking in the CP scheme.

Acknowledgements

Over the years I have benefitted greatly from discussions, comparisons and collaborations with colleagues too numerous to enumerate here. I would not have entered the arena of self-consistent pseudopotential theory, however, were it not for an IBM Postdoctoral Fellowship at the University of California at Berkeley in 1976–7 with the very active group headed by M.L. Cohen. I also want to acknowledge close interaction with, and the unselfish sharing of expertise and codes by S.G. Louie, K.-M. Ho, G.P. Kerker and J. Ihm during this Fellowship at Berkeley. My continuing involvement in this field of research has been encouraged and supported by the Office of Naval Research.

Appendix A. Norm-conserving pseudopotential transformation

As discussed in section 3.4, the published parameters of Bachelet, Hamann and Schlüter [25] must be transformed to another set of constants for use in computer codes and formal expressions. If the transformation matrix is inverted numerically, unanticipated results may arise, especially if low numerical precision is used. A subroutine for carrying out this transformation

directly, without matrix inversion, is supplied here. It follows the expressions given in eqs. (3.26)–(3.30) and should be quite reliable. Note however that double (64 bit) precision should always be used in this routine and for the resulting parameters.

```
      SUBROUTINE C2A (NKIND)
      IMPLICIT REAL*8 (A-H,O-Z)
C*********************************************************************
C--->  READS PUBLISHED ("C") PARAMETERS OF BACHELET-HAMANN-SCHLUETER *
C--->   RETURNS THE CONSTANTS ("A") NEEDED IN THEIR FORMULA FOR THE  *
C--->   NON-LOCAL PSEUDOPOTENTIALS                                   *
C*********************************************************************
      COMMON/VNONLC/A(3,3,6),ALF(3,3,3)
      DIMENSION Q(6,6),S(6,6),C(6)
      DATA PI/3.14159265D0/
C     ================================================================
    1 FORMAT(9F8.4)
    2 FORMAT(' ATOM',I2,'  L=',I1,3F8.2,6F10.4)
    3 FORMAT(1X,2I2,6F15.5)
    4 FORMAT(6F20.14)
C     ================================================================
C
C--->  SUM OVER ATOM KINDS *NKIND*
C
      DO 200 N=1,NKIND
C
C--->  SUM OVER ANGULAR MOMENTUM VALUES 0,1,2
C
      DO 100 L=1,3
      READ(5,1) (ALF(N,L,J),J=1,3),(C(I),I=1,6)
      WRITE(6,2) N,L,(ALF(N,L,J),J=1,3),(C(I),I=1,6)
        DO 10 I=1,6
        DO 10 J=1,6
        Q(I,J)=0.D0
        II=I
        JJ=J
        IF(II.GT.3) II=I-3
        IF(JJ.GT.3) JJ=J-3
        ALFINV=1.D0/( ALF(N,L,II)+ALF(N,L,JJ) )
        S(I,J)=0.25D0*ALFINV*DSQRT(PI*ALFINV)
        IF( (I.GT.3 .AND. J.LE.3) .OR. (I.LE.3 .AND. J.GT.3) )
     +       S(I,J)=S(I,J)*1.5D0*ALFINV
        IF(I.GT.3 .AND. J.GT.3) S(I,J)=S(I,J)*3.75D0*ALFINV*ALFINV
   10   CONTINUE
C
C--->  OVERLAP MATRIX S(I,J) IS NOW CONSTRUCTED
C--->  NOW CONSTRUCT TRANSFORMATION MATRIX Q(I,J)
C
      Q(1,1)=DSQRT(S(1,1))
        DO 20 J=2,6
   20   Q(1,J)=S(1,J)/Q(1,1)
      Q(2,2)=DSQRT( S(2,2)-Q(1,2)**2 )
      Q(2,3)=( S(2,3)-Q(1,2)*Q(1,3) )/Q(2,2)
      Q(3,3)=DSQRT( S(3,3)-Q(1,3)**2-Q(2,3)**2 )
      Q(2,4)=( S(2,4)-Q(1,2)*Q(1,4) )/Q(2,2)
      Q(3,4)=( S(3,4)-Q(1,3)*Q(1,4)-Q(2,3)*Q(2,4) )/Q(3,3)
      Q(4,4)=DSQRT(S(4,4)-Q(1,4)**2-Q(2,4)**2-Q(3,4)**2)
      Q(2,5)=( S(2,5)-Q(1,2)*Q(1,5) )/Q(2,2)
      Q(3,5)=( S(3,5)-Q(1,3)*Q(1,5)-Q(2,3)*Q(2,5) )/Q(3,3)
      Q(4,5)=(S(4,5)-Q(1,4)*Q(1,5)-Q(2,4)*Q(2,5)-Q(3,4)*Q(3,5))/Q(4,4)
      Q(5,5)=DSQRT(S(5,5)-Q(1,5)**2-Q(2,5)**2-Q(3,5)**2-Q(4,5)**2)
```

```
      Q(2,6)=( S(2,6)-Q(1,2)*Q(1,6) )/Q(2,2)
      Q(3,6)=( S(3,6)-Q(1,3)*Q(1,6)-Q(2,3)*Q(2,6) )/Q(3,3)
      Q(4,6)=(S(4,6)-Q(1,4)*Q(1,6)-Q(2,4)*Q(2,6)-Q(3,4)*Q(3,6))/Q(4,4)
      Q(5,6)=(S(5,6)-Q(1,5)*Q(1,6)-Q(2,5)*Q(2,6)-Q(3,5)*Q(3,6)
     +                -Q(4,5)*Q(4,6))/Q(5,5)
      Q(6,6)=DSQRT(S(6,6)-Q(1,6)**2-Q(2,6)**2-Q(3,6)**2-Q(4,6)**2
     +                -Q(5,6)**2)
      WRITE(6,4) ((S(I,J),J=1,6),I=1,6)
      WRITE(6,4)
      WRITE(6,4) ((Q(I,J),J=1,6),I=1,6)
      WRITE(6,4)
C
C--->  BACK-SUBSTITUTE TO OBTAIN WANTED PARAMETERS
C
      A(N,L,6)=-C(6)/Q(6,6)
      A(N,L,5)=-(C(5)+Q(5,6)*A(N,L,6))/Q(5,5)
      A(N,L,4)=-(C(4)+Q(4,5)*A(N,L,5)+Q(4,6)*A(N,L,6))/Q(4,4)
      A(N,L,3)=-(C(3)+Q(3,4)*A(N,L,4)+Q(3,5)*A(N,L,5)
     +               +Q(3,6)*A(N,L,6))/Q(3,3)
      A(N,L,2)=-(C(2)+Q(2,3)*A(N,L,3)+Q(2,4)*A(N,L,4)
     +               +Q(2,5)*A(N,L,5)+Q(2,6)*A(N,L,6))/Q(2,2)
      A(N,L,1)=-(C(1)+Q(1,2)*A(N,L,2)+Q(1,3)*A(N,L,3)
     + +Q(1,4)*A(N,L,4)+Q(1,5)*A(N,L,5)+Q(1,6)*A(N,L,6))/Q(1,1)
  100 CONTINUE
  200 CONTINUE
C
C---> WRITE OUT TRANSFORMED PSEUDOPOTENTIAL PARAMETERS
C
      DO 300 N=1,NKIND
      DO 300 L=1,3
  300 WRITE(6,3) N,L,(A(N,L,I),I=1,6)
C
C--->  CHECK PROCEDURE BY CALCULATING IN REAL SPACE FOR SILICON
C--->   COMPARE WITH TABLE IN PAPER OF BACHELET, HAMANN, SCHLUETER
C
C     DO 605 IR=1,26
C     RR=0.1*(IR-1)
C     IF(IR.EQ.1) RR=0.0001
C     VCORE=1.6054*DERF(DSQRT(2.16D0)*RR)-0.6054*DERF(DSQRT(0.86D0)*RR)
C     VCORE=-4.*VCORE/RR
C     DO 604 L=1,3
C     C(L)=0.
C     DO 603 I=1,3
C 603 C(L)=C(L)+(A(1,L,I)+RR*RR*A(1,L,I+3))*DEXP(-ALF(1,L,I)*RR*RR)
C 604 C(L)=C(L)+VCORE
C 605 WRITE(6,606) RR,C(1),C(2),C(3)
C 606 FORMAT(1X,'R=',F8.4,3F15.4)
C
      RETURN
      END
```

Appendix B. Exchange-correlation functionals

To make this review more nearly self-contained, in this appendix we provide a few expressions for the exchange-correlation (xc) energy density and potential which are in common use. The xc energy is defined by eq. (2.4), that is, it is what remains beyond (a) the kinetic energy $T_0[n]$ of a non-interacting system with the same density $n(\boldsymbol{r})$, (b) the energy of interaction with the external potential (viz, the ionic pseudopotentials), and (c) the Hartree energy E_n.

The xc energy is written in terms of an xc energy density (eq. (4.61)) ϵ_{xc} as

$$E_{xc}[n] = \int d^3r'\, n(\boldsymbol{r}')\epsilon_{xc}(\boldsymbol{r}';\, n), \tag{B.1}$$

in which case the xc potential V_{xc} is given by

$$V_{xc}(\boldsymbol{r};\ n) = \frac{\delta E_{xc}[n]}{\delta n(\boldsymbol{r})} = \epsilon_{xc}(\boldsymbol{r};\ n) + \int d^3r' n(\boldsymbol{r}') \frac{\delta \epsilon_{xc}(\boldsymbol{r}';\ n)}{\delta n(\boldsymbol{r})}. \tag{B.2}$$

Here we deal only with the *local density approximation* (LDA), for which the xc energy density at $\boldsymbol{r}$ is taken as that of a homogeneous electron gas with the same local density:

$$\epsilon_{xc}(\boldsymbol{r};\ n) \to \epsilon_{xc}^{LD}(n(\boldsymbol{r})). \tag{B.3}$$

Here $\epsilon_{xc}^{LD}(n)$ is a simple function of n. Then from (B.2) we find

$$V_{xc}(\boldsymbol{r};\ n) \to V_{xc}^{LD}(n(\boldsymbol{r})), \tag{B.4}$$

where

$$V_{xc}^{LD}(n) = \epsilon_{xc}^{LD}(n) + n \frac{d\epsilon_{xc}^{LD}(n)}{dn}. \tag{B.5}$$

One of the most commonly used approximations is that of Hedin and Lundqvist [329]. The xc energy density is separated into exchange and correlation parts, with the known electron gas exchange term given by

$$\epsilon_x^{LD}(r_x) = -\frac{3e^2}{4\pi}(3\pi^2 n)^{1/3} = -\frac{3e^2}{4\pi\alpha r_s}, \tag{B.6}$$

where r_s is the electron gas parameter which gives the mean distance between electrons:

$$\frac{4\pi}{3}(r_s a_B)^3 = \frac{1}{n}. \tag{B.7}$$

To obtain energies and potentials in atomic (Hartree) units, set $e^2 = 1$; for Rydberg units set $e^2 = 2$. The exchange potential form (B.6) is

$$V_x^{LD}(r_s) = \epsilon_x^{LD}(r_s) - \frac{r_s}{3}\frac{d\epsilon_x^{LD}(r_s)}{dr_s} = \tfrac{4}{3}\epsilon_x^{LD}(r_s). \tag{B.8}$$

The Hedin–Lundqvist result for the correlation energy was given in the form

$$\epsilon_c^{HL}(r_s) = -\frac{Ce^2}{2}\left[(1+x^3)\log\left(1+\frac{1}{x}\right) + \frac{x}{2} - x^2 - \tfrac{1}{3}\right], \tag{B.9}$$

where $A = 21$, $C = 0.045$, and $x = r_s/A$. The correlation potential is

$$V_c^{HL}(r_s) = \epsilon_c^{HL}(r_s) - \frac{r_s}{3}\frac{d\epsilon_c^{HL}(r_s)}{dr_s} = -\frac{Ce^2}{2}\log\left(1+\frac{1}{x}\right). \tag{B.10}$$

More accurate, essentially exact results for the homogeneous electron gas were obtained by quantum Monte Carlo techniques by Ceperley and Alder [330]. Their correlation energies were fit by Vosko, Wilk and Nusiar [331,332] to the form

$$\epsilon_c^{\mathrm{VWN}}(r_s) = \frac{Ae^2}{2}\left\{\log\left(\frac{y^2}{Y(y)}\right) + \frac{2b}{Q}\tan^{-1}\left(\frac{Q}{2y+b}\right)\right.$$

$$\left. - \frac{by_0}{Y(y_0)}\left[\log\left(\frac{(y-y_0)^2}{Y(y)}\right) + \frac{2(b+2y_0)}{Q}\tan^{-1}\left(\frac{Q}{2y+b}\right)\right]\right\}. \tag{B.11}$$

Here $y = r_s^{1/2}$, $Y(y) = y^2 + by + c$, $Q = (4c - b^2)^{1/2}$, $y_0 = -0.10498$, $b = 3.72744$, $c = 12.93532$ and $A = 0.0621814$. The corresponding potential can be obtained from

$$V_c^{\mathrm{VWN}}(r_s) = \epsilon_c^{\mathrm{VWN}}(r_s) - \frac{r_s}{3}\frac{\mathrm{d}\epsilon_c^{\mathrm{VWN}}(r_s)}{\mathrm{d}r_s} \tag{B.12}$$

with

$$r_s\frac{\mathrm{d}\epsilon_c^{\mathrm{VWN}}(r_s)}{\mathrm{d}r_s} = A\frac{e^2}{2}\frac{c(y-y_0) - by_0 y}{(y-y_0)(y^2+by+c)}. \tag{B.13}$$

This last expression was not given explicitly by VWN.

VWN chose the form (B.13) because it satisfied certain limiting forms which are known. They chose to fit the parameters to the calculated values of Ceperley and Alder at $r_s = 10$, 20, 50 and 100, where their calculations are extremely accurate (errors of 0.05 mRy or less in ϵ_c). The fit was within 0.1 mRy for these points. They then checked the "predictive" power of the parametrization for $r_s = 2$ and 5, obtaining values within the statistical error in the Ceperley–Alder results (which were 0.4 and 0.1 mRy, respectively. For practical purposes it seems the VWN form can be considered to give the "exact" LDA correlation energy and potential, and there can hardly be any excuse for not employing it in LDA calculations. Perdew and Zunger [332a] have presented a parametrization which is very similar (numerically) to that of VWN.

Another form which is still sometimes used is the correlation energy of Wigner [333] (corrected by Pines [334]), which can be written

$$\epsilon_c^{\mathrm{W}}(r_s) = -e^2\frac{W_1}{r_s + W_2}, \tag{B.14}$$

$$V_c^{\mathrm{W}}(r_s) = -e^2 W_1\frac{(2/3)r_s + W_2}{(r_s + W_2)^2}, \tag{B.15}$$

where $W_1 = 0.88$ and $W_2 = 7.79$. This form was chosen simply as an interpolation between his high density (small r_s) estimate of ~ 0.11 Ry and the low density form corresponding to a Wigner solid.

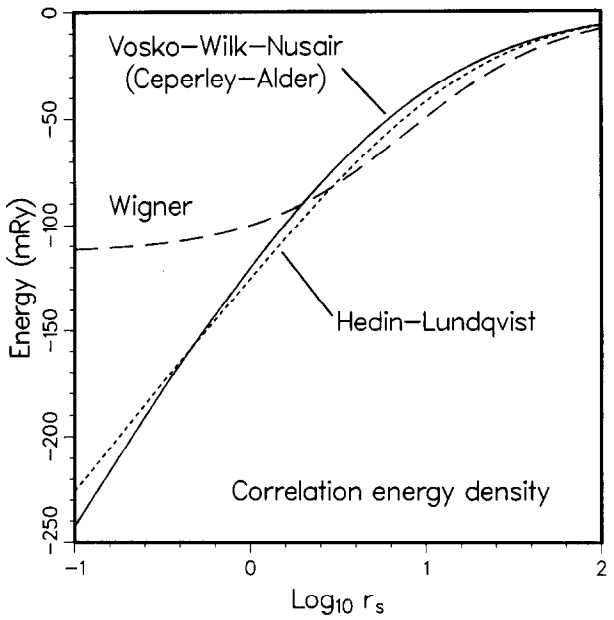

Fig. 22. Correlation-energy density versus r_s from the Vosko–Wilk–Nusair, Hedin–Lundqvist, and Wigner parametrizations. The first two are similar at all densities, while the Wigner form deviates from the better results at high density (low r_s).

These forms for the correlation energy and potential are plotted in figs. 22 and 23. While the Hedin–Lundqvist form is rather close to that of VWN, the Wigner interpolation formula deviates rather strongly for r_s less than 1 a.u. Although the difference appears rather severe, densities larger than this occur only near and in the core region of atoms. Within the pseudopotential approximation, densities of this magnitude rarely arise. Even in all-electron calculations the Wigner approximation for correlation is not so bad, since in the core region the total potential is 10–10^6 rydbergs in magnitude, which makes the correlation potential a small fraction of the total. While there is evidence that the Wigner correlation energy leads to somewhat better work functions in surface calculations than alternative forms, it is certainly not clear that it does so for the right reasons.

A related property which arises in linear response applications is the second functional derivative of the correlation energy with respect to the density,

$$u_c(\boldsymbol{r}, \boldsymbol{r}'; n) = \frac{\delta^2 E_c[n]}{\delta n(\boldsymbol{r})\delta n(\boldsymbol{r}')}. \tag{B.16}$$

This quantity arises in parallel with the Coulomb interaction $e^2/|\boldsymbol{r}-\boldsymbol{r}'|$ (together with an analogous exchange interaction) and should be thought of as a *correlation interaction*. (See the discussion near eq. (4.57).) In LDA it is diagonal (local), and can be written

$$u_c(\boldsymbol{r}, \boldsymbol{r}'; n) = u_c^{\mathrm{LD}}(n(\boldsymbol{r}))\delta(\boldsymbol{r}-\boldsymbol{r}'), \tag{B.17}$$

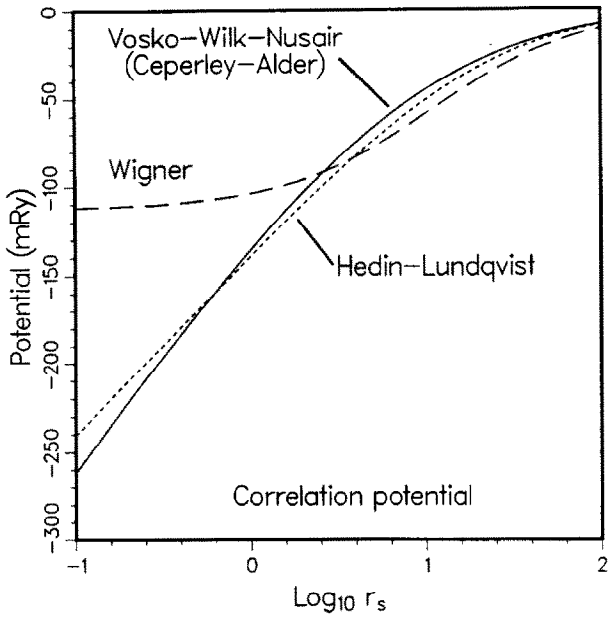

Fig. 23. Correlation potential versus r_s from the Vosko–Wilk–Nusair, Hedin–Lundqvist and Wigner parametrizations. Note that the correlation potential follows the correlation energy density (fig. 22) closely for each form.

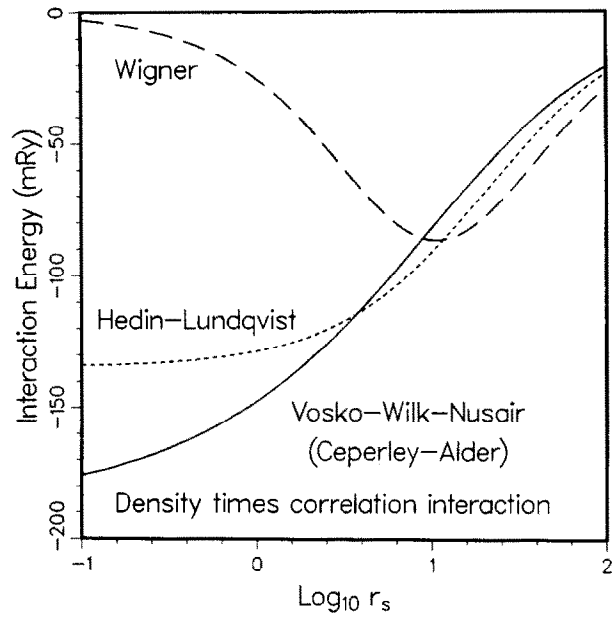

Fig. 24. Correlation interaction strength (multiplied by the density) versus r_s from the Vosko–Wilk–Nusair, Hedin–Lundqvist and Wigner parametrizations. The Wigner form is badly in error at high density (small r_s). The Vosko–Wilk–Nusair form should be the most accurate (see text).

with

$$u_c^{LD}(n) = \frac{dV_c(n)}{dn}. \tag{B.18}$$

In fig. 24 this quantity is plotted in the form $nu_c^{LD}(n)$ versus $\log_{10} r_s$. The Wigner form is found to differ strongly from the Hedin–Lundqvist and VWN forms already at metallic densities. Since applications of linear screening theory are noticeably affected by the exchange and correlation interactions, it is clearly advisable to use the more accurate forms.

Note added in proof

An extensive discussion of the Broyden mixing formalism and numerical tests and comparison with other procedures has been given by Blügel [335].

References

[1] C. Herring, Phys. Rev. 57 (1940) 1169.
[2] J.C. Phillips and L. Kleinman, Phys. Rev. 116 (1959) 287.
[3] M.L. Cohen and V. Heine, in: Solid State Physics, vol. 24, eds. H. Ehrenreich, F. Seitz and D. Turnbull (Academic, New York, 1970) pp. 37–248.
[4] Theory of the Inhomogeneous Electron Gas, eds. S. Lundqvist and N.H. March (Plenum, New York, 1983).
[5] Many-Body Phenomena at Surfaces, eds. D. Langreth and H. Suhl (Academic, New York, 1984) secs. I and II.
[6] Density Functional Methods in Physics, eds. R.M. Dreizler and J. da Providência (Plenum, New York, 1985).
[7] The Electronic Structure of Complex Systems, eds. P. Phariseau and W.M. Temmerman (Plenum, New York, 1984).
[8] W.E. Pickett, Comm. Solid State Phys. 12 (1985) 1, ibid. 12 (1986) 57.
[9] T.L. Gilbert, Phys. Rev. B12 (1975) 2111.
[10] M. Berrondo and O. Goscinski, Intl. J. Quant. Chem. Symp. 9 (1975) 67.
[11] M. Levy, Proc. Natl. Acad. Sci. USA 76 (1979) 6062.
[12] J.F. Janak, Solid State Commun. 20 (1976) 151.
[13] U. von Barth and C.D. Gelatt, Phys. Rev. B21 (1980) 2222.
[14] B.J. Austin, V. Heine and L.J. Sham, Phys. Rev. 127 (1962) 276.
[15] J.D. Weeks, A. Hazi and S.A. Rice, Adv. Can. Chem. Phys. 16 (1969) 283.
[16] L.R. Kahn and W.A. Goddard, J. Chem. Phys. 56 (1972) 2685.
[17] J.D. Weeks and S.A. Rice, J. Chem. Phys. 49 (1968) 2741.
[18] S. Topiol, A. Zunger and M.A. Ratner, Chem. Phys. Lett. 49 (1977) 367.
[19] A. Zunger and M.L. Cohen, Phys. Rev. B18 (1978) 5449.
[20] P.K. Lam, M.L. Cohen and A. Zunger, Phys. Rev. B22 (1980) 1698.
[21] D.R. Hamann, M. Schlüter and C. Chiang, Phys. Rev. Lett. (1979) 1494.
[22] J.A. Applebaum and D.R. Hamann, Phys. Rev. B8 (1973) 1777.
[23] M. Schlüter, J.R. Chelikowsky, S.G. Louie and M.L. Cohen, Phys. Rev. Lett. 34 (1975) 1385.
[24] W.C. Topp and J.J. Hopfield, Phys. Rev. B7 (1974) 1295.
[25] T. Starkloff and J.D. Joannopoulos, Phys. Rev. B16 (1977) 5212.
[26] G.B. Bachelet, D.R. Hamann and M. Schlüter, Phys. Rev. B26 (1982) 4199.
[27] L. Kleinman, Phys. Rev. B21 (1980) 2630.
[28] G.B. Bachelet and M. Schlüter, Phys. Rev. B25 (1982) 2103.

[29] A.H. MacDonald, W.E. Pickett and D.D. Koelling, J. Phys. C13 (1980) 2675.
[30] H.S. Greenside and M. Schlüter, Phys. Rev. B28 (1983) 535.
[31] G.P. Kerker, J. Phys. C13 (1980) L189.
[32] D. Vanderbilt, Phys. Rev. B32 (1985) 8412.
[33] P.C. Pattnaik, G. Fletcher and J.L. Fry, Phys. Rev. B28 (1983) 3564.
[34] L. Kleinman and D.M. Bylander, Phys. Rev. Lett. 48 (1982) 1425.
[35] D.M. Bylander and L. Kleinman, Phys. Rev. B29 (1984) 2274.
[36] S.G. Louie, S. Froyen and M.L. Cohen, Phys. Rev. B26 (1982) 1738.
[37] J.R. Gardner and N.A.W. Holzwarth, Phys. Rev. B33 (1986) 7139.
[38] P.-O. Löwdin, J. Chem. Phys. 19 (1951) 1396.
[39] J.W. Cooley and J.W. Tukey, Math. Comput. 19 (1965) 297.
[40] A. Baldereschi, Phys. Rev. B7 (1973) 5212.
[41] D.J. Chadi and M.L. Cohen, Phys. Rev. B8 (1973) 5747.
[42] D.J. Chadi, Phys. Rev. B16 (1977) 1746.
[43] L. Kleinman and J.C. Phillips, Phys. Rev. 116 (1959) 880.
[44] H.J. Monkhorst and J.D. Pack, Phys. Rev. B13 (1976) 5188.
[45] A.H. MacDonald, Phys. Rev. B18 (1978) 5897.
[46] L.G. Yaffe and W.A. Goddard III, Phys. Rev. A13 (1976) 1682.
[47] J. Harris (unpublished), referenced by Dederichs and Zeller, ref. [49].
[48] F. Manghi and E. Molinari, J. Phys. C15 (1982) 3627.
[49] P.H. Dederichs and R. Zeller, Phys. Rev. B28 (1983) 5462.
[50] L.G. Ferreira, J. Comp. Phys. 36 (1980) 198.
[51] D.G. Anderson, J. Assoc. Comput. Mach. 12 (1964) 547.
[52] D.R. Hamann, unpublished.
[53] L.F. Mattheiss and D.R. Hamann, Phys. Rev. B33 (1986) 823.
[54] K.-C. Ng, J. Chem. Phys. 61 (1974) 2681.
[55] K.-M. Ho, J. Ihm and J.D. Joannopoulos, Phys. Rev. B25 (1982) 4260.
[56] G.P. Kerker, Phys. Rev. B23 (1981) 3082.
[57] J. Arponen, P. Hautojärvi, R. Nieminen and E. Pajanne, J. Phys. F3 (1973) 2092.
[58] M. Manninen, R. Nieminen, P. Hautojärvi and J. Arponen, Phys. Rev. B12 (1975) 4012.
[59] C.G. Broyden, Math. Comput. 19 (1965) 577.
[60] P. Bendt and A. Zunger, Phys. Rev. B26 (1982) 3114.
[61] G.P. Srivastara, J. Phys. A17 (1984) L317.
[62] J.E. Dennis, Jr. and J.J. Moré, SIAM (Soc. Ind. Appl. Math.) Review 19 (1977) 46.
[63] D. Vanderbilt and S.G. Louie, Phys. Rev. B30 (1984) 6118.
[64] D.D. Johnson, Phys. Rev. B38 (1988) 12807.
[65] D. Singh, H. Krakauer and C.S. Wang, Phys. Rev. B34 (1986) 8391.
[66] S.G. Louie, K.-M. Ho and M.L. Cohen, Phys. Rev. B19 (1979) 1774.
[67] E.E. Lafon, R.C. Chaney and C.C. Lin, in: Computational Methods in Band Theory, eds. P.M. Marcus, J.F. Janak and A.R. Williams (Plenum, New York, 1971) p. 284.
[68] K.-M. Ho, W.E. Pickett and M.L. Cohen, Phys. Rev. Lett. 41 (1978) 580.
[69] W.E. Pickett, K.-M. Ho and M.L. Cohen, Phys. Rev. B19 (1979) 1734.
[70] K.-M. Ho, W.E. Pickett and M.L. Cohen, Phys. Rev. B19 (1979) 1751.
[71] P.B. Allen, W.E. Pickett, K.-M. Ho and M.L. Cohen, Phys. Rev. Lett. 40 (1978) 1532.
[72] K.-M. Ho, W.E. Pickett and M.L. Cohen, Phys. Rev. Lett. 41 (1978) 580.
[73] S.G. Louie, Phys. Rev. Lett. 40 (1978) 1525.
[74] D.J. Chadi, Phys. Rev. B16 (1977) 3572.
[75] G.A. Baraff and M. Schlüter, Phys. Rev. B19 (1979) 4656.
[76] G.A. Baraff and M. Schlüter, in: Proc. of the XIth Intern. Conf. on Defects and Radiation Effects in Semiconductors, Oiso, Japan, 1980, ed. R.R. Hasiguti (IOP, Bristol, 1981) p. 287.
[77] J. Bernholc and S.T. Pantelides, Phys. Rev. B18 (1978) 1780.
[78] J. Bernholc, N.O. Lipari and S.T. Pantelides, Phys. Rev. 21 (1980) 3545.

[79] S.G. Louie, M. Schlüter, J.R. Chelikowsky and M.L. Cohen, Phys. Rev. B13 (1976) 1654.
[80] W.E. Pickett, M.L. Cohen and C. Kittel, Phys. Rev. B20 (1979) 5050.
[81] W.E. Pickett, Phys. Rev. B23 (1981) 6603.
[82] W.E. Pickett, Phys. Rev. B26 (1982) 5650.
[83] B. Chakraborty, R.W. Seigel and W.E. Pickett, Phys. Rev. B24 (1981) 5445.
[84] J.R. Chelikowsky and S.G. Louie, Phys. Rev. B29 (1984) 3470.
[85] D. Vanderbilt and S.G. Louie, Phys. Rev. B30 (1984) 6118.
[86] D. Vanderbilt and S.G. Louie, J. Comput. Phys. 56 (1984) 259.
[86a] C.T. Chan, D. Vanderbilt and S.G. Louie, Phys. Rev. B33 (1986) 2455.
[87] R.W. Jansen and O.F. Sankey, Phys. Rev. B36 (1987) 6520.
[88] P.J. Feibelman, Phys. Rev. B35 (1987) 2626.
[89] P.J. Feibelman, Phys. Rev. Lett. 54 (1985) 2627.
[90] M.H. Kang, R.C. Tatar, E.J. Mele and P. Soven, Phys. Rev. B35 (1987) 5457.
[91] S. Itoh and K. Nakao, J. Phys. Soc. Japan 54 (1985) 4648.
[92] R.E. Watson, Phys. Rev. 111 (1958) 1108.
[93] S. Itoh and K. Nakao, J. Phys. Soc. Japan 54 (1985) 4657.
[94] W. Kohn and L.J. Sham, Phys. Rev. 140 (1965) A1133.
[95] O. Gunnarsson and B.I. Lundqvist, Phys. Rev. B13 (1976) 4274.
[96] K. Fuchs, Proc. Roy, Soc. 151 (1935) 585.
[97] R.A. Coldwell-Horsfall and A.A. Maradudin, J. Math. Phys. 1 (1960) 395.
[98] J. Ihm, A. Zunger and M.L. Cohen, J. Phys. C12 (1979) 4409.
[99] M.T. Yin and M.L. Cohen, Phys. Rev. B26 (1982) 5668.
[100] M.T. Yin and M.L. Cohen, Phys. Rev. B26 (1982) 3259.
[101] A. Zunger and M.L. Cohen, Phys. Rev. B19 (1979) 568.
[102] M. Weinert, R.E. Watson and J.W. Davenport, Phys. Rev. B32 (1985) 2115.
[103] J. Harris, Phys. Rev. B31 (1985) 1770.
[104] J.R. Chelikowsky, Phys. Rev. B21 (1980) 3074.
[105] M.T. Yin and M.L. Cohen, Solid State Commun. 38 (1981) 625.
[106] J. Ihm and M.L. Cohen, Phys. Rev. B21 (1980) 1527.
[107] J. Ihm and M.L. Cohen, Solid State Commun. 29 (1979) 711.
[108] G.P. Srivistava, J. Phys. C15 (1982) L739.
[109] E. Holzschuh, Phys. Rev. B28 (1983) 7346.
[110] G.B. Bachelet, H.S. Greenside, G.A. Baraff and M. Schlüter, Phys. Rev. B24 (1981) 4745.
[111] E.O. Kane, Phys. Rev. B21 (1980) 4600.
[112] P.J.H. Denteneer and W. van Haeringen, Solid State Commun. 59 (1986) 829.
[113] P.J.H. Denteneer and W. van Haeringen, J. Phys. C18 (1985) 4127.
[114] M.T. Yin and M.L. Cohen, Phys. Rev. B26 (1982) 5675.
[115] R. Biswas and D.R. Hamann, Phys. Rev. Lett. 55 (1985) 2001.
[116] S. Froyen and M.L. Cohen, Solid State Commun. 43 (1982) 447.
[117] J. Ihm and J.D. Joannopoulos, Phys. Rev. B24 (1981) 4191.
[118] K. Kunc and R.M. Martin, Phys. Rev. B24 (1981) 2311.
[119] R.J. Needs, R.M. Martin and O.H. Nielsen, Phys. Rev. B33 (1986) 3778.
[120] D. Vanderbilt and J.D. Joannopoulos, Solid State Commun. 35 (1980) 535.
[121] K.C. Hass, H. Ehrenreich and B. Velicky, Phys. Rev. B27 (1983) 1088.
[122] G.P. Srivistava, J.L. Martins and A. Zunger, Phys. Rev. B31 (1985) 2561.
[123] P. Boguslawski and A. Baldereschi, in: Proc. 17th Intern. Conf. Phys. Semicond., eds. D.J. Chadi and W.A. Harrison (Springer, New York, 1984) p. 939.
[124] K.C. Hass and D. Vanderbilt, in Proc. 18th Intern. Conf. Phys. Semicond., eds. O. Engström (World Scientific, Singapore, 1987) p. 1181.
[125] K.M. Ho, S.G. Louie, J.R. Chelikowsky and M.L. Cohen, Phys. Rev. B15 (1977) 1755.
[126] K.M. Ho, C.L. Fu, B.N. Harmon, W. Weber and D.R. Hamann, Phys. Rev. Lett. 49 (1982) 673.
[127] C.T. Chan, D. Vanderbilt, S.G. Louie and J.R. Chelikowsky, Phys. Rev. B33 (1986) 7941.

[128] A. Zunger and M.L. Cohen, Phys. Rev. B19 (1979) 568.
[129] C.T. Chan, D. Vanderbilt and S.G. Louie, Phys. Rev. B33 (1986) 2455.
[130] D.M. Bylander and L. Kleinman, Phys. Rev. B29 (1984) 1534.
[131] J.R. Chelikowsky, C.T. Chan and S.G. Louie, Phys. Rev. B34 (1986) 6656.
[132] M.Y. Chou, P.K. Lam and M.L. Cohen, Phys. Rev. B28 (1983) 4179.
[133] E. Jensen, R.A. Bartynski, T. Gustafsson, E.W. Plummer, M.Y. Chou, M.L. Cohen and G.B. Hofland, Phys. Rev. B30 (1984) 5500.
[134] K.J. Chang, S. Froyen and M.L. Cohen, Phys. Rev. B28 (1983) 4736.
[135] S. Froyen and M.L. Cohen, Phys. Rev. B28 (1983) 3258.
[136] D.M. Wood, A. Zunger and R. de Groot, Phys. Rev. B31 (1985) 2570.
[137] F. Manghi, G. Riegler, C.M. Bertoni and G.B. Bachelet, Phys. Rev. B31 (1985) 3680.
[138] W. Andreoni, L. Maschke and M. Schlüter, Phys. Rev. B26 (1982) 2314.
[139] J.R. Chelikowsky and J.K. Burdett, Phys. Rev. Lett. 56 (1986) 961.
[140] J.R. Chelikowsky, Phys. Rev. B35 (1987) 1174.
[141] C.T. Chan and S.G. Louie, Phys. Rev. B27 (1983) 3325.
[142] B.K. Bhattacharyya, D.M. Bylander and L. Kleinman, Phys. Rev. B32 (1985) 7973.
[143] J. Ihm, Rep. Prog. Phys. 51 (1988) 105.
[144] P.J.H. Denteneer, Ph.D. Thesis, Technische Universiteit Eindhoven, 1987 (unpublished). This thesis is available from its author on request.
[145] J.A. Applebaum and D.R. Hamann, Rev. Mod. Phys. 48 (1976) 479.
[146] M.L. Cohen, Physics Today 32 (1979) 40.
[147] M.L. Cohen, in: Electronics and Electron Physics, vol. 51, eds. L. Marton and C. Marton (Academic, New York, 1980) p. 1.
[148] M. Schlüter, Festkörperprobleme (Adv. Solid State Physics) 18 (1978) 155.
[149] M. Schlüter, J. Vac. Sci. Technol. 16 (1979) 1331.
[150] J. Pollmann, Festkörperprobleme (Adv. Solid State Physics) 20 (1980) 117.
[151] S.G. Louie, in: Electronic Structure, Dynamics, and Quantum Structural Properties of Condensed Matter, eds. D.T. Devreese and P. van Camp (Plenum, New York, 1985) p. 335.
[152] J.R. Chelikowsky and M.Y. Chou, Phys. Chem. Minerals 14 (1987) 308.
[153] J.A. Applebaum and D.R. Hamann, Phys. Rev. B6 (1972) 2166.
[154] J.A. Applebaum and D.R. Hamann, Phys. Rev. Lett. 31 (1973) 106.
[155] J.A. Appelbaum and D.R. Hamann, Phys. Rev. Lett. 32 (1974) 225.
[156] J.A. Appelbaum and D.R. Hamann, Phys. Rev. B12 (1975) 1410.
[157] J.A. Appelbaum and D.R. Hamann, Phys. Rev. Lett. 34 (1975) 806.
[158] J.A. Applebaum, G.A. Baraff and D.R. Hamann, Phys. Rev. B12 (1975) 3822.
[159] J.A. Applebaum, G.A. Baraff and D.R. Hamann, Phys. Rev. B12 (1975) 5749.
[160] J.A. Applebaum, G.A. Baraff and D.R. Hamann, Phys. Rev. Lett. 35 (1975) 729.
[161] J.A. Applebaum, G.A. Baraff and D.R. Hamann, Phys. Rev. B14 (1976) 1623.
[162] J.R. Chelikowsky and M.L. Cohen, Phys. Rev. B13 (1976) 826.
[163] S.G. Louie and M.L. Cohen, Phys. Rev. Lett. 35 (1975) 866.
[164] S.G. Louie, J.R. Chelikowsky and M.L. Cohen, J. Vac. Sci. Technol. 13 (1976) 790.
[165] M. Schlüter, K.-M. Ho and M.L. Cohen, J. Vac. Sci. Technol. 13 (1976) 779.
[166] J. Ihm, M.L. Cohen and D.J. Chadi, Phys. Rev. B21 (1980) 4592.
[167] J.E. Northrup, in: Proc. 17th Intern. Conf. on the Physics of Semiconductors, eds. D.J. Chadi and W.A. Harrison (Springer-Verlag, New York, 1985) p. 95.
[168] J.R. Chelikowsky, Phys. Rev. B15 (1977) 3236.
[169] J. Ihm, S.G. Louie and M.L. Cohen, Phys. Rev. B17 (1978) 769.
[170] J. Northrup, J. Ihm and M.L. Cohen, Phys. Rev. B27 (1983) 6553.
[171] D. Vanderbilt and S.G. Louie, Phys. Rev. B30 (1984) 6118.
[172] G.P. Srivastava, J. Phys. C15 (1982) 699.
[173] J.R. Chelikowsky and M.L. Cohen, Phys. Rev. B20 (1979) 4150.
[174] A. Zunger, Phys. Rev. B22 (1980) 959.

[175] F. Manghi, Phys. Rev. B33 (1986) 2554.
[176] S.B. Zhang and M.L. Cohen, Surf. Sci. 172 (1986) 754.
[177] E. Kaxiras, Y. Bar-Yam, J.D. Joannopoulos and K.C. Pandey, Phys. Rev. B35 (1987) 9625.
[178] E. Kaxiras, Y. Bar-Yam, J.D. Joannopoulos and K.C. Pandey, Phys. Rev. B35 (1987) 9636.
[179] J.R. Chelikowsky and M.L. Cohen, J. Vac. Sci. Technol. 16 (1979) 1307.
[180] C.M. Bertoni, F. Manghi and C. Calandra, J. Phys. Soc. Japan 49 (Supp. A) (1980) 1097.
[181] F. Manghi, C.M. Bertoni, C. Calandra and E. Molinari, Phys. Rev. B24 (1981) 6029.
[182] G.P. Srivastava, I. Singh, V. Montgomery and R.H. Williams, J. Phys. C16 (1983) 3627.
[183] F. Manghi, E. Molinari, C.M. Bertoni and C. Calandra, J. Phys. C15 (1982) 1099.
[184] A.C. Ferraz and G.P. Srivastava, Phys. Rev. B34 (1986) 5621.
[185] K.M. Ho, M.L. Cohen and M. Schlüter, Phys. Rev. B15 (1977) 3888.
[186] H.I. Zhang and M. Schlüter, Phys. Rev. B18 (1978) 1923.
[187] J.R. Chelikowsky, D.J. Chadi and M.L. Cohen, Phys. Rev. B23 (1981) 4013.
[188] G.A. Baraff, E.O. Kane and M. Schlüter, Phys. Rev. B25 (1982) 548.
[189] F. Manghi, C. Calandra and E. Molinari, Surf. Sci. 184 (1987) 449.
[190] F. Manghi, C.M. Bertoni, C. Calandra and E. Molinari, J. Vac. Sci. Technol. 21 (1982) 371.
[191] S.G. Louie, K.-M. Ho, J.R. Chelikowsky and M.L. Cohen, Phys. Rev. Lett. 37 (1976) 1289.
[192] S.G. Louie, K.-M. Ho, J.R. Chelikowsky and M.L. Cohen, Phys. Rev. B15 (1977) 5627.
[193] S.G. Louie, Phys. Rev. Lett. 40 (1978) 1525.
[194] G.P. Kerker, K.-M. Ho and M.L. Cohen, Phys. Rev. Lett. 40 (1978) 1593.
[195] G.P. Kerker, K.-M. Ho and M.L. Cohen, Phys. Rev. B18 (1978) 5473.
[196] C.T. Chan and S.G. Louie, Phys. Rev. B33 (1986) 2861.
[197] M.Y. Chou and J.R. Chelikowsky, Phys. Rev. B35 (1987) 2124.
[198] K.-M. Ho and K.P. Bohnen, Phys. Rev. Lett. 56 (1986) 934.
[199] S.G. Louie, Phys. Rev. Lett. 42 (1979) 476.
[200] G.P. Kerker, M.T. Yin and M.L. Cohen, Solid State Commun. 32 (1979) 433.
[201] G.P. Kerker, M.T. Yin and M.L. Cohen, Phys. Rev. B20 (1979) 4940.
[202] C.T. Chan and S.G. Louie, Phys. Rev. B30 (1984) 4153.
[203] W.E. Pickett, S.G. Louie and M.L. Cohen, Phys. Rev. Lett. 39 (1977) 109.
[204] W.E. Pickett, S.G. Louie and M.L. Cohen, Phys. Rev. B17 (1978) 815.
[205] G.A. Baraff, J.A. Appelbaum and D.R. Hamann, Phys. Rev. Lett. 38 (1977) 237.
[206] G.A. Baraff, J.A. Appelbaum and D.R. Hamann, J. Vac. Sci. Technol. 14 (1977) 999.
[207] C.G. Van de Walle and R.M. Martin, J. Vac. Sci. Technol. B4 (1986) 1055.
[208] C.G. Van de Walle and R.M. Martin, Phys. Rev. B35 (1987) 8154.
[209] A.C. Ferraz and G.P. Srivastava, Semicond. Sci. and Technol. 1 (1986) 169.
[210] S. Ciraci and I.P. Batra, Phys. Rev. B36 (1987) 1225.
[211] D.M. Bylander and L. Kleinman, Phys. Rev. B36 (1987) 3229.
[212] D.M. Bylander and L. Kleinman, Phys. Rev. B34 (1986) 5280.
[213] M.A. Gell, D. Ninno, M. Jaros, D.J. Wolford, T.F. Keuch and J.A. Bradley, Phys. Rev. B35 (1987) 1196.
[214] W.E. Pickett and M.L. Cohen, Phys. Rev. B18 (1978) 939.
[215] K. Kunc and R.M. Martin, J. Phys. Soc. Japan 49 Suppl. A (1980) 1117.
[216] K. Kunc and R.M. Martin, Phys. Rev. B24 (1981) 3445.
[217] W.E. Pickett and M.L. Cohen, Solid State Commun. 25 (1978) 225.
[218] J. Ihm and M.L. Cohen, Phys. Rev. B20 (1979) 729.
[219] J. Ihm, P.K. Lam and M.L. Cohen, Phys. Rev. B20 (1979) 4120.
[220] J. Ihm, P.K. Lam and M.L. Cohen, J. Vac. Sci. Technol. 16 (1979) 1512.
[221] M.A. Gell, K.B. Wong, D. Ninno and M. Jaros, J. Phys. C19 (1986) 3821.
[222] C.G. Van de Walle and R.M. Martin, J. Vac. Sci. Technol. B3 (1985) 1256.
[223] C.G. Van de Walle and R.M. Martin, Phys. Rev. B34 (1986) 5621.
[224] C. Calandra, F. Manghi and C.M. Bertoni, Surf. Sci. 162 (1985) 605.
[225] A. Taguchi and T. Ohno, Phys. Rev. B36 (1987) 1696.
[226] J.E. Northrup, J. Ihm and M.L. Cohen, Phys. Rev. B22 (1980) 2060.

[227] G.A. Baraff and M. Schlüter, Phys. Rev. Lett. 41 (1978) 892.
[228] J. Bernholc, N.O. Lipari and S.T. Pantelides, Phys. Rev. Lett. 41 (1978) 895.
[229] J. Bernholc, N.O. Lipari, S.T. Pantelides and M. Scheffler, Phys. Rev. B26 (1982) 5706.
[230] G.B. Bachelet, G.A. Baraff and M. Schlüter, Phys. Rev. B24 (1981) 915.
[231] U. Lindefelt and A. Zunger, Phys. Rev. B24 (1981) 5913.
[232] M.J. Kirton and P.W. Banks, J. Phys. C17 (1984) 2475.
[233] M.J. Kirton, P.W. Banks, Lu Da Lian and M. Jaros, J. Phys. C17 (1984) 2487.
[234] S. Itoh and K. Nakao, J. Phys. Soc. Japan 54 (1985) 4657.
[235] E. Falck, H. Stoll and L.O. Schwan, Solid State Commun. 41 (1982) 565.
[236] M.G. Ramsey and P.V. Smith, J. Phys. F12 (1982) 1697.
[237] J. Kanamori, K. Terakura and K. Yamada, Prog. Theor. Phys. Suppl. 46 (1970) 221.
[238] F. Gautier, J. Phys. F1 (1971) 382.
[239] G.A. Baraff and M. Schlüter, Phys. Rev. B28 (1983) 2296.
[240] G.A. Baraff and M. Schlüter, Phys. Rev. B30 (1984) 1853.
[241] R.W. Jansen, D.S. Wolde-Kidane and O.F. Sankey, J. Appl. Phys. (1988, in press).
[242] A. Zunger, in: Solid State Physics, vol. 39, eds. H. Ehrenreich and D. Turnbull (Academic, New York, 1986) p. 275.
[243] K. Kunc, in: Electronic Structure, Dynamics, and Quantum Structural Properties of Condensed Matter, eds. J.T. Devreese and P. Van Camp (Plenum, New York, 1985) p. 227.
[244] H. Wendel and R.M. Martin, Phys. Rev. B19 (1979) 5251.
[245] H. Wendel and R.M. Martin, Feskörperprobleme (Adv. Solid State Physics) 19 (1979) 23.
[246] J. Ihm, M.T. Yin and M.L. Cohen, Solid State Commun. 37 (1981) 491.
[247] M.T. Yin and M.L. Cohen, Phys. Rev. Lett. 45 (1980) 1004.
[248] K. Kunc and R.M. Martin, Phys. Rev. B24 (1981) 2311.
[249] K.-M. Ho, C.-L. Fu and B.N. Harmon, Phys. Rev. B29 (1984) 1575.
[250] Y. Chen, C.-L. Fu, K.-M. Ho and B.N. Harmon, Phys. Rev. B31 (1985) 6775.
[251] Y.-Y. Ye, Y. Chen, K.-M. Ho, B.N. Harmon and P.-A. Lindgard, Phys. Rev. Lett. 58 (1987) 1769.
[252] H.-J. Tao, K.-M. Ho and X.-Y. Zhu, Phys. Rev. B34 (1986) 8394.
[253] M.T. Yin and M.L. Cohen, Phys. Rev. B26 (1982) 3259.
[254] H. Hellman, Einführung in die Quanten Theorie (Deuticke, Leipzig, 1937) p. 285.
[255] R.P. Feynman, Phys. Rev. 56 (1939) 340.
[256] W.E. Pickett, Phys. Rev. B26 (1982) 5650.
[257] P. Pulay, Mol. Phys. 17 (1969) 197.
[258] P. Bendt and A. Zunger, Phys. Rev. Lett. 50 (1983) 1684.
[259] K. Kunc and R.M. Martin, Phys. Rev. Lett. 48 (1982) 406.
[260] M. Cardona, K. Kunc and R.M. Martin, Solid State Commun. 44 (1982) 1205.
[261] P.K. Lam, M.M. Dacorogna and M.L. Cohen, Phys. Rev. B34 (1986) 5065.
[262] M.M. Dacorogna, M.L. Cohen and P.K. Lam, Phys. Rev. B34 (1986) 4865.
[263] M.M. Dacorogna, M.L. Cohen and P.K. Lam, Phys. Rev. Lett. 55 (1985) 837.
[264] K.J. Chang and M.L. Cohen, Phys. Rev. B34 (1986) 4552.
[265] M.M. Dacorogna, K.J. Chang and M.L. Cohen, Phys. Rev. B32 (1985) 1853.
[266] K.J. Chang, M.M. Dacorogna, M.L. Cohen, J.M. Mignot, C. Chouteau and G. Martinez, Phys. Rev. Lett. 54 (1985) 2375.
[267] D. Erskine, P.Y. Yu, K.J. Chang and M.L. Cohen, Phys. Rev. Lett. 57 (1986) 274.
[268] J.L. Martins and M.L. Cohen, Phys. Rev. B37 (1988) 3304.
[269] M.T. Yin, Phys. Rev. B27 (1983) 7769.
[270] O.H. Nielsen and R.M. Martin, Phys. Rev. B32 (1985) 3780.
[271] O.H. Nielsen and R.M. Martin, Phys. Rev. B32 (1985) 3792.
[272] O.H. Nielsen and R.M. Martin, Phys. Rev. Lett. 50 (1983) 697.
[273] O.H. Nielsen, Phys. Rev. B34 (1986) 5808.
[274] R.J. Needs, Phys. Rev. Lett. 58 (1987) 53.
[274a] D. Vanderbilt, Phys. Rev. Lett. 59 (1987) 1456.

[275] P. Gomes Dacosta, O.H. Nielsen and K. Kunc, J. Phys. C19 (1986) 3163.
[276] J.P. Walter and M.L. Cohen, Phys. Rev. B2 (1970) 1821.
[277] R. Car and A. Selloni, Phys. Rev. Lett. 42 (1979) 1365.
[278] Z.H. Levine and S.G. Louie, Phys. Rev. B25 (1982) 6310.
[279] R. Car, E. Tosatti, S. Baroni and S. Leelaprute, Phys. Rev. B24 (1981) 985.
[280] A. Milchev, Phys. Stat. Sol. B90 (1978) 679.
[281] A.C. Sharma and S. Auluck, J. Phys. C16 (1984) L1233.
[282] S.G. Louie, J.R. Chelikowksy and M.L. Cohen, Phys. Rev. Lett. 34 (1975) 155.
[283] A. Balderseschi and E. Tosatti, Solid State Commun. 29 (1979) 131.
[284] R. Resta and A. Baldereschi, Phys. Rev. B23 (1981) 6615.
[285] M.S. Hybertsen and S.G. Louie, Phys. Rev. B35 (1987) 5585.
[286] A. Baldereschi and E. Tosatti, Phys. Rev. B17 (1978) 4710.
[287] S. Baroni and R. Resta, Phys. Rev. B33 (1986) 7017.
[288] M.S. Hybertsen and S.G. Louie, Phys. Rev. B35 (1987) 5602.
[289] J.B. McKitterick, Phys. Rev. B28 (1983) 7384.
[290] K. Kunc and R.M. Martin, Phys. Rev. Lett. 48 (1982) 406.
[291] K. Kunc and R. Resca, Phys. Rev. Lett. 51 (1983) 686.
[292] K. Kunc and E. Tosatti, Phys. Rev. B29 (1984) 7045.
[293] A. Fleszar and R. Resta, Phys. Rev. B31 (1985) 5305.
[294] S. Baroni, P. Giannozzi and A. Testa, Phys. Rev. Lett. 58 (1987) 1861.
[294a] S. Baroni, P. Giannozzi and A. Testa, Phys. Rev. Lett. 59 (1987) 2662.
[295] J.P. Perdew and M. Levy, Phys. Rev. Lett. 51 (1983) 1884.
[296] L.J. Sham and M. Schlüter, Phys. Rev. Lett. 51 (1983) 1888.
[297] P.A. Sterne and W.E. Pickett (unpublished).
[298] G.B. Bachelet and N.E. Christensen, Phys. Rev. B31 (1985) 879.
[299] W.E. Pickett and C.S. Wang, Phys. Rev. B30 (1984) 4719.
[300] C.S. Wang and W.E. Pickett, Phys. Rev. Lett. 51 (1983) 597.
[301] W.E. Pickett and C.S. Wang, Int. J. Quant. Chem.: QC Symp. 20 (1986) 299.
[302] P.A. Sterne and C.S. Wang, Phys. Rev. B27 (1988) 10436.
[303] M.S. Hybertsen and S.G. Louie, Phys. Rev. B34 (1986) 5390.
[304] M.S. Hybertsen and S.G. Louie, in Proc. 17th Intern. Conf. on the Physics of Semiconductors, San Francisco, 1984, eds. D.J. Chadi and W.A. Harrison (Springer, New York, 1985), p. 1001.
[305] M.S. Hybertsen and S.G. Louie, Phys. Rev. Lett. 55 (1985) 1418.
[306] M.S. Hybertsen and S.G. Louie, Phys. Rev. B32 (1985) 7005.
[307] J.E. Northrup, M.S. Hybertsen and S.G. Louie, Phys. Rev. Lett. 59 (1987) 819.
[308] M.S. Hybertsen and S.G. Louie, Phys. Rev. B37 (1988) 2733.
[309] R.W. Godby, M. Schlüter and L.J. Sham, Phys. Rev. B36 (1987) 6497.
[310] L. Guttman and C.Y. Fong, Phys. Rev. B26 (1982) 6756.
[311] J.S. Nielsen, C.Y. Fong, L. Guttman and I.P. Batra, Phys. Rev. B37 (1988) 2622.
[312] M.Y. Chou, M.L. Cohen and S.G. Louie, Phys. Rev. B32 (1985) 7979.
[313] P.J.H. Denteneer and W. van Haeringen, J. Phys. C20 (1987) L883.
[314] D.P. DiVincenzo, O.L. Alerhand, M. Schlüter and J.W. Wilkins, Phys. Rev. Lett. 56 (1986) 1925.
[315] A. Qteish and R. Resta, Phys. Rev. B37 (1988) 1308.
[316] L. Goodwin, R.J. Needs and V. Heine, Phys. Rev. Lett. 60 (1988) 2050.
[317] R. Car and M. Parrinello, Phys. Rev. Lett. 55 (1985) 2471.
[318] S. Kirkpatrick, C.D. Gelatt, Jr. and M.P. Vecchi, Science 220 (1983) 671.
[319] R. Car and M. Parrinello, Solid State Commun. 62 (1987) 403.
[320] E.R. Davidson, J. Comp. Phys. 17 (1975) 87.
[321] E.R. Davidson, in: Methods in Computational Molecular Physics, eds. G.H.F. Diercksen and S. Wilson (Reidel, Dordrecht, 1983) p. 95.
[322] D.M. Wood and A. Zunger, J. Phys. A18 (1985) 1343.
[322a] C.M.M. Nex, J. Comp. Phys. 70 (1987) 138.

[322b] R. Natarajan and D. Vanderbilt, J. Comp. Phys. (in press).
[323] M.C. Payne, J.D. Joannopoulos, D.C. Allan, M.P. Teter and D.H. Vanderbilt, Phys. Rev. Lett. 56 (1986) 2656.
[324] D.C. Allan and M.P. Teter, Phys. Rev. Lett. 59 (1987) 1136.
[325] M.C. Payne, P.D. Bristowe and J.D. Joannopoulos, Phys. Rev. Lett. 58 (1987) 1348.
[326] G.B. Bachelet and G. De Lorenzi, Physica Scripta T19A (1987) 311.
[327] M. Needels, M.C. Payne and J.D. Joannopoulos, Phys. Rev. Lett. 58 (1987) 1765.
[328] M.C. Payne, M. Needels and J.D. Joannopoulos, Phys. Rev. B37 (1988) 8138.
[329] L. Hedin and B.I. Lundqvist, J. Phys. C4 (1971) 2064.
[330] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. 45 (1980) 566.
[331] S.H. Vosko, L. Wilk and M. Nusair, Can. J. Phys. 58 (1980) 1200.
[332] L. Wilk and S.H. Vosko, J. Phys. C15 (1982) 2139.
[332a] J.P. Perdew and A. Zunger, Phys. Rev. B23 (1981) 5048.
[333] E.P. Wigner, Phys. Rev. 46 (1934) 1002.
[334] D. Pines, Solid State Phys. 1 (1955) 367.
[335] S. Blügel, Ph.D. Thesis, Kernforschungsanlage Jülich GmbH (1988), unpublished.