Real-Space Approach to Calculation of Electric Polarization and Dielectric Constants

R. W. Nunes and David Vanderbilt

Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08855-0849

We describe a real-space approach to the calculation of the properties of an insulating crystal in an applied electric field, based on the iterative determination of the Wannier functions (WF's) of the occupied bands. It has been recently shown that a knowledge of the occupied WF's allows the calculation of the spontaneous (zero-field) electronic polarization. Building on these ideas, we describe a method for calculating the electronic polarization and dielectric constants of a material in nonzero field. The method is demonstrated for a one-dimensional tight-binding Hamiltonian.

Modern electronic band-structure and total-energy methods, such as those based on the local-density approximation (LDA) to density-functional theory [1], are not well suited for studying the properties of materials in an applied electric field. These approaches typically rely on solving for the eigenfunctions of the effective one-electron Hamiltonian. Unfortunately, the electric field acts as a singular perturbation, so that the eigenstates of the Hamiltonian are no longer Bloch states, and the electron band structure is destroyed, even for arbitrarily small applied fields. Linear-response methods [2] are capable of computing derivatives of various quantities with respect to the applied field, but cannot be used to study the electronic structure of a crystal in a nonzero electric field directly. Extensions of these methods beyond linear response [3] consist of rather involved expressions which must be carefully handled in order to avoid divergences in the static limit. Formal studies of the structure of the Wannier-Stark ladder [4] are instructive, but are not of much practical help from a computational point of view.

Recently, several groups have introduced new real-space approaches to the solution of the electronic structure problem. Based either on the locality of the real-space density matrix [5,6] or the use of a localized, Wannier-like representation of the occupied subspace [7,8], these methods were motivated largely by the search for so-called "order-N methods" (for which the computational effort scales only linearly with system size N). However, these methods also hold promise for application to the electric-field problem. The methods based on a Wannier-like representation [7,8] appear particularly promising in view of recent work showing that the electric polarization of a solid can be directly related to the centers of charge of the Wannier functions (WF's) [9].

In this paper, we show that the real-space method of Mauri, Galli, and Car (MGC) [7] can be developed naturally into a practical method for calculating the response of an insulator to an electric field. Both the spontaneous and induced electric polarization **P** are easily calculated, as are the perturbed charge density and polarization energy, and dielectric constants can be obtained by finite differences. The WF's are expanded in a local basis and

truncated to zero beyond a real-space cutoff radius R_c, these being the only approximations involved. We apply the method to a one-dimensional (1D) tight-binding (TB) Hamiltonian and find that both the spontaneous polarization and dielectric constant converge quickly, with respect to R_c, to the values obtained by standard k-space techniques [2,9]. While the method is demonstrated in a simple tight-binding context, we see no obvious obstacles to its implementation in a fully self-consistent *ab initio* LDA calculation.

We first summarize the MGC scheme using a density-matrix description. The exact density matrix of an insulating crystal is

$$\rho_{\text{exact}} = \sum_{n\mathbf{k}} |\psi_{n\mathbf{k}}\rangle\langle\psi_{n\mathbf{k}}|, \tag{1}$$

where the $\{\psi_{n\mathbf{k}}\}$ are occupied Bloch eigenstates of the one-electron (e.g., LDA) Hamiltonian. For insulators, we can define a unitary transformation from the $\{\psi_{n\mathbf{k}}\}$ to a set of localized (Wannier-like) functions $\{\psi_j\}$ such that the occupied subspace is invariant under the transformation. In terms of these localized orbitals the density matrix is

$$\rho_{\text{exact}} = \sum_{j}^{M} |\psi_j\rangle\langle\psi_j|, \tag{2}$$

where j runs over a set of localized orbitals that span the M-dimensional occupied subspace of Bloch eigenstates. The density matrix is spatially localized, i.e., $\rho(\mathbf{r},\mathbf{r}') \to 0$ as $|\mathbf{r} - \mathbf{r}'| \to \infty$, the decay being exponential for an appropriate choice of phases of the Bloch functions [10].

In the MGC scheme, Eq. (2) is taken as an ansatz for a *trial* density matrix, ρ. Each ψ_j is free to vary within a real-space localization region (LR) of radius R_c, and is set to zero outside the LR. No orthonormalization constraint is imposed. The *physical* density matrix is then defined by

$$\tilde{\rho} = \rho \sum_{k=0}^{K} (I - \rho)^k, \tag{3}$$

where K is odd. The expectation value of any operator A is then given by $\text{tr}[\tilde{\rho}A]$. In the limit where ρ is idempotent we have $\tilde{\rho} = \rho$, as can be seen from Eq. (3). In general, the variational ρ is not exactly idempotent,

due to the approximations involved. The parameter K controls the accuracy of the idempotency requirement [7]. Note that the set $\{\psi_j\}$ in Eq. (2) spans an M-dimensional subspace and that $\tilde{\rho}$ spans the same subspace, since the function $f_K(x) = 1 - (1 - x)^{K+1}$, corresponding to Eq. (3), is such that $f_K(0) = 0$. Moreover, $f_K(x)$ has its absolute maximum at $x = 1$. From these properties it follows that: (i) when $\rho = \rho_{\text{exact}} + O(\delta)$ then $\tilde{\rho} = \rho_{\text{exact}} + O(\delta^2)$ ($|\delta| \ll 1$); (ii) the total energy, $\text{tr}[\tilde{\rho}H]$, is variational provided that all eigenvalues of the Hamiltonian H are negative.

In what follows, we restrict ourselves to $K = 1$, so that $\tilde{\rho} = 2\rho - \rho^2$. The variational total energy is given by $E[\{\psi\}] = \text{tr}[\rho(2 - \rho)H]$, where H has been shifted to make all eigenvalues negative [11]. This functional is minimized with respect to the coefficients of the wave functions $\{\psi_j\}$ in the given basis. In the limit $R_c \to \infty$, minimization of this functional yields a set of orthonormal orbitals that exactly span the occupied subspace of H, and consequently $E[\{\psi\}]_{\min} = E_{\text{GS}}$ (the exact ground-state energy). For a finite R_c the resulting $\{\psi_j\}$ are not exactly orthonormal and $E[\{\psi\}]_{\min} \geq E_{\text{GS}}$, as a result of the variational nature of the functional.

We next propose a generalized formulation of the MGC scheme which is more natural and efficient in the case of periodic systems with small unit cells. In its original formulation, the MGC scheme is well suited to solving problems involving large supercells, where a minimization of a $\mathbf{k}$-dependent energy functional, $E^{\mathbf{k}}[\{\psi\}]$, is typically performed only at $\mathbf{k} = 0$. When applied to a periodic system with a small unit cell, one would have to minimize $E^{\mathbf{k}}[\{\psi\}]$ independently on a mesh of $\mathbf{k}$ points. In either case, the orbitals $\{\psi_j\}$ are not truly localized, but rather are Bloch functions of the cell or supercell. Instead, we propose to let the wave functions be truly localized in the manner of Wannier functions (WF's), $\psi_j = w_{l,n}$ (l and n are cell and occupied-band indices, respectively), and work directly with the original Hamiltonian. In this formulation, $\mathbf{k}$ plays no role. The electronic degrees of freedom are the coefficients of the WF's $\{w_{0,n}\}$ in one cell. The periodicity of the system is now taken into account by introducing the periodic images of these WF's in neighboring cells,

$$|w_{l,n}\rangle = T_l|w_{0,n}\rangle, \tag{4}$$

where T_l is the translation operator corresponding to lattice vector $\mathbf{R}_l$. In terms of the density matrix, $\rho = \sum_{l,n} |w_{l,n}\rangle\langle w_{l,n}|$, the periodicity of the electronic state is expressed as

$$\rho(\mathbf{r}, \mathbf{r}') = \rho(\mathbf{r} + \mathbf{R}_l, \mathbf{r}' + \mathbf{R}_l). \tag{5}$$

The total energy per unit cell can be written explicitly as

$$E_0[\{w\}] = 2\sum_{n=1}^{M} \langle w_{0,n}|H_0|w_{0,n}\rangle$$
$$- \sum_{n,m=1}^{M} \sum_l \langle w_{0,n}|w_{l,m}\rangle\langle w_{l,m}|H_0|w_{0,n}\rangle, \tag{6}$$

where M is the number of occupied bands. E_0 and H_0 denote the total energy and the Hamiltonian at zero electric field. Since we are dealing with localized orbitals, only a finite number of orbitals will contribute to the second sum in Eq. (6). Minimization of E_0 yields M approximate WF's, for a given choice of R_c. Note that Eq. (6) reduces to the original MGC scheme if the sum in the last term is restricted to $l = 0$ only.

It is thus seen that, when applied to a periodic system, the generalized MGC scheme is equivalent to the direct determination of its WF's. This brings us to the work of Ref. [9], where it is shown that the electronic contribution to the spontaneous polarization of a periodic system can be written as $\mathbf{P}_e = -(e/\Omega)\sum_n \mathbf{r}_n$, where $\mathbf{r}_n$ is the center of charge of the WF of band n in zero field, e is the electronic charge, and Ω is the unit cell volume. In the present formalism, this becomes

$$\mathbf{P}_e = -\frac{e}{\Omega}\left[2\sum_{n=1}^{M} \langle w_{0,n}|\mathbf{r}|w_{0,n}\rangle \right.$$
$$\left. - \sum_{n,m=1}^{M}\sum_l \langle w_{0,n}|w_{l,m}\rangle\langle w_{l,m}|\mathbf{r}|w_{0,n}\rangle \right]. \tag{7}$$

At this point we have already completed the necessary steps for a real-space computation of the spontaneous polarization of a crystalline insulator, by combining Eq. (7) with our modified MGC scheme.

Now, we extend the relation between polarization and the centers of charge of the WF's to a periodic system in an electric field $\mathbf{F}$. The Hamiltonian becomes

$$H = H_0 + e\mathbf{F} \cdot \mathbf{r}. \tag{8}$$

Replacing H_0 by H in Eq. (6) leads to the total-energy functional

$$E[\{w^F\}] = E_0[\{w^F\}] - e\mathbf{F} \cdot \mathbf{P}_e[\{w^F\}]. \tag{9}$$

We retain Eq. (4) and hence Eq. (5), and minimize this functional subject only to the constraint of localization of the field-dependent WF's $\{w_n^F\}$ to the LR, as before. Because of the locality of the WF's, the additional terms which enter Eq. (9) through Eq. (7) do not add appreciably to the computational effort.

We emphasize that our solution does not correspond to the true ground state of the system. (There is no true ground state in finite field, as the energy can always be lowered by transferring charge from "valence-band" states in one region to lower-energy "conduction-band" states in a distant region.) We can think of our solution heuristically as the one which is generated from the zero-field state by adiabatically turning on $\mathbf{F}$, and keeping the periodicity of the electronic state expressed in Eq. (5). This should be very closely related to what is done when the field-dependent response of the crystal is measured experimentally. In order to measure "static" properties, the field must be turned on (or allowed to oscillate) on a time scale that is slow compared to usual electronic processes, but fast compared to the

characteristic electronic tunneling rate at the maximum field encountered. Thus, the experimental object of study is also really an excited state (more properly, a very narrow resonance) of the finite-field Hamiltonian. Our solution has a very similar interpretation.

One might object that the existence of a functional $E_0[\{w\}]$, as in Eq. (6), relies on the identification of bands. In the presence of an electric field, where all gaps disappear [4], the existence of bands is not obvious. Nevertheless, Nenciu [4] has shown that one can define a sequence of periodic Hamiltonians $\{H_q\}$, constructed from H by projecting out the nonperiodic part of $\mathbf{F} \cdot \mathbf{r}$, such that the subspace of occupied states of a given H_q reduces to that of H_0 in the limit $\mathbf{F} \to 0$. It is argued that the "bands" defined by the Hamiltonian H_q provide an increasingly accurate description of the finite-field electronic state as the integer index q is increased. (Eventually, as q gets too large, the behavior diverges, in the usual manner of asymptotic perturbation theory. In other words, the radius of convergence F_q tends to zero as $q \to \infty$.) At least at small $\mathbf{F}$, our $\{w^F\}$ are presumably very similar to the Wannier functions that would be constructed from the bands of H_q, and thus should give a good description of the experimental electronic state of the system. We return to this point below.

Before turning to our tests and results, we mention one technicality. The minimization of the functional of Eq. (9) can be preformed directly at fixed $\mathbf{F}$ using steepest-descent or conjugate-gradient techniques. This appears to work quite well at weak fields, but can become unstable for strong fields. Alternatively, one can perform the minimization with a constraint of fixed $\langle \mathbf{r} \rangle$ (i.e., fixed $\mathbf{P}$), treating $\mathbf{F}$ as an adjustable Lagrange multiplier. Since $\langle H \rangle - \mathbf{F} \cdot \langle \mathbf{r} \rangle$ defines a Lagrange transformation from $\langle H \rangle$ to $\langle H_0 \rangle$, it follows that $\nabla_{\langle \mathbf{r} \rangle} \langle H_0 \rangle = -\mathbf{F}$. In this way, the function $\mathbf{F}(\mathbf{P})$ can be mapped out, and inverted numerically to give $\mathbf{P}(\mathbf{F})$. The latter approach is more appropriate for investigating the strong-field behavior of the solutions.

We apply our scheme to a 1D tight-binding three-band Hamiltonian in which each unit cell consists of three atoms with one orbital per atom,

$$H(\alpha) = \sum_j \left\{ \epsilon_j(\alpha) c_j^\dagger c_j + t \left[c_j^\dagger c_{j+1} + \text{H.c.} \right] \right\}, \quad (10)$$

with the site energy given by $\epsilon_{3m+k}(\alpha) = \Delta \cos(\alpha - \beta_k)$. Here m is the cell index, $k = \{-1, 0, 1\}$ is the site index, and $\beta_k = 2\pi k/3$. This is a simple model of a sliding commensurate charge-density wave which slides by one period as the parameter α evolves through 2π. For our tests we set $e = 1$ and $t = \Delta = -1$, and use $x = \sum_j x_j c_j^\dagger c_j$ with $x_j = j/3$ for the TB position operator. We discuss the results obtained with only the lowest band filled; the discussion applies equally to the case of two filled bands (the three-filled-band case is trivial).

We first consider the TB Hamiltonian with no external field, and calculate the spontaneous polarization as a function of α, comparing with the results obtained by the method of Ref. [9] which we take as exact for the sake of comparison. The results are shown in Fig. 1(a). In the inset we show the exact polarization as a function of α in the interval $[0°, 120°]$; also plotted are the results for $|P_{\text{exact}}(\alpha) - P(\alpha)|$ in the same interval, for $R_c = 1.5$ (9 sites within LR) and $R_c = 2.5$ (15 sites). The LR was kept centered at the origin for all values of α. Convergence is already very good for $R_c = 1.5$ with a maximum percentage error of $\sim 1.5\%$, dropping to $\sim 0.5\%$ for $R_c = 2.5$.

We now apply an external electric field to the system. We minimize Eq. (9) for six fixed values of F between 0.01 and 0.06. In this field region the polarization P is linear with F, to a very good approximation. In Fig. 1(b) we show the linear dielectric susceptibility χ as a function of α for $R_c = 1.5$, $R_c = 2.5$, and $R_c = 3.5$ (21 sites). Also shown are the linear-response results we obtained using the method of Ref. [2]. χ converges less rapidly than P, but the maximum percentage error for $\alpha \in [0°, 120°]$ is already $\sim 30\%$ for $R_c = 3.5$. Convergence is systematically worse around $\alpha = 60°$ where the WF's are least localized, due to the fact that the gap between the
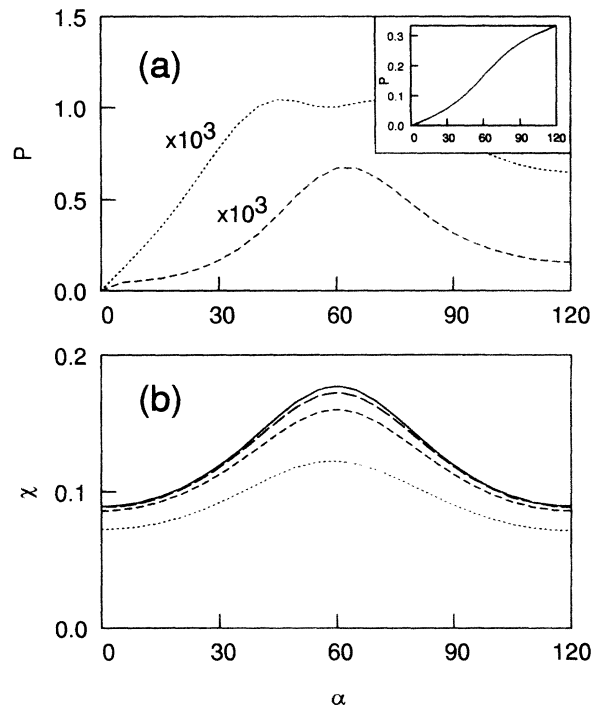

FIG. 1. Polarization $P(\alpha)$ and linear dielectric susceptibility $\chi(\alpha)$ for the tight-binding model of the text. (a) $|P(\alpha) - P_{\text{exact}}(\alpha)|$ for $R_c = 2.5$ (dashed) and $R_c = 1.5$ (dotted line). Inset: $P_{\text{exact}}(\alpha)$. (b) $\chi(\alpha)$ from linear response (solid line) and from current method with $R_c = 3.5$ (long dashed), $R_c = 2.5$ (dashed), and $R_c = 1.5$ (dotted line).
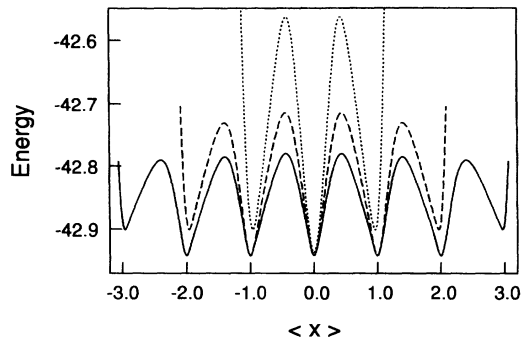
FIG. 2. $\langle H_0 \rangle$ as a function of $\langle x \rangle$, for $\alpha = 0$. Solid line, $R_c = 3.5$; dashed line, $R_c = 2.5$; dotted line, $R_c = 1.5$.

two lowest bands reaches its minimum value of 0.814 at $\alpha = 60°$.

Next, we minimize Eq. (9) keeping $\langle x \rangle$ fixed. For a given value of R_c we explore the interval $[-R_c, R_c]$ of possible values of $\langle x \rangle$. Figure 2 shows $E_0 = \langle H_0 \rangle$ as a function of $P = \langle x \rangle$ for $\alpha = 0$ and $R_c = 1.5, 2.5,$ and 3.5. For unconstrained WF's we expect E_0 to be periodic with P. Our results reproduce that behavior remarkably well, except near the boundaries of the LR where the variational solution becomes poor. Note also that for a given value of P, the value of E_0 diminishes with increasing R_c, reflecting the increasing quality of the variational solution as R_c increases.

We point out that it should be straightforward to obtain the higher-order dielectric constants from the $E_0(\mathbf{P})$ curve, by performing careful finite difference calculations. Moreover, the generalization of our approach to an *ab initio* LDA calculation should be easily implemented. A localized-orbital basis would be ideally suited, although a plane-wave basis might also be used [7,8]. As in the original MGC scheme, self-consistency can be included in a straightforward manner, since the density $n(\mathbf{r}) = \bar{\rho}(\mathbf{r}, \mathbf{r})$ remains periodic. Thus the Hartree and exchange-correlation terms can be computed as usual and do not contribute to the nonperiodic part of the self-consistent potential.

A final word of caution is in order. It should not be imagined that there is a well-defined curve $E_0(P)$, periodic in P, which can be obtained by taking the limit $R_c \to \infty$ of our procedure. On the contrary, as the LR grows very large, it becomes possible to construct a solution for w having arbitrary $\langle x \rangle$ (i.e., arbitrary P), and energy arbitrarily close to $E_0(F = 0)$. This can be done by starting with the zero-field WF of the occupied band and admixing a small amplitude (of order $l^{-1/2}$) of a zero-field unoccupied-band WF at a distance l; its energy approaches $E_0(F = 0)$ as $l \to \infty$. Thus, we have the pathological situation that $E_0(P)$ becomes perfectly flat in the limit $R_c \to \infty$. [When working at fixed F, this pathology shows up in the form of a growing number of false local minima of the functional of Eq. (9) as $R_c \to$

∞.] However, the Taylor coefficients of $E_0(P)$ (expanded about the minimum) are well behaved in the limit $R_c \to \infty$, even as its radius of convergence is decreasing to zero. Underlying this behavior is the asymptotic nature of the expansion, which is also the case for the Nenciu [4] construction of the "polarized" subspaces. In both cases, full convergence as $R_c \to \infty$ or $q \to \infty$ is obtained only in the limit $\mathbf{F} \to 0$. We do not claim that our proposed method has superior convergence or gives more physical results than that of Ref. [4] as $\mathbf{F}$ get large. But both approaches must have the same small-field behavior, and hence yield the correct perturbation coefficients (e.g., linear and nonlinear dielectric constants). The proposed method also has the advantages of being computationally tractable and convenient to implement.

In conclusion, we propose a method for calculating the response of an insulator to an applied electric field based on a Wannier-function-like representation of the electronic orbitals. In this approach the spontaneous polarization, the perturbed charge density, and the polarization energy are easily obtained, and dielectric constants can be calculated by finite differences. The method is variational, and therefore is well suited to solution by iterative techniques such as conjugate gradients. The computational effort scales only linearly with system size and the method becomes exact as the cutoff radius used to truncate the Wannier functions is increased. The method is demonstrated in a simple tight-binding context, but is also well suited to implementation in a fully self-consistent *ab initio* LDA calculation.

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

[2] H. Ehrenreich and M. H. Cohen, Phys. Rev. **115**, 786 (1959).

[3] Z. H. Levine, Phys. Rev. B **42**, 3567–3577 (1990); **44**, 5981 (1991).

[4] G. Nenciu, Rev. Mod. Phys. **63**, 91 (1991), and references therein.

[5] X.-P. Li, R. W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).

[6] M. Daw, Phys. Rev. B **47**, 10895 (1993).

[7] F. Mauri, G. Galli, and R. Car, Phys. Rev. B **47**, 9973 (1993); F. Mauri and G. Galli (unpublished).

[8] P. Ordejon, D. A. Drabold, M. P. Grumbach, and R. Martin, Phys. Rev. B **48**, 14646 (1993).

[9] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993); D. Vanderbilt and R. D. King-Smith, *ibid.* **48**, 4442 (1993).

[10] W. Kohn, Phys. Rev. **115**, 809 (1959); Phys. Rev. B **7**, 4388 (1973).

[11] In practice it is sufficient to make all *occupied* eigenvalues negative, in which case the exact solution is a local but not a global minimum of $\mathrm{tr}[\bar{\rho}H]$.