Macroscopic polarization in crystalline dielectrics: the geometric phase approach

Raffaele Resta

Scuola Internazionale Superiore di Studî Avanzati (SISSA), via Beirut 4, I-34014 Trieste, Italy

The macroscopic electric polarization of a crystal is often defined as the dipole of a unit cell. In fact, such a dipole moment is ill defined, and the above definition is incorrect. Looking more closely, the quantity generally measured is *differential* polarization, defined with respect to a "reference state" of the same material. Such differential polarizations include either derivatives of the polarization (dielectric permittivity, Born effective charges, piezoelectricity, pyroelectricity) or finite differences (ferroelectricity). On the theoretical side, the differential concept is basic as well. Owing to continuity, a polarization difference is equivalent to a macroscopic current, which is directly accessible to the theory as a bulk property. Polarization is a quantum phenomenon and cannot be treated with a classical model, particularly whenever delocalized valence electrons are present in the dielectric. In a quantum picture, the current is basically a property of the *phase* of the wave functions, as opposed to the charge, which is a property of their modulus. An elegant and complete theory has recently been developed by King-Smith and Vanderbilt, in which the polarization difference between any two crystal states—in a null electric field—takes the form of a geometric quantum phase. The author gives a comprehensive account of this theory, which is relevant for dealing with transverse-optic phonons, piezoelectricity, and ferroelectricity. Its relation to the established concepts of linear-response theory is also discussed. Within the geometric phase approach, the relevant polarization difference occurs as the circuit integral of a Berry connection (or "vector potential"), while the corresponding curvature (or "magnetic field") provides the macroscopic linear response.

CONTENTS

I. INTRODUCTION

Macroscopic electric polarization is a fundamental concept in the physics of matter, upon which the phenomenological description of dielectrics is based (Landau and Lifshitz, 1984). Notwithstanding, this concept has long evaded even a precise microscopic definition. A typical incorrect statement—often found in textbooks—is that the macroscopic polarization of a solid is the dipole of a unit cell. It is easy to realize that such a quantity is neither measurable nor model-independent: the dipole of a periodic charge distribution is in fact ill defined (Martin, 1974), except in the extreme Clausius-Mossotti model, in which the total charge is unambiguously decomposed into an assembly of *localized* and *neutral* charge distributions.

One can adopt an alternative viewpoint by considering a macroscopic and finite piece of matter and defining its polarization $\mathbf{P}$ as the dipole per unit volume:

$$\mathbf{P} = \frac{1}{\mathcal{V}} \left[-e \sum_l Z_l \mathbf{R}_l + \int d\mathbf{r}\, \mathbf{r} \rho(\mathbf{r}) \right], \qquad (1)$$

where e is the electron charge, $\mathcal{V}$ is the sample volume, the l summation is over the ionic sites, $-eZ_l$ are the bare ionic charges, and $\rho(\mathbf{r})$ is the electronic charge density. Although such a dipole is in principle well defined, $\mathbf{P}$ is *not* a bulk property, being dependent upon truncation and shape of the sample. The key point is that the *variations* of $\mathbf{P}$ are indeed measured as bulk material properties in several circumstances.

Some macroscopic physical properties are just derivatives of $\mathbf{P}$ with respect to suitably chosen perturbations. This is the case for dielectric permittivity, piezoelectricity, effective charges (for lattice dynamics), and pyroelectricity, which are phenomenologically measured as bulk material tensors. As for ferroelectric materials, they are known to sustain a spontaneous polarization $\mathbf{P}$, which persists at null field; but again the quantity measured—via hysteresis cycles—is only the *difference* $\Delta\mathbf{P}$ between two enantiomorphous metastable states of the crystal (see, for example, Lines and Glass, 1977). From the theoretical side, I wish to stress three fundamental concepts. First: Any Clausius-Mossotti-like approach does not apply, particularly in materials where delocalized covalent charge is present (see Sec. II). Second: It is the occurrence of *differences*—at the very level of definition—that makes polarization accessible to quantum-mechanical calculations, as shown in the work of Posternak *et al.* (1990; for subsequent discussions, see also Resta *et al.*, 1990; Tagantsev, 1991, 1992;

 0034-6861/94/66(3)/899(17)/$06.70

Baldereschi *et al.*, 1992; Resta, 1992). Third: The electronic *wave functions*—as opposed to the charge—of the polarized crystal *do* contain the relevant information, as is demonstrated from several linear-response calculations of macroscopic tensor properties which have been performed over the years (see Sec. VIII for a review).

I present here a comprehensive account of a modern theory of macroscopic polarization in crystalline dielectrics, which elucidates the fundamental quantum nature of the phenomenon. The scope of this work is limited to cases in which the polarization is due to a source *other* than an "external" electric field; a zero-temperature framework is furthermore adopted, in which the ionic positions are "frozen." The present formulation applies therefore mainly to lattice dynamics, piezoelectricity, and ferroelectricity. Even when the polarization of the solid is not *due to* an electric field—as in the above-mentioned cases—the polarization may (or may not) be *accompanied by* a field, depending on the boundary conditions chosen for the macroscopic sample. The formulation given here concerns the polarization *in a null field*. In the case of lattice dynamics the theory applies to transverse-optic zone-center phonons, whose polarization is measured by the Born (or transverse) effective charge tensors.

According to the present viewpoint, the basic quantity of interest is the difference $\Delta\mathbf{P}$ in polarization between two different states of the same solid; this quantity is obtained from a formulation whose only ingredients are the ground-state electronic wave functions of the crystal in the two states. The first step towards a theory of polarization was made by Resta (1992), who cast $\Delta\mathbf{P}$ as an integrated macroscopic current. New avenues were then opened by the historic contribution of King-Smith and Vanderbilt (1993), who identified in $\Delta\mathbf{P}$ a geometric quantum phase (Berry, 1984, 1989). Besides being very elegant, such an approach is extremely powerful on computational grounds, as has been demonstrated in some calculations for real materials (King-Smith and Vanderbilt, 1993; Dal Corso *et al.*, 1993b; Resta *et al.*, 1993a, 1993b). I present these recent findings from a slightly different perspective, developing the formulation along a different logical path from that of the original King-Smith and Vanderbilt paper. In full analogy with other geometric phase problems (Berry, 1984; Jackiw, 1988), I define a "connection" (gauge-dependent, nonobservable) and its generalized curl, the "curvature" (gauge-invariant, observable). These two quantities play the same role as the ordinary vector potential and magnetic field in the theory of the Aharonov-Bohm (1959) effect, which is the archetypical geometric phase in quantum mechanics. I then cast the physical observable $\Delta\mathbf{P}$ as a circuit integral of the connection. An outline of the present formulation has been presented elsewhere (Resta, 1993).

In Sec. II I discuss the nature of polarization and screening as quantum phenomena; I then outline some analogies between the present case and other known occurrences of geometric phases in quantum mechanics. In Sec. III I establish the main formalism, arriving at the basic definition of $\Delta\mathbf{P}$, Eqs. (2) and (12), assumed throughout this work. In Sec. IV I prove that these equations define a macroscopic physical observable. In Sec. V I show the equivalence of Eq. (12) with the geometric phase formulation. In Sec. VI I prove that $\Delta\mathbf{P}$ originates from the circuit integral of a Berry connection, in a four-dimensional parameter space; its curvature yields straightforwardly the macroscopic linear response of the system. In Sec. VII I discuss the general strategy for numerical computation of Berry phases, and in particular of those leading to $\Delta\mathbf{P}$. In Sec. VIII I show the equivalence of the geometric phase approach with the well-established perturbative approach as far as the macroscopic linear response of the crystalline solid is concerned. Then I outline briefly the main features of linear response in the presence of macroscopic fields and review the most recent calculations of the Born effective charge tensors and of the piezoelectric effect. In Sec. IX I define the concept of spontaneous polarization in ferroelectrics and illustrate the first quantum calculation of such polarization. In Sec. X I offer some conclusions.

II. POLARIZATION AND QUANTUM MECHANICS

Macroscopic polarization is a manifestation of screening. Quite generally, screening can be defined as the effect of competition between electrical forces and some hindering mechanism of a different kind. Within the present context (zero-temperature electronic screening) the restoring forces are provided by quantum mechanics; roughly speaking by the Pauli principle. In some special cases, a purely classical modeling of the quantum forces is possible. Within the popular Clausius-Mossotti picture, one schematizes the dielectric solid as an assembly of well separated and independently polarizable units. All of the quantum mechanics of the problem is then integrated out in a single parameter, the dipolar polarizability of a single unit. I wish to stress that the Clausius-Mossotti picture safely applies only to extreme cases, such as ionic or molecular crystals. At the other extreme are covalent materials, in which the electronic charge is delocalized and no local-dipole picture is acceptable. In this case the dipole of a unit cell is completely ill defined (Martin, 1974). Well studied covalent materials are the simplest semiconductors, in which the behavior of valence electrons is known to be strongly nonclassical. Covalent bonding is a purely quantum phenomenon, and the consequent dielectric behavior is a quantum phenomenon as well (this viewpoint is emphasized, for example, by Phillips, 1973). Even oversimplified model screening theories for covalent materials must explicitly invoke quantum mechanics in some approximate form. This is the case for the popular screening models of Penn (1962; Grimes and Cowley, 1975) and Resta (1977). The latter is based on the Thomas-Fermi approximation. Within both these models, the valence electrons are schematized as a "semiconducting electron gas." Polarization is due to a uniform current flowing across the sample, while the role of local dipoles is totally ignored. In real materials the two extreme mechanisms—uniform polarization and local dipoles—coexist (for a thorough discussion, see Resta and Kunc, 1986).

The dipole of a macroscopic sample, Eq. (1), is generally neutralized at equilibrium by electrically active defects and/or surface charges (Landau and Lifshitz, 1984), whose relaxation times may nonetheless be extremely long (e.g., hours). Therefore even a very slow perturbation may induce a measurable $\Delta\mathbf{P}$. The important point is that $\Delta\mathbf{P}$ is phenomenologically known to be a *bulk* property, i.e., independent—in the thermodynamic limit—of the surface conditions of the sample. The basic quantity addressed in this work is therefore the difference $\Delta\mathbf{P}$ in macroscopic polarization between two different states of the same solid. We consider this difference within the adiabatic approximation at zero temperature, and we separate its ionic and electronic terms as in Eq. (1):

$$\Delta\mathbf{P} = \Delta\mathbf{P}_{\text{ion}} + \Delta\mathbf{P}_{\text{el}}, \tag{2}$$

$$\Delta\mathbf{P}_{\text{el}} = \frac{1}{\mathcal{V}} \int d\mathbf{r}\; \mathbf{r}\; \Delta\rho(\mathbf{r}). \tag{3}$$

Using this definition, $\Delta\mathbf{P}$ is a property of the charge of the finite sample. To define a bulk property requires taking the thermodynamic limit: $\Delta\mathbf{P}$ has contributions from both the bulk and the surface regions, which in general cannot be disentangled. A successful strategy for arriving at a bulk definition is to switch from *charge* to *current* (Resta, 1992). While the former is the squared modulus of the wave function, the latter is fundamentally related to its *phase*. Within a finite system, two alternate descriptions are equivalent, owing to the continuity equation: the charge that piles up at the surface during the continuous transformation is related to the current that flows through the bulk region. This link is lost for an infinite crystal in the thermodynamic limit: the charge and the current (alias the wave function's modulus and phase) then carry quite distinct pieces of information. In this same limit, macroscopic polarization is a property of the current, *not* of the charge (contrary to a rather common belief, found in many textbooks).

Therefore, in order to evaluate $\Delta\mathbf{P}$ in an infinite periodic crystal, one has to monitor the macroscopic current flowing through the unit cell. The geometric phase performs precisely this task in an elegant and effective way. An adiabatic macroscopic current was previously identified with a geometric phase in quite different contexts—such as the quantum Hall effect (e.g., Prange and Girvin, 1987; Morandi, 1988) or sliding charge-density waves (Thouless, 1983; Kunz, 1986)—owing to the work of Thouless (1983); this work in fact inspired the original King-Smith and Vanderbilt derivation. This is not the approach taken here: I follow instead an independent proof of the main King-Smith and Vanderbilt result (Resta, 1993).

The occurrence of nontrivial geometric phases in the band theory of solids was first discovered by Zak (1989) and attributed to the breaking of crystal inversion symmetry. The Zak phase is an essential ingredient of the present approach to macroscopic polarization, and in fact a nonvanishing value of $\Delta\mathbf{P}$ is allowed only if the crystal transformation breaks inversion symmetry. Needless to say, the breaking of the same symmetry within a finite system does *not* produce any geometric phase, while instead the most common occurrence of a geometric phase is due to breaking of time-reversal symmetry, as in a magnetic field. Some formal analogies of the magnetic case with the present electrostatic one can be found at the level of the Hamiltonian (8) below, having discrete eigenstates, where a very peculiar ($\mathbf{r}$-independent, $\mathbf{q}$-dependent) vector potential appears. Some precursor considerations on this point can be found in an early (1964) paper of Kohn.

The geometric phase approach—in its present status—is basically a one-electron theory, in the same sense as is the whole band theory of solids (Blount, 1962). The main results can therefore be stated in terms of any mean-field theoretical framework. I have chosen here to formulate the theory within the familiar language of the density-functional theory (see for example, Lundqvist and March, 1983) of Kohn and Sham, which has at least two main advantages: it is a formally *exact* theory of the electronic ground state, and is currently implemented—within the local-density approximation—in numerical work. It is a trivial exercise to rephrase all of the results of the present paper within the language of the Hartree-Fock theory of solids (Pisani *et al.*, 1988), or that of any other mean-field theory. Finally I observe that no *genuine* many-body generalization is available so far, that is able to cope with highly correlated dielectrics: in these cases one expects unphysical features in the Kohn-Sham potential, and therefore density-functional theory—despite being formally exact—is probably useless.

Density-functional theory is quite appropriate for dealing with $\Delta\mathbf{P}$, which is an adiabatic observable of the electronic ground state. I shall show that $\Delta\mathbf{P}$ is a property of the manifold of the occupied Kohn-Sham orbitals as a whole, as is the crystal density; but at variance with the density—where any phase information is deleted—$\Delta\mathbf{P}$ depends in a gauge-invariant way on the *phases* of the Kohn-Sham orbitals.

III. MICROSCOPICS AND MACROSCOPICS

We start from the basic definitions of Eqs. (2) and (3), and we address the thermodynamic limit $\mathcal{V} \to \infty$. The basic assumption of the present theory is the existence of a continuous adiabatic transformation of the Kohn-Sham Hamiltonian connecting the two crystal states. It must fulfill two important hypotheses: (i) the transformation is performed at null electric field, and (ii) the system remains an insulator—in the sense that its Kohn-Sham gap does not close—throughout the transformation. For the sake of simplicity we parametrize the transformation with a variable λ, chosen to have the values of 0 and 1 at the initial and final states, respectively (Resta, 1992):

$$\Delta \mathbf{P} = \int_0^1 d\lambda \, \mathbf{P}'(\lambda). \tag{4}$$

If one identifies the variable λ with time (in appropriate units), then Eq. (4) can be spelled out by saying that $\Delta \mathbf{P}$ is the integrated current flowing through the sample during the adiabatic transformation. This current is the very quantity that is phenomenologically measured.

As for the physical nature of the transformation, we remain quite general about it. As an example, λ could be taken to be an internal coordinate. In this case the transformation is a relative displacement of sublattices in the periodic crystal. This example is relevant for the polarization induced by zone-center transverse-optic phonon modes (in polar crystals) and for ferroelectric polarization. To start with, only transformations that conserve the volume and the shape of the unit cell are explicitly considered, but the approach applies with no major change to cell-nonconserving transformations as well, to cope with piezoelectric polarization. The discussion on this point is deferred to the end of Sec. V.

Since the crystalline solid is in a null electric field, periodic boundary conditions can be used at any λ: the Kohn-Sham orbitals $\psi_n^{(\lambda)}(\mathbf{q},\mathbf{r})$ then have the Bloch form. For an insulating system with $\bar{n}$ doubly occupied bands, the electronic charge density is

$$\rho^{(\lambda)}(\mathbf{r}) = \frac{2e}{(2\pi)^3} \sum_{n=1}^{\bar{n}} \int_{\mathrm{BZ}} d\mathbf{q} \, |\psi_n^{(\lambda)}(\mathbf{q},\mathbf{r})|^2, \tag{5}$$

where BZ is the Brillouin zone, and a plane-wave-like normalization is assumed for the Bloch functions. Any phase information about the Kohn-Sham orbitals is lost in Eq. (5).

An alternative expression is obtained via a band-by-band Wannier transformation (Blount, 1962):

$$a_n^{(\lambda)}(\mathbf{r}) = \frac{\sqrt{\Omega}}{(2\pi)^3} \int_{\mathrm{BZ}} d\mathbf{q} \, \psi_n^{(\lambda)}(\mathbf{q},\mathbf{r}), \tag{6}$$

where Ω is the cell volume. Wannier functions displaced by lattice vectors $\mathbf{R}_l$ are orthogonal to each other and define a unitary transformation; the inverse transformation is

$$\psi_n^{(\lambda)}(\mathbf{q},\mathbf{r}) = e^{i\mathbf{q}\cdot\mathbf{r}} u_n^{(\lambda)}(\mathbf{q},\mathbf{r}) = \sqrt{\Omega} \sum_l e^{i\mathbf{q}\cdot\mathbf{R}_l} a_n^{(\lambda)}(\mathbf{r}-\mathbf{R}_l). \tag{7}$$

The periodic functions $u_n^{(\lambda)}(\mathbf{q},\mathbf{r})$ will be a basic ingredient of the present theory. At a given $\mathbf{q}$, they are discrete eigenstates of the Kohn-Sham Hamiltonian

$$H^{(\lambda)}(\mathbf{q}) = \frac{1}{2m_e}(\mathbf{p}+\hbar\mathbf{q})^2 + V^{(\lambda)}(\mathbf{r}), \tag{8}$$

where m_e is the electron mass. They obey—owing to Eq. (7)—the important phase relationship

$$u_n^{(\lambda)}(\mathbf{q}+\mathbf{G},\mathbf{r}) = e^{-i\mathbf{G}\cdot\mathbf{r}} u_n^{(\lambda)}(\mathbf{q},\mathbf{r}), \tag{9}$$

where $\mathbf{G}$ is any reciprocal lattice vector. The Wannier-transformed form of the electronic charge density is

$$\rho^{(\lambda)}(\mathbf{r}) = 2e \sum_{n=1}^{\bar{n}} \sum_l |a_n^{(\lambda)}(\mathbf{r}-\mathbf{R}_l)|^2. \tag{10}$$

We are interested in the periodic charge $\Delta\rho = \rho^{(1)} - \rho^{(0)}$, which occurs in the thermodynamic limit of Eq. (3):

$$\Delta\rho(\mathbf{r}) = 2e \sum_{n=1}^{\bar{n}} \sum_l [\,|a_n^{(1)}(\mathbf{r}-\mathbf{R}_l)|^2 - |a_n^{(0)}(\mathbf{r}-\mathbf{R}_l)|^2\,]. \tag{11}$$

Since the periodic density difference is now decomposed into a sum of *localized* and *neutral* charge distributions—as in simple Clausius-Mossotti models—its dipole moment per cell is well defined and given by

$$\Delta \mathbf{P}_{\mathrm{el}} = \frac{2e}{\Omega} \sum_{n=1}^{\bar{n}} \int d\mathbf{r} \, \mathbf{r} \, [\,|a_n^{(1)}(\mathbf{r})|^2 - |a_n^{(0)}(\mathbf{r})|^2\,]. \tag{12}$$

The convergence of the integrals follows from the results of Blount (1962).

The above derivation is mathematically correct; nonetheless Eqs. (2) and (12) cannot be accepted as the basic definition of a physical observable without further analysis. In fact the phases of Bloch functions entering Eq. (6) are arbitrary, thus making the Wannier transformation strongly nonunique; in a three-dimensional system with composite bands, further nonuniqueness comes from separating the states in overlapping energy regions into different bands. Ever since the pioneering work of Kohn (1959, 1973) and des Cloizeaux (1964), it has been well known that different choices provide different shapes, symmetries, and even asymptotic behaviors for the Wannier functions. From a more fundamental density-functional point of view, the individual Kohn-Sham orbitals carry no physical meaning: electronic ground-state properties are in fact a *global* property of the occupied manifold as a whole. I shall therefore consider a quite general unitary transformation of the occupied u's amongst themselves at a given $\mathbf{q}$. Such a gauge transformation is defined by the unitary $\bar{n}\times\bar{n}$ matrix $U(\mathbf{q})$. Any physical electronic ground-state property must be gauge invariant.

Any nonpathological gauge transformation ensures convergence of the first moments appearing in Eq. (12). Higher moments could be more problematic (Blount, 1962). To ensure that Eqs. (2) and (12) define $\Delta\mathbf{P}$ as a macroscopic observable of the system, it remains to prove gauge invariance and translational invariance; these steps are accomplished below.

IV. GAUGE AND TRANSLATION INVARIANCES

It proves useful to transform $\Delta\mathbf{P}_{\rm el}$ back in terms of the u wave functions. We introduce, following Blount (1962), the $\bar{n}\times\bar{n}$ overlap matrix $S^{(\lambda)}(\mathbf{q},\mathbf{q}')$, whose elements are

$$S^{(\lambda)}_{mn}(\mathbf{q},\mathbf{q}') = \langle u^{(\lambda)}_m(\mathbf{q})|u^{(\lambda)}_n(\mathbf{q}')\rangle = \frac{1}{\Omega}\int_{\rm cell} d\mathbf{r}\; u^{(\lambda)*}_m(\mathbf{q},\mathbf{r})\, u^{(\lambda)}_n(\mathbf{q}',\mathbf{r}). \quad (13)$$

The overlap matrix is obviously gauge dependent; when $\mathbf{q}'-\mathbf{q}$ equals a reciprocal vector $\mathbf{G}$, it fulfills the relationship

$$S^{(\lambda)}_{mn}(\mathbf{q},\mathbf{q}+\mathbf{G}) = \langle u^{(\lambda)}_m(\mathbf{q})|e^{-i\mathbf{G}\cdot\mathbf{r}}|u^{(\lambda)}_n(\mathbf{q})\rangle, \quad (14)$$

owing to Eq. (9). Straightforward manipulations transform Eq. (12) into the equivalent form

$$\Delta\mathbf{P}_{\rm el} = i\frac{2e}{(2\pi)^3}\int_{\rm BZ} d\mathbf{q}\;{\rm tr}\{\nabla_{\mathbf{q}'}S^{(1)}(\mathbf{q},\mathbf{q}') - \nabla_{\mathbf{q}'}\;S^{(0)}(\mathbf{q},\mathbf{q}')\}\Big|_{\mathbf{q}'=\mathbf{q}}, \quad (15)$$

where tr indicates the trace, and we consider only the gauges in which $S^{(\lambda)}$ is a differentiable function of its arguments. Despite the integrand's being gauge dependent, Eq. (15) is gauge invariant, as well as Eq. (12). The proof is reported in the venerable paper of Blount (1962), although for the case of nonoverlapping bands only. More recently, Zak (1989) recognized that expressions such as those on the right-hand side of Eq. (15) are geometric phases, but again he focused on properties of the individual bands only. These geometric phases were not related to any physical observable of the crystalline solid until the major contribution of King-Smith and Vanderbilt, who identified their fundamental link to the macroscopic electric polarization.

I generalize the gauge-invariance proof of Zak (1989; Michel and Zak, 1992) to the multiband case upon considering the most general gauge transformation which changes the matrix $S^{(\lambda)}(\mathbf{q},\mathbf{q}')$ into $\tilde{S}^{(\lambda)}(\mathbf{q},\mathbf{q}') = U^{-1}(\mathbf{q})S^{(\lambda)}(\mathbf{q},\mathbf{q}')U(\mathbf{q}')$. The integrands in Eq. (15) then become

$$\begin{aligned}{\rm tr}\;\{\nabla_{\mathbf{q}'}\tilde{S}^{(\lambda)}(\mathbf{q},\mathbf{q}')\}\Big|_{\mathbf{q}'=\mathbf{q}} &= {\rm tr}\;\{\nabla_{\mathbf{q}'}S^{(\lambda)}(\mathbf{q},\mathbf{q}')\}\Big|_{\mathbf{q}'=\mathbf{q}} + {\rm tr}\;\{U^{-1}(\mathbf{q})\nabla_{\mathbf{q}}U(\mathbf{q})\},\end{aligned} \quad (16)$$

where I have used the cyclic invariance of the trace and the fact that $S^{(\lambda)}$ coincides with the unit matrix at $\mathbf{q}=\mathbf{q}'$. I then transform the last term using the well-known matrix identity (Schiff, 1968)

$$\det\exp A = \exp{\rm tr}\;A, \quad (17)$$

which, applied to $A=\ln U$, yields

$${\rm tr}\;\{U^{-1}\nabla U\} = \nabla\;\ln\det U = i\nabla\vartheta, \quad (18)$$

where ϑ is the phase of the determinant of U. Although otherwise arbitrary, U must conserve Eqs. (9) and (14); this implies that U is periodic in reciprocal space, yielding

$$e^{i\vartheta(\mathbf{q}+\mathbf{G})} = e^{i\vartheta(\mathbf{q})}. \quad (19)$$

The general form for ϑ is then

$$\vartheta(\mathbf{q}) = \alpha(\mathbf{q}) + \mathbf{q}\cdot\mathbf{R}_l, \quad (20)$$

where $\alpha(\mathbf{q})$ is a periodic function and $\mathbf{R}_l$ is any lattice vector: the gradient of this phase—when integrated over the Brillouin zone—contributes to Eq. (15) the constant term

$$\mathbf{P}_l = \frac{2e}{\Omega}\mathbf{R}_l. \quad (21)$$

The final value of $\Delta\mathbf{P}_{\rm el}$ is therefore gauge invariant and well defined, modulo the "quantum" $\mathbf{P}_l$. One often expects $|\Delta\mathbf{P}_{\rm el}|$ and—most important—$|\Delta\mathbf{P}|$ itself to be much smaller than such quanta, in which case no ambiguity arises. Otherwise $\Delta\mathbf{P}$ *cannot* be determined as a function of the initial and final states only, as in Eq. (15). Additional intermediate points in the λ interval [0,1] have to be considered to resolve the ambiguity. In the latter case, when λ is in a multiparameter space, the value of $\Delta\mathbf{P}_{\rm el}$ may depend on the actual path joining the initial and final states. It is worth pointing out that—in *both* cases—the hypothesis that λ transforms the crystal with continuity *and* via insulating states for all λ's is essential to get an unambiguous gauge-invariant result.

We have proved that the two terms in Eq. (15), originating from $S^{(1)}$ and $S^{(0)}$, are separately gauge invariant. It is therefore tempting to identify each with an "absolute" electronic polarization of a specific crystal state. Such a concept is ill defined. To realize this, consider Eq. (12), where we see that each of these two terms is the dipole of a *non-neutral* charge distribution. Accordingly, neither is separately translationally invariant. To better illustrate this point, let us consider a uniform rigid translation of the crystal as a whole by an amount $\lambda\mathbf{r}_0$, where $\mathbf{r}_0$ is fixed and λ is between 0 and 1, as usual. One gets the $S^{(1)}$ matrix elements simply by multiplying those of $S^{(0)}$ by the phase $\exp i(\mathbf{q}-\mathbf{q}')\mathbf{r}_0$, whence the rigid translation induces a change in polarization:

$$\Delta \mathbf{P}_{\rm el} = \frac{2\bar{n}e}{\Omega}\mathbf{r}_0. \qquad (22)$$

Since a bulk macroscopic property must be translationally invariant, there is no way of defining the absolute electronic polarization of the crystal in a given state.

The translational invariance is of course recovered when we consider the contribution of the ions as well, starting from Eq. (1). The ionic contribution exactly cancels that of Eq. (22), due to the overall charge neutrality of the crystal cell. The relationship between charge neutrality and translational invariance of the macroscopic polarization is indeed fundamental, as is emphasized in the classic work of Pick, Cohen, and Martin (1970). In the present formulation the total $\Delta\mathbf{P}$ correctly vanishes for a rigid translation of the crystal as a whole.

V. GEOMETRIC QUANTUM PHASES

We start by introducing the scalar function $\varphi^{(\lambda)}$ as the phase of the determinant of $S^{(\lambda)}$:

$$\varphi^{(\lambda)}(\mathbf{q},\mathbf{q}') = \mathrm{Im}\ln\det S^{(\lambda)}(\mathbf{q},\mathbf{q}'), \qquad (23)$$

defined modulo 2π, which measures the "phase difference" between the Kohn-Sham orbitals at $\mathbf{q}'$ and those at $\mathbf{q}$, once the Bloch phase is removed. In the jargon of geometric phases, Eq. (23) would be a Pancharatnam (1956) phase. Here it is a property of the occupied Kohn-Sham manifold as a whole and is of course gauge dependent; its infinitesimal variation is expressed as

$$d\varphi = \nabla_{\mathbf{q}'}\varphi^{(\lambda)}(\mathbf{q},\mathbf{q}')\Big|_{\mathbf{q}'=\mathbf{q}}\cdot d\mathbf{q}. \qquad (24)$$

The differential phase can be equivalently expressed in terms of the trace of $S^{(\lambda)}$, since

$$\nabla_{\mathbf{q}'}\varphi^{(\lambda)}(\mathbf{q},\mathbf{q}')\Big|_{\mathbf{q}'=\mathbf{q}} = -i\,\mathrm{tr}\,\{\nabla_{\mathbf{q}'}S^{(\lambda)}(\mathbf{q},\mathbf{q}')\}\Big|_{\mathbf{q}'=\mathbf{q}} = -i\sum_{n=1}^{\bar{n}}\langle u_n^{(\lambda)}(\mathbf{q})|\nabla_{\mathbf{q}}u_n^{(\lambda)}(\mathbf{q})\rangle, \qquad (25)$$

as is easily proven by applying the same identity in Eq. (18), to $S^{(\lambda)}$. One then exploits the fact that, at $\mathbf{q}'{=}\mathbf{q}$, $S^{(\lambda)}(\mathbf{q},\mathbf{q}')$ coincides with the identity, while the trace of its $\mathbf{q}'$ gradient is purely imaginary, owing to orthonormality. Equation (25) leads to an alternative expression for $\Delta\mathbf{P}_{\rm el}$, since substituting it in Eq. (15) yields

$$\Delta\mathbf{P}_{\rm el} = \frac{2e}{(2\pi)^3}\int_{\rm BZ} d\mathbf{q}\,[-\nabla_{\mathbf{q}'}\varphi^{(1)}(\mathbf{q},\mathbf{q}') + \nabla_{\mathbf{q}'}\,\varphi^{(0)}(\mathbf{q},\mathbf{q}')]\Big|_{\mathbf{q}'=\mathbf{q}}. \qquad (26)$$

In numerical implementations the determinant form of Eq. (23) is essential in order to yield gauge-invariant results. This point will be further elaborated in Sec. VII.

Let us take the two points $\mathbf{q}_0$ and $\mathbf{q}_0+\mathbf{G}$ in reciprocal space. Their phase difference $\varphi^{(\lambda)}(\mathbf{q}_0,\mathbf{q}_0+\mathbf{G})$ is easily proven to be gauge invariant using the results of the previous section. If we now consider a continuous path C joining these two points, the line integral of the differential phase

$$\gamma^{(\lambda)}(C) = -\int_C d\varphi \qquad (27)$$

is gauge invariant as well and has the properties of a geometric phase. This result is a simple generalization of the work of Zak (1989; Michel and Zak, 1992), in which a similar result is proved for a single band. Nonetheless, the present generalization to the occupied manifold as a whole is essential to cope with valence-band crossings in real solids. A standard Berry phase is a *circuit* integral of the differential phase in a parameter space (Berry, 1984, 1989; Jackiw, 1988). In the following I shall call a "Zak phase" the peculiar form of Eq. (27), in which the line integral is evaluated over a *special open path* in $\mathbf{q}$ space, and I reserve the name of Berry phases for line integrals evaluated along *arbitrary closed paths* in *arbitrary* parameter spaces.

The three-dimensional Brillouin-zone integral in Eqs. (15) and (26) can be evaluated upon performing two Zak phase calculations—such as the line integral of Eq. (27)—and a surface integration in succession: this is in fact the approach followed in King-Smith and Vanderbilt. In the following I explicitly illustrate such integral reduction in the most general case of an arbitrary Bravais lattice.

First of all we observe that the Brillouin-zone integrals in Eqs. (15) and (26) can be equivalently performed upon a unit cell of the reciprocal lattice, since this amounts to a simple gauge transformation. We then map the (nonrectangular) unit reciprocal cell into a unit cube via a linear change of variables. We call the basic translations of the reciprocal lattice $\mathbf{G}_j$ (with $j{=}1,\ldots,3$) and those of the direct lattice $\mathbf{R}_j$. The transformation to the dimensionless ξ_j variables is

$$\mathbf{q} = \xi_1\mathbf{G}_1 + \xi_2\mathbf{G}_2 + \xi_3\mathbf{G}_3\ ,\ \ \xi_j = \frac{1}{2\pi}\mathbf{q}\cdot\mathbf{R}_j. \qquad (28)$$

One gets a more compact form upon defining $\xi_4{=}\lambda$ and considering the four-dimensional vector $\boldsymbol{\xi}$ as a single parameter. The notation $|u_n(\boldsymbol{\xi})\rangle = |u_n^{(\lambda)}(\mathbf{q})\rangle$ is adopted for the state vectors. The differential phase for infinitesimal variation of *both* $\mathbf{q}$ and λ is provided by the linear differential form

$$d\varphi = -i \sum_{n=1}^{\bar{n}} \langle u_n(\boldsymbol{\xi}) | \nabla_{\boldsymbol{\xi}} u_n(\boldsymbol{\xi}) \rangle \cdot d\boldsymbol{\xi}. \tag{29}$$

The change of variables transforms Eq. (26) into three equations, which provide the components of $\Delta \mathbf{P}_{\rm el}$ along the $\mathbf{G}_j$ directions. Let us focus for the sake of simplicity upon one of these components, say j=3. The result is cast as

$$\mathbf{G}_3 \cdot \Delta \mathbf{P}_{\rm el} = \frac{2e}{\Omega} \int d\xi_1 d\xi_2 \left[\int_{C_0} d\varphi - \int_{C_1} d\varphi \right] , \tag{30}$$

where the two-dimensional integral is over the unit square $[0,1]\times[0,1]$, and the two Zak phases are evaluated along appropriate unit segments. The points of C_0 are defined by $\boldsymbol{\xi} = (\xi_1, \xi_2, x, 0)$, $0 \le x \le 1$, and those of C_1 by $\boldsymbol{\xi} = (\xi_1, \xi_2, x, 1)$, $0 \le x \le 1$. The derivation of Eq. (30) is given in the Appendix. Analogous expressions can be obtained for the remaining $\mathbf{G}_j$ components.

Whenever crystal symmetry restricts the polarization to be along $\mathbf{R}_3$—and the two other basic translations to be orthogonal to it—Eq. (30) is most conveniently written

$$\Delta \mathbf{P}_{\rm el} = \frac{e\mathbf{R}_3}{\pi\Omega} \int d\xi_1 d\xi_2 \left[\int_{C_0} d\varphi - \int_{C_1} d\varphi \right] , \tag{31}$$

which coincides with the main result of King-Smith and Vanderbilt.

In both Eq. (30) and Eq. (31) it proves better to use the alternative determinant form, as in Eqs. (23) and (24). This is obtained via the obvious generalization of the overlap matrix,

$$S_{mn}(\boldsymbol{\xi}, \boldsymbol{\xi}') = \langle u_m(\boldsymbol{\xi}) | u_n(\boldsymbol{\xi}') \rangle. \tag{32}$$

The Pancharatnam phase, Eq. (23), is generalized to this augmented parameter space as

$$\varphi(\boldsymbol{\xi}, \boldsymbol{\xi}') = \text{Im ln det } S(\boldsymbol{\xi}, \boldsymbol{\xi}'), \tag{33}$$

and the differential phase, Eq. (29), has the alternative equivalent expression

$$d\varphi = \nabla_{\boldsymbol{\xi}'} \varphi(\boldsymbol{\xi}, \boldsymbol{\xi}')\big|_{\boldsymbol{\xi}'=\boldsymbol{\xi}} \cdot d\boldsymbol{\xi}. \tag{34}$$

The expressions given above, Eqs. (30) to (34), are those used in practical applications of the approach to real materials (King-Smith and Vanderbilt, 1993; Dal Corso *et al.*, 1993b; Resta *et al.*, 1993a, 1993b), which have been performed within the local-density approximation (e.g., Lundqvist and March, 1983) to density-functional theory.

So far, I have considered only transformations at constant volume and shape of the unit cell, i.e., transformations in which the $\mathbf{R}_j$ vectors do not vary with λ (alias ξ_4). Thanks to the present scaled formulation, this restriction can be eliminated with no harm. The expression is particularly simple for the special case of Eq. (31), which is generalized to

$$\Delta \mathbf{P}_{\rm el} = \frac{e}{\pi} \int d\xi_1 d\xi_2 \left[\frac{\mathbf{R}_3^{(0)}}{\Omega^{(0)}} \int_{C_0} d\varphi - \frac{\mathbf{R}_3^{(1)}}{\Omega^{(1)}} \int_{C_1} d\varphi \right] . \tag{35}$$

At this point we may look back at our starting definition of $\Delta \mathbf{P}_{\rm el}$, Eq. (12), to notice that it does not apply as it stands to the cell-nonconserving cases. This is not a serious problem, since a fully satisfactory generalized definition, based on Eq. (12), can be obtained upon performing a two-step transformation on the solid: first a pure scaling of the charge of one of the two crystal states, and then a suitable cell-conserving transformation of the electronic Hamiltonian. Further elaboration on this point is unnecessary, since the geometric phase approach provides an equivalent, and more useful, formulation, e.g., in Eq. (35).

VI. CONNECTION AND CURVATURE

The four-dimensional formulation—introduced in the previous section for the purpose of simplifying notation—is more than just cosmetic, in that it allows us to look at the problem from a quite general viewpoint and offers a deep insight into the fundamental quantum nature of the macroscopic polarization as a "standard" Berry phase in $\boldsymbol{\xi}$ space.

The state vectors $|u_n(\boldsymbol{\xi})\rangle$ are discrete eigenstates of the parametric Kohn-Sham Hamiltonian $H(\boldsymbol{\xi}) = H^{(\lambda)}(\mathbf{q})$, Eq. (8). We define the Berry connection of the problem in the usual way (Jackiw, 1988):

$$\boldsymbol{\mathcal{X}}(\boldsymbol{\xi}) = i \sum_{n=1}^{\bar{n}} \langle u_n(\boldsymbol{\xi}) | \nabla_{\boldsymbol{\xi}} u_n(\boldsymbol{\xi}) \rangle. \tag{36}$$

At the most elementary level, the connection is defined for a single state; the generalization to the set of the $\bar{n}$ lowest states is trivial, provided these $\bar{n}$ states are *not* degenerate with the higher ones at any point of the domain (Jackiw, 1988). This is indeed the case, since we have assumed the solid to be an insulator for all λ's. It is no surprise that we have met this very same Berry connection before, in the expressions for the differential phase, Eqs. (29) and (34). The circuit integral of the connection along any *closed* path C in $\boldsymbol{\xi}$-space is just a "nonexotic" Berry phase,

$$\gamma(C) = -\oint_C d\varphi = \oint_C \boldsymbol{\mathcal{X}}(\boldsymbol{\xi}) \cdot d\boldsymbol{\xi}, \tag{37}$$

whose gauge invariance is by now almost obvious (Berry, 1984, 1989; Jackiw, 1988). I shall show that the quantity of interest, $\Delta \mathbf{P}_{\rm el}$, can be expressed in terms of such circuit integrals.

Let us consider only the $\Delta \mathbf{P}_{\rm el}$ component along $\mathbf{G}_3$. Equation (30) is then equivalent to

$$\mathbf{G}_3 \cdot \Delta \mathbf{P}_{\rm el} = \frac{2e}{\Omega} \int d\xi_1 d\xi_2 \, \gamma(C), \tag{38}$$

when the closed path C is the contour of the unit square in the plane parallel to the ξ_3 and ξ_4 axes, at given values of ξ_1 and ξ_2 (this is illustrated in Fig. 1). The proof of the equivalence is straightforward. The path consists of four straight-line segments. Two of them coincide by construction with C_0 and C_1 in Eq. (30), and their contributions to $\gamma(C)$ are exactly the Zak phases of Eq. (30), with the appropriate sign; the contributions of the remaining two segments cancel. This is most easily seen in the $(\mathbf{q},\lambda)$ variables, since the points of these two segments differ by the reciprocal vector $\mathbf{G}_3$, and Eq. (9) implies

$$\langle u_m^{(\lambda)}(\mathbf{q}+\mathbf{G}_3)|\frac{\partial}{\partial\lambda}u_n^{(\lambda)}(\mathbf{q}+\mathbf{G}_3)\rangle = \langle u_m^{(\lambda)}(\mathbf{q})|\frac{\partial}{\partial\lambda}u_n^{(\lambda)}(\mathbf{q})\rangle. \tag{39}$$

Incidentally, Eq. (38) proves the gauge invariance of $\Delta\mathbf{P}_{\rm el}$ in an alternative—and more elegant—way with respect to the proof given in Sec. IV; in both derivations, the role of Eq. (9) is pivotal.

I stress once more that the connection is gauge dependent and nonobservable, while its circuit integral is gauge invariant and provides the relevant physical quantity $\Delta\mathbf{P}_{\rm el}$. The connection $\boldsymbol{\mathcal{X}}(\boldsymbol{\xi})$, therefore, plays the same role as the ordinary vector potential in the theory of the Aharonov-Bohm (1959) effect, which is the archetype of geometric quantum phases.

The appropriate generalization of Stokes's theorem (Arnold, 1989) transforms Eq. (37) into the surface integral of the curl of $\boldsymbol{\mathcal{X}}$, i.e., using Berry's (1984, 1989) notations,

$$\gamma(C) = -\text{Im}\sum_{n=1}^{\bar{n}}\int d\sigma\ \langle\nabla_{\boldsymbol{\xi}}u_n(\boldsymbol{\xi})|\times|\nabla_{\boldsymbol{\xi}}u_n(\boldsymbol{\xi})\rangle, \tag{40}$$

where $d\sigma$ denotes the area element in $\boldsymbol{\xi}$ space, and the integral is performed over any surface enclosed by the contour C. The integrand itself is now gauge invariant, as opposed to the connection, which is not. The curvature $\boldsymbol{\mathcal{Y}}$ is defined as the generalized curl of the connection (Jackiw, 1988),

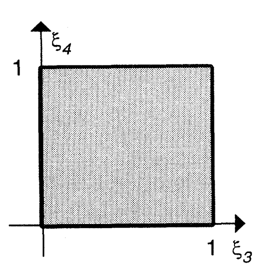

FIG. 1. Projection over the (ξ_3,ξ_4) plane of the contours where the Berry phase, Eq. (37), is evaluated as a line integral of the connection (thick solid line). By Stokes's theorem, the Berry phase equals the integral of the curvature over a surface whose projection is also shown (shaded area).

$$\begin{aligned}\mathcal{Y}_{ij}(\boldsymbol{\xi}) &= \frac{\partial}{\partial\xi_i}\mathcal{X}_j(\boldsymbol{\xi}) - \frac{\partial}{\partial\xi_j}\mathcal{X}_i(\boldsymbol{\xi}) \\ &= 2\ \text{Im}\sum_{n=1}^{\bar{n}}\langle\frac{\partial}{\partial\xi_i}u_n(\boldsymbol{\xi})|\frac{\partial}{\partial\xi_j}u_n(\boldsymbol{\xi})\rangle.\end{aligned} \tag{41}$$

The surface integral over the unit square in Fig. 1 provides the Berry phase as

$$\gamma(C) = \int d\xi_3 d\xi_4\ \mathcal{Y}_{34}(\boldsymbol{\xi}); \tag{42}$$

therefore Eq. (38)—and its analogs—are equivalent to

$$\mathbf{G}_j\cdot\Delta\mathbf{P}_{\rm el} = \frac{2e}{\Omega}\int d\boldsymbol{\xi}\ \mathcal{Y}_{j4}(\boldsymbol{\xi})\ ,\ j=1,\ldots,3\ , \tag{43}$$

where the four-dimensional $\boldsymbol{\xi}$ integral is performed over the unit hypercube $[0,1]\times[0,1]\times[0,1]\times[0,1]$.

Written in the form of Eq. (41), $\boldsymbol{\mathcal{Y}}$ is not explicitly gauge invariant. Following Berry (1984), I insert a complete set of states in Eq. (41). Straightforward manipulations lead to the equivalent form

$$\begin{aligned}\mathcal{Y}_{ij}(\boldsymbol{\xi}) = 2\ \text{Im}\sum_{n=1}^{\bar{n}}\sum_{m=\bar{n}+1}^{\infty}&\langle u_n(\boldsymbol{\xi})|\frac{\partial}{\partial\xi_i}u_m(\boldsymbol{\xi})\rangle \\ &\times\langle\frac{\partial}{\partial\xi_j}u_m(\boldsymbol{\xi})|u_n(\boldsymbol{\xi})\rangle,\end{aligned} \tag{44}$$

which explicitly shows invariance under unitary transformations of the occupied u's amongst themselves.

Comparison of Eq. (43) with Eq. (4) shows immediately that the curvature provides the $\mathbf{G}_j$ components of the electronic term in the polarization derivative with respect to λ (alias ξ_4):

$$\mathbf{G}_j\cdot\mathbf{P}'_{\rm el}(\xi_4) = \frac{2e}{\Omega}\int d\xi_1 d\xi_2 d\xi_3\ \mathcal{Y}_{j4}(\boldsymbol{\xi}). \tag{45}$$

The equivalence of Eq. (45) with the established results of linear-response theory (Vogl, 1978; Giannozzi *et al.*, 1991; Resta, 1992) is discussed in Sec. VIII.A. The formal analogy with the Aharonov-Bohm (1959) phase—in which the curvature is just the ordinary magnetic field—sheds new light on dielectric polarization as a fundamental quantum phenomenon. In Eq. (45) we get the macroscopic linear response as a basic phase feature of the electronic ground state.

VII. NUMERICAL CONSIDERATIONS

The application of the geometric phase approach to actual calculations requires the evaluation of $\mathbf{q}$ gradients of the Kohn-Sham eigenstates, as in Eq. (25), or equivalently of $\boldsymbol{\xi}$ gradients, as in Eqs. (29), (34), or (36). At first sight, first-order $\mathbf{q}\cdot\mathbf{p}$ perturbation theory (see, for example, Bassani and Pastori Parravicini, 1975) would appear as the most natural tool. In fact, this is not the case. On quite general grounds, perturbation theory can be safely used in evaluating gauge-invariant quantities, but it is useless for the gauge-dependent ones (Mead and Truhlar, 1979), as is the differential phase $d\varphi$. This can be shown as follows. Let us consider the nth eigenstate of $H(\boldsymbol{\xi})$, Eq. (8), whose eigenvalue is $E_n(\boldsymbol{\xi})$; standard first-order perturbation theory yields

$$|u_n(\boldsymbol{\xi}+\Delta\boldsymbol{\xi})\rangle \simeq |u_n(\boldsymbol{\xi})\rangle + \sum_{m\neq n} |u_m(\boldsymbol{\xi})\rangle \frac{\langle u_m(\boldsymbol{\xi})|H(\boldsymbol{\xi}+\Delta\boldsymbol{\xi})-H(\boldsymbol{\xi})|u_n(\boldsymbol{\xi})\rangle}{E_n(\boldsymbol{\xi})-E_m(\boldsymbol{\xi})} \tag{46}$$

$$|\nabla_{\boldsymbol{\xi}} u_n(\boldsymbol{\xi})\rangle = \sum_{m\neq n} |u_m(\boldsymbol{\xi})\rangle \frac{\langle u_m(\boldsymbol{\xi})|\nabla_{\boldsymbol{\xi}} H(\boldsymbol{\xi})|u_n(\boldsymbol{\xi})\rangle}{E_n(\boldsymbol{\xi})-E_m(\boldsymbol{\xi})}. \tag{47}$$

Use of this gradient in Eqs. (29) and (36) provides a vanishing connection (and differential phase) at any $\boldsymbol{\xi}$. The apparent paradox is solved by recognizing that Eq. (46) fixes a particular gauge, corresponding to the so-called "parallel transport" (Berry, 1989). Within this gauge, the phase of the $|u_n(\boldsymbol{\xi})\rangle$ state is in general *multiple valued* for a cyclic evolution in parameter space. At a given $\boldsymbol{\xi}+\Delta\boldsymbol{\xi}$, the perturbed state is undetermined by an arbitrary phase. A continuous single-valued behavior can be recovered upon multiplying the right-hand member of Eq. (46) by a nonintegrable phase (linear in $\Delta\boldsymbol{\xi}$). This phase provides the only nonvanishing contribution to the connection, but perturbation theory is useless in determining it.

The successful numerical strategy for coping with geometric phases is direct discretization of the line integrals. By this I mean performing *both* the gradient and the integration entering the geometric phase expression over a discrete mesh. I illustrate discretization of the Berry phase in Eq. (37), and for the most general closed path, schematically shown in Fig. 2. We take a discrete set of N contiguous points $\boldsymbol{\xi}_s$ on the path, with $s=0,\ldots,N-1$; we further define $\boldsymbol{\xi}_N=\boldsymbol{\xi}_0$, whereas it is understood that the eigenstates at $\boldsymbol{\xi}_N$ and at $\boldsymbol{\xi}_0$ are *the same* (same phases, same n ordering). A simple-minded discretization yields

$$\oint d\varphi \simeq \sum_{s=0}^{N-1} \Delta\varphi_s, \tag{48}$$

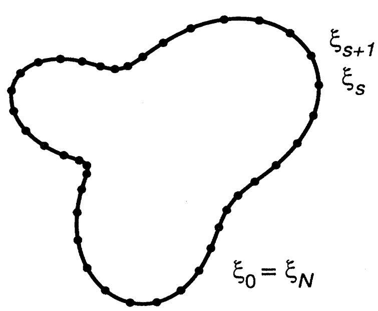

FIG. 2. Discretization of the circuit integral for numerical evaluation of Berry phases: an arbitrary path in $\boldsymbol{\xi}$ space is shown.

where $\Delta\varphi_s$ is the phase difference between $\boldsymbol{\xi}_{s+1}$ and $\boldsymbol{\xi}_s$. Such discretization is safe only if the phase varies smoothly from point to point. This is far from being the case. The approximated eigenstates are in fact usually obtained from numerical diagonalization of the Hamiltonian, Eq. (8), over a finite basis. The gauge is thus arbitrarily chosen by the diagonalization routine, and the behavior of the phase is erratic; valence-band crossings along the path are a further source of nonsmoothness. A stable algorithm must therefore be *numerically* gauge invariant, in the sense that arbitrary fluctuations of the gauge phase do not affect the result; this is the case for the algorithm proposed by King-Smith and Vanderbilt, based on the Pancharatnam phase in its determinant form, Eq. (33). When we use this equation, the discretization becomes

$$\Delta\varphi_s = \text{Im}\ln\det S(\boldsymbol{\xi}_s,\boldsymbol{\xi}_{s+1}), \tag{49}$$

$$\oint d\varphi \simeq \text{Im}\ln \prod_{s=0}^{N-1} \det S(\boldsymbol{\xi}_s,\boldsymbol{\xi}_{s+1}), \tag{50}$$

which is obviously correct if the diagonalization routine is gentle enough to provide a smooth phase. A nasty routine—or even an ordinary one—will instead provide the overlap matrix $\tilde{S}(\boldsymbol{\xi}_s,\boldsymbol{\xi}_{s+1}) = U^{-1}(\boldsymbol{\xi}_s)S(\boldsymbol{\xi}_s,\boldsymbol{\xi}_{s+1})U(\boldsymbol{\xi}_{s+1})$, where the U's are unitary random matrices. The effect of these matrices on the determinant of $S(\boldsymbol{\xi}_s,\boldsymbol{\xi}_{s+1})$ is a multiplication by the overall gauge phase $\exp i(\vartheta_{s+1}-\vartheta_s)$. One can immediately verify that the gauge phases cancel in the cyclic product of Eq. (50). Therefore, despite wild fluctuations of the factors in Eq. (50) from point to point, their cyclic product is *numerically* gauge invariant and the discretization of the circuit integral stable. This basic property is *not* shared by a discretization of the trace expression, Eq. (29).

The calculation of a given component of $\Delta\mathbf{P}$, starting from Eq. (38), proceeds as follows. One evaluates the surface average over the (ξ_1,ξ_2) unit square on a finite mesh. For each of the chosen (ξ_1,ξ_2) points, one then considers the loop integral over the circuit shown in Fig. 1. As explained in the previous section, the vertical sides do not contribute. In several cases of interest, the bottom horizontal side does not contribute either. This occurs whenever the crystal Hamiltonian is centrosymmetric at $\xi_4=0$. It is therefore enough to evaluate the line integral over the top horizontal side at $\xi_4=1$, i.e., using the self-consistent Kohn-Sham Hamiltonian of the final state, which is evaluated within the local-density approximation. This Hamiltonian is diagonalized over a discrete mesh on the relevant segment, i.e., at the N points

$$\boldsymbol{\xi}_s = (\xi_1, \xi_2, s/N, 1) \ , \ s = 0, \ldots, N-1 \ , \tag{51}$$

and then the overlap matrices $S(\boldsymbol{\xi}_s, \boldsymbol{\xi}_{s+1})$ between the $\bar{n}$ occupied u_n orbitals are evaluated. The important point is that the orbitals at $s=0$ and $s=N$ (whose $\mathbf{q}$ vectors differ by $\mathbf{G}_3$) must *not* be obtained from independent diagonalizations. The basic relationship of Eq. (9) must be used instead to get the orbitals at $s=N$ from the corresponding ones at $s=0$. The line integral is finally computed from Eq. (50), as discussed above.

VIII. INDUCED POLARIZATION

A. Linear-response theory

The curvature $\mathcal{Y}$, Eqs. (41) and (44), is a gauge-invariant quantity. Therefore the $\boldsymbol{\xi}$ derivatives entering it can be safely evaluated via perturbation theory and in a given gauge. Following again the derivation of Berry (1984), and using Eq. (47), one gets

$$\mathcal{Y}_{ij}(\boldsymbol{\xi}) = 2 \,\mathrm{Im} \sum_{n=1}^{\bar{n}} \sum_{m=\bar{n}+1}^{\infty} \frac{\langle u_n(\boldsymbol{\xi}) | \partial H(\boldsymbol{\xi})/\partial \xi_i | u_m(\boldsymbol{\xi}) \rangle \langle u_m(\boldsymbol{\xi}) | \partial H(\boldsymbol{\xi})/\partial \xi_j | u_n(\boldsymbol{\xi}) \rangle}{[E_m(\boldsymbol{\xi}) - E_n(\boldsymbol{\xi})]^2} . \tag{52}$$

It has already been observed that the curvature provides, after Eq. (45), the λ derivative of the macroscopic polarization. Within density-functional theory (Lundqvist and March, 1983), the linear response is a property of the electronic ground state, involving the occupied Kohn-Sham orbitals only. This feature is evident in Eq. (41). In contrast, this same feature is somewhat obscured in the equivalent expression, Eq. (52), which apparently depends on the empty orbitals as well.

Expressions for evaluating polarization derivatives have been known for several years, having been obtained in other ways than the present geometric phase approach. In order to make contact with the more traditional linear-response theory and to show the equivalence explicitly, it proves better to switch back to the $(\mathbf{q},\lambda)$ variables. Using Eqs. (8) and (28) one gets

$$\frac{\partial H(\boldsymbol{\xi})}{\partial \xi_j} = \frac{\hbar}{m_e} \mathbf{G}_j \cdot (\mathbf{p} + \hbar \mathbf{q}) \ , \ j = 1, \ldots, 3, \tag{53}$$

$$\frac{\partial H(\boldsymbol{\xi})}{\partial \xi_4} = \frac{\partial V^{(\lambda)}(\mathbf{r})}{\partial \lambda}, \tag{54}$$

whence Eqs. (45) and (52) read

$$\mathbf{P}'_{\mathrm{el}}(\lambda) = \frac{4\hbar e}{(2\pi)^3 m_e} \mathrm{Im} \sum_{n=1}^{\bar{n}} \sum_{m=\bar{n}+1}^{\infty} \int_{\mathrm{BZ}} d\mathbf{q} \, \frac{\langle u_n^{(\lambda)}(\mathbf{q}) | \mathbf{p} | u_m^{(\lambda)}(\mathbf{q}) \rangle \langle u_m^{(\lambda)}(\mathbf{q}) | \partial V^{(\lambda)}/\partial \lambda | u_n^{(\lambda)}(\mathbf{q}) \rangle}{[E_m^{(\lambda)}(\mathbf{q}) - E_n^{(\lambda)}(\mathbf{q})]^2}, \tag{55}$$

which indeed coincides with the standard linear-response expression for the macroscopic polarization, as reported, for example, by Resta (1992). This same expression was previously derived from a first-order perturbative expansion of the occupied orbitals. Owing to time-reversal ($\mathbf{q} \to -\mathbf{q}$) symmetry, the Brillouin-zone integral in Eq. (55) is purely imaginary. Macroscopic linear-response tensors involve explicitly the Kohn-Sham orbitals (as opposed to the density). In the context of the present geometric phase approach, these tensors are obtained as Brillouin-zone integrals of the curvature, and therefore assume the meaning of a gauge-invariant phase feature of the Kohn-Sham orbitals.

An expression like Eq. (55) was first proposed by Vogl (1978) in order to deal with the polarization induced by zone-center transverse-optic phonons in polar crystals; the more general case of an arbitrary—albeit cell-conserving—transformation of the Hamiltonian is considered by Resta (1992), who gives a straightforward proof. Here I give an alternate proof, which emphasizes the meaning of Eq. (4) as the integrated macroscopic current induced by the adiabatic transformation. Suppose we add a small time-dependent (and lattice-periodic) perturbation to the Hamiltonian of Eq. (8), i.e.,

$$H^{(\lambda)}(\mathbf{q}) \longrightarrow H^{(\lambda)}(\mathbf{q}) + \mathrm{Re}\, \delta V(\omega) e^{i\omega t}. \tag{56}$$

The induced electronic current is then

$$\mathbf{j}_{\mathrm{el}}(\omega) \simeq \frac{2\hbar e}{(2\pi)^3 m_e} \times \mathrm{Re} \sum_{n=1}^{\bar{n}} \int_{\mathrm{BZ}} d\mathbf{q} \, \langle u_n^{(\lambda)}(\mathbf{q}) | \, (\mathbf{p} + \hbar \mathbf{q}) \, | \delta u_n^{(\lambda)}(\mathbf{q}, \omega) \rangle. \tag{57}$$

First-order perturbation theory (see Landau and Lifshitz, 1977), followed by straightforward manipulations, yields

$$\mathbf{j}_{\rm el}(\omega) \simeq \frac{2\hbar e}{(2\pi)^3 m_e} {\rm Re} \sum_{n=1}^{\bar{n}} \sum_{m=\bar{n}+1}^{\infty} \int_{\rm BZ} d\mathbf{q} \left[\frac{\langle u_n^{(\lambda)}(\mathbf{q})|\mathbf{p}|u_m^{(\lambda)}(\mathbf{q})\rangle \langle u_m^{(\lambda)}(\mathbf{q})|\delta V(\omega)|u_n^{(\lambda)}(\mathbf{q})\rangle}{E_m^{(\lambda)}(\mathbf{q}) - E_n^{(\lambda)}(\mathbf{q}) + \hbar\omega} + \frac{\langle u_m^{(\lambda)}(\mathbf{q})|\mathbf{p}|u_n^{(\lambda)}(\mathbf{q})\rangle \langle u_n^{(\lambda)}(\mathbf{q})|\delta V(\omega)|u_m^{(\lambda)}(\mathbf{q})\rangle}{E_m^{(\lambda)}(\mathbf{q}) - E_n^{(\lambda)}(\mathbf{q}) - \hbar\omega} \right]. \tag{58}$$

Taking then the static ($\omega \to 0$) limit of $\delta\mathbf{P}_{\rm el}(\omega) = \mathbf{j}_{\rm el}(\omega)/i\omega$, and identifying $\delta V(0)$ with $\delta\lambda\ \partial V^{(\lambda)}/\partial\lambda$, one gets immediately Eq. (55).

In practical implementations with modern nonlocal pseudopotentials (Pickett, 1989) an extra term appears in the expression for the current, Eq. (57), and hence in Eq. (55) as well. The velocity in this case is in fact

$$\mathbf{v} = \frac{i}{\hbar}[H^{(\lambda)}(\mathbf{q}),\mathbf{r}] = \frac{1}{m_e}(\mathbf{p}+\hbar\mathbf{q}) + \frac{i}{\hbar}[V^{(\lambda)},\mathbf{r}]. \tag{59}$$

It has been demonstrated by Baroni and Resta (1986) that the matrix elements of this extra term are well defined and do not cause any harm (see also Hybertsen and Louie, 1987; Giannozzi *et al.*, 1991).

An alternative linear-response method, due to Baroni, Giannozzi, and Testa (1987), has become fashionable recently. This is usually called density-functional perturbation theory, and its applications to semiconductor physics are performed within the local-density approximation to density-functional theory (Lundqvist and March, 1983), in a pseudopotential framework (Pickett, 1989). A somewhat different implementation of this approach has been developed by Gonze *et al.* (1992). The basic idea is the same as in the "direct" self-consistent methods, which are well known in atomic (Sternheimer, 1954, 1957, 1959, 1969, 1970; Mahan, 1980) and molecular (Dalgarno, 1962; Amos, 1987) physics. The density-functional perturbation theory directly provides the self-consistent λ derivatives of the occupied Kohn-Sham orbitals. Upon transforming Eqs. (41) and (45) into the $(\mathbf{q},\lambda)$ variables one gets

$$\mathbf{P}'_{\rm el}(\lambda) = \frac{4e}{(2\pi)^3} {\rm Im} \sum_{n=1}^{\bar{n}} \int_{\rm BZ} d\mathbf{q}\ \langle \nabla_{\mathbf{q}} u_n^{(\lambda)}(\mathbf{q})| \frac{\partial}{\partial\lambda} u_n^{(\lambda)}(\mathbf{q})\rangle. \tag{60}$$

The $\mathbf{q}$ gradient could be evaluated via perturbation theory, but it is preferable to avoid the occurrence of slowly convergent perturbation sums. One writes the projector over the empty states as

$$P_c(\mathbf{q}) = 1 - \sum_{n=1}^{\bar{n}} |u_n^{(\lambda)}(\mathbf{q})\rangle\langle u_n^{(\lambda)}(\mathbf{q})|, \tag{61}$$

and in terms of it Eq. (60) is easily transformed to

$$\mathbf{P}'_{\rm el}(\lambda) = \frac{4\hbar e}{(2\pi)^3 m_e} {\rm Im} \sum_{n=1}^{\bar{n}} \int_{\rm BZ} d\mathbf{q}\ \langle u_n^{(\lambda)}(\mathbf{q})|\mathbf{p}\ P_c(\mathbf{q})\ [H(\mathbf{q}) - E_n(\mathbf{q})]^{-1}\ P_c(\mathbf{q})| \frac{\partial}{\partial\lambda} u_n^{(\lambda)}(\mathbf{q})\rangle. \tag{62}$$

This expression coincides with the finding of Baroni *et al.* (1987) for the macroscopic response. The Green's function appearing in Eq. (62) is *not* explicitly calculated, and its relevant matrix elements are evaluated via solution of linear systems; for a detailed account, see Giannozzi *et al.* (1991).

B. Macroscopic electric fields

Whenever a macroscopic electric field is present inside the dielectric, the Kohn-Sham orbitals no longer have the Bloch form, and the whole geometric phase approach does not apply. On the other hand, all of the various implementations of linear-response theory *do* allow the study of the polarization induced by a macroscopic field, or even induced by a different source and accompanied—because of the chosen boundary conditions—by a field. This has been well known since the early work with dielectric matrices—reviewed, for example, by Baldereschi and Resta (1983)—in which appropriate $\mathbf{q} \to 0$ limits of nonanalytic dielectric-matrix elements solve the problem.

The way in which linear-response theory copes with macroscopic fields can be easily illustrated starting from the formulation given above. Let us consider Eq. (55), where we identify the parameter λ with a field $\boldsymbol{\mathcal{E}}$. In this case $\partial V/\partial\boldsymbol{\mathcal{E}}$ includes a macroscopic term equal to $-e\mathbf{r}$, in addition to a periodic (so-called local-field) term. Although $\mathbf{r}$ is *not* a lattice-periodical operator, its off-diagonal matrix elements appearing in Eq. (55) can be easily evaluated in boundary-insensitive form—using the velocity operator, Eq. (59)—at the price of an extra energy factor in the denominator of Eq. (55), or equivalently of an extra Green's function in Eq. (62). It is further worth pointing out that density-functional perturbation theory—in its most recent implementations (Giannozzi *et al.*, 1991)—exploits an additional appealing feature: the (screened) macroscopic field $\boldsymbol{\mathcal{E}}$ may be used as an *explicitly adjustable* boundary condition for solving Poisson's equation. Therefore one may choose to perform the iterative calculation to self-consistency (say for a zone-center optic phonon) either in a null field or in a depolarizing field. The former case is transverse, and the latter is longitudinal. A further possible choice is to assign a nonzero constant field, which does not vary during the iteration process, and to calculate the electronic ground state in this field self-consistently (to linear order in the field magnitude).

The theory presented here allows us to evaluate polarization differences—due to adiabatic variations of a parameter λ in the crystal Hamiltonian—in a null field. Suppose instead that we are interested in the same crystal transformation, but in a field. The key quantity to consider (Landau and Lifshitz, 1984) is then the thermodynamic potential $\tilde{F}(\lambda, \boldsymbol{\mathcal{E}})$, in which the field $\boldsymbol{\mathcal{E}}$ is regarded as an independent variable (or boundary condition). For instance, if λ is identified with macroscopic strain, then $\tilde{F}$ coincides with the (zero-temperature) electric enthalpy defined, for example, in Chapter 3 of Lines and Glass (1977). The most general expansion of $\tilde{F}$ to second order in $\boldsymbol{\mathcal{E}}$, and to all orders in λ, reads

$$\tilde{F}(\lambda, \boldsymbol{\mathcal{E}}) \simeq \tilde{F}(\lambda, 0) - \mathbf{P}(\lambda)\, \boldsymbol{\mathcal{E}} - \frac{1}{8\pi} \boldsymbol{\mathcal{E}}\, \varepsilon(\lambda)\, \boldsymbol{\mathcal{E}}, \qquad (63)$$

where $\varepsilon(\lambda)$ is the macroscopic dielectric tensor and $\mathbf{P}(\lambda)$ is the macroscopic polarization in zero field. The latter is defined only modulo the arbitrary additive constant vector $\mathbf{P}(0)$, which depends on sample termination and does not affect any bulk property.

The generalized force f and the electric displacement $\boldsymbol{\mathcal{D}}$ are obtained from Eq. (63) as conjugate variables:

$$f(\lambda, \boldsymbol{\mathcal{E}}) = -\frac{\partial}{\partial \lambda} \tilde{F}(\lambda, \boldsymbol{\mathcal{E}})$$

$$= -\frac{\partial}{\partial \lambda} \tilde{F}(\lambda, 0) + \mathbf{P}'(\lambda)\, \boldsymbol{\mathcal{E}} + \frac{1}{8\pi} \boldsymbol{\mathcal{E}}\, \varepsilon'(\lambda)\, \boldsymbol{\mathcal{E}}, \qquad (64)$$

$$\boldsymbol{\mathcal{D}}(\lambda, \boldsymbol{\mathcal{E}}) = -4\pi \nabla_{\mathcal{E}} \tilde{F}(\lambda, \boldsymbol{\mathcal{E}}) = \varepsilon(\lambda)\, \boldsymbol{\mathcal{E}} + 4\pi \mathbf{P}(\lambda)\ . \qquad (65)$$

The second expression relates the macroscopic polarization in a field to the one in zero field as

$$\mathbf{P}(\lambda, \boldsymbol{\mathcal{E}}) = \mathbf{P}(\lambda) + \chi(\lambda)\, \boldsymbol{\mathcal{E}}, \qquad (66)$$

where the macroscopic polarizability tensor $\chi = (\varepsilon - 1)/4\pi$ has been used. In a bulk solid, the macroscopic field does *not* depend on the local charge density. On the contrary, it is an arbitrary boundary condition for the Poisson equation, which can often be controlled by the experimental setup. Throughout this work we have used the "transverse" boundary conditions, i.e., $\boldsymbol{\mathcal{E}}=0$; another interesting case of Eq. (66) is when the adiabatic transformation of the Hamiltonian is performed imposing "longitudinal" boundary conditions on the sample, i.e., $\Delta\boldsymbol{\mathcal{E}} = -4\pi\Delta\mathbf{P}$.

Insofar as the second-order expansion in $\boldsymbol{\mathcal{E}}$—Eq. (63)—is justified, the geometric phase approach can be used even to study polarization in macroscopic fields (to all orders in λ), provided the macroscopic polarizability tensor $\chi(\lambda)$ of the dielectric is available by other means (typically from linear-response theory).

C. Born effective charges

The Born (or transverse) effective charge tensors measure by definition the macroscopic polarization linearly induced by a unit sublattice displacement in a null electric field (Pick *et al.*, 1970; see also Pick and Takemori, 1986). These tensors represent therefore the simplest application of the formal results discussed in this work. When λ is identified with a suitable phonon coordinate, the Born effective charge tensors are obtained from the polarization derivative $\mathbf{P}'(\lambda_{\text{eq}})$, where λ_{eq} is the equilibrium value, i.e., the minimum of $\tilde{F}(\lambda, 0)$.

In the past, these tensors have been evaluated either from Eq. (55) or from more complex linear-response techniques, typically involving the calculation of dielectric matrices in the small-$\mathbf{q}$ limit (Baldereschi and Resta, 1983). On a few occasions, supercell calculations have also been performed in order to evaluate the effective charges (Kunc, 1985). In more recent times, most calculations of the effective charge tensors in semiconductors are performed within the density-functional perturbation theory of Baroni, Giannozzi, and Testa (1987), using the local-density approximation. For systematic applications to lattice-dynamical problems see de Gironcoli *et al.* (1989, 1990), Giannozzi *et al.* (1991), Gonze *et al.* (1992), and Dal Corso *et al.* (1993a, 1993b). Within such an approach, the effective charges can be evaluated (and have indeed been evaluated) in several alternative ways. One choice is to calculate the perturbed ground state in zero field and to compute Eq. (62) after such perturbed wave functions. This gives directly the Born effective charges. A second choice—in fact the original one of Baroni *et al.*—is to perform the self-consistent calculation for the perturbed crystal in a depolarizing field. One calculates in this way the *longitudinal* polarization; a similar calculation provides the macroscopic dielectric tensor, whence the Born (alias transverse) effective charges are easily evaluated.

A third choice is to exploit Eq. (64), where the Born effective charge tensors appear as the forces linearly induced on the ions by a macroscopic field, at vanishing phonon amplitude (λ=λ_{eq}). One then performs the self-consistent calculation for the perturbed solid in a given field and with no ionic displacements: the forces on the ions are finally evaluated from the Hellmann-Feynman theorem (Feynman, 1939; Deb, 1973; Kunc, 1985).

Linear response is a powerful tool, but it requires specialized computer codes and, furthermore, is easily implemented only in a pseudopotential scheme (Pickett, 1989), using a plane-wave basis set. This fact has hindered first-principle calculations of the effective charge tensors in many interesting materials, where different basis sets are typically used in order to get state-of-the-art results. In contrast, the geometric phase approach requires standard ground-state calculations for the solid with "frozen-in" phonons. All that must be evaluated additionally are the overlap matrices between occupied orbitals at the neighboring points of a suitable grid in reciprocal space, as explained at the end of Sec. VII. The polarization derivatives are obtained as finite differences. When the problem can be studied in both ways—and all *technical* ingredients are kept the same—the two approaches provide identical results within computational noise. Some examples have been published by Dal Corso *et al.* (1993b).

The materials in which first-principles access to the effective charge tensors is most badly needed are probably the perovskites oxides, whose cubic paraelectric phase is illustrated in Fig. 3. Since the early work of Slater (1950) the effective charge tensors have been expected to be relevant for understanding the ferroelectric instability in these materials; classical models (e.g., Axe, 1967) predict highly nontrivial values of the effective charge tensors. In ferroelectric perovskites delocalized electrons are present (Cohen, 1992). For the reasons given in Sec. II, no estimate—even rough—of the effective charges is possible without a quantum treatment of the electronic system. Experiment is not very informative either, since only partial data are available via Raman spectroscopy (for a recent outline of the problems, see Dougherthy *et al.*, 1992).

It happens that the constituents of ferroelectric perovskites are "unfriendly" atoms (in a computational physics sense), such as oxygen and transition metals. Several very informative first-principles studies of these materials exist in the literature, using basis sets more complex than the plane waves. I cite here a paper of Cohen (1992) as a single example. Nonetheless, no quantum calculation of the Born effective charge tensors in a ferroelectric perovskite was available until the advent of the geometric phase approach. The first such calculation, performed for $KNbO_3$, is due to Resta *et al.* (1993a); a few technical details are given below in Sec. IX, when dealing with spontaneous polarization. In the paraelectric phase, K and Nb sites have cubic symmetry, and the effective charge tensors are isotropic; their calculated values are $Z^*_K = 0.8$ and $Z^*_{Nb} = 9.1$. The O ions sit at

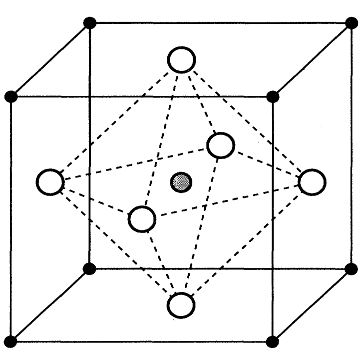

FIG. 3. Cubic perovskite structure, with general formula ABO_3, where A is a mono- or divalent metal (solid circles) and B is a tetra- or pentavalent metal (shaded circle). The oxygens (empty circles) form octahedral cages, with B at their centers, and arranged in a simple cubic pattern. The calculations reviewed here are for $KNbO_3$.

noncubic sites, and the effective charge tensor has two independent components: one (Z^*_{O1}) relative to displacements pointing towards the Nb ion, and the other (Z^*_{O2}) for displacements in the orthogonal plane. The calculated values are $Z^*_{O1} = -6.6$ and $Z^*_{O2} = -1.7$. These first-principles data demonstrate strong asymmetry of the O effective charge tensor and large absolute values of Z^*_{Nb} and Z^*_{O1}. The latter fact indicates that relative displacements of neighboring O and Nb ions against each other trigger highly polarizable electrons. Roughly speaking, a large, nonrigid, delocalized charge is responsible for both Z^*_{Nb} and Z^*_{O1}.

D. Piezoelectricity

The piezoelectric tensor is defined as the polarization derivative with respect to strain, when the macroscopic field is kept vanishing. In a milestone paper, Martin (1972a) proved that piezoelectricity is a well defined bulk property, independent of surface termination. Notwithstanding, Martin's proof was challenged, and the debate lasted until recent times (Martin, 1972b, Woo and Landauer, 1972, Landauer, 1981, 1987; Kallin and Halperin, 1984; Tagantsev, 1991).

Using the formulation of the present paper, the main reason why piezoelectricity looks like a difficult problem is that Eq. (55) does not apply. Indeed, $\partial V^{(\lambda)}/\partial\lambda$ is *not* a lattice-periodical operator when λ is identified with the macroscopic strain. In 1989, de Gironcoli *et al.* found an alternative path for *ab initio* studies of piezoelectricity in real materials (the case studied was III-V semiconductors). The calculations performed therein are lattice-periodical and boundary-insensitive, therefore providing further evidence (if any was needed) that piezoelectricity is a bulk effect. The key idea—using the electric enthalpy $\tilde{F}(\lambda, \boldsymbol{\mathcal{E}})$ of Eq. (63)—is to exploit Eq. (64). Since the conjugate variable to strain is macroscopic stress, the piezoelectric response appears therein as the stress linearly induced by unit field at zero strain ($\lambda = \lambda_{eq}$). Starting with this definition, de Gironcoli *et al.* use density-functional perturbation theory to evaluate the linear change in the eigenfunctions induced by a macroscopic field, as outlined in Sec. VIII.B; they then compute the linear change in macroscopic stress, using the stress theorem of Nielsen and Martin (1983, 1985).

Within the geometric phase approach the (linear and nonlinear) piezoelectric coefficients are accessible—via finite differences—much in the same way as are the Born effective charges. It is enough to compare ground-state calculations performed at different shapes and volumes of the unit cell. This poses no problem, and the approach applies almost as it stands, as discussed here at the end of Sec. V. Indeed, King-Smith and Vanderbilt in their original paper use the linear-response piezoelectric constant of GaAs—calculated by de Gironcoli *et al.* (1989)—as a benchmark. Since the technical ingredients are not the same, they find a 20% disagreement, which is not a serious drawback. The final figure results in fact from a large cancellation of two terms, which are separately computed. A. Dal Corso (unpublished calculation) has performed an independent check: the calculated values of the piezoelectric constant of GaAs—via the two different approaches—disagree by no more than 3% when all technical ingredients are kept the same. Other examples have been published by Dal Corso *et al.* (1993b), who also perform the first *ab initio* study of nonlinear piezoelectricity (the case study is CdTe, which has experimental interest for strained-layer superlattices).

IX. SPONTANEOUS POLARIZATION IN FERROELECTRICS

The geometric phase approach, as formulated throughout this work, deals with the polarization difference $\Delta\mathbf{P}$ for a couple of arbitrary initial and final states, in a general crystal. Suppose now that the initial (λ=0) state corresponds to a highly symmetric crystal structure, such as the typical prototype (or aristotype) structure of a ferroelectric material (Lines and Glass, 1977). In this structure any bulk vector property is symmetry forbidden, as is the case with centrosymmetric and tetrahedral solids. The polarization $\mathbf{P}(0)$ is then zero. This looks like a useful convention (on crystal termination) more than a physical statement, since the "absolute" bulk electric polarization has never been measured. A typical experiment—performed via a hysteresis cycle—measures in fact only an integrated current, which coincides with the polarization difference between two enantiomorphous ferroelectric crystal states. The present approach provides theoretical access to precisely this kind of observable.

Once the above symmetry-based convention is assumed, the prototype structure can be taken as a reference, and the spontaneous polarization of the low-symmetry structures can be defined through the difference. This is unambiguously possible under two conditions: (i) there must exist a continuous adiabatic transformation of the Kohn-Sham Hamiltonian which relates the initial and final states in such a way that the crystal remains insulating throughout the transformation; and (ii) the difference in polarization between the final and initial states must be smaller than the polarization quanta, Eq. (21). Under these hypotheses, the polarization of the final state is—according to the expressions given in this work—independent of the particular path chosen in parameter space.

The wave functions of the reference state can be eliminated from the formalism, through a choice of origin and phases such that the λ=0 contribution to Eqs. (15) and (26) vanishes. For centrosymmetric prototype crystals this is realized by choosing real u_n wave functions, which imply a vanishing geometric phase. Strictly speaking—as remarked by Zak (1989)—the geometric phase in the centrosymmetric case is either 0 or π (modulo 2π). The latter occurrence has never been found in the cases studied so far and would anyhow have little practical effect within the present approach. Incidentally, it is worth mentioning that the occurrence of the value of π for the geometric phase in a system having real wave functions is well known in molecular physics (Mead and Truhlar, 1979; Mead, 1992). After eliminating the reference state, one gets an expression for the spontaneous electronic polarization, which can be evaluated using the wave functions of the low-symmetry structure as the only ingredients. Since the reference state can be eliminated from the formalism, it looks as if the polarization difference $\Delta\mathbf{P}_{\mathrm{el}}$ was measured with respect to an "internal" reference, no longer depending on any explicit choice of reference system. Such a viewpoint is incorrect: only the *total* difference $\Delta\mathbf{P}$ is a macroscopic (i.e., translationally invariant) observable, owing to charge neutrality. Since the partition of $\Delta\mathbf{P}$ into an electronic and an ionic term is nonunique—notably when the prototype crystal has several centrosymmetric sites in the cell—one must always consider both terms together.

The paradigmatic materials in which it is relevant to investigate spontaneous polarization are the ferroelectric perovskites, having a cubic prototype phase above the Curie temperature and displaying a series of structural transitions to low-symmetry ferroelectric phases when temperature is lowered. Typically, the first transition is to a tetragonal phase, characterized by a small uniaxial macroscopic strain accompanied by microscopic displacements of the ions out of their high-symmetry sites. The latter distortion—henceforth called internal strain—determines a preferred polarity of the tetragonal axis and is responsible for the occurrence of spontaneous polarization. This is illustrated in Fig. 4 for the specific example of $KNbO_3$, which has been studied by Resta *et al.* (1993a, 1993b) via the geometric phase approach. The main features of this calculation, and some of the results are discussed in the remainder of this section. No study of the spontaneous polarization of a ferroelectric material—based on quantum mechanics in any form—has been available up until now.

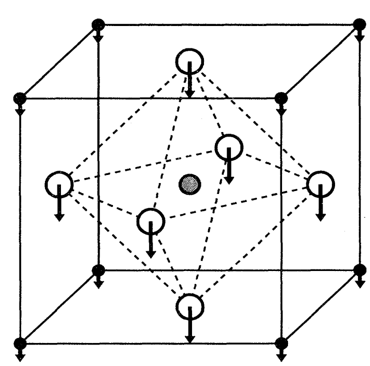

FIG. 4. Centrosymmetric tetragonal structure of $KNbO_3$, with c/a = 1.017, taken as the (λ=0) reference structure. Solid, shaded, and empty circles represent K, Nb, and O atoms, respectively. Internal displacements (indicated by arrows, and magnified by a factor of 4) transform the reference structure into the ferroelectric (λ=1) structure.

Within the present approach, the material is studied in a "frozen-ion" structure. The parameters of the ferroelectric (λ=1) structure are taken from the experimental crystallographic data, measured at finite temperature. As for the reference (λ=0) structure, the obvious choice is a tetragonal structure in which the internal strain is taken as vanishing, and whose primitive cell is the same as for the ferroelectric structure. In this material, the internal strain leaves the oxygen cage almost undistorted, while the two cation sublattices undergo different displacements with respect to it; this is shown in Fig. 4, where the origin has been conventionally fixed at the Nb site. The adiabatic transformation of the Hamiltonian is cell conserving by construction for all λ's. The reciprocal cell is rectangular. Therefore the King-Smith and Vanderbilt expression, Eq. (31), can be used to evaluate $\Delta\mathbf{P}_{el}$, where $\mathbf{R}_3$ is chosen along the polarization axis. Since the λ=0 reference structure is centrosymmetric, only the line integrals along C_1 are explicitly needed, as explained above. Both the line integral and the two-dimensional (ξ_1,ξ_2) average are performed on a discrete mesh, as explained in Sec. VII. The Kohn-Sham occupied wave functions entering the overlap matrix, Eq. (32), are obtained by Resta *et al.* (1993a, 1993b) within the local-density approximation from the full-potential linearized augmented-plane-wave (FLAPW) method, as implemented by Jansen and Freeman (1984).

The calculation provides for the (ξ_1,ξ_2)-averaged Berry phase the value of -3.95, modulo 2π. Indeed this value, shown in Fig. 5(a), solid line, is definitely *not* much smaller than 2π and seems to leave much ambiguity. One has to bear in mind, however, that the genuine macroscopic observable is $\Delta\mathbf{P}$ rather than $\Delta\mathbf{P}_{el}$. The ionic term $\Delta\mathbf{P}_{ion}$ can be converted in phase units using the obvious recipe $\gamma_{ion} = \Omega\mathbf{G}_3\Delta\mathbf{P}_{ion}/2e$, analogous to Eq. (38), and then added to the Berry phase. Amongst the possible quantized values of the *total* (electronic plus ionic) phase, the one leading to the minimum $|\Delta\mathbf{P}|$ is shown in Fig 5(a), shaded sector. Its value is -1.11, i.e., -63.5 degrees, which can be considered much smaller than 2π. As a check of the correct choice of the quantized phase, Resta *et al.* have performed independent calculations with the internal strain scaled to smaller values, obtaining a total phase that monotonically decreases towards zero. It is also worth recalling that the partition in electronic and ionic terms is nonunique: if the origin is kept fixed at the K site instead of at Nb, the corresponding phases are those shown in Fig 5(b).

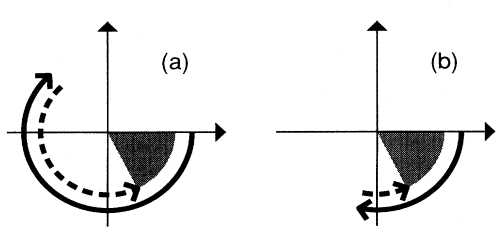

FIG. 5. Berry phase in ferroelectric $KNbO_3$ (solid line); classical ionic contribution (dashed line); total phase, due to electrons and ions (shaded sector). (a) The internal distortion is performed while keeping the origin at the Nb site, as in Fig. 4. (b) The origin is fixed at the K site.

The Berry phase calculation provides for the spontaneous polarization of $KNbO_3$ the value $|\Delta\mathbf{P}| = 0.35$ C/m^2, to be compared with the most recent experimental figure of 0.37 by Kleeman *et al.* (1984). This kind of agreement could appear embarrassing, particularly given the fact (Edwardson, 1989; Dougherty *et al.*, 1992) that a real ferroelectric at finite temperature looks rather different from the frozen-ion schematization of the theoretical approach. Indeed, the agreement is *not* embarrassing at all, since Resta *et al.* have demonstrated that the polarization in this material is *linear* in the ferroelectric distortion (i.e., in λ). This fact implies that the time-averaged polarization can be safely computed from a frozen crystal structure, where time-averaged crystallographic data are used. Linearity is a nontrivial finding, given that ferroelectricity is essentially a nonlinear phenomenon; furthermore, it is worth recalling that the accepted theory of the pyroelectric effect, due to Born (1945), crucially depends on the assumption that the polarization is *nonlinear* in the ionic displacements.

X. CONCLUSIONS

This paper describes a modern theory of macroscopic polarization in crystalline solids. The dielectric behavior of a solid is essentially a quantum phenomenon. A model-independent microscopic approach to bulk macroscopic polarization involves the current operator, that is, the *phases* of the wave functions. I present here several recent advances, amongst which the most significant is the King-Smith and Vanderbilt approach to the problem. The formal derivation of the whole theory is given in such a way as to show very naturally the links with previously established concepts and results, and in particular with state-of-the-art linear-response theory. The main message of the present work is that macroscopic polarization—both induced and spontaneous—is a gauge-invariant phase feature of the electronic wave function, and bears in general *no relationship* to the periodic charge distribution of the polarized dielectric. The geometric phase viewpoint leads to definition of the observed bulk quantities (such as $\Delta\mathbf{P}$ and $\mathbf{P}'$) in terms of a Berry connection (or "vector potential") and of a curvature (or "magnetic field"). In addition to being important in terms of formulation, the geometric phase approach provides an extremely powerful computational tool for dealing with Born effective charges, linear and nonlinear piezoelectricity, and—last but not least—spontaneous polarization in ferroelectric materials.

Note added. After this work was completed, the many-body generalization of the present theory was obtained by Ortíz and Martin [Phys. Rev. B **49**, 14 202 (1994)].

ACKNOWLEDGMENTS

I thank A. Baldereschi, S. Baroni, A. Dal Corso, R.M. Martin, M. Posternak, E. Richter, and D. Vanderbilt for useful discussions.

APPENDIX

I provide here the transformation from Eq. (26) to Eq. (30). The change of variables of Eq. (28) transforms the $\mathbf{q}$ gradient of an arbitrary function f into

$$\nabla_{\mathbf{q}} f = \frac{1}{2\pi} \sum_{j=1}^{3} \mathbf{R}_j \frac{\partial f}{\partial \xi_j}, \tag{A1}$$

which is equivalent to

$$\mathbf{G}_j \cdot \nabla_{\mathbf{q}} f = \frac{\partial f}{\partial \xi_j}. \tag{A2}$$

The basic expressions for $\Delta \mathbf{P}_{\rm el}$, Eqs. (15) and (26), are then equivalent to the set of three equations

$$\mathbf{G}_j \cdot \Delta \mathbf{P}_{\rm el} = i \frac{2e}{\Omega} \int d\xi_1 d\xi_2 d\xi_3 \left[\sum_{n=1}^{\bar{n}} \langle u_n^{(0)}(\mathbf{q}) | \frac{\partial}{\partial \xi_j} u_n^{(0)}(\mathbf{q}) \rangle - \sum_{n=1}^{\bar{n}} \langle u_n^{(1)}(\mathbf{q}) | \frac{\partial}{\partial \xi_j} u_n^{(1)}(\mathbf{q}) \rangle \right], \tag{A3}$$

where the integral is performed over the unit cube $[0,1]\times[0,1]\times[0,1]$. One then recovers from Eq. (A3) the components of $\Delta \mathbf{P}_{\rm el}$ along the $\mathbf{G}_j$ directions. Let us focus for the sake of simplicity upon one of these components, say $j=3$. We reduce Eq. (A3) as a line integral and a surface integral performed in succession:

$$\mathbf{G}_3 \cdot \Delta \mathbf{P}_{\rm el} = \frac{2e}{\Omega} \int d\xi_1 d\xi_2 \left[\gamma_3^{(0)}(\xi_1, \xi_2) - \gamma_3^{(1)}(\xi_1, \xi_2) \right], \tag{A4}$$

where the two-dimensional integral is over the unit square $[0,1]\times[0,1]$, and the γ's are the Zak phases:

$$\gamma_3^{(\lambda)}(\xi_1, \xi_2) = i \int_0^1 d\xi_3 \sum_{n=1}^{\bar{n}} \langle u_n^{(\lambda)}(\mathbf{q}) | \frac{\partial}{\partial \xi_3} u_n^{(\lambda)}(\mathbf{q}) \rangle. \tag{A5}$$

Upon defining $\xi_4=\lambda$, as in Sec. V, and considering $\boldsymbol{\xi}$ as a single four-dimensional parameter, we find that the geometric phase of Eq. (A5) coincides with the line integral of the differential phase $d\varphi$, Eq. (29), over a unit segment parallel to the ξ_3 axis, at constant values of ξ_1, ξ_2, and ξ_4. We therefore arrive at Eq. (30):

$$\mathbf{G}_3 \cdot \Delta \mathbf{P}_{\rm el} = \frac{2e}{\Omega} \int d\xi_1 d\xi_2 \left[\int_{C_0} d\varphi - \int_{C_1} d\varphi \right], \tag{A6}$$

where the integration domains are those given in the main text.

REFERENCES

Aharonov, Y., and D. Bohm, 1959, Phys. Rev. **115**, 485. [Reprinted in: *Geometric Phases in Physics*, edited by A. Shapere and F. Wilczek (World Scientific, Singapore, 1989), p. 104].
Amos, R.D., 1987, in *Ab-Initio Methods in Quantum Chemistry - I*, edited by K.P. Lawley (Wiley, New York), p. 99.
Arnold, V.I., 1989, *Mathematical Methods of Classical Mechanics* (Springer, New York).
Axe, J.D., 1967, Phys. Rev. **157**, 429.
Baldereschi, A., M. Posternak, and R. Resta, 1992, Phys. Rev. Lett. **69**, 390.
Baldereschi, A., and R. Resta, 1983, in *Ab-Initio Calculation of Phonon Spectra*, edited by J.T. Devreese, V.E. Van Doren, and P.E. Van Camp (Plenum, New York), p. 1.
Baroni, S., and R. Resta, 1986, Phys. Rev. B **33**, 7017.
Baroni, S., P. Giannozzi, and A. Testa, 1987, Phys. Rev. Lett. **58**, 1861.
Bassani F., and G. Pastori Parravicini, 1975, *Electronic States and Optical Transitions in Solids* (Pergamon, Oxford).
Berry, M.V., 1984, Proc. R. Soc. London, Ser. A **392**, 45.
Berry, M.V., 1989, in *Geometric Phases in Physics*, edited by A. Shapere and F. Wilczek (World Scientific, Singapore), p. 7.
Blount, E.I., 1962, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull, Vol. 13 (Academic, New York), p. 305.
Born, M., 1945, Rev. Mod. Phys. **17**, 245.
Cohen, R.E, 1992, Nature **358**, 136.
Dal Corso, A., S. Baroni, R. Resta, and S. de Gironcoli, 1993a, Phys. Rev. B **47**, 3588.
Dal Corso, A., R. Resta, and S. Baroni, 1993b, Phys. Rev. B **47**, 16252.
Dalgarno, A., 1962, Adv. Phys. **11**, 281.
Deb, B.M., 1973, Rev. Mod. Phys. **43**, 22.
de Gironcoli, S., S. Baroni, and R. Resta, 1989, Phys. Rev. Lett. **62**, 2853.
de Gironcoli, S., S. Baroni, and R. Resta, 1990, Ferroelectrics **111**, 19.
Des Cloizeaux, J., 1964a, Phys. Rev. **135**, 685A.
Des Cloizeaux, J., 1964b, Phys. Rev. **135**, 698A.
Dougherty T.P., G.P. Wiederrecht, K.A. Nelson, M.H. Garret, H.P. Jensen, and C. Warde, 1992, Science **258**, 770.
Edwardson, P.J., 1989, Phys. Rev. Lett. **63**, 55.
Feynman, R.P., 1939, Phys. Rev. **56**, 340.
Giannozzi, P., S. de Gironcoli, P. Pavone, and S. Baroni, 1991, Phys. Rev. B **43**, 7231.
Gonze, X., D.C. Allan, and M.P. Teter, 1992, Phys. Rev. Lett. **68**, 3603.
Grimes, R.D., and E.R. Cowley, 1975, Can. J. Phys. **53**, 2549.
Hybertsen, M.S., and S.G. Louie, 1987, Phys. Rev. B **35**, 5585 (Part I); 5602 (Part II).
Jackiw, R., 1988, Int. J. Mod. Phys. A **3**, 285 [Reprinted in *Geometric Phases in Physics*, edited by A. Shapere and F.

Wilczek (World Scientific, Singapore, 1989), p. 29].
Jansen, H.J.F., and A. J. Freeman, 1984, Phys. Rev. B. **30**, 561.
Kallin, C., and B.I. Halperin, 1984, Phys. Rev. B **29**, 2175.
King-Smith, R.D., and D. Vanderbilt, 1993, Phys. Rev. B **47**, 1651.
Kleemann, W., F. J. Schäfer, and M. D. Fontana, 1984, Phys. Rev. B **30**, 1148.
Kohn, W., 1959, Phys. Rev. **115**, 809.
Kohn, W., 1964, Phys. Rev. **133**, 171A.
Kohn, W., 1973, Phys. Rev. B **7**, 4388.
Kunc, K., 1985, in *Electronic Structure, Dynamics and Quantum Structural Properties of Condensed Matter*, edited by J.T. Devreese and P. Van Camp (Plenum, New York), p. 227.
Kunz, H., 1986, Phys. Rev. Lett. **57**, 1095.
Landau, L.D., and E.M. Lifshitz, 1977, *Quantum Mechanics* (Pergamon, Oxford).
Landau, L.D., and E.M. Lifshitz, 1984, *Electrodynamics of Continuous Media* (Pergamon, Oxford).
Landauer, R., 1981, Solid State Commun. **40**, 971.
Landauer, R., 1987, Ferroelectrics **73**, 41.
Lines, M.E., and A.M. Glass, 1977, *Principles and Applications of Ferroelectrics and Related Materials* (Clarendon, Oxford).
Lundqvist, S., and N.H. March, 1983, Eds., *Theory of the Inhomogeneous Electron Gas* (Plenum, New York).
Mahan, G.D., 1980, Phys. Rev. A **22**, 1780.
Martin, R.M., 1972a, Phys. Rev. B **5**, 1607.
Martin, R.M., 1972b, Phys. Rev. B **6**, 4874.
Martin, R.M., 1974, Phys. Rev. B **9**, 1998.
Mead, C.A., 1992, Rev. Mod. Phys. **64**, 51.
Mead, C.A., and D.G. Truhlar, 1979, J. Chem. Phys. **70**, 2284; **78**, 6344(E) [Reprinted in *Geometric Phases in Physics*, edited by A. Shapere and F. Wilczek (World Scientific, Singapore, 1989), p. 90].
Michel, L., and J. Zak, 1992, Europhys. Lett. **18**, 239.
Morandi, G., 1988, *Quantum Hall Effect* (Bibliopolis, Napoli).
Nielsen, O.H., and R.M. Martin, 1983, Phys. Rev. Lett. **50**, 697.
Nielsen, O.H., and R.M. Martin, 1985, Phys. Rev. B **32**, 3780.
Pancharatnam, S., 1956, Proc. Indian Acad. Sci. **44**, Sec. A, 247 [Reprinted in *Geometric Phases in Physics*, edited by A. Shapere and F. Wilczek (World Scientific, Singapore, 1989), p. 51].
Penn, D.R., 1962, Phys. Rev. **128**, 2093.
Phillips, J.C., 1973, *Bonds and Bands in Semiconductors* (Academic, New York).
Pick, R., M.H. Cohen, and R.M. Martin, 1970, Phys. Rev. B **1**, 910.
Pick, R., and T. Takemori, 1986, in *Crystalline Semiconducting Materials and Devices*, edited by P. N. Butcher, N. H. March, and M. P. Tosi (Plenum, New York), p. 103.
Pickett, W.E., 1989, Comput. Phys. Rep. **9**, 115.
Pisani, C., R. Dovesi, and C. Roetti, 1988, *Hartree-Fock Ab-Initio Treatment of Crystalline Systems* (Springer, Berlin).
Posternak, M., A. Baldereschi, A. Catellani, and R. Resta, 1990, Phys. Rev. Lett. **64**, 1777.
Prange, R.P., and S.M. Girvin, 1987, Eds., *The Quantum Hall Effect* (Springer, Berlin).
Resta, R., 1977, Phys. Rev. B **16**, 2717.
Resta, R., 1992, Ferroelectrics **136**, 51.
Resta, R., 1993, Europhys. Lett. **22**, 133.
Resta, R., and K. Kunc, 1986, Phys. Rev. B **34**, 7146.
Resta, R., M. Posternak, A. Baldereschi, and A. Catellani, 1990, Ferroelectrics **111**, 15.
Resta, R., M. Posternak, and A. Baldereschi, 1993a, Phys. Rev. Lett. **70**, 1010.
Resta, R., M. Posternak, and A. Baldereschi, 1993b, Mat. Res. Soc. Symp. Proc. **291**, 647.
Schiff, L.I., 1968, *Quantum Mechanics* (Mc-Graw-Hill, New York).
Slater, J.C., 1950, Phys. Rev. **78**, 748.
Sternheimer, R.M., 1954, Phys. Rev. **96**, 951.
Sternheimer, R.M., 1957, Phys. Rev. **107**, 1565.
Sternheimer, R.M., 1959, Phys. Rev. **115**, 1198.
Sternheimer, R.M., 1969, Phys. Rev. **183**, 112.
Sternheimer, R.M., 1970, Phys. Rev. A **1**, 321.
Tagantsev, A.K., 1991, Phase Transitions **35**, 119.
Tagantsev, A.K., 1992, Phys. Rev. Lett. **69**, 389.
Thouless, D.J., 1983, Phys. Rev. B **27**, 6083.
Vogl, P., 1978, J. Phys. C **11**, 251.
Woo, W.F., and R. Landauer, 1972, Phys. Rev. B **6**, 4876.
Zak, J., 1989, Phys. Rev. Lett. **62**, 2747.