PHYSICAL REVIEW B VOLUME 48, NUMBER 7 15 AUGUST 1993-I

# Electric polarization as a bulk quantity and its relation to surface charge

David Vanderbilt and R. D. King-Smith

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

(Received 26 January 1993)

A definition of the electric polarization of an insulating crystalline solid is given in terms of the centers of charge of the Wannier functions of the occupied bands. The change of this quantity under an adiabatic evolution of the Hamiltonian has previously been shown to correspond to the physical change in polarization. Here, we show that the polarization as defined above also has a direct and predictive relationship to the surface charge which accumulates at an insulating surface or interface.

## I. INTRODUCTION

There has been considerable controversy over the years as to whether electric polarization effects in crystalline solids are well defined in terms of bulk properties. An early controversy over the piezoelectric response (i.e., the strain derivative of the polarization)[1-4] has been resolved in favor of the view that the piezoelectric coefficients are indeed well-defined bulk quantities, independent of surface termination. In fact, Resta has argued[5] that any first derivative of the bulk polarization with respect to a parameter $\lambda$ of the Hamiltonian is well defined and is given by

$$\frac{\partial \mathbf{P}_e}{\partial\lambda} = \frac{ie\hbar}{N\Omega m_e}\sum_{\mathbf{k}}\sum_{n=1}^{M}\sum_{m=M+1}^{\infty}\frac{\langle\psi_{n\mathbf{k}}^{(\lambda)}\mid\mathbf{p}\mid\psi_{m\mathbf{k}}^{(\lambda)}\rangle\langle\psi_{m\mathbf{k}}^{(\lambda)}\mid\partial V^{(\lambda)}/\partial\lambda\mid\psi_{n\mathbf{k}}^{(\lambda)}\rangle}{(\epsilon_{n\mathbf{k}}^{(\lambda)}-\epsilon_{m\mathbf{k}}^{(\lambda)})^2} + \text{c.c.} \qquad (1)$$

where $m_e$ and $-e$ are the electron mass and charge ($e > 0$), $N$ is the number of unit cells in the crystal, $\Omega$ is the volume of a unit cell, $M$ is the number of occupied bands (counting spin), and $\mathbf{p}$ is the momentum operator. We limit ourselves here to an independent-electron description of the solid within Kohn-Sham density-functional theory,[6] so that $V^{(\lambda)}$ is to be interpreted as the Kohn-Sham potential $V_{\text{KS}}^{(\lambda)}$. (Typically, $\lambda$ parametrizes displacements of atoms in the unit cell.) Equation (1) can be regarded as expressing the current which is induced in the solid by a slow variation of $\lambda$, and can be derived from the adiabatic limit of a Kubo formula.[7] By the same token, Resta points out, the change in the polarization under a *finite* adiabatic change of the Hamiltonian is well defined and is given by

$$\Delta\mathbf{P}_e = \int_{\lambda_1}^{\lambda_2}\frac{\partial\mathbf{P}_e}{\partial\lambda}\,d\lambda\ , \qquad (2)$$

where the scalar $\lambda$ is to be thought of as parametrizing a path in the space of Kohn-Sham Hamiltonians. Of course it is required that the system remain insulating everywhere along the path.

However, the previous work is still ambiguous as to whether the polarization $\mathbf{P}_e$ itself is well defined as a bulk quantity. An obvious but ultimately fruitless approach is to define

$$\mathbf{P}_e(\Omega) = \frac{1}{\Omega}\int_{\Omega}\mathbf{r}\,\rho_e(\mathbf{r})\,d\mathbf{r}\ \ , \qquad (3)$$

where $\rho$ is the electronic charge density and $\Omega$ represents some particular choice of unit cell. The total polarization $\mathbf{P}(\Omega)$, defined similarly in terms of the total (electronic plus ionic) charge density, is then independent of choice of origin. However, it is *not* independent of the choice of unit-cell boundaries [hence the notation $\mathbf{P}(\Omega)$], and can be made to take on any value by a sufficiently pathological choice of cell. For this reason, Eq. (3) is not a useful definition.

In searching for an alternative definition, it is desirable that the polarization should obey an equation of the form

$$\sigma = \mathbf{P}\cdot\hat{\mathbf{n}}\ \ , \qquad (4)$$

where $\sigma$ is the "bound charge" which accumulates at a surface or interface of orientation $\hat{\mathbf{n}}$. Turning this idea around, Posternak *et al.* recently reported *ab initio* calculations of the interface charge arising at wurtzite–zinc-blende boundaries in BeO using a supercell technique,[8] and interpreted the result as a calculation of $\mathbf{P}$ for the wurtzite crystal. However, the justification for such an interpretation is not immediately clear, and has recently been challenged.[9,10]

In this paper, we show that it *is* possible to give precise definitions of the polarization $\mathbf{P}_e$ and of the surface "bound charge" $\sigma$ such that an equation of the form (4) is satisfied. We take as our definition of the "bound charge" the excess areal surface charge present when the surface is *insulating*, i.e., when the Fermi level lies in a gap common to both the bulk and surface, and all surface bands are completely full or completely empty. The starting point of our definition of $\mathbf{P}_e$ is Eq. (2), which is there-

0163-1829/93/48(7)/4442(14)/$06.00

fore automatically satisfied. However, unlike Eq. (3), our definition of $\mathbf{P}_e$ cannot be written as a functional of the charge density alone, nor indeed as an expectation value of any operator. Instead, it is related to a global phase property ("Berry phase"[11]) of the manifold of occupied Bloch bands of the crystal as a whole.

Our proposed definition of $\mathbf{P}_e$ is based upon recent work[12] in which we showed that an integral of the form of Eq. (2), when carried around a closed loop in parameter space, necessarily results in a polarization change which takes the form

$$\oint \frac{\partial \mathbf{P}_e}{\partial \lambda} d\lambda = \frac{e\mathbf{R}}{\Omega} , \tag{5}$$

where $\mathbf{R}$ is a lattice vector. This suggests that the polarization $\mathbf{P}_e$ might be well-defined modulo $e\mathbf{R}/\Omega$ (i.e., $\Omega\mathbf{P}_e/e$ would be well-defined modulo a lattice vector). In fact, we also showed that the change in polarization for an arbitrary path can be computed (modulo $e\mathbf{R}/\Omega$) from only a knowledge of the system at the end points,

$$\int_{\lambda_1}^{\lambda_2} \frac{\partial \mathbf{P}_e}{\partial \lambda} d\lambda = \mathbf{P}_e(\lambda_2) - \mathbf{P}_e(\lambda_1) , \tag{6}$$

where $\mathbf{P}_e(\lambda)$ is given in terms of the cell-periodic functions, $u_{n\mathbf{k}}^{(\lambda)}$, by

$$\mathbf{P}_e(\lambda) = -\frac{ie}{(2\pi)^3} \sum_{n=1}^{M} \int_{\mathrm{BZ}} d\mathbf{k}\, \langle u_{n\mathbf{k}}^{(\lambda)} \mid \nabla_{\mathbf{k}} \mid u_{n\mathbf{k}}^{(\lambda)} \rangle . \tag{7}$$

It is understood that the phase relation between $u_{n\mathbf{k}}$ and $u_{n,\mathbf{k}+\mathbf{G}}$ is fixed by requiring $\psi_{n\mathbf{k}} = \psi_{n,\mathbf{k}+\mathbf{G}}$. Alternatively, this "Berry phase" expression for $\mathbf{P}_e$ may be reexpressed, following Blount,[15] in terms of the centers of charge of the Wannier functions $W_n^{(\lambda)}(\mathbf{r})$ of the occupied bands:

$$\mathbf{P}_e(\lambda) = -\frac{e}{\Omega} \sum_{n=1}^{M} \int \mathbf{r} \left| W_n^{(\lambda)}(\mathbf{r}) \right|^2 d\mathbf{r} . \tag{8}$$

In either case, the remaining phase freedom in the choice of the $u_{n\mathbf{k}}$ was shown to leave $\mathbf{P}_e$ invariant modulo $e\mathbf{R}/\Omega$.

This leads us to propose Eq. (7) or (8) as a *definition* of the electronic contribution to the polarization of a crystalline solid. (The total $\mathbf{P}$ also contains an ionic contribution.) We emphasize that this $\mathbf{P}_e$ is only well-defined modulo $e\mathbf{R}/\Omega$; however, the same kind of arbitrariness is associated with the ionic contribution in any case. With this definition, the polarization can be assigned physical significance in two ways. First, as implied by our previous work, the *difference* in the polarization of two crystals is correctly given by this definition *provided* the two structures can be connected by a continuous path in the parameter space of insulating Hamiltonians; cf., Eq. (6). Second, as discussed in this paper, the surface charge which accumulates at an insulating surface of an insulating crystal is predicted, modulo $e/A_{\mathrm{surf}}$, to be just $\mathbf{P} \cdot \hat{\mathbf{n}}$ (where $A_{\mathrm{surf}}$ is the surface cell area). The demonstration of this pair of results, we argue, suggests that Eq. (7) or (8) is indeed a physically reasonable definition.

Several implications of this formulation suggest themselves. For example, an immediate consequence of Eq. (4) is that any finite perturbation applied to an insulating surface of an insulating crystal can have no effect at all on the areal surface charge density. This remarkable result has been previously discussed in special cases by several authors[13,14] and in more generality by Kallin and Halperin.[3] Also, the present formulation provides an alternate derivation of electron counting rules which can be used to determine when surfaces of semiconductors and insulators can be simultaneously neutral and insulating. Finally, a similar argument can be formulated for interfaces, and implies that the polarization of Eq. (7) should give the same value for the spontaneous polarization of a pyroelectric material as that deduced from a supercell calculation of the type carried out in Ref. 8.

In the case of spin-degenerate insulators, $\mathbf{P}$ is well-defined modulo $2e\mathbf{R}/\Omega$, not just $e\mathbf{R}/\Omega$. This allows us to predict the surface charge modulo $2e/A_{\mathrm{surf}}$, not just modulo $e/A_{\mathrm{surf}}$, provided a proper accounting is made of any odd-integer ionic charges in the vicinity of the surface.

The paper is organized as follows. In Sec. II, the formulation of Eq. (7) and its relation to the Wannier functions of the occupied bands is reviewed. Section III, which discusses the relation between the polarization and the surface charges, contains the main results of the paper. Section IV contains some general remarks about the definition of the polarization of crystalline solids, emphasizing in particular a viewpoint in which the charge density of the real quantum-mechanical system is mapped onto a system of quantized classical point charges. In Sec. V, we give several examples of the counting of surface charge, mostly within the context of tight-binding models. The generalization to the case of interacting-electron systems is discussed briefly in Sec. VI. Finally, we conclude in Sec. VII.

## II. BACKGROUND

### A. Polarization

Here, we give a brief review of the demonstration, already given in Ref. 12, of the fact that Eq. (8) is invariant with respect to the choice of phase of the Bloch functions. The derivation relies on the fact that the center of charge $\int \mathbf{r}|W_n(\mathbf{r})|^2 d\mathbf{r}$ of a Wannier function is invariant, modulo a lattice vector, with respect to the phase choice.[15] This motivates Eq. (8) as a physical definition of the polarization, and allows us to give an elementary proof of the quantization of Eq. (2) for closed paths in parameter space. Initially, we limit the treatment to the case where the occupied bands are distinct, i.e., they remain nondegenerate everywhere in the Brillouin zone. (Spin-up and spin-down bands will also be considered "distinct" unless they become mixed by spin-orbit interactions.) We discuss the more general case of composite bands at the end of this subsection.

The forward and inverse relations between the Wannier and Bloch functions are

$$W_n(\mathbf{r}) = \frac{\sqrt{N}\Omega}{(2\pi)^3}\int_{\mathrm{BZ}} d\mathbf{k}\,\psi_{n\mathbf{k}}(\mathbf{r})\ , \tag{9}$$

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}W_n(\mathbf{r}-\mathbf{R})\ , \tag{10}$$

where the sum over $\mathbf{R}$ runs over all real-space lattice vectors. Equation (10) is only consistent with a choice of phase such that

$$\psi_{n,\mathbf{k}+\mathbf{G}} = \psi_{n\mathbf{k}}\ . \tag{11}$$

Now we construct a new set of Bloch functions

$$\overline{\psi}_{n\mathbf{k}} = e^{i\theta_n(\mathbf{k})}\psi_{n\mathbf{k}} \tag{12}$$

with a different phase choice. But note that Eq. (11) applied to both $\psi$ and $\overline{\psi}$ implies that $\theta_n$ must return to itself, modulo $2\pi$, as the Brillouin zone is crossed, i.e.,

$$\theta_n(\mathbf{k}+\mathbf{G}) = \theta_n(\mathbf{k}) + \mathbf{G}\cdot\mathbf{R}_n \tag{13}$$

for some lattice vector $\mathbf{R}_n$.

Recall that $\mathbf{P}_e$ can be written, Eq. (8),

$$\mathbf{P}_e = -\frac{e}{\Omega}\sum_{n=1}^{M}\mathbf{r}_n, \tag{14}$$

where

$$\mathbf{r}_n = \langle W_n|\mathbf{r}|W_n\rangle, \tag{15}$$

and let similar relations hold for the polarization $\overline{\mathbf{P}}_e$, Wannier centers $\bar{\mathbf{r}}$, and Wannier functions $\overline{W}$ associated with the $\overline{\psi}$. Then, using Eq. (7), we have[15]

$$\begin{aligned}\bar{\mathbf{r}}_n &= \frac{\Omega}{N(2\pi)^3}\sum_{\mathbf{RR}'}\int_{\mathrm{BZ}} d\mathbf{k}\,\overline{W}^*_n(\mathbf{r}-\mathbf{R})e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})} \\ &\qquad\times i\nabla_{\mathbf{k}}\overline{W}_n(\mathbf{r}-\mathbf{R}')e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R}')} \\ &= \langle W_n|\mathbf{r}|W_n\rangle - \frac{\Omega}{N(2\pi)^3}\sum_{\mathbf{R}}\int_{\mathrm{BZ}} d\mathbf{k}\,\nabla_{\mathbf{k}}\theta_n(\mathbf{k}) \\ &= \mathbf{r}_n - \mathbf{R}_n\end{aligned} \tag{16}$$

from which it immediately follows that $\overline{\mathbf{P}}_e = \mathbf{P}_e + e\mathbf{R}/\Omega$, where $\mathbf{R}$ is the sum of $\mathbf{R}_n$ over occupied bands.

This analysis indicates that, for the purposes of understanding polarization effects, we may think of the electronic charge as being localized into point charges $-e$ located at the Wannier centers associated with the occupied bands in each unit cell. This point of view, in which the true quantum-mechanical system is mapped onto an effective classical system of quantized point charges, is discussed further in Sec. IV. With this picture in mind, Eq. (5) follows immediately: if a Wannier center of a given band is followed during the evolution of $\lambda$ around a closed loop, it must either return to itself, or to the same position in a neighboring unit cell. In either case the change in polarization vanishes modulo $e\mathbf{R}/\Omega$.

In the discussion above, we have assumed that each occupied band is distinct. However, in most crystals of interest (e.g., tetrahedral semiconductors), symmetries give rise to interband degeneracies at certain points or along certain lines in the Brillouin zone. In this case of "composite bands," it is natural to divide the valence bands into distinct groups. In GaAs, for example, the lowest two bands (which would be completely degenerate but for spin-orbit splitting) would form one group, and the next six bands would form a second group. The more general analysis of Ref. 12 indicates that the $\alpha$th group containing $N_\alpha$ bands can be characterized by a common center $\mathbf{r}_\alpha$ having charge $-N_\alpha e$. Moreover, the location of $\mathbf{r}_\alpha$ is only well-defined modulo $\mathbf{R}/N_\alpha$, so that again there is an overall indeterminacy of the polarization modulo $e\mathbf{R}/\Omega$.

Finally, the total polarization is

$$\mathbf{P} = \mathbf{P}_{\mathrm{ion}} + \mathbf{P}_e, \tag{17}$$

where $\mathbf{P}_e$ is the electronic contribution discussed above, and the ionic contribution is

$$\mathbf{P}_{\mathrm{ion}} = \frac{e}{\Omega}\sum_{j=1}^{\mathcal{N}} Z_j\mathbf{u}_j\ . \tag{18}$$

Here $\mathcal{N}$ is the number of atoms in the primitive unit cell, and $Z_j$ and $\mathbf{u}_j$ are the atomic number and the position vector of the $j$th basis atom. (In a pseudopotential or frozen-core context, $Z_j$ is understood to be the *valence* atomic number, e.g., $Z = 4$ for Si. Alternatively, the full atomic number can be used, provided that core bands are included in the sums over occupied bands for the electronic contributions.) Note that there is some arbitrariness about the choice of ionic basis; in GaAs, for example, one could equally well choose a nearest-neighbor pair of atoms oriented along [111] or along $[1\bar{1}\bar{1}]$ to represent the unit cell. However, a change in the choice of ionic basis merely corresponds to a translation of an integer charge by a lattice vector, so $\mathbf{P}_{\mathrm{ion}}$ is in fact well-defined modulo $e\mathbf{R}/\Omega$. Also, note that while $\mathbf{P}_e$ and $\mathbf{P}_{\mathrm{ion}}$ individually depend upon the choice of origin, the total polarization $\mathbf{P}$ is independent of origin.

### B. Localization properties of Wannier functions

In the discussion which follows, we will need to use some localization properties of the Wannier functions and the band projection operator (density matrix) in one dimension. These are briefly reviewed here.

It is well known[15–17] that the Wannier function for band $n$ can be chosen to be exponentially localized in space, with a decay length $\kappa_n^{-1}$ on the order of a lattice constant. $\kappa_n$ is related to the maximum imaginary part of $k$ in the "complex band structure" (complex $k$ associated with real $E$) in the gaps above or below the band $n$; alternatively, it may be regarded as the half-width of the strip of analyticity of the Bloch function $\psi_{nk}(x)$ regarded as a function of complex $k$.[16] (The choice of phases which minimizes the spread of the Wannier function is that for which $\langle u_k|\ d/dk\ |u_k\rangle$ is constant.) These localization properties have been demonstrated for the case of noncentrosymmetric[15] as well as centrosymmetric[16] potentials.

The localization of the Wannier functions also implies the localization of the band projection operator

$$\rho_n(x,x') = \sum_k \psi^*_{nk}(x)\psi_{nk}(x') \quad (19)$$

as a function of $|x-x'|$, with the same exponential decay length $\kappa_n^{-1}$. This follows easily from the representation of $\rho_n$ in terms of the Wannier functions:

$$\rho_n(x,x') = \sum_l W^*_n(x - X_l)W_n(x' - X_l) \quad (20)$$

($X_l$ is the $l$th lattice vector). The density operator is just the sum of the band projection operators of the occupied states,

$$\rho(x,x') = \sum_{n=1}^{M} \rho_n(x,x') \ , \quad (21)$$

and is therefore also exponentially localized in $|x - x'|$ with a decay constant $\kappa \geq \min\{\kappa_n\}$. Equation (20) reminds us that the density matrix is just diagonal in the basis of Wannier functions.

Strictly speaking, the above discussion applies only to the localization properties in the bulk of a crystalline material. However, the work of Kohn and Onffroy,[18] Rehr and Kohn,[19] and Kallin and Halperin[3] indicates that these properties also survive in the vicinity of defects and surfaces. In general, there exist also exponential decay lengths associated with electrostatic perturbations and lattice distortions in the vicinity of the surface; if one of these should happen to be of longer range than $\kappa^{-1}$, then it should be understood that $\kappa$ as used below will represent the inverse of the longest of these decay lengths.

## III. SURFACE THEOREM

In this section, we demonstrate a connection between the polarization of an insulating crystal as defined via Eq. (17) and Eq. (7) or Eq. (8), and the areal surface charge density $\sigma$ on an insulating face of the crystal. We limit ourselves here to an independent-electron treatment within the local spin-density approximation (LSDA). Initially, we will assume that the spin-up and spin-down bands can be treated as distinct; the more general case of spin-orbit interactions will be discussed briefly in Sec. III C.

Note that the bulk polarization is assumed to have been calculated under boundary conditions of vanishing macroscopic electric field $\overline{E}$ in the crystal. Thus, it is to be understood that the crystal surface of interest is under boundary conditions of $\overline{E} = 0$ in the bulk and $E = 4\pi\sigma$ in the vacuum outside the surface.

### A. One-dimensional case

For simplicity, we begin our discussion with the simplest possible case, that of a one-dimensional (1D) crystal in which each occupied valence band of the solid is distinct, so that the Wannier centers $x_n = \langle W_n|x|W_n\rangle$ are uniquely defined modulo a lattice vector $a$. As shown schematically in Fig. 1, we consider a geometry in which a thick slab of the 1D crystal is bounded by a pair of "surfaces" $A$ and $B$. The two surfaces need not be identical, but they are both assumed to be insulating, in the sense that the Fermi level falls into a gap common to the bulk and to both surfaces. Our claim is that a relation of the form $\sigma_B = \mathbf{P}\cdot\hat{\mathbf{n}}$ holds between the bulk polarization as defined in Eq. (8) and the surface charge at $B$; for the 1D case, this is just

$$\sigma_B = P \quad (\text{modulo } e) \ . \quad (22)$$

A similar relation $\sigma_A = -P$ holds for surface $A$.

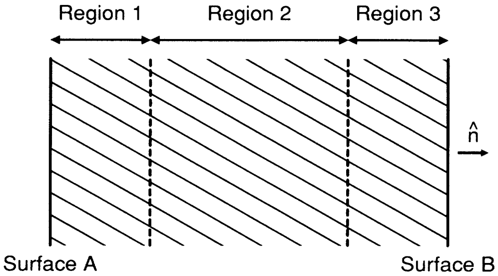

FIG. 1. Sketch of the surface slab configuration discussed in the text.

To prove this assertion, it is useful to divide the slab conceptually into three regions, as shown in Fig. 1. The widths of the two surface regions 1 and 3 are taken large enough so that the influence of the surface is negligible in the central region 2. Region 2 is assumed to consist of precisely $N_2$ bulk unit cells.

Now we construct a basis set to describe the occupied states of this system as follows. First, we take the Wannier functions $W_{nl} = W_n(x - X_l)$, where $n$ runs over the occupied bands and the $X_l$ are the lattice vectors contained between the surfaces $A$ and $B$. (The precise choice of the first and last $X_l$ included is not important, but some definite choice is assumed to have been made.) Next, we choose some additional set of localized basis orbitals $\varphi_i^B$ in the vicinity of surface $B$; we choose enough of these so that, together with the Wannier functions introduced above, they span the space of occupied states of the Hamiltonian in the vicinity of surface $B$. The latter orbitals are required to be localized to within a depth $d$ of the surface, where $d$ is much less than the width $L_3$ of region 3. We then Graham-Schmidt orthonormalize the $\varphi_i^B$ against all of the Wannier functions (and against each other) to yield a set $\phi_i^B$. It is important to note that the $\phi_i^B$ remain exponentially localized in the vicinity of surface $B$ even after this orthonormalization procedure. This result follows from the exponential localization of the Wannier functions discussed in Sec. II B; in the orthogonalization contribution $-\sum_l\langle W_{nl}|\varphi_i^B\rangle W_{nl}(x)$, both the dot product and the Wannier function on the right decay exponentially away from surface $B$ at least as fast as $\exp(\kappa x)$. Finally, a similar procedure generates a set $\phi_i^A$. Thus, by construction, the $W_{nl}$ and $\phi_i^B$ and $\phi_i^A$ taken together constitute an orthonormal set that spans the space of occupied eigenstates of the slab Hamiltonian.

Now consider the form of the density operator, Eq. (21), when written in the above basis. In the bulk, the density matrix is diagonal in the basis of Wannier functions; thus, we must have $\langle W_{nl}|\rho|W_{n'l'}\rangle = \delta_{nn'}\delta_{ll'}$ except in the close vicinity of the surfaces. Moreover, the $\langle W_{nl}|\rho|\varphi_i^B\rangle$, and hence $\langle W_{nl}|\rho|\phi_i^B\rangle$, decay exponentially with the distance of $l$ from surface $B$, by virtue of the spatial localization of $W_{nl}$ and $\varphi_i^B$. In light of the above, it is useful to repartition the basis orbitals into three groups: orbitals $\phi_i^{(1)}$ comprising all of the $\phi_i^A$ together with the $W_{nl}$ in region 1; orbitals $\phi_i^{(3)}$ comprising all of the $\phi_i^B$ together with the $W_{nl}$ in region 3; and orbitals $\phi_i^{(2)}$ comprising the remaining $W_{nl}$ in region 2. Then the above discussion implies that the density matrix is *block diagonal* when written in this representation:

$$\rho = \begin{pmatrix} \rho^{(1)} & 0 & 0 \\ 0 & \mathbb{1} & 0 \\ 0 & 0 & \rho^{(3)} \end{pmatrix} . \tag{23}$$

What is really meant by the zero entries is that they can be made to vanish exponentially by increasing the widths of regions 1 and 3. This is a critical result.

It follows from this block-diagonal form of $\rho$ that the electronic density of the system can be written

$$n(x) = \sum_{\mu=1}^{3} n_\mu(x), \tag{24}$$

where the contribution to the density from region $\mu$ is

$$n_\mu(x) = \sum_{ij} \rho_{ij}^{(\mu)} \phi_j^{(\mu)*}(x) \phi_i^{(\mu)}(x) . \tag{25}$$

Moreover, it also follows that the idempotency of the density matrix, $\rho^2 = \rho$, carries over to each block, $[\rho^{(\mu)}]^2 = \rho^{(\mu)}$; and since the trace of an idempotent matrix must be integral, we have

$$\int dx\, n_\mu(x) = N_\mu \tag{26}$$

independently for each region $\mu$, where $N_\mu$ is an integer.

Now note that the total charge density (electronic plus ionic) associated with region 2 can be written as a superpositon,

$$\rho_2(x) = \sum_{l \in \text{region } 2} \rho_{\text{unit}}(x - X_l) \tag{27}$$

of neutral entities

$$\rho_{\text{unit}}(x) = -e \sum_{n=1}^{M} |W_n(x)|^2 + \sum_{j=1}^{\mathcal{N}} Z_j e\, \delta(x - u_j) . \tag{28}$$

Moreover, the dipole moment associated with $\rho_{\text{unit}}$ is just $\Omega P$, with $P$ defined in Eqs. (17) and (14). Thus, the "surface charge" associated with the charge density $\rho_2(x)$, accumulating at the boundary between regions 2 and 3, is just $P$. To get the total surface charge $\sigma_B$ at surface $B$, we must add to this all the electronic and ionic charges associated with region 3. However, from Eq. (26) it follows that *both* species of charges come *only* in units of the charge quantum $e$. This proves the assertion (22). Obviously, a similar argument applies to surface $A$.

### B. Three-dimensional case

We now generalize to the case of a three-dimensional (3D) crystal. We again initially assume that each of the occupied valence bands is distinct; however, the generalized case is discussed at the end of the subsection. In this and the following subsection, the surface is assumed to be unreconstructed, i.e., to have the smallest possible surface cell area $A_{\text{surf}}$ consistent with the bulk periodicity. This condition is relaxed later in Sec. III D.

We again adopt the geometry of Fig. 1, and again assume that the Fermi level falls in a gap common to the bulk and both surfaces. Our claim is that

$$\sigma_B = P_\perp \quad (\text{modulo } e/A_{\text{surf}}) , \tag{29}$$

where $\sigma_B$ is the areal surface charge density at surface $B$, $P_\perp = \mathbf{P} \cdot \hat{\mathbf{n}}$, and $A_{\text{surf}}$ is the primitive surface cell area.

Let $\mathbf{G}_\perp$ be the reciprocal-lattice vector of minimum length $G_\perp = 2\pi A_{\text{surf}}/\Omega$ aligned in the direction perpendicular to the surface. We choose the bulk Brillouin zone to be a prism of height $G_\perp$ in the direction perpendicular to the surface, and having a base $\mathcal{A}$ of area $(2\pi)^2/A_{\text{surf}}$ in the directions parallel to the surface. It follows from Eq. (7) that the projection of the electronic polarization onto the surface normal, $P_{e,\perp} = \mathbf{P}_e \cdot \hat{\mathbf{n}}$, can be written

$$P_{e,\perp} = \frac{-ie}{(2\pi)^3} \int_{\mathcal{A}} d\mathbf{k}_\parallel \sum_{n=1}^{M} \int_0^{G_\perp} dk_\perp \left\langle u_{n\mathbf{k}} \middle| \frac{\partial}{\partial k_\perp} \middle| u_{n\mathbf{k}} \right\rangle \tag{30}$$

in an obvious notation.

Now observe that for a given $\mathbf{k}_\parallel$, the inner integral over $k_\perp$ is in a form which allows us to make close contact with the one-dimensional theory of the previous subsection. Moreover, $\mathbf{k}_\parallel$ is a good quantum number at the surface, so we can also decompose the contributions to the surface charge density $\sigma$ into those arising from electronic eigenstates of different $\mathbf{k}_\parallel$. This will allow us to prove assertion (29). To be more explicit, we write

$$P_{e,\perp} = \frac{1}{(2\pi)^2} \int_{\mathcal{A}} d\mathbf{k}_\parallel\, P_{e,\perp}^{(\mathbf{k}_\parallel)} , \tag{31}$$

$$P_{\text{ion},\perp} = \frac{1}{(2\pi)^2} \int_{\mathcal{A}} d\mathbf{k}_\parallel\, P_{\text{ion},\perp}^{(\mathbf{k}_\parallel)} , \tag{32}$$

and

$$\sigma_B = \frac{1}{(2\pi)^2} \int_{\mathcal{A}} d\mathbf{k}_\parallel\, \sigma_B^{(\mathbf{k}_\parallel)} . \tag{33}$$

Here $P_{\text{ion},\perp}^{(\mathbf{k}_\parallel)} = A_{\text{surf}}\, P_{\text{ion},\perp}$ is independent of $\mathbf{k}_\parallel$, and

$$\begin{aligned} P_{e,\perp}^{(\mathbf{k}_\parallel)} &= \frac{-ie}{2\pi} \sum_{n=1}^{M} \int_0^{G_\perp} dk_\perp \left\langle u_{nk_\perp}^{(\mathbf{k}_\parallel)} \middle| \frac{\partial}{\partial k_\perp} \middle| u_{nk_\perp}^{(\mathbf{k}_\parallel)} \right\rangle \\ &= -\frac{e}{a_\perp} \sum_{n=1}^{M} \int x_\perp \left| W_n^{(\mathbf{k}_\parallel)}(x_\perp) \right|^2 dx_\perp , \end{aligned} \tag{34}$$

where $a_\perp = \Omega/A_{\rm surf}$ is the periodic repeat distance normal to the surface. The treatment of the bulk Bloch functions $u_{nk_\perp}^{(\mathbf{k}_\parallel)}$, the Wannier functions $W_n^{(\mathbf{k}_\parallel)}(x_\perp)$, the slab basis functions $\phi_i^{(\mathbf{k}_\parallel)A}$ and $\phi_i^{(\mathbf{k}_\parallel)B}$, and the density matrix $\rho^{(\mathbf{k}_\parallel)}(x_\perp, x'_\perp)$ follows precisely the same lines as in the preceding subsection. In the present context, Eq. (22) becomes

$$\sigma_B^{(\mathbf{k}_\parallel)} = P_\perp^{(\mathbf{k}_\parallel)} + Je \; , \tag{35}$$

where $J$ is an integer. Using Eqs. (31)–(33), this leads directly to the desired result, Eq. (29), provided only that the integer $J$ is independent of $\mathbf{k}_\parallel$. But an integer function of a continuous variable can only change by sudden jumps, and this certainly cannot occur in the present case. In particular, the continuity of $u_{n\mathbf{k}}$ as a function of $\mathbf{k}$ prevents any sudden jumps in $P_{\rm e,\perp}^{(\mathbf{k}_\parallel)}$; and the assumption that the surface is insulating precludes the possibility of any surface bands crossing the Fermi level, and thus of any jumps in $\sigma^{(\mathbf{k}_\parallel)}$. This proves the theorem (29).

The above derivation assumes that each of the occupied bands is distinct. If instead there are degeneracies between bands at certain $k$ points, we refer to the bands connected in this way as "composite bands." In such a case, the degeneracies will generally only occur at special symmetry *points* or *lines* in the Brillouin zone. (At symmetry *planes*, such as those associated with mirror symmetries, the even and odd irreducible representations are both one dimensional.) Thus, in carrying out the integrals over $k_\perp$ needed to evaluate Eq. (34), such a degeneracy will only be encountered for a set of measure zero of $\mathbf{k}_\parallel$ values, and will thus not affect the integral (32). In any case, if a need should arise to evaluate Eq. (34) at such a special $\mathbf{k}_\parallel$, the method of Ref. 12 (in which the sum over bands and integral over $k_\perp$ are converted into a log of a product of determinants of $M\times M$ matrices) can be used to obtain a convergent and well-defined result.

### C. Spin degeneracy and spin-orbit interaction

In this subsection, we consider the effects of spin degeneracy and spin-orbit interactions. We focus first on the case in which the spin-orbit interaction is absent, and there is no magnetic ordering or external magnetic field present to break the spin degeneracy in the bulk or at the surface. In this case, spin-up and spin-down states are degenerate and uncoupled, and it is natural to take the phases of the spatial wave functions $\psi_{n\mathbf{k}\sigma}(\mathbf{r})$ to be independent of spin $\sigma$. With this convention, the Wannier functions are the same for spin-up and spin-down bands, so that we may associate a charge $-2e$ with each spatial Wannier center.

In this case, it is possible to formulate a stronger version of the surface charge theorem which gives an expression for the surface charge modulo $2e/A_{\rm surf}$, instead of just modulo $e/A_{\rm surf}$. In the case of a system composed entirely of atoms of even $Z$, we have just $\sigma_B = P_\perp$ modulo $2e/A_{\rm surf}$. This result follows immediately from the arguments of the preceding section, using the fact that both electron and ion charges are quantized in units of $2e$. (The integrated electron density $N_3$ for region 3 must be an even integer, since the contributions from spin-up and spin-down electrons must be equal and integral.) In the general case where odd-$Z$ atoms are present, the appropriate generalization is

$$\sigma_B = P_\perp + \frac{e}{A_{\rm surf}} \sum_{j\in {\rm surf}} Z_j \quad ({\rm modulo}\ 2e/A_{\rm surf}) \; , \tag{36}$$

where the sum is over surface atoms only, in a sense to be specified shortly. In a system of even-$Z$ atoms, the last term drops out, as anticipated.

We now explain what is meant by "surface atoms" in Eq. (36). As pointed out in Sec. II A, the choice of an ionic basis is not unique. The polarization $\mathbf{P}$ is invariant modulo $e\mathbf{R}/\Omega$, but not necessarily modulo $2e\mathbf{R}/\Omega$, with respect to the choice of basis. We assume that a definite choice of ionic basis has been made, and that $\mathbf{P}$ is defined consistently. We then imagine tiling up close to the surface by replicating unit cells containing the specified basis; the atoms which remain (i.e., which are not identified with a replicated atom) are identified as "surface atoms." The procedure is illustrated schematically in Fig. 2.

Again, the proof of the claim (36) follows immediately from the considerations of the previous subsections. The "surface charge" associated with charge density $\rho_2(\mathbf{r})$ at the boundary between regions 2 and 3 is again just $P_\perp$, and the integrated electronic density in region 3 is an even integer. The ionic charges to be counted in region 3 are only those that were not included in the charge density of region 2, Eq. (27), the ionic part of which corresponds to replication of the ionic basis, Eq. (28). Additional entire unit cells of ionic charges have no influence on the counting in Eq. (36), since the total ionic charge in the unit cell of a spin-degenerate insulator must be even. Thus, the counting does not depend on precisely where the tiling is stopped. This completes the demonstration of Eq. (36).

We emphasize that a different choice of ionic basis may result in a different definition of $\mathbf{P}$, but it will also result in a different identification of "surface atoms," as illustrated in Fig. 2. The changes are correlated in such a way that the surface charge $\sigma_B$ of Eq. (36) is invariant with respect to the choice of ionic basis. For example, let us assume that the open and shaded atoms in Fig. 2 carry ionic charges $Z_1 = +e$ and $Z_2 = +3e$, respectively, and that the band structure consists of two spin-degenerate bands, one with its Wannier center on the open-atom site and one with its Wannier center on the shaded-atom site. Then the polarization is $\mathbf{P} = \mp(\hat{\mathbf{x}} + \hat{\mathbf{z}})e/2a$ for the choice of Fig. 2(a) or 2(b), respectively ($\hat{\mathbf{z}}$ = surface normal). Meanwhile, $(e/A_{\rm surf})\sum_{j\in{\rm surf}} Z_j = 0$ or $e/a$, respectively, so that in either case $\sigma = P_z + (e/a)\sum_{j\in{\rm surf}} Z_j = -e/2a = 3e/2a$ modulo $2e/a$.

As we shall discuss further in Sec. III E, it is sometimes convenient to make a choice of ionic basis in such a way that one or more of the ionic charges is divided among more than one unit cell; this will be called a "split basis,"

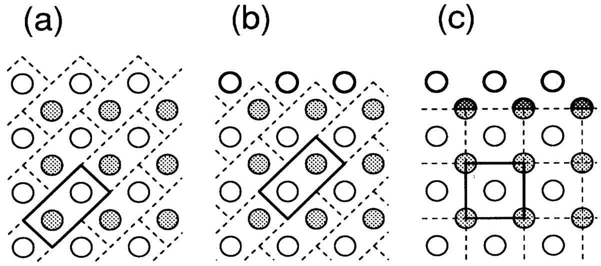

FIG. 2. Possible choices of ionic basis for a system composed of two types of atoms having ionic charges $Z_1 = +e$ (open circles) and $Z_2 = +3e$ (shaded circles). Atoms shown in bold are "surface atoms" which remain after tiling and contribute to the sum in Eq. (36). (a) and (b) Unit cell is specified by two complete basis ions, but in different relative orientations. (c) Unit cell is specified by "split basis" consisting of one complete $+e$ charge and four charges of $+3e/4$ in a symmetric arrangement.

and is illustrated in Fig. 2(c). With this choice of basis, $\mathbf{P} = \mathbf{P}_e = (\hat{\mathbf{x}} + \hat{\mathbf{z}})e/a$ and $\sum_{j\in\mathrm{surf}} Z_j = (+1) + \frac{1}{2}(+3) = 5/2$, thus leading to the same conclusion $\sigma = 7e/2a = 3e/2a$ modulo $2e/a$.

We now turn to the consideration of spin-orbit interactions, which cause the spin-up and spin-down bands to mix and form composite bands. In this case, the surface theorem in its weaker form, Eq. (29), clearly holds, but what about the stronger version expressed in Eq. (36)? As long as the spin-orbit interaction is weak, the arbitrariness modulo $e\mathbf{R}/\Omega$ in $\mathbf{P}$ can be resolved in practice (modulo $2e\mathbf{R}/\Omega$) by comparison to a reference system without spin-orbit interaction, and the stronger theorem applies. To be more precise, we assign to $\mathbf{P}$ the value that it acquires as the spin-orbit interaction is adiabatically turned on from zero strength. Then this polarization will be related to the surface charge via the stronger (36), provided the Fermi level does not lie in a gap between spin-orbit-split surface bands (i.e., provided $\sigma$ also evolves adiabatically as the spin-orbit interaction is turned on).

### D. Reconstructed surfaces

So far, the discussion has been limited to surfaces with primitive $1 \times 1$ periodicity. Suppose now that the surface is reconstructed, so that the surface cell area is increased to $SA_{\mathrm{surf}}$ ($S$ is an integer). Then regarding the bulk crystal as being composed of supercells of volume $S\Omega$, the considerations of the previous subsections imply that the surface charge $\sigma$ equals $P_\perp$ modulo $e/SA_{\mathrm{surf}}$ (or modulo $2e/SA_{\mathrm{surf}}$ in the spin-degenerate case). Thus the surface theorem survives, but in a weakened form.

### E. Remarks on symmetry

It is natural to expect that in crystals of high symmetry, the polarization vector will be highly constrained by the symmetry. This is true, but in a way which is sometimes counterintuitive. For example, with the definitions introduced above, the polarization of a centrosymmetric crystal need not vanish. Conventional vector quantities such as the magnetic polarization $M$ are required to be absolutely invariant under the point-group symmetry operations, and therefore must vanish in a centrosymmetry crystal, or in GaAs for example, whose point group is $T_d$. However, the electric polarization $P$ need only be invariant modulo $e\mathbf{R}/\Omega$ under point-group operations, and need *not* vanish in the above cases. Thus, the role of symmetry deserves special comment.

Recall that the electronic contribution to the polarization is related to the location of the Wannier center, Eq. (15). Symmetry restrictions on the locations of the Wannier centers have been discussed by Zak[20,21] and Michel and Zak.[22] Because the Wannier center is indeterminate modulo a lattice vector, the condition imposed by a symmetry is only that a space-group element must translate the Wannier center by a lattice vector. For example, in a 1D centrosymmetric crystal of lattice constant $a$, with one of the centers of inversion at the origin, the Wannier center can be located either at the origin or at $a/2$.[16,20–22] (It is important to realize that there is no freedom here; once the choice of origin has been made, the Wannier center associated with a given band is either definitely at 0 or definitely at $a/2$.) In the case of Si, for which the four valence bands form a composite group, it is natural to define a common center $\bar{\mathbf{r}}_c = -\Omega\mathbf{P}_e/8e$ (essentially the average of the positions of the four Wannier centers) which is indeterminate modulo $\mathbf{R}/4$. Symmetry considerations alone would allow $\bar{\mathbf{r}}_c$ to be located either at an atomic site or at a midbond site; a calculation is needed to choose between these possibilities. (As we shall see in Sec. V B, $\bar{\mathbf{r}}_c$ turns out to be located on an atomic site in Si or GaAs; *which* site is immaterial, as the two atomic sites are related by a translation vector of the form $\mathbf{R}/4$.)

Let us now consider the value of the total polarization $\mathbf{P} = \mathbf{P}_{\mathrm{ion}} + \mathbf{P}_e$, and suppose that we only want the value of $\mathbf{P}$ modulo $e\mathbf{R}/\Omega$ and not $2e\mathbf{R}/\Omega$. Then, as long as we insist that a split basis (see Sec. III C) *not* be used, $\mathbf{P}$ is uniquely defined modulo $e\mathbf{R}/\Omega$. However, this unique value does not always agree with naive expectations. For example, we shall see that with these conventions $\mathbf{P}$ *does not vanish* in GaAs.

Let us see how this works out for GaAs. To be definite, we take the origin on a Ga atom, and the ionic basis to consist of a nearest-neighbor pair of atoms with $\mathbf{R}_{\mathrm{As}} - \mathbf{R}_{\mathrm{Ga}} = a/4(111)$. Then $\mathbf{P}_e = 0$, and $\mathbf{P} = \mathbf{P}_{\mathrm{ion}} = ea/\Omega(5/4, 5/4, 5/4) = ea/\Omega(1/4, 1/4, 1/4)$ modulo $e\mathbf{R}/\Omega$. The point-group operation $C_2^z$ carries $\mathbf{P} = ea/\Omega(1/4, 1/4, 1/4)$ into $\mathbf{P}' = ea/\Omega(-1/4, -1/4, 1/4) = \mathbf{P} - ea/\Omega(1/2, 1/2, 0) = \mathbf{P}$, since $a(1/2, 1/2, 0)$ is a lattice vector. It is easily verified that all other point-group operations also leave $\mathbf{P}$ invariant modulo $e\mathbf{R}/\Omega$.

So far we have shown that if we insist that a split basis *not* be used, then $\mathbf{P}$ is uniquely defined (modulo $e\mathbf{R}/\Omega$), but it does not vanish in certain high-symmetry cases where we might expect it should. An alternative approach in such cases is to insist on using a split basis which itself reflects the point-group symmetry, e.g., a Ga ion and one-quarter of each of the four neighboring As ions. In this case, $\mathbf{P}_{\mathrm{ion}} = \mathbf{P}_e = \mathbf{P} = 0$ for GaAs.

However, this convention has the drawback that $\mathbf{P}$ may not be uniquely defined in low-symmetry crystals. In any case, we emphasize that one cannot say whether $\mathbf{P} = 0$ or $\mathbf{P} = ea/\Omega(1/4, 1/4, 1/4)$ is the "right" answer without specifying the rules for choosing the ionic basis.

If we want to know $\mathbf{P}$ modulo $2e\mathbf{R}/\Omega$ (not just $e\mathbf{R}/\Omega$), and the system contains odd-$Z$ atoms, then the value of $\mathbf{P}$ may depend on the choice of ionic basis even if we insist that a split basis not be used. Returning to the GaAs example, we would have $\mathbf{P} = \mathbf{P}_{\text{ion}} = 2ea/\Omega(5/8, 5/8, 5/8)$ for the choice of ionic basis specified above, or $\mathbf{P} = 2ea/\Omega(-5/8, -5/8, 5/8) = 2ea/\Omega(3/8, 3/8, 5/8)$ for the choice $\mathbf{R}_{\text{As}} - \mathbf{R}_{\text{Ga}} = a(-1/4, -1/4, 1/4)$. Now, $\mathbf{P}$ need not be invariant modulo $2e\mathbf{R}/\Omega$ under point-group operations; it only need be carried onto a value that would have resulted from a different choice of (nonsplit) ionic basis. This requirement is clearly satisfied in the above example.

Finally, we emphasize that the surface charge is correctly given by Eq. (36) regardless of which of the above definitions is chosen for the ionic basis, as long as the counting of "surface atoms" is done consistently. We will illustrate this for the case of GaAs surfaces in Sec. V B.

### F. Interface theorem

We now discuss the generalization of the surface theorem, Eq. (29) or (36) to the case of an interface between two crystalline insulators. We assume that the crystals are aligned epitaxially in such a way that there is a well-defined interface unit cell of area $A_{\text{int}}$ characterizing the periodicity in the plane of the interface, and that the Fermi level lies in a gap common to both crystals and to the interface. Then one is tempted to generalize Eq. (29) to become

$$\sigma_{\text{int}} = (\mathbf{P}_2 - \mathbf{P}_1) \cdot \hat{\mathbf{n}} \quad (\text{modulo } e/A_{\text{int}}) \,, \tag{37}$$

where $\mathbf{P}_1$ and $\mathbf{P}_2$ are the polarizations on either side of the interface, and $\hat{\mathbf{n}}$ is the interface unit normal.

However, there is a problem with the formulation given in Eq. (37): we have only defined the polarization of the crystals in the absence of a macroscopic electric field. But the presence of $\sigma_{\text{int}}$ implies the existence of a macroscopic electric field on at least one side of the interface. (It also becomes impossible for the Fermi level to lie in a gap common to the entire system.) To overcome these difficulties, we can resort to a formulation in which we imagine that an external planar charge density $\sigma_{\text{ext}}$ has been imposed upon the interface in such a way that the macroscopic electric field does vanish in both crystals. Then we claim that

$$\sigma_{\text{ext}} = -(\mathbf{P}_2 - \mathbf{P}_1) \cdot \hat{\mathbf{n}} \quad (\text{modulo } e/A_{\text{int}}) \tag{38}$$

provided that the Fermi level lies in a gap common to both crystals and the interface.

The demonstration of the interface theorem of Eq. (38) follows easily using arguments similar to those developed in the previous sections. In brief, we construct dividing surfaces on each side of the interface at a distance $L$ sufficiently large to insure the absence of influence from the interface. Now there will be five contributions to the excess planar charge density in the vicinity of the interface; these must sum to zero to be consistent with the absence of a discontinuity in the macroscopic electric field. These are (i) $\sigma_{\text{ext}}$; (ii) $\mathbf{P}_1 \cdot \hat{\mathbf{n}}$, the surface charge coming from the bulklike region of crystal 1 to the left of the dividing surface at $-L$; (iii) $-\mathbf{P}_2 \cdot \hat{\mathbf{n}}$, the corresponding contribution from crystal 2 to the right of $+L$; (iv) an integral number of ionic charges in the interface region; and (v) an integral number of electron charges in the interface region. The quantization of the latter electronic charge contribution again follows from the block-diagonal form and idempotency of the density matrix, which implies that the trace of the interface block must be an integer. The requirement that (i)–(v) sum to zero leads directly to Eq. (38).

The derivation of a stronger version of Eq. (38) in the spin-degenerate case, analogous to Eq. (36), is straightforward.

Of course, the introduction of an external planar charge to cancel the polarization charge is somewhat unphysical. However, in many cases of interest where the polarization difference is small, useful physical information can be extracted from a perturbation analysis. For example, suppose a supercell calculation has been carried out on a structure consisting of periodically alternating slabs of materials 1 and 2, without unphysical external charges, and the self-consistent macroscopic fields in regions 1 and 2 have been found to be $E_1$ and $E_2$, respectively. Then the interfaces carry charges $\sigma = \pm \Delta E/4\pi$, where $\Delta E = E_2 - E_1$. Expanding

$$P_1(E) = P_1 + \chi_1 E + \mathcal{O}(E^2) \tag{39}$$

(and similarly for material 2) and dropping terms of order $E^2$, we find

$$-4\pi \Delta P = \epsilon_2 E_2 - \epsilon_1 E_1 \,. \tag{40}$$

Here $\Delta P = P_2 - P_1$ is the difference in the zero-field or "spontaneous" polarizations of the media, and $\epsilon_1 = 1 + 4\pi\chi_1$ and $\epsilon_2 = 1 + 4\pi\chi_2$ are the dielectric constants of the two materials. If the slab thickness is equal for the two materials, we expect $E_2 = -E_1$ (since the electrostatic potential must be periodic in the supercell), so that $-4\pi\Delta P = \bar{\epsilon}\,\Delta E$ [where $\bar{\epsilon} = (\epsilon_1 + \epsilon_2)/2$] can be used to determine $\Delta P$. In general,

$$-\,4\pi\,\Delta P = \bar{\epsilon}\,\Delta E + \Delta\epsilon\,\overline{E} \,, \tag{41}$$

where $\Delta\epsilon = \epsilon_2 - \epsilon_1$ and $\overline{E} = (E_1 + E_2)/2$. If $\Delta\epsilon$ is not accurately known, it is difficult to determine $\Delta P$ accurately from this calculation alone. However, a second calculation of the same type, with different thicknesses of materials 1 and 2 and thus different $\Delta\epsilon$, can be used to eliminate $\Delta\epsilon$; one finds

$$-\,4\pi\,\Delta P = \bar{\epsilon}\,[f\Delta E + (1-f)\Delta E'], \tag{42}$$

where $f = \overline{E}'/(\overline{E}' - \overline{E})$ and the primes refer to the second slab configuration.

Thus, in practice it should be possible to determine the difference in zero-field polarizations of two materials without resorting to the introduction of unphysical external planar charges at the interface. Moreover, the difference calculated this way should agree with that calculated using Eq. (7) or (8).

Finally, we should like to emphasize that the interface theorem discussed here provides a way of attaching physical significance to the polarization difference calculated between two similar materials. Recall that in our earlier work, Ref. 12, we argued that the difference in the polarization of two insulators, as defined via Eq. (7) or (8), could sometimes be given physical meaning as the change in polarization resulting from an adiabatic evolution of one system into the other, Eq. (6). However, this is only possible if there exists a path for the conversion, such that the system remains insulating along the entire path. This will not always be the case. For example, imagine that we have calculated the polarization difference $\Delta\mathbf{P}$ between ZnO and ZnS in the identical wurtzite structure using an all-electron method. Clearly it will be impossible to construct an insulating path in the Hamiltonian parameter space to convert ZnO into ZnS, since O and S contain different numbers of core electrons, and the number of electrons per unit cell would have to change continuously during the conversion. However, we *can* now attach a meaning to this $\Delta\mathbf{P}$ via Eq. (38), since it is almost certain to be the case that an insulating interface can be arranged between two such similar materials. Thus, we can say that the polarization difference calculated above would be the same as that deduced from supercell calculations on structures composed of alternating slabs of ZnO and ZnS.

These considerations also apply to the calculations of Posternak *et al.*,[8] who reported *ab initio* calculations of the spontaneous electric polarization of wurtzite BeO using a supercell composed of alternating slabs of wurtzite and zinc-blende BeO. Recently, these workers have repeated the calculation of the polarization difference using our Eq. (7), and have found entirely consistent results.[23] This provides strong empirical evidence (if any was needed) that the bulk and interface definitions of the polarization difference are consistent.

## IV. MAPPING ONTO A SYSTEM OF QUANTIZED POINT CHARGES

In this work, we have argued that a useful *bulk* definition of the electric polarization $\mathbf{P}$ can be given. However, we have noted that $\mathbf{P}$ is only well-defined modulo $e\mathbf{R}/\Omega$ or $2e\mathbf{R}/\Omega$, and that the value of $\mathbf{P}$ depends in some cases on the choice of ionic basis. Moreover, some authors have previously argued that no such bulk definition of $\mathbf{P}$ can be given.[9,10] As these considerations may give rise to some confusion, we attempt here to clarify the meaning of $\mathbf{P}$ and the sense in $\mathbf{P}$ is well defined by emphasizing a point of view in which the true quantum-mechanical electron density of the crystal is mapped onto a periodically repeated set of quantized point charges located at the Wannier centers.

For clarity, we restrict ourselves here to the case of insulators in which all of the occupied valence bands are distinct. In such a case, the electronic charge density of each valence band can be replaced, for the purposes of calculating the polarization, by a periodically repeated set of point charges $-e$ located at the corresponding Wannier center in each unit cell. (These positions are the "band centers" in the terminology of Zak.[20,21]) Thus, one arrives at a picture of a fictitious crystal composed of periodically repeated point charges $+Ze$ for each nucleus or core of atomic number $Z$, and $-e$ for each occupied band. The procedure is illustrated in Figs. 3(a) and 3(b). It is crucial that both the positive and negative charges are quantized in units of the elementary charge $e$.

We emphasize that the arbitrariness in our definition of $\mathbf{P}$ is precisely the same as the arbitrariness which would characterize $\mathbf{P}$ for the classical system of point charges, e.g., that of Fig. 3(b). For example, the point-charge system also has an ambiguity about which charges to assign to the unit cell; this leads to an arbitrariness modulo $e\mathbf{R}/\Omega$ in the cell dipole moment per unit volume $\mathbf{P}$. (If the system is composed of even-$Z$ atoms, and the bands are spin degenerate so that the Wannier charges are all $-2e$, then of course $\mathbf{P}$ is well-defined modulo $2e\mathbf{R}/\Omega$.) The comments of Sec. III E about symmetries, e.g., the statement that $\mathbf{P}$ need not be invariant under point-group operations, are also equally relevant for the point-charge system. In short, the subtleties in our definition of $\mathbf{P}$ are neither more nor less problematic than those that arise for any system of periodically repeated, quantized point charges.

Tagantsev has argued explicitly that a bulk definition of the electric polarization is not possible.[9] However, his arguments are based on an implicit assumption that the

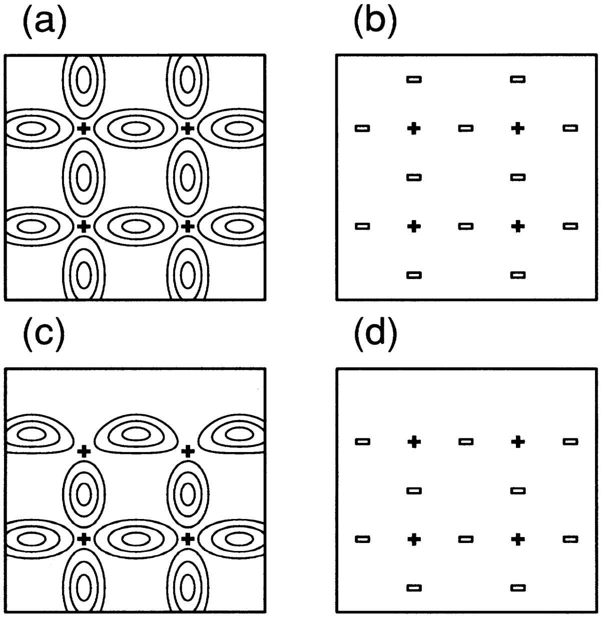

FIG. 3. Illustration of mapping from physical charge density onto a system of point charges quantized in units of $e$. Heavy (+) symbols indicate $+4e$ (ionic) point charges; open (−) symbols indicate $-2e$ (effective electronic) point charges. (a) True charge density of bulk crystal, showing contours of electronic density. (b) Reference point-charge system for bulk crystal. (c) As in (a), but for an insulating surface, top. (d) Reference point-charge system for surface.

electron charges are *not* quantized. Tagantsev points out that a classical system of point charges of *arbitrary* magnitude would *not* have a well-defined polarization vector. This is true. However, a classical system of point charges quantized in multiples of $e$ *does* have a well-defined polarization modulo $e\mathbf{R}/\Omega$. Thus, Tagantsev failed to anticipate that an argument could be given leading to effective quantization of the electron charges.

If we wish, the mapping of the full system onto quantized point charges also allows us to determine the surface charge of an insulating surface, as it would be given by the surface theorem of Eq. (29), without ever defining $\mathbf{P}$ explicitly. The procedure is illustrated in Figs. 3(c) and 3(d). The positive and negative point charges representing the ions and the Wannier centers are tiled up to the surface in their ideal positions, as in Fig. 3(d). Now $\sigma_{\text{surf}}$, the surface charge of this point-charge system of Fig. 3(d), can be determined by any one of the many procedures available for infinite-series summation. For example, we can use the window convolution method[24] to define $\sigma_{\text{surf}} = \int_{-\infty}^{\infty} dz\, \overline{\rho}(z)$, where

$$\overline{\rho}(z) = \frac{1}{\Omega} \int_{A_{\text{surf}}} dx\, dy \int_{z-c/2}^{z+c/2} dz\, \rho(x,y,z) \qquad (43)$$

(here $c$ is the lattice constant along the surface normal direction $\hat{\mathbf{z}}$). For the system of Fig. 3(d), we get $\overline{\rho} = 0$ except for $\overline{\rho}(z) = 2e/\Omega$ inside $0 \le z \le c/2$, or $\sigma_{\text{surf}} = e/A_{\text{surf}}$. Alternatively, we may imagine splitting each charge $-2e$ into two contributions of $-e$ which are then moved symmetrically apart (so as not to disturb the dipole moment) until they coincide with and cancel the $+4e$ ionic charges; this leaves all charges zero except on the surface ions, which carry a net charge $+e$, leading again to a surface charge $e/A_{\text{surf}}$. Thus, we would claim that the real insulating surface of Fig. 3(c) should have the same surface charge as is deduced in this way from the point-charge system of Fig. 3(d), modulo $2e/A_{\text{surf}}$ (in the spin-degenerate case). Alternatively, we would claim that there is a way of truncating the tiling of the $-2e$ electron charges at the surface such that the surface charge of the reference point-charge system exactly equals $\sigma_{\text{surf}}$ of the real insulating surface. This point of view allows one to avoid defining $\mathbf{P}$ altogether, so that any potential confusion about the choice of ionic basis (see Sec. III E) is circumvented.

In the discussion above, the ionic and electronic point charges at the surface were not displaced from their ideal (bulk-related) locations. Such displacements would not affect $\sigma_{\text{surf}}$, but *would* affect the surface dipole density. Thus, while the reference system of quantized point charges of Fig. 3(d) gives the correct $\sigma_{\text{surf}}$, it *cannot* be expected to give the right surface dipole. However, it appears likely that if the ionic charges are displaced to their correct surface locations, and if the electronic point charges $-2e$ are placed at the centers of charge of the *surface Wannier functions* defined in Ref. 19, then the resulting system of point charges *would* give the correct surface dipole as well as surface charge.

## V. EXAMPLES OF SURFACE CHARGE COUNTING

In this section, we consider several examples of tight-binding models for which the calculation of the polarization is relatively elementary. In some cases, we explicitly calculate the surface charge and find that it does indeed agree with the prediction based on the bulk polarization.

### A. One-dimensional tight-binding models

#### *1. Alternating sites model*

First consider a one-dimensional tight-binding chain with one $s$ orbital per site and having alternating site energies and nearest-neighbor hopping matrix elements as follows:

$$H = \sum_j \left\{ \epsilon_j\, c_j^\dagger c_j \;+\; V_{j,j+1}\, [c_j^\dagger c_{j+1} + \text{H.c.}] \right\}, \qquad (44)$$

where

$$\begin{aligned} \epsilon_{2m} &= -\Delta\,, \qquad & V_{2m-1,2m} &= -t-\delta\,,\\ \epsilon_{2m+1} &= \;\;\Delta\,, \qquad & V_{2m,2m+1} &= -t+\delta\,. \end{aligned} \qquad (45)$$

Here $m$ is an integer, and atom $j$ is located at $x_j = ja/2$, where $a$ is the lattice constant. We consider the model at half filling, with each band having twofold spin degeneracy, and assign charges $Z = +e$ to both kinds of ions to maintain neutrality. A model of this kind has been considered previously[25] in the context of solitons in polyenes.

The chain is metallic for $\Delta = \delta = 0$, but is otherwise insulating. For $\delta = 0$, the Wannier center of the lower band is found to be located on an atomic site, at $x = 0$ or $x = a/2$ for $\Delta > 0$ or $\Delta < 0$, respectively, consistent with the reflection symmetry about the atomic sites. For $\Delta = 0$, on the other hand, the Wannier center is at a midbond site, at $x = a/4$ or $x = 3a/4$ for $\delta > 0$ or $\delta < 0$, respectively, again reflecting the symmetry. It is instructive to let the Hamiltonian be parametrized according to

$$\begin{aligned} \Delta &= \Delta_0 \cos(\theta)\,,\\ \delta &= \delta_0 \sin(\theta)\,, \end{aligned} \qquad (46)$$

with $\Delta_0 > 0$ and $\delta_0 > 0$. Then, if $\theta$ is increased from 0 to $2\pi$, the Wannier center is found to shift continuously by a lattice vector $+a$, so that the polarization $P$ decreases continuously by $2e$, as shown in Fig. 4. These results are thus consistent with the quantization of charge transport[26,27] expected, e.g., for a sliding charge-density wave.

Figure 4 also shows that a knowledge of the bulk quantity $\mathbf{P}(\theta)$ is indeed sufficient to predict the variation of the surface charge with $\theta$. The surface charge predicted from the bulk using Eq. (29) and 120 $k$ points, and that obtained directly from analyzing the occupied states of a 101-site chain, are found to agree to more that six significant figures.

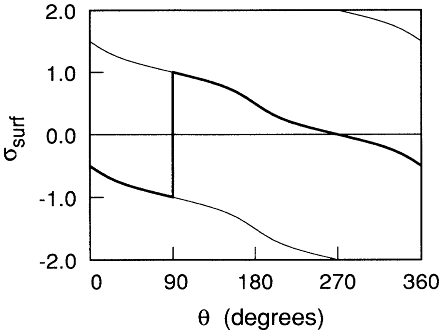

FIG. 4. Surface charge (in units of $e$) for the one-dimensional tight-binding model of Eqs. (44) and (45). Light line, surface charge predicted from bulk polarization, Eq. (7), arbitrary modulo $2e$. Heavy line, actual charge calculated to reside on one end of a long but finite chain. The Hamiltonian of Eqs. (44) and (45) is parametrized by $t = 1.0$, $\Delta = 0.6\cos(\theta)$, and $\delta = 0.6\sin(\theta)$. A surface state crosses the Fermi level ($\mu = 0$) at $\theta = 90°$.

#### *2. Coupled s and p bands*

We now consider another two-band model, this one having only one site per cell (ionic charge $+2e$, lattice constant $a$), but having one $s$ and one $p_x$ orbital per site. The Hamiltonian can be written

$$H = \sum_{j\alpha} \epsilon_\alpha c^\dagger_{j\alpha} c_{j\alpha} + \sum_{j\alpha\beta} V_{\alpha\beta} \, [\, c^\dagger_{j\alpha} c_{j+1,\beta} + \text{H.c.}\,] \,, \qquad (47)$$

where $\alpha = \{s,p\}$, and $V_{sp} = -V_{ps}$. The model is parametrized by $V_{sp}$, $V_{ss}$, $V_{pp}$, and $\epsilon_{sp} = \epsilon_p - \epsilon_s$. We take $\epsilon_{sp} > 0$, $V_{ss} < 0$, $V_{pp} > 0$, and $V_{sp} \neq 0$.

The interesting feature of this model is that it has two regimes of behavior. When $\epsilon_{sp} > 2V_{pp} - 2V_{ss}$ the Wannier center is found to be localized on an atomic site ($P = 0$), whereas it is located on the midbond site ($P = e$) when $\epsilon_{sp} < 2V_{pp} - 2V_{ss}$. (Because of the symmetry of the model, these are the only two possibilities. When $\epsilon_{sp} = 2V_{pp} - 2V_{ss}$, there is an accidental degeneracy at $k = \pi/a$ and the system is metallic.) When $\epsilon_{sp}$ is larger than the critical value, we can think of the lower (upper) band as being $s$-like ($p$-like); otherwise, we can interpret the lower (upper) band as having the character of $sp$-hybrid bonding (antibonding) orbitals. These assignments correspond to the symmetry of the Wannier functions. The former case corresponds to an "ionic" picture, while the latter is "covalent."

Again, the surface theorem of Eq. (29) predicts that the distinction between these two regimes should be reflected in the surface charge assigned to any insulating surface. We have checked this by calculating the surface charge explicitly on long but finite chains for many different choices of parameters, and find that the surface charge is indeed always 0 (modulo $2e$) when $\epsilon_{sp} > 2V_{pp} - 2V_{ss}$, and $e$ (modulo $2e$) otherwise, as expected.

On the other hand, the distinction between the two regimes of behavior would *not* be reflected in any way in the symmetry of the charge density. The discussion above is presented in the context of a tight-binding model, but it would not be difficult to exhibit a 1D or 3D continuous Hamiltonian showing the same qualitative behavior. The charge density of such a model would be centrosymmetric about *both* the atomic site and the midbond site, in *both* regimes. Thus, a formulation of the form of Eq. (3) would be incapable of distinguishing between the regimes. In fact, the most natural approach based on Eq. (3) would be to take the unit cell $\Omega$ to be centered about an atomic site, in which case one would expect vanishing surface charge regardless of regime, a conclusion which is *incorrect.*

### B. Tetrahedral semiconductors

In Sec. III E, we claimed that for tetrahedral semiconductors such as Si or GaAs in the diamond or zinc-blende structure, the average Wannier center $\bar{\mathbf{r}}_c = -\Omega \mathbf{P}_e/8e$ for the four valence bands (arbitrary modulo $\mathbf{R}/4$) can be taken to be located at an atomic site. (Whether it is a cation or anion site is irrelevant as these are related by a displacement of the form $\mathbf{R}/4$.) This claim is not as trivial as it sounds; one might have thought that it should be located at a bond-center site, which is a site of inversion symmetry in homopolar semiconductors. Here, we will substantiate the claim that $\bar{\mathbf{r}}_c$ lies on an atomic site, and discuss the consequences of this fact for the surface charges of semiconductor surfaces of (100), (110), and (111) orientations.

First, we note the work of Kohn,[17] in which a variational procedure is described for constructing four "bond-orbital" Wannier functions centered at midbond sites, for homopolar tetrahedral semiconductors. Such an arrangement was also suggested by Zak.[20] The four bonds can be chosen to be neighbors of a given atomic site, so that $\bar{\mathbf{r}}_c$ is indeed located on an atomic site.

Second, we have calculated $\mathbf{P}_e$ explicitly for GaAs, using a plane-wave pseudopotential density-functional approach to calculate the wave functions, and using the method of Ref. 12 to evaluate Eq. (7). Our calculation used a plane-wave cutoff of 20 Ry for the wave functions, with the Wigner form for exchange and correlation, and ignored spin-orbit effects. The self-consistent potential was computed using 10 $k$ points in the irreducible wedge of the zone. For the evaluation of the polarization we used a Brillouin zone with its axis along the (111) direction, which allows the projection of $\Omega \mathbf{P}_e/2e$ along the (111) direction to be determined modulo $a/\sqrt{3}$. We used a total of 16 $k$-point strings with 20 $k$ points per string. With this $k$-point sampling, and with the origin on the Ga atom, we found that $\Omega \mathbf{P}_e/2e$ does indeed vanish to within $0.000\,175a$; better $k$-point sampling would reduce this value even further.

Third, in order to check that the above result is not unique to GaAs, we have investigated the electronic polarization of the zinc-blende II-VI, III-V, and IV-IV semiconductors systematically within the framework of Harrison's tight-binding model.[28] Harrison proposed that the tight-binding parameters for all $sp$-bonded semiconductors could be taken in the ratio $V_{ss\sigma} : V_{sp\sigma} : V_{pp\sigma} : V_{pp\pi} = -1.32 : 1.42 : 2.22 : -0.63$ with a prefactor $\hbar^2/md^2$ depending on the bond length $d$, and the free-atom eigenvalues $\epsilon_s$ and $\epsilon_p$ taken from Hartree-Fock term values. We make the approximation that anion and cation share a common $s$-$p$ splitting $E_{sp} = \epsilon_p - \epsilon_s$, so that a given semiconductor is characterized by two dimensionless parameters which we take to be the hopping $t$ (in units such that $t = 1$ gives the hopping prefactor 1.380 appropriate for Si), and anion-cation splitting $\alpha$ (in units of the $s$-$p$ splitting $E_{sp} = 7.21$ eV appropriate for Si).

We have systematically explored the behavior of the bands, the location of the Wannier centers, and the resulting electronic polarization, as a function of $t$ and $\alpha$. The grouping of the bands is shown in the "phase diagram" of Fig. 5. The vertical axis corresponds to the homopolar case, for which we find the expected band ordering when $|t| > 0.797$: at $\Gamma$ the ordering is $\Gamma_1$, $\Gamma_{25'}$, $\Gamma_{2'}$, and $\Gamma_{15}$, from bottom to top. When degeneracies at the $X$ point are taken into account, this means that one finds two groups of four bands, for which we introduce the notation "4/4" (ordering is from lower to higher energy). Since four bands are occupied, the system is insulating. For $|t| < 0.797$ the $\Gamma_{2'}$ level falls below $\Gamma_{25'}$, so that the band grouping becomes 2/6 and the system is metallic. Exploring the rest of the $\alpha - t$ plane, we find a metallic region near the origin (band ordering 1/1/3/3) and an insulating region outside (ordering 1/3/1/3).

Even without an explicit calculation, it is possible to deduce the electronic polarization for the insulating region of Fig. 5. For $\alpha > 1$ and $t = 0$ the basis states are eigenstates of the Hamiltonian, and the occupied Wannier functions are just the $s$ and $p$ states on the anion. Thus, the common Wannier center $\bar{\mathbf{r}}_c$ of the four occupied bands is located on the anion site, or equivalently, $\mathbf{P}_e = 0$ (modulo $2e\mathbf{R}/\Omega$) when referred to an origin on either atomic site. Now any other point in the insulating region of the $\alpha - t$ plane can be reached by a path which remains in the insulating region. Since the system evolves smoothly and adiabatically in this insulating region, $\mathbf{P}_e$ can only change continuously. However, the system also retains tetrahedral ($T_d$) symmetry throughout, which implies (see Sec. III E) that only a discrete set of values of $\mathbf{P}_e$ are possible, one of which is $\mathbf{P}_e = 0$. Thus, $\mathbf{P}_e$ must vanish everywhere in the insulating region. We have verified this conclusion by calculating $\mathbf{P}_e$ explicitly, again using the method of Ref. 12 to evaluate Eq. (7), and we do indeed find that it vanishes everywhere in the insulating region. (As $\alpha$ passes through zero at $|t| > 0.797$, the Wannier center of the bottom band shifts discontinuously from anion to cation, but a corresponding change in the common Wannier center of the other three occupied bands leaves $\mathbf{P}_e$ invariant.)

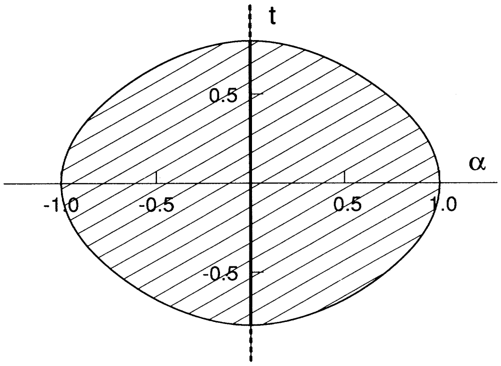

FIG. 5. Phase diagram showing the behavior of the bands for a tight-binding model of a tetrahedral semiconductor characterized by a dimensionless hopping strength $t$ (such that $t = 1$ for Si), and anion-cation self-energy difference $\alpha$ (in units of the $s$-$p$ splitting $E_{sp} = \epsilon_p - \epsilon_s$ common to both anions and cations). Unhashed region: bands are grouped as 1/3/1/3; heteropolar insulator. Hashed region: bands are grouped as 1/1/3/3; metal. Heavy solid line: bands are grouped as 2/6; metal. Heavy dashed line: bands are grouped as 4/4; homopolar insulator. Si lies at $\alpha = 0$, $t = 1$.

We expect that real II-VI, III-V, and IV-IV semiconductors will be fairly well represented by a point in the insulating portion of the $\alpha - t$ plane of the tight-binding model. This, together with the *ab initio* calculation discussed above for the case of GaAs, provides strong circumstantial evidence that all such semiconductors have vanishing electronic polarization $\mathbf{P}_e$, when referred to an origin on either atomic site.

The consequences of this assignment for semiconductor surface charge densities are summarized in Table I. We illustrate the derivation of the entries in this table by discussing one case, that of the Ga-terminated GaAs (111) surface. We ask what surface charge must be present (modulo $2e/A_{\text{surf}}$) if the surface has $1 \times 1$ translational symmetry and is insulating. Taking the origin on an As site, and choosing a *split* ionic basis consisting of one As ion (charge $+5e$) and four neighboring fractional ions (each 1/4 of a Ga ion, charge $+3e/4$), we have $\mathbf{P}_{\text{ion}} = \mathbf{P}_e = 0$ and thus $P_\perp = 0$. From Eq. (36) it follows that the surface charge density $\sigma$ is obtained just by counting the ionic charge which is unaccounted for at the surface by a tiling of the split ionic basis up close to the surface. For example, if the last replication included in the tiling is the one centered on the second-layer As atom, then just 1/4 of a Ga ion (or charge $+3e/4$ per surface cell) is unaccounted for, so that $\sigma = +3e/4A_{\text{surf}}$.

Use of a nonsplit ionic basis is a little more complicated, but leads to the same conclusions. For example, let us repeat the above derivation, but using an ionic basis consisting of a nearest-neighbor pair of atoms with $\mathbf{R}_{\text{As}} - \mathbf{R}_{\text{Ga}} = a/4(111)$. Then $\mathbf{P} = \mathbf{P}_{\text{ion}} = 2ea/\Omega(5/8, 5/8, 5/8)$, or $A_{\text{surf}} P_\perp = 15e/4 = 7e/4$ (mod-

TABLE I. Electron counting rules for surfaces of semiconductors of zinc-blende or diamond structure. Listed are the excess numbers of electrons per $1 \times 1$ cell $A_{\text{surf}}\sigma_{\text{surf}}/e$, arbitrary modulo 2, which would have to be present for an insulating $1 \times 1$ surface.

| Surface orientation | Surface layer | II-VI | III-V | IV-IV |
|---|---|---|---|---|
| (100) | cation | 1 | 3/2 | 0 |
| (100) | anion | 1 | 1/2 | 0 |
| (110) | mixed | 0 | 0 | 0 |
| (111) | cation | 1/2 | 3/4 | 1 |
| $(\bar{1}\bar{1}\bar{1})$ | anion | 3/2 | 5/4 | 1 |

ulo $2e$). We can let the tiling of the ionic basis include up to the last As layer, leaving the surface Ga layer unaccounted for, so that from Eq. (36) we obtain $A_{\text{surf}}\sigma = 7e/4 + 3e = 3e/4$ (modulo $2e$), in agreement with the conclusion of the previous paragraph. Other choices of ionic basis would lead to the same conclusion.

The results of Table I are consistent with counting rules already discussed elsewhere in the literature;[29,30] in fact, they can be taken as a new derivation of these rules. Such rules are frequently useful for guessing low-energy surface structures, which are usually both insulating and neutral. Returning to the case of GaAs (111), for example, it is immediately apparent that a 1 × 1 surface could not be both neutral and insulating. On the other hand, a 2 × 2 reconstruction with one surface Ga atom removed could be. A reconstruction of precisely this type is believed to have been observed[31] and to be the energetically favored structure[32,33] of the GaAs (111) surface.

## VI. GENERALIZATION TO THE MANY-BODY CASE

So far, we have focused primarily on a discussion of polarization effects using a single-particle description of the electrons. In this section, we briefly discuss the generalization of this work to the cases where many-body effects are important.

We begin by pointing out that by application of density-functional theory (DFT) we already have a many-particle theory of polarization effects. The key point here is that ground-state charge densities and thus electrical polarizations are in principle obtained exactly within DFT.[6] For example, if we know the Kohn-Sham wave functions of an insulating solid with an insulating surface, then we can certainly construct a set of Wannier functions from these orbitals and compute the surface charges following the methods of Sec. III. This is an exact method for computing surface charges given that the many-body and DFT charge densities are identical; the fact that the DFT Wannier functions may bear no relation to the many-particle ground-state is immaterial. Similarly Eq. (1) will correctly predict the polarization current of Kohn-Sham electrons in the bulk of a large but finite insulator where the potential is allowed to undergo an adiabatic evolution, provided $V^{(\lambda)}$ is interpreted as the self-consistent Kohn-Sham potential. The polarization current calculated from the many-body wave functions and the Kohn-Sham polarization current must be identical, if the Kohn-Sham and many-body surface charges are to agree for each value of $\lambda$.

It should be borne in mind that the above arguments may rely heavily on the nonlocal behavior of the exact Kohn-Sham functional, and may even break down in certain circumstances. For example, there is evidence[34] that some insulating materials may have metallic Kohn-Sham band structures, which would invalidate the use of Eq. (1). Thus, it is also important to explore more direct approaches that work directly with the $N$-particle wave functions. An important advance has been made in this regard by Niu and Thouless,[35] who have shown quite generally that charge transport on a ring is quantized in units of $e$ for adiabatic changes in the Hamiltonian where $H^{(\lambda=0)} = H^{(\lambda=1)}$, provided there is no closing of the gap between the ground and first excited states along the path, and that there are no long-range correlations in the wave function.

Finally, it should be pointed out that symmetry-related properties of $\mathbf{P}$ are expected to survive in the many-body case, provided that no phase transition or gap closure occurs as the many-body part of the Hamiltonian is turned on. For example, the exact many-body polarization of GaAs presumably still vanishes (using a symmetric split basis), just as it does in the independent-electron approximation.

## VII. CONCLUSIONS

In Ref. 12, we showed that the quantity defined in Eq. (7), or equivalently Eq. (8), can be assigned a physical significance as follows: the *change* in this quantity as the system adiabatically follows an insulating path in the parameter space of the Hamiltonian correctly predicts the resulting charge transport (i.e., the integrated polarization current). This has also been the viewpoint in a recent review by Resta[36] of our earlier development.[12] In the present work, we show that this quantity can be assigned another meaning which does not involve differences: it is simply related to the areal charge density which accumulates at an insulating surface or interface bounding the crystal, modulo the natural unit $e/A_{\text{surf}}$ or $2e/A_{\text{surf}}$.

In view of these two connections, we believe it is reasonable and natural to take Eq. (7) or (8) as a *definition* of the bulk electronic polarization of an insulating crystalline solid.

We acknowledge that there may be situations for which this definition is not useful. For example, one must admit the possible existence of an insulating crystalline solid for which no insulating surface can be constructed, and which cannot be connected via an insulating path in parameter space to any significantly different crystal (i.e., the given crystal is located in a small "island" of insulating Hamiltonians entirely surrounded by metallic ones). In such a case, it is indeed difficult to see how the polarization $\mathbf{P}$ calculated via Eq. (7) can be assigned any physical significance. However, it appears unlikely that such cases will be common.

## ACKNOWLEDGMENTS

We would like to thank Raffaele Resta for useful discussions. This work was supported by the Office of Naval Research under Contract No. N00014-91-J-1184, and by NSF Grant No. DMR-91-15342.

[1] R. Landauer, Solid State Commun. **40**, 971 (1981).

[2] R. M. Martin, Phys. Rev. B **5**, 1607 (1972); **6**, 4874 (1972).

[3] C. Kallin and B. I. Halperin, Phys. Rev. B **29**, 2175 (1984).

[4] R. Landauer, Ferroelectrics **73**, 41 (1987).

[5] R. Resta, Ferroelectrics **136**, 51 (1992).

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

[7] E. Fradkin, *Field Theories of Condensed Matter Systems* (Addison Wesley, Redwood City, 1991).

[8] M. Posternak, A. Baldereschi, A. Catellani, and R. Resta, Phys. Rev. Lett. **64**, 1777 (1990).

[9] A. K. Tagantsev, Phase Trans. **35**, 119 (1991).

[10] A. K. Tagantsev, Phys. Rev. Lett. **69**, 389 (1992).

[11] M. V. Berry, Proc. R. Soc. London, Ser. A **392**, 45 (1984).

[12] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).

[13] J. A. Applebaum and D. R. Hamann, Phys. Rev. B **10**, 4973 (1974).

[14] F. Claro, Phys. Rev. B **17**, 699 (1978).

[15] E. I. Blount, in *Solid State Physics,* edited by F. Seitz and D. Turnbull (Academic, New York, 1962), Vol. 13, p. 305.

[16] W. Kohn, Phys. Rev. **115**, 809 (1959).

[17] W. Kohn, Phys. Rev. B **7**, 4388 (1973).

[18] W. Kohn and J. R. Onffroy, Phys. Rev. B **8**, 2485 (1973).

[19] J. J. Rehr and W. Kohn, Phys. Rev. B **10**, 448 (1974).

[20] J. Zak, Phys. Rev. Lett. **48**, 359 (1982).

[21] J. Zak, Phys. Rev. Lett. **62**, 2747 (1989).

[22] L. Michel and J. Zak, Europhys. Lett. **18**, 239 (1992).

[23] R. Resta, M. Posternak, and A. Baldereschi, Phys. Rev. Lett. **70**, 1010 (1993); in *Materials Theory and Modelling,* edited by P. D. Bristowe and J. Broughton, MRS Symposia Proceedings No. 291 (Materials Research Society, Pittsburgh, 1993).

[24] A. Baldereschi, S. Baroni, and R. Resta, Phys. Rev. Lett. **61**, 734 (1988).

[25] M. J. Rice and E. J. Mele, Phys. Rev. Lett. **49**, 1455 (1982).

[26] D. J. Thouless, Phys. Rev. B **27**, 6083 (1983).

[27] H. Kunz, Phys. Rev. Lett. **57**, 1095 (1986).

[28] W. A. Harrison, Phys. Rev. B **24**, 5835 (1981); **27**, 3592 (1983).

[29] D. J. Chadi, J. Vac. Sci. Technol. A **5**, 834 (1987).

[30] H. H. Farrell, J. P. Harbison, and L. D. Peterson, J. Vac. Sci. Technol. B **5**, 1482 (1987); H. H. Farrell and C. J. Palmstrom, *ibid.* **8**, 903 (1990).

[31] S. Y. Tong, G. Xu, and W. N. Mei, Phys. Rev. Lett. **52**, 1693 (1984).

[32] D. J. Chadi, Phys. Rev. Lett. **52**, 1911 (1984).

[33] E. Kaxiras, Y. Bar-Yam, J. D. Joannopoulos, and K. C. Pandey, Phys. Rev. B **33**, 4406 (1986).

[34] R. W. Godby and R. J. Needs, Phys. Rev. Lett. **62**, 1169 (1989).

[35] Q. Niu and D. J. Thouless, J. Phys. A **17**, 2453 (1984).

[36] R. Resta, Europhys. Lett. **22**, 133 (1993); Rev. Mod. Phys. (to be published).