

Why are insulators insulating and metals conducting?

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2002 J. Phys.: Condens. Matter 14 R625

(<http://iopscience.iop.org/0953-8984/14/20/201>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.104

The article was downloaded on 18/05/2010 at 06:41

Please note that [terms and conditions apply](#).

TOPICAL REVIEW

Why are insulators insulating and metals conducting?

Raffaele RestaINFN–Dipartimento di Fisica Teorica, Università di Trieste, Strada Costiera 11,
I-34014 Trieste, Italy¹

and

Max-Planck-Institut für Festkörperforschung, D-70506 Stuttgart, Germany

Received 16 November 2001

Published 9 May 2002

Online at stacks.iop.org/JPhysCM/14/R625**Abstract**

At variance with what happens in metals, the electronic charge in insulators cannot flow freely under an applied dc field, and undergoes instead static polarization. These two features arise from the difference in nature of the excitation spectra, but also from the difference in organization of the electrons in their *ground state*: electrons are localized in insulators and delocalized in metals. Such localization, however, is hidden in a rather subtle way in the many-body wavefunction. We review the theory of the insulating state, on the basis of electron localization, addressing on the same basis *all* insulators: either independent electron or correlated, either crystalline or disordered. The starting point is a 1964 milestone paper by Kohn. Significant advances occurred from 1999 onwards. These advances are deeply rooted in the modern theory of polarization: localization and polarization can be regarded as two aspects of the same phenomenon, and stem from essentially the same formalism. Starting from the many-body ground wavefunction, one defines a dimensionless complex number which vanishes in metals and is finite in insulators; in the latter case, its phase (the Berry phase) yields the macroscopic polarization, while its modulus measures localization. Conductivity features are addressed within the same theoretical scheme.

1. Introduction

For the present purposes, materials are conveniently divided into just two classes: insulators and metals. We use the terms here in a loose sense, as synonyms for nonconducting and conducting: an insulator is distinguished from a metal by its vanishing conductivity at low temperature and low frequency. In contrast to what happens in metals, the electronic charge in insulators cannot flow freely under an applied dc field, and instead undergoes static polarization. We focus here on purely electronic phenomena, where the ions are considered ‘clamped’ and the temperature is strictly zero. The properties of interest, therefore, only relate to electronic wavefunctions.

¹ Permanent address.

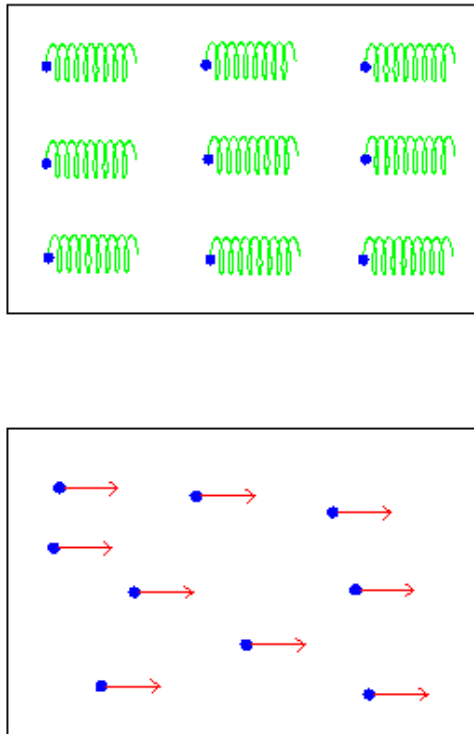


Figure 1. A schematic view of insulators and metals in classical physics. Top sketch: the Lorentz model for insulators, where each electron is tied (by a harmonic force) to a particular centre. Bottom sketch: the Drude model for metals, where electrons roam freely over macroscopic distances, hindered only by atomic scattering potentials.

(This figure is in colour only in the electronic version)

Within classical physics, the qualitative difference between insulators and metals is attributed to the natures of the electronic charge, as sketched in figure 1: either ‘bound’ (the Lorentz model for insulators) or ‘free’ (the Drude model for metals). In other words, electrons are *localized* in insulators and *delocalized* in metals. Switching to quantum physics, this kind of sharp distinction is apparently lost. Most textbooks [1, 2] address the insulating/metallic behaviour by means of band structure theory, focusing on the position of the Fermi level of the given material: either in a band gap (insulators), or across a band (metals), as in figure 2. This picture explains the insulating/metallic behaviour by means of the spectrum of the system, considering the nature of the low-lying electronic *excitations*. But the qualitative difference in the dc conductivity also reflects a difference in the organization of the electrons in their *ground state*, as first emphasized by Kohn in 1964 [3, 4]. The outstanding message of this article is that even within quantum mechanics the *cause* for insulating behaviour is electron localization. Such localization, however, manifests itself in a very subtle way; in fact the electrons in a condensed system appear, from several viewpoints, about equally delocalized in insulators and in metals.

The challenge is to show how electron localization can be detected and measured in the ground wavefunction of a many-electron system. The difference between localized and delocalized must be, in the thermodynamic limit, a sharp one. The first solution to this problem

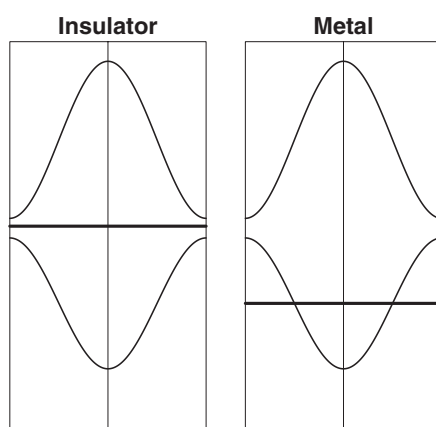


Figure 2. Electronic energy bands in a mean-field potential for a crystalline system. Insulators and metals are distinguished by the position of the Fermi level; this is the traditional textbook view of their qualitative difference.

was provided by Kohn in his original 1964 paper. Much later, starting since 1999, the problem was reconsidered and a solution different from Kohn's, and complementary to it, was found; this is the main theme of the present review.

Why do we need a theory of the insulating state different—and formally more complex—than the familiar one sketched in figure 2? The point is that such a picture applies only to a crystalline material, within the independent-electron approximation [1]—a limited class of insulators indeed. Noncrystalline insulators do in fact exist, and the electron–electron interaction is a fact of nature: in some materials the insulating behaviour is dominated by disorder (Anderson insulators); in some other materials the insulating behaviour is dominated by electron correlation (Mott insulators). Even more exotic insulators are also known, e.g. the excitonic insulators [4].

For a large number of insulators the band picture is then grossly inadequate. Instead, the theory of the insulating state reviewed here, and based on electron localization, addresses *all* insulators on the same basis. Even in the very special case of figure 2 the ground state is indeed localized (in a manner defined below), despite the fact that the eigenfunctions of the independent-particle Hamiltonian underlying figure 2 are of the (delocalized) Bloch form both for insulators and for metals—and qualitatively rather similar. The kind of localization addressed in the present work is *not* a feature characterizing the individual single-particle orbitals; it is, instead, a global property of the many-body ground wavefunction *as a whole*.

The modern approach to electron localization is deeply rooted in the theory of polarization, developed from 1992 onwards [5–13], and based on Berry's phases [14–16]: there is in fact an outstanding phenomenological link between macroscopic polarization and the insulating state of matter. Suppose we expose a finite macroscopic sample to an electric field, say by inserting it in a charged capacitor. Then the induced macroscopic polarization is qualitatively different in metals and insulators. In the former materials, polarization is trivial: universal, material independent, due to surface phenomena only (screening by free carriers). Therefore polarization in metals is *not* a bulk phenomenon. The opposite is true for insulators: macroscopic polarization is a nontrivial, material-dependent, bulk phenomenon. We can therefore characterize an insulator, in very general terms, as a material whose ground wavefunction sustains a bulk macroscopic polarization whenever the electronic Hamiltonian is

noncentrosymmetric. If the Hamiltonian is instead centrosymmetric, the polarization vanishes but remains a well-defined bulk property, at variance with the metallic case. On the theoretical side, the concept itself of macroscopic polarization in quantum physics has long evaded even a correct definition; most textbooks contain incorrect statements. The modern theory revolutionizes both the very definition of the relevant bulk observable, and the ways to compute it in real solids. Much novel and deep physical insight has been gathered in recent years from calculations based on the Berry phase theory, both for real materials and for model systems [17].

The modern theory of electron localization in the insulating state was developed in 1999 in a short paper by Resta and Sorella [18]. Its major message is that, since polarization and localization are two manifestations of the same phenomenon, they both stem from essentially the same formalism. This work was followed in 2000 by a comprehensive paper by Souza *et al* [19], where important additional results are found, in particular relating localization to conductivity. In the following, we will refer to these papers as RS and SWM, respectively. Other related papers have appeared as well [20–24]. The major part of the present review is based on RS and SWM, although the presentation follows a different logical path to both of them: a more general perspective is adopted, and additional unpublished results are included.

The plan of this work is as follows. In section 2 I introduce some notation, definitions, and general results which will be used throughout. The formalism addresses all condensed systems (either crystalline or disordered) with periodic boundary conditions (PBC), and deals with both correlated- and independent-electron wavefunctions. In section 3 I outline the drawback which has long plagued the theory of macroscopic polarization, and the modern solution of the problem, based on a Berry's phase. The same main ingredient as was used therein, namely, the many-body phase operator, is used in the following to address electron localization as well. In section 4 I address the other property that, in addition to macroscopic polarization, discriminates between insulators and metals: conductivity. In section 5, after reviewing the historic work of Kohn, I illustrate the modern approach to electron localization in insulators, as it emerges from RS and SWM. In section 6 I discuss in detail how the RS localization is measured in a crystalline system of independent electrons (finite in insulators, diverging in metals) and in an explicitly correlated model system.

2. Wavefunction, orbitals, and density matrices

2.1. Periodic boundary conditions

When dealing with condensed matter physics, one is generally interested in evaluating physical quantities—either intensive or extensive—in the thermodynamic limit. Such quantities will also be called ‘bulk properties’ in the following. They are defined by taking the limit of a very large system, keeping the density (and any other relevant parameter) constant. On practical grounds, it is enough to watch that the size of the system is larger than the relevant correlation lengths.

Suppose we address a crystalline solid: the thermodynamic limit could in principle be approached by studying a ‘molecule’, or cluster, of increasing size and suitably chosen shape, such that the surface-to-volume ratio goes to zero and the genuinely bulk properties emerge in the limit. While any N -electron wavefunction becomes an ill-defined mathematical object in the $N \rightarrow \infty$ limit, the reduced density matrices [25] do converge to well-defined quantities in the inner region of the cluster, thus yielding the bulk electronic properties of interest. Notice that the ‘molecules’ are thought of as isolated in space; at any finite N the ground wavefunction is therefore bound, that is square integrable and exponentially decaying; the density matrices are bound as well.

However, the finite-molecule approach is definitely not the most convenient one. When addressing the bulk properties of any condensed system (either crystalline, amorphous, or liquid) one adopts almost invariably Born–von Kármán PBC. The scope of PBC is much wider than electronic structure, since they apply to classical simulations, lattice dynamics, and other fields as well. As a general feature, it is postulated that any genuine bulk property of a condensed system cannot depend on the boundary conditions adopted, and therefore in the thermodynamic limit the finite-molecule approach and the PBC approach lead in principle to the same result. There are however very good reasons for preferring the latter to the former, in order both (i) to derive analytical results and (ii) to perform actual calculations. As regards (i), let me mention that PBC are mandatory, at the most elementary level [1, 2], e.g. to define the Fermi surface or to prove the Bloch theorem; as regards (ii), let me just observe that with PBC, even a system of *finite* size has no surface by construction, therefore being a better approximant to its corresponding infinite-size limit.

Choosing any kind of boundary conditions for the wavefunction is tantamount to choosing the Hilbert space where the solutions of Schrödinger's equation are defined. The choice of the Hilbert space, in turn, has very remarkable consequences for some quantum-mechanical observables and for the corresponding operators. Much of the focus of the present work will be on the position (or dipole) operator and higher powers thereof, which indeed has many nontrivial and somewhat counterintuitive features with PBC. The matter has remained rather obscure until very recently. The light came from the development of the modern theory of polarization, dating from 1992 [5–10], and whose ultimate developments are very recent [12, 13, 16, 18, 19].

2.2. Wavefunction and orbitals

In this work we adopt PBC throughout, with the only exception of section 4.2. For the sake of simplicity we assume that we have N three-dimensional electrons, with even N , in a cubic box of size L and volume L^3 . This choice makes the formulation simplest, and is also the most practical one for amorphous and liquid materials. As for crystalline solids, our choice is the most convenient one for cubic solids, while for crystals of lower symmetry other choices are more convenient (see below).

We call $\mathbf{x}_i \equiv (\mathbf{r}_i, \sigma_i)$ the space and spin coordinates of the i th electrons. The many-body ground wavefunction Ψ_0 is an antisymmetric function of the \mathbf{x}_i -variables, and normalizes to one when integrated in the hypercube:

$$\int d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N |\Psi_0(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2 = 1. \quad (1)$$

Here integration over each $d\mathbf{x}_i$ means summing over the spin coordinates and integrating over the periodicity cell of volume L^3 . In the following, this cell will be called the 'supercell'.

Choosing PBC is equivalent to imposing that the wavefunction is periodic, with period L , for translations of the electronic coordinates in each spatial direction and for each electron independently. All the physical quantities are obtained by integration in the hypercube; it is sometimes convenient to imagine that the many-electron system is periodically repeated and that the external potential is periodic as well. Notice that for any finite value of L the whole spectrum is *discrete*.

In the special case of noninteracting electrons, the exact wavefunction is the antisymmetrized product (Slater determinant) of N single-particle spin orbitals. In the simple cases of interest here, each of them is the product of a spatial orbital and a spinor (either up or down): we use for this the notation $\psi_j(\mathbf{r})$ (spin up) and $\bar{\psi}_j(\mathbf{r})$ (spin down). We only consider the so-called restricted case, where up- and down-spin orbitals share the same spatial orbital.

For convenience, and also for the sake of uniformity with other published work [10, 16, 23], we adopt a plane-wave-like normalization for the orbitals:

$$\langle \psi_j | \psi_{j'} \rangle = \frac{1}{L^3} \int d\mathbf{r} \psi_j^*(\mathbf{r}) \psi_{j'}(\mathbf{r}) = \delta_{jj'}, \quad (2)$$

where the integration is performed over the cube of volume L^3 . The normalized wavefunction is therefore

$$|\Psi_0\rangle = \frac{1}{\sqrt{N!L^{3N}}} |\psi_1 \bar{\psi}_1 \psi_2 \bar{\psi}_2 \cdots \psi_{N/2} \bar{\psi}_{N/2}| = A \prod_{j=1}^{N/2} \frac{1}{L^3} \psi_j \bar{\psi}_j, \quad (3)$$

where A is the antisymmetrizer and each orbital is periodic in the supercell.

As stated above, a wavefunction of the form of equation (3) is exact for noninteracting electrons in an external potential. However, such form also defines the approximate wavefunction within a mean-field theory (e.g. Hartree–Fock [26] or Kohn–Sham (KS) [27]; see also section 3.3), where each electron moves in a potential due to the external potential and to the mean-field interaction with all the other electrons.

I have already stressed the value of PBC both in leading to analytical results (letting $L \rightarrow \infty$) and in making actual calculations feasible (keeping L finite): in the latter case, it is enough to choose L reasonably larger than the relevant correlation lengths. Indeed, many breakthrough results in recent years have been obtained by means of Car–Parrinello [28] simulations (also called first-principles molecular dynamics) for liquid and disordered systems. These calculations are of the mean-field KS [27] type, where PBC are imposed on a supercell which typically contains $N \simeq 10^2$ – 10^3 valence electrons.

2.3. Reduced density matrices

The many-body density matrix is by definition $\hat{\rho} = |\Psi_0\rangle\langle\Psi_0|$; the spinless reduced matrices are obtained from $\hat{\rho}$ by taking partial traces over all the spin coordinates, and over $N-1$, $N-2$, ... space coordinates, respectively. Explicitly, the one-body reduced density matrix is

$$\rho^{(1)}(\mathbf{r}_1, \mathbf{r}'_1) = N \sum_{\sigma_1} \int d\mathbf{x}_2 d\mathbf{x}_3 \cdots d\mathbf{x}_N \Psi_0(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \Psi_0^*(\mathbf{x}'_1, \mathbf{x}_2, \dots, \mathbf{x}_N). \quad (4)$$

Its diagonal in coordinate space is just the number density:

$$n(\mathbf{r}) = \rho^{(1)}(\mathbf{r}, \mathbf{r}), \quad (5)$$

and integrates to N over the volume L^3 .

In the special case of a determinantal wavefunction of the form of equation (3), the reduced density matrix has a meaningful form in terms of the single-particle orbitals. It is in fact easy to prove that [25]

$$\rho^{(1)}(\mathbf{r}, \mathbf{r}') = 2 P(\mathbf{r}, \mathbf{r}') = 2 \sum_{j=1}^{N/2} \psi_j(\mathbf{r}) \psi_j^*(\mathbf{r}'). \quad (6)$$

The orbitals in equation (3) enjoy a substantial nonuniqueness: any unitary transformation of the occupied orbitals amongst themselves in fact leaves the Slater determinant unchanged, and hence defines the same wavefunction Ψ_0 . We call these unitary transformations ‘gauge transformations’; what defines the wavefunction uniquely (apart from an overall phase) is the $(N/2)$ -dimensional manifold spanned by the occupied orbitals. Indeed, the operator P in equation (6) is precisely the projector over the occupied manifold, and is manifestly gauge invariant. Since the wavefunction is a function of P , so are all the (higher-order) reduced density matrices and all the physical observables.

2.4. The crystalline case

While PBC can be applied quite generally to any condensed system, the case of a crystalline system deserves additional comments. Suppose the crystal is simple cubic with lattice parameter a ; we then choose the Born–von Kármán period as an integer multiple of the lattice parameter: $L = Ma$, where the thermodynamic limit obtains for $M \rightarrow \infty$. In the case of a low-symmetry crystal it is obviously more convenient to choose a noncubic supercell as a multiple of the elementary crystal cell. Apart from some burdening of notation, the generalization of the present results to low-symmetry crystals is straightforward and will not be considered here. Details on this issue can be found in [10, 29], and SWM.

The lattice vectors are \mathbf{R}_l , where $l \equiv (l_1, l_2, l_3)$ and

$$\mathbf{R}_l = a(l_1, l_2, l_3), \quad l_\alpha = 0, 1, \dots, M - 1. \quad (7)$$

Since the single-particle potential is lattice periodic, a suitable gauge choice leads to orbitals of the Bloch form:

$$\psi_{n\mathbf{q}}(\mathbf{r}) = e^{i\mathbf{q}\cdot\mathbf{r}} u_{n\mathbf{q}}(\mathbf{r}), \quad (8)$$

where \mathbf{q} is the Bloch quasimomentum, and the u s are lattice periodic:

$$u_{n\mathbf{q}}(\mathbf{r} + \mathbf{R}_l) = u_{n\mathbf{q}}(\mathbf{r}). \quad (9)$$

Owing to PBC, there are M^3 allowed Bloch vectors \mathbf{q}_s , which lie on a regular mesh in the unit reciprocal cell, where $s \equiv (s_1, s_2, s_3)$ and

$$\mathbf{q}_s = \frac{2\pi}{Ma} (s_1, s_2, s_3), \quad s_\alpha = 0, 1, \dots, M - 1. \quad (10)$$

The orbitals $\psi_{n\mathbf{q}_s}$ belong then to the Hilbert space of the periodic functions on the *supercell* of size L , while the orbitals $u_{n\mathbf{q}}$ belong to the space of the periodic functions on the *elementary crystal cell* of size a (at any \mathbf{q}). The Bloch orbitals $\psi_{n\mathbf{q}_s}$ at different s are orthogonal by symmetry, where the scalar product is the same as that defined in equation (2).

If the system is insulating with n_b doubly occupied bands, the Slater determinant of equation (3) can be written as

$$|\Psi_0\rangle = A \prod_{n,s} \frac{1}{L^3} \psi_{n\mathbf{q}_s} \bar{\psi}_{n\mathbf{q}_s}, \quad (11)$$

where the product runs over all occupied bands and all mesh points. If, instead, the system is metallic, then the many-body wavefunction Ψ_0 can still be written in the form of equation (11), but where not all the Bloch vectors of a given band are included in the product. We anticipate that this *qualitative* difference has outstanding implications for the main theme of this work.

A gauge transformation allows us to write the insulating wavefunction in equation (11) as a Slater determinant of $N = 2n_b M^3$ localized functions. We define the Wannier function (WF) of the n th band in the central cell as [30, 31]

$$w_n(\mathbf{r}) = \frac{1}{M^3 a^{3/2}} \sum_s \psi_{n\mathbf{q}_s}, \quad (12)$$

and the generic WF as

$$\langle \mathbf{r} | w_{n\mathbf{R}_l} \rangle = w_n(\mathbf{r} - \mathbf{R}_l). \quad (13)$$

Equation (11) is then identically recast as

$$|\Psi_0\rangle = A \prod_{n,l} w_{n\mathbf{R}_l} \bar{w}_{n\mathbf{R}_l}. \quad (14)$$

We stress that the definition of equation (12) still leaves a large gauge freedom, since the phases of the Bloch functions at different s can be chosen arbitrarily.

2.5. The thermodynamic limit

By definition, the thermodynamic limit obtains for $N \rightarrow \infty$, while the density N/L^3 is kept constant. In practical calculations, the limit is approximated using finite- N systems, and choosing L much larger than the atomic dimensions; more precisely, L must be larger than the correlation lengths of interest. It is important to stress, however, that the spectrum of the finite system is *qualitatively* different from the limiting case: while the former is always discrete, the latter is continuous as e.g. for the usual band spectrum of crystalline solids [1, 2]. Also the analytic properties of the quantities of interest are qualitatively different. At finite L , the density matrix $\rho^{(1)}(\mathbf{r}, \mathbf{r}')$ is a periodic function of either \mathbf{r} or \mathbf{r}' , independently. In the thermodynamic limit it goes to zero for $|\mathbf{r} - \mathbf{r}'| \rightarrow \infty$.

Although, as observed above, the many-body wavefunction is an ill-defined mathematical object in the thermodynamic limit, the single-particle orbitals of an independent-electron crystal do indeed go to a well-defined limit. The Bloch orbitals retain the form of equations (8) and (9), where the \mathbf{q} -vector becomes a continuous variable. The single-particle density matrix, equation (6), becomes then a reciprocal-space integral:

$$P(\mathbf{r}, \mathbf{r}') = \frac{1}{(2\pi)^3} \sum_{n=1}^{n_b} \int d\mathbf{q} \psi_{n\mathbf{q}}(\mathbf{r}) \psi_{n\mathbf{q}}^*(\mathbf{r}'). \quad (15)$$

For insulators, the integral is performed over the whole reciprocal cell or, equivalently, over the whole Brillouin zone; for large $|\mathbf{r} - \mathbf{r}'|$ the projector $P(\mathbf{r}, \mathbf{r}')$ then decays exponentially [32–35]. In the metallic case the integral in equation (15) is performed over the \mathbf{q} -vectors within the Fermi surface; as a consequence, the decay for large $|\mathbf{r} - \mathbf{r}'|$ is only polynomial. There are reasons suggesting that such an analytic difference remains in the presence of electron–electron interaction and/or disorder [36].

2.6. Optimally localized Wannier functions

The WFs also go to a well-defined thermodynamic limit, which can be expressed as a reciprocal-cell integral:

$$w_n^{(\infty)}(\mathbf{r}) = \frac{a^{3/2}}{(2\pi)^3} \int d\mathbf{q} \psi_{n\mathbf{q}}(\mathbf{r}), \quad (16)$$

where we have introduced the explicit superscript (∞) in order to distinguish the WFs of equation (16) from those of equation (12). In fact they belong to *completely different* Hilbert spaces: while the WFs of equation (12) obey PBC over the supercell, and are orthonormal when the integration is performed over the volume L^3 , the ones of equation (16) are square integrable and orthonormal over the whole three-dimensional space:

$$\langle w_{n\mathbf{R}_l}^{(\infty)} | w_{n'\mathbf{R}_{l'}}^{(\infty)} \rangle = \int_{\text{all space}} d\mathbf{r} w_n^{(\infty)*}(\mathbf{r} - \mathbf{R}_l) w_{n'}^{(\infty)}(\mathbf{r} - \mathbf{R}_{l'}) = \delta_{nn'} \delta_{ll'}. \quad (17)$$

Such a dramatic change is a direct consequence of the above-mentioned change in the analytic properties of the spectrum.

The moments of the $w^{(\infty)}$ s are well-defined quantities. Let us focus on the spherical second moments, i.e. on the quantity

$$\Omega = \sum_{n=1}^{n_b} \left(\int_{\text{all space}} d\mathbf{r} |\mathbf{r}|^2 |w_n^{(\infty)}(\mathbf{r})|^2 - \left| \int_{\text{all space}} d\mathbf{r} \mathbf{r} |w_n^{(\infty)}(\mathbf{r})|^2 \right|^2 \right), \quad (18)$$

where we are dealing with an insulator with n_b doubly occupied bands. Owing to the nonuniqueness of the WFs stressed above, Ω is a gauge-dependent quantity [31]; there is

then a particular gauge such that Ω assumes its minimum value. This problem was thoroughly studied in 1997 by Marzari and Vanderbilt [37], hereafter quoted as MV, whose key result will be useful in the remainder of the present review.

Besides the choice of the phases in equation (12), an even larger gauge freedom exists, since any unitary transformation of the Bloch orbitals of the n_b occupied bands at the same \mathbf{q} leaves the many-body wavefunction, equation (11), invariant. Following MV, it is understood throughout that we refer to multi-band WFs in this generalized sense. The WFs which actually minimize equation (18) are called ‘optimally localized’; we will assume in the following that such WFs are exponentially decaying for large r . This can be analytically proved in one dimension [32, 35], and is a very plausible conjecture in three dimensions.

MV prove that Ω is no smaller (in three dimensions indeed, *strictly* larger) than a *gauge-invariant* quantity, for which they provide an explicit expression. Using notation which will be useful below, we define the tensor $\langle r_\alpha r_\beta \rangle_c$, where the subscript c stands for ‘cumulant’, as

$$\langle r_\alpha r_\beta \rangle_c = \frac{a^3}{n_b (2\pi)^3} \int d\mathbf{q} \left(\sum_{n=1}^{n_b} \left\langle \frac{\partial}{\partial q_\alpha} u_{n\mathbf{q}} \middle| \frac{\partial}{\partial q_\beta} u_{n\mathbf{q}} \right\rangle - \sum_{n,n'=1}^{n_b} \left\langle u_{n\mathbf{q}} \middle| \frac{\partial}{\partial q_\alpha} u_{n'\mathbf{q}} \right\rangle \left\langle \frac{\partial}{\partial q_\beta} u_{n'\mathbf{q}} \middle| u_{n\mathbf{q}} \right\rangle \right). \quad (19)$$

This tensor is gauge invariant and intensive (e.g. invariant on cell doubling). The MV inequality is then expressed by saying that the average quadratic spherical spread is larger than the trace of $\langle r_\alpha r_\beta \rangle_c$, i.e.,

$$\Omega/n_b > \sum_{\alpha} \langle r_\alpha r_\alpha \rangle_c. \quad (20)$$

It was later shown by Sgiarovello *et al* [23] that equation (19) can be identically recast, using equation (15), as

$$\langle r_\alpha r_\beta \rangle_c = \frac{1}{2n_b} \int_{\text{cell}} d\mathbf{r} \int_{\text{all space}} d\mathbf{r}' (\mathbf{r} - \mathbf{r}')_\alpha (\mathbf{r} - \mathbf{r}')_\beta |P(\mathbf{r}, \mathbf{r}')|^2. \quad (21)$$

This has the simple meaning of a second moment of the (squared) density matrix in the coordinate $\mathbf{r} - \mathbf{r}'$; the gauge invariance of equation (21) is clear.

3. Dipole, phase operator, and macroscopic polarization

3.1. The main problem

For a molecule (or cluster) isolated in space the ground wavefunction is bound. The electronic term in the electric dipole is then obviously defined as the first moment of the single-particle density:

$$\mathbf{d} = e \int d\mathbf{r} \mathbf{r} n(\mathbf{r}), \quad (22)$$

where e is the electron charge (negative throughout this work). On introducing the position operator as

$$\hat{\mathbf{R}} = \sum_{i=1}^N \mathbf{r}_i, \quad (23)$$

equation (22) is equivalently recast in a many-body form:

$$\mathbf{d} = e \langle \Psi_0 | \hat{\mathbf{R}} | \Psi_0 \rangle. \quad (24)$$

Unfortunately, both the above expressions for the dipole are devoid of any meaning when PBC are adopted, as is common practice in condensed matter physics. This drawback has plagued

the theory of macroscopic polarization for a long time; most textbooks contain incorrect statements on the subject. Equally devoid of any meaning are higher moments of the electron distribution.

The drawback is a very serious one, and has to do with the Hilbert space defined by PBC. Any operator, by its very definition, maps a vector belonging to the given Hilbert space into another vector of the same space. But when we apply the multiplicative operator \hat{R} to a function which obeys PBC, say on a cube of side L , we get a function which is no longer periodic, i.e. *does not belong* to the space. Hence the position operator, as defined by equation (23), is a ‘forbidden’ operator with PBC, while in fact is a quite acceptable operator in the Hilbert space of square-integrable functions.

The macroscopic polarization of a condensed system is a crucial concept in any phenomenological description of dielectric media [38]; this quantity is ideally defined as the dipole of a macroscopic sample, divided by its volume. But the simple definition of equation (22) is useless, and a different strategy must be used to attack the polarization problem. Since 1992 [5–8, 10] it has been well established that bulk macroscopic polarization is a physical observable *completely independent* from the charge distribution of the polarized crystalline dielectric. Instead, and perhaps surprisingly to several readers, it has much to do with the *phase* of the wavefunction.

3.2. Many-body phase operator

The key quantity in dealing with the dipole (alias polarization) problem, as well as some problems arising from higher moments, is the many-body phase operator, first introduced in [12]. We start by defining the unitary many-body operator

$$\hat{U}(\mathbf{k}) = \exp(i\mathbf{k} \cdot \hat{\mathbf{R}}), \quad (25)$$

where \mathbf{k} is an arbitrary vector, having the dimensions of an inverse length. Even this operator is local in the Schrödinger representation (i.e. multiplicative), and suffers from the same drawback as \hat{R} with regard to PBC, *except for special values of the vector \mathbf{k}* .

In fact, whenever the components of the vector \mathbf{k} are integer multiples of $2\pi/L$, the operator $\hat{U}(\mathbf{k})$ maps a periodic function (over the supercell) into another periodic function; it is therefore an admissible operator in the PBC Hilbert space. We introduce the notation $\kappa^{(\alpha)}$ for three useful vectors \mathbf{k} in this class, choosing them proportional to the three Cartesian unit vectors:

$$\kappa_{\beta}^{(\alpha)} = \frac{2\pi}{L} \delta_{\alpha\beta}. \quad (26)$$

We also introduce special notation for the three ground-state expectation values:

$$z_N^{(\alpha)} = \langle \Psi_0 | \hat{U}(\kappa^{(\alpha)}) | \Psi_0 \rangle. \quad (27)$$

These are dimensionless complex numbers whose modulus is no larger than one. We show later that, in insulators, the *phase* of $z_N^{(\alpha)}$ is indeed the α -component of macroscopic polarization, measured in suitable units [12, 13, 16]. Explicitly, this phase is defined as

$$z_N^{(\alpha)} = |z_N^{(\alpha)}| e^{i\gamma_N^{(\alpha)}}; \quad (28)$$

$$\gamma_N^{(\alpha)} = \text{Im} \log z_N^{(\alpha)}. \quad (29)$$

For reasons explained elsewhere [16], $\gamma_N^{(\alpha)}$ as defined by equations (28) and (29) is a Berry’s phase [14, 15].

There is an important feature of the many-body phase operator which is worth stressing: unlike \hat{R} , the operator $\hat{U}(\mathbf{k})$ is a *genuine* many-body one. Looking at equation (23), we

notice that \hat{R} is the *sum* of N identical operators, each acting on the coordinates of one electron. In other words, it is a one-body operator written in many-body notation: the many-body expectation value in equation (24) simplifies to (22). Since the wavefunction contains redundant and cumbersome information, this is an enormous simplification—and notice that it applies to correlated- and independent-electron wavefunctions on the same basis. More generally, the expectation value of any one-body operator can be expressed using solely the one-body density matrix $\rho^{(1)}$. Having said this, it is then clear in which sense our operator $\hat{U}(\mathbf{k})$ is a genuinely many-body one. In fact the integrations in equation (27), i.e.

$$z_N^{(\alpha)} = \int d\mathbf{x}_1 d\mathbf{x}_2 \dots d\mathbf{x}_N e^{i\kappa^{(\alpha)} \cdot \sum_{i=1}^N \mathbf{r}_i} |\Psi_0(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2, \quad (30)$$

do not factor and require treatment of the full wavefunction; reduced density matrices are not enough to get the result. The latter statement has one obvious exception: in the special case of a Slater-determinant wavefunction of the form of equation (3), the wavefunction itself is a unique function of $\rho^{(1)}$; hence $z_N^{(\alpha)}$ is also a unique function of $\rho^{(1)}$.

3.3. Implications for the foundations of density-functional theory

As previously stated, $z_N^{(\alpha)}$ is the key quantity which defines the macroscopic polarization of a condensed system, as well as the localization properties of its ground wavefunction: the former issue is discussed in section 3.5 and the latter one in section 5.3.

Since the concept of polarization is intuitively related to the dipole, it may appear surprising that polarization is *not* a one-body property, expressible in terms of the one-body density matrix $\rho^{(1)}$. This fact stems at a very fundamental level from PBC, and can be regarded as an unavoidable trade-off. In a cluster, where equation (22) applies, an important contribution to the dipole comes from the cluster surface; with PBC there is no surface by definition, and the information about the dipole goes instead into the *phase* of the many-body wavefunction. This trade-off also has outstanding implications for the foundations of density-functional theory [27].

By definition, the density of the fictitious noninteracting KS system equals the density of the interacting one; equation (22) then obviously implies that, in a molecule or in a cluster of any size, the dipole of the KS system equals the exact one. Matters change dramatically when PBC are adopted. Since the macroscopic polarization of a condensed system is *not* a function of the density, the KS system and the interacting one have *different* polarizations. This quite important and unexpected feature was first discovered in 1995 by Gonze *et al* [39], and discussed in several subsequent papers [40]. Within the formalism of the present review, a discussion can be found in [13]; numerical evidence on this issue has been given in [41, 42].

Below we provide explicit formulae for dealing with an independent-electron system, as regards macroscopic polarization (sections 3.6 and 3.7) and electron localization (section 6.1). These formulae are exact for a genuinely noninteracting system. When applied within density-functional theory, they provide polarization and localization, respectively, of the noninteracting KS system; both are different, in principle, from the corresponding quantities for the interacting system.

Within any electronic structure calculations, one unavoidably introduces systematic errors due to the physical approximations. Within the first-principles KS framework, usually the *only* source of systematic error is the specific choice made for the exchange–correlation functional [27]: as a general feature, one cannot compute the exact KS orbitals, but only some approximations to them. When such KS orbitals are used in evaluating either polarization or localization, one then introduces a further systematic error due to the above issue. The actual importance of the latter feature—and the relative importance of the two kinds of error—is

currently a matter of debate [41, 42]. We proceed by heuristically assuming that, in weakly correlated systems at least, $z_N^{(\alpha)}$ for the noninteracting KS system is a good approximation to the $z_N^{(\alpha)}$ for the interacting system. This assumption is supported by the success of many first-principles KS calculations of the macroscopic polarization, in several different materials [17].

3.4. Hamiltonian modified with a flux

Our N -body stationary states Ψ_n are eigenstates of the Hamiltonian

$$\hat{H} = \frac{1}{2m} \sum_{i=1}^N |\mathbf{p}_i|^2 + \hat{V}, \quad (31)$$

with eigenvalues E_n . In equation (31) m is the electron mass and the potential \hat{V} includes a one-body term (the external potential) and a two-body one (the electron–electron interaction). Following Kohn [3, 4], it is expedient to define a more general class of Hamiltonians:

$$\hat{H}(\mathbf{k}) = \frac{1}{2m} \sum_{i=1}^N (\mathbf{p}_i - \hbar \mathbf{k})^2 + \hat{V}, \quad (32)$$

where \mathbf{k} is arbitrary. Obviously, $\hat{H}(0)$ coincides with our original Hamiltonian \hat{H} , equation (31). It is understood that, for any \mathbf{k} , the eigenstates $\Psi_n(\mathbf{k})$ of equation (32) are found by imposing the *ordinary* PBC on the supercell. Owing to the PBC, the eigenvalues depend on \mathbf{k} ; they will be called $E_n(\mathbf{k})$ in the following. The modified Hamiltonian, equation (32), is generally called the ‘Hamiltonian with a flux’ [43]. This is because its one-dimensional analogue describes the motion of an electron on a ring threaded by an Aharonov–Bohm magnetic flux [44, 45].

A perturbative expansion of the ground eigenstate of $\hat{H}(\mathbf{k})$ yields, to leading order in \mathbf{k} ,

$$|\Psi_0(\mathbf{k})\rangle \simeq |\Psi_0\rangle + \frac{\hbar}{m} \mathbf{k} \cdot \sum_{n \neq 0}' |\Psi_n\rangle \frac{\langle \Psi_n | \hat{\mathbf{P}} | \Psi_0 \rangle}{E_n - E_0}, \quad (33)$$

where $\hat{\mathbf{P}}$ is the momentum operator

$$\hat{\mathbf{P}} = \sum_{i=1}^N \mathbf{p}_i = -i\hbar \sum_{i=1}^N \nabla_i. \quad (34)$$

We notice that $\hat{\mathbf{P}}$ is an admissible operator for PBC; it is a one-body and nonlocal operator. We now consider the states

$$|\Phi_n(\mathbf{k})\rangle = \hat{U}(\mathbf{k}) |\Psi_n\rangle, \quad (35)$$

where \hat{U} is the many-body phase operator. It is easy to prove that they satisfy

$$\hat{H}(\mathbf{k}) |\Phi_n(\mathbf{k})\rangle = E_n |\Phi_n(\mathbf{k})\rangle. \quad (36)$$

Despite this identity, the $\Phi_n(\mathbf{k})$ are in general *not* eigenstates of $\hat{H}(\mathbf{k})$, since they do not obey PBC. In fact, the $\Phi_n(\mathbf{k})$ wavefunctions are *quasiperiodic* over the supercell; this fact is also expressed by saying that they obey ‘twisted’ boundary conditions [19, 43]. There exist, however, special vectors \mathbf{k} such that the states $\Phi_n(\mathbf{k})$ obey PBC and are indeed genuine eigenstates of $\hat{H}(\mathbf{k})$ with eigenvalues E_n ; for instance, this is the case when \mathbf{k} equals one of the $\boldsymbol{\kappa}^{(\alpha)}$ vectors, equation (26).

We focus now on $\Psi_0(\boldsymbol{\kappa}^{(\alpha)})$, the ground eigenstate of $\hat{H}(\boldsymbol{\kappa}^{(\alpha)})$; what was said above guarantees that it coincides with either $\Phi_0(\boldsymbol{\kappa}^{(\alpha)})$, or with $\Phi_n(\boldsymbol{\kappa}^{(\alpha)})$ where $n \neq 0$. We address here only the case of a ‘nonexotic’ insulator, having a finite gap in the excitation spectrum;

more about the issue will appear below, section 4.4. We therefore assume for the time being that the ground state Ψ_0 is nondegenerate, and that the lowest excitation energy goes to a finite value ε_g in the thermodynamic limit. It is easy to prove that, for large enough L , one has $|E_0(\kappa^{(\alpha)}) - E_0| < \varepsilon_g$; ergo $\Psi_0(\kappa^{(\alpha)})$ coincides with $\Phi_0(\kappa^{(\alpha)})$, apart for an overall phase.

The above findings, together with equations (28) and (33), yield then

$$\hat{U}(\kappa^{(\alpha)})|\Psi_0\rangle = |\Phi_0(\kappa^{(\alpha)})\rangle \simeq e^{i\gamma_N^{(\alpha)}} \left(|\Psi_0\rangle + \frac{2\pi\hbar}{mL} \sum'_{n \neq 0} |\Psi_n\rangle \frac{\langle \Psi_n | \hat{P}_\alpha | \Psi_0 \rangle}{E_n - E_0} \right) \quad (37)$$

to leading order in $\kappa^{(\alpha)} \propto L^{-1}$.

3.5. Macroscopic polarization

For centrosymmetric materials the expectation values $z_N^{(\alpha)}$ are real (provided that the origin is chosen at a centrosymmetric site), while in noncentrosymmetric materials they are in general complex: their phases $\gamma_N^{(\alpha)}$ define then the Cartesian components of the macroscopic polarization in suitable units. As will be discussed below, in the metallic case the $z_N^{(\alpha)}$ vanish (fast enough with N) and the polarization is ill defined. For insulators, the expectation values of the many-body phase operator do not vanish and the macroscopic polarization is well defined.

When discussing macroscopic polarization, it should be kept in mind that the ‘absolute’ polarization of a given sample has never been measured as a bulk property: one invariably measures derivatives (dielectric constants, piezoelectric coefficients, dynamical charges. . .) or finite differences (polarization reversal in ferroelectrics) [10, 11, 16]. Furthermore, the quantity actually measured is the macroscopic current j which traverses the sample while the given perturbation is adiabatically switched on. One is therefore interested in the quantity

$$\Delta \mathbf{P} = \int_0^{\Delta t} dt j(t); \quad (38)$$

in the adiabatic limit, Δt goes to infinity and $j(t)$ goes to zero, while $\Delta \mathbf{P}$ depends on the initial and final states. There is no need to define what the absolute polarization of the initial and final states is. Let me however mention that there are indeed ways of defining absolute polarization [7], but one pays the price of getting a rather unusual vector property, which we are not going to discuss here.

Macroscopic polarization is generally understood as a *static* property. Within PBC it is instead more convenient to address only polarization *differences*, and to regard them as *dynamical properties in the adiabatic limit*, as in equation (38). Indeed, this is a very general feature of most observables having the nature of a Berry’s phase [14–16].

We suppose that the Hamiltonian, equation (31), includes a slowly varying time-dependent term. We call Ψ_n the adiabatic instantaneous eigenstates at time t , and $\tilde{\Psi}$ the exact time-evolved ground state. To leading order in the adiabaticity parameter we have

$$|\tilde{\Psi}\rangle \langle \tilde{\Psi}| \simeq |\Psi_0\rangle \langle \Psi_0| + i\hbar \sum'_{n \neq 0} \left(|\Psi_n\rangle \frac{\langle \Psi_n | \dot{\Psi}_0 \rangle}{E_n - E_0} \langle \Psi_0| - \text{H.c.} \right), \quad (39)$$

where H.c. indicates the Hermitian conjugate. Equation (39) is generally attributed to Niu and Thouless [46]. The current flowing through the system is then

$$j = \frac{e}{mL^3} \langle \tilde{\Psi} | \hat{P} | \tilde{\Psi} \rangle \simeq -\frac{ie\hbar}{mL^3} \sum'_{n \neq 0} \frac{\langle \dot{\Psi}_0 | \Psi_n \rangle \langle \Psi_n | \hat{P} | \Psi_0 \rangle}{E_n - E_0} + \text{c.c.}, \quad (40)$$

where c.c. indicates the complex conjugate.

Following [12] we are going to show that the macroscopic polarization is

$$P_\alpha = \frac{e\gamma_N^{(\alpha)}}{2\pi L^2} = \frac{e}{2\pi L^2} \text{Im} \log z_N^{(\alpha)}. \quad (41)$$

To this end, it is enough to evaluate the Berry phase $\gamma_N^{(\alpha)}$ using the adiabatic instantaneous eigenstates, and to take its time derivative. From equations (27) and (29) we get

$$\dot{\gamma}_N^{(\alpha)} = \text{Im} \left(\frac{\langle \dot{\Psi}_0 | \hat{U}(\boldsymbol{\kappa}^{(\alpha)}) | \Psi_0 \rangle}{\langle \Psi_0 | \hat{U}(\boldsymbol{\kappa}^{(\alpha)}) | \Psi_0 \rangle} + \frac{\langle \Psi_0 | \hat{U}(\boldsymbol{\kappa}^{(\alpha)}) | \dot{\Psi}_0 \rangle}{\langle \Psi_0 | \hat{U}(\boldsymbol{\kappa}^{(\alpha)}) | \Psi_0 \rangle} \right) = -i \frac{\langle \dot{\Psi}_0 | \hat{U}(\boldsymbol{\kappa}^{(\alpha)}) | \Psi_0 \rangle}{\langle \Psi_0 | \hat{U}(\boldsymbol{\kappa}^{(\alpha)}) | \Psi_0 \rangle} + \text{c.c.} \quad (42)$$

Using then equation (37) we have, to leading order in L^{-1} ,

$$\dot{\gamma}_N^{(\alpha)} \simeq -i \frac{2\pi\hbar}{mL} \sum_{n \neq 0} \frac{\langle \dot{\Psi}_0 | \Psi_n \rangle \langle \Psi_n | \hat{P} | \Psi_0 \rangle}{E_n - E_0} + \text{c.c.} \quad (43)$$

Comparison with equations (38) and (40) immediately leads to the proof of equation (41).

3.6. Independent electrons

In the special case of a Slater-determinant wavefunction, the ground-state expectation value $z_N^{(\alpha)}$ is easily expressed in terms of the occupied spin orbitals ψ_j . Rewriting equation (27) as

$$z_N^{(\alpha)} = \langle \Psi_0 | \Phi_0(\boldsymbol{\kappa}^{(\alpha)}) \rangle, \quad (44)$$

we notice that whenever the wavefunction Ψ_0 is a Slater determinant, even $\Phi_0(\boldsymbol{\kappa}^{(\alpha)})$ is a Slater determinant, whose spin orbitals ϕ s are obtained from the corresponding ψ s of equation (3) as

$$\phi_j(\mathbf{r}) = e^{i\boldsymbol{\kappa}^{(\alpha)} \cdot \mathbf{r}} \psi_j(\mathbf{r}), \quad (45)$$

and analogously for the $\bar{\phi}$ s (spin-down) ones. According to a well-known theorem, the overlap between two Slater determinants is equal to the determinant of the $N \times N$ overlap matrix built out of the occupied spin orbitals. Since the overlaps between different-spin spin orbitals vanish, and those between equal-spin ones are identical in pairs, we can factor the determinant as

$$z_N^{(\alpha)} = (\det S^{(\alpha)})^2, \quad (46)$$

where $S^{(\alpha)}$ is the overlap matrix between the occupied spatial orbitals, having size $N/2$. Its elements are

$$S_{jj'}^{(\alpha)} = \int d\mathbf{r} \psi_j^*(\mathbf{r}) e^{i\boldsymbol{\kappa}^{(\alpha)} \cdot \mathbf{r}} \psi_{j'}(\mathbf{r}). \quad (47)$$

The expressions of equations (46) and (47) as they stand are of great practical value, particularly for disordered systems. In fact they have been extensively used [47] in evaluating macroscopic polarization and infrared spectra of liquid and disordered systems by means of Car-Parrinello simulations [28].

3.7. King-Smith and Vanderbilt's formula

For crystalline systems, we may exploit the translational symmetry by choosing the occupied orbitals in the Bloch form; such a choice leads to a powerful factoring of the determinant of the overlap matrix $S^{(\alpha)}$. When we identify the orbital index j with the band index n and the Bloch index \mathbf{q}_s together, as in equation (11), the $S^{(\alpha)}$ elements become

$$S_{n\mathbf{q}_s, n'\mathbf{q}_{s'}}^{(\alpha)} = \frac{1}{L^3} \int d\mathbf{r} \psi_{n\mathbf{q}_s}^*(\mathbf{r}) \phi_{n'\mathbf{q}_{s'}}(\mathbf{r}) = \frac{1}{L^3} \int d\mathbf{r} u_{n\mathbf{q}_s}^*(\mathbf{r}) u_{n'\mathbf{q}_{s'}}(\mathbf{r}) e^{i(\boldsymbol{\kappa}^{(\alpha)} + \mathbf{q}_{s'} - \mathbf{q}_s) \cdot \mathbf{r}}. \quad (48)$$

The matrix $S^{(\alpha)}$ is therefore very sparse, since the supercell integrals in equation (48) vanish unless $\kappa^{(\alpha)} + \mathbf{q}_{s'} - \mathbf{q}_s = 0$. We define the small overlap matrix $S(\mathbf{q}, \mathbf{q}')$ as

$$S_{nn'}(\mathbf{q}, \mathbf{q}') = \langle u_{n\mathbf{q}} | u_{n'\mathbf{q}'} \rangle = \frac{1}{a^3} \int_{\text{cell}} d\mathbf{r} u_{n\mathbf{q}}^*(\mathbf{r}) u_{n'\mathbf{q}'}(\mathbf{r}), \quad (49)$$

where now the integral is performed over the crystalline cell of volume a^3 . The small overlap matrix $S(\mathbf{q}, \mathbf{q}')$, first introduced by Blount [31], allows a compact formulation of the modern theory of polarization [10].

For an insulating crystal with n_b doubly occupied bands the determinant of the large overlap matrix $S^{(\alpha)}$, of size $N/2 = n_b M^3$, factors into the product of M^3 small determinants of size n_b . From equation (46) we have then

$$z_N^{(\alpha)1/2} = \det S^{(\alpha)} = \prod_s \det S(\mathbf{q}_s + \kappa^{(\alpha)}, \mathbf{q}_s), \quad (50)$$

and the macroscopic polarization, equation (41), becomes

$$P_\alpha = -\frac{e}{\pi L^2} \text{Im} \log \prod_s \det S(\mathbf{q}_s, \mathbf{q}_s + \kappa^{(\alpha)}). \quad (51)$$

It is straightforward to prove [10, 16] that the thermodynamic limit of equation (51) is the reciprocal-cell integral:

$$P_\alpha = \frac{2ie}{(2\pi)^3} \sum_{n=1}^{n_b} \int d\mathbf{q} \left\langle u_{n\mathbf{q}} \left| \frac{\partial}{\partial q_\alpha} u_{n\mathbf{q}} \right. \right\rangle. \quad (52)$$

This is indeed the—by now famous—Berry's phase formula for the macroscopic polarization of a crystalline system of independent electrons, derived by King-Smith and Vanderbilt in 1993 [6].

4. Linear response and conductivity

4.1. Kubo formulae and sum rules

So far, we have only discussed ground-state properties of our N -electron system. Suppose now that it is subject to a small time-dependent perturbation contributing to the Hamiltonian the term

$$\delta \hat{H}(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega f(\omega) \frac{1}{2} (\hat{A} e^{-i\omega t} + \hat{A}^\dagger e^{i\omega t}), \quad (53)$$

where \hat{A} determines the 'shape' of the perturbation and f its amplitude. In order to get a Hermitian $\delta \hat{H}$, we assume $f(\omega) = f(-\omega)$. We wish to measure the response to such perturbation by means of the expectation value of some observable \hat{B} , i.e.

$$\delta \langle \hat{B} \rangle = \langle \tilde{\Psi} | \hat{B} | \tilde{\Psi} \rangle - \langle \Psi_0 | \hat{B} | \Psi_0 \rangle, \quad (54)$$

where $\tilde{\Psi} = \Psi_0 + \delta \Psi(t)$ is the perturbed time-evolved ground state. If we limit ourselves to studying terms which are linear in the response, it is enough to consider the single oscillatory perturbation

$$\hat{H}'(\omega) = \frac{1}{2} (\hat{A} e^{-i\omega t} + \hat{A}^\dagger e^{i\omega t}), \quad (55)$$

whose response can be written, using the compact notation due to Zubarev [25, 48, 49], as

$$\delta \langle \hat{B} \rangle = \frac{1}{2} (\langle \langle \hat{B} | \hat{A} \rangle \rangle_\omega e^{-i\omega t} + \langle \langle \hat{B} | \hat{A} \rangle \rangle_{-\omega} e^{i\omega t}). \quad (56)$$

The quantity $\langle\langle\hat{B}|\hat{A}\rangle\rangle_\omega$ is by definition the linear response induced by the perturbation \hat{A} at frequency ω on the expectation value $\langle\hat{B}\rangle$. Straightforward first-order perturbation theory provides its explicit expression:

$$\langle\langle\hat{B}|\hat{A}\rangle\rangle_\omega = \frac{1}{\hbar} \lim_{\eta \rightarrow 0^+} \sum'_{n \neq 0} \left(\frac{\langle\Psi_0|\hat{B}|\Psi_n\rangle\langle\Psi_n|\hat{A}|\Psi_0\rangle}{\omega - \omega_{0n} + i\eta} - \frac{\langle\Psi_0|\hat{A}|\Psi_n\rangle\langle\Psi_n|\hat{B}|\Psi_0\rangle}{\omega + \omega_{0n} + i\eta} \right), \quad (57)$$

where $\omega_{0n} = (E_n - E_0)/\hbar$ are the excitation frequencies of the unperturbed system, and the positive infinitesimal η ensures causality. Expressions of the kind of equation (57) go under the name of Kubo formulae.

Denoting by \mathcal{P} the principal part of an integral, one has the identity

$$\lim_{\eta \rightarrow 0^+} \frac{1}{\omega + i\eta} = \mathcal{P} \frac{1}{\omega} - i\pi \delta(\omega). \quad (58)$$

The imaginary part of the response has therefore the spectral decomposition

$$\begin{aligned} \text{Im} \langle\langle\hat{B}|\hat{A}\rangle\rangle_\omega &= -\frac{\pi}{\hbar} \sum'_{n \neq 0} [\langle\Psi_0|\hat{B}|\Psi_n\rangle\langle\Psi_n|\hat{A}|\Psi_0\rangle\delta(\omega - \omega_{0n}) \\ &\quad - \langle\Psi_0|\hat{A}|\Psi_n\rangle\langle\Psi_n|\hat{B}|\Psi_0\rangle\delta(\omega + \omega_{0n})]. \end{aligned} \quad (59)$$

Integrating over the positive frequencies, and exploiting the completeness of the Hilbert space, we get the sum rule

$$\int_0^\infty d\omega \text{Im} \langle\langle\hat{B}|\hat{A}\rangle\rangle_\omega = -\frac{\pi}{\hbar} (\langle\Psi_0|\hat{B}\hat{A}|\Psi_0\rangle - \langle\Psi_0|\hat{B}|\Psi_0\rangle\langle\Psi_0|\hat{A}|\Psi_0\rangle). \quad (60)$$

This exact relationship looks like the zero-temperature limit of a fluctuation-dissipation theorem, several forms of which have been known in statistical physics [50, 51] from about one century ago [52]: equation (60) equates in fact the imaginary part of the response—and hence dissipation—to some equilibrium expectation values. In the special case of $\hat{B} = \hat{A}$ the right-hand side of equation (60) is precisely the quantum fluctuation of the observable \hat{A} in the ground state. Finally, we stress that the sum rule holds for any pair of many-body operators \hat{A} and \hat{B} , *provided that* both are admissible operators in the Hilbert space defined by the given boundary conditions.

4.2. Dipole response and fluctuations

We now outline the case where the operators \hat{A} and \hat{B} are both identified with the position operator \hat{R} , equation (23). As previously stressed in section 3.1, this is *not* an admissible operator for PBC. The forthcoming discussion, therefore, is limited to the case of a molecule (or cluster), where the ground wavefunction is bound; it is customary to refer to this case as that of ‘open’ boundary conditions, in contrast to PBC.

An oscillating longitudinal electric field $\mathbf{E} \cos \omega t$ couples to the N -electron system, in the dipole gauge, via the position operator

$$\hat{H}'(\omega) = -e \cos(\omega t) \mathbf{E} \cdot \hat{\mathbf{R}}. \quad (61)$$

The induced dipole, owing to equation (24), is by definition the expectation value of the same operator. The dipolar polarizability tensor $\alpha_{\beta\gamma}$ at frequency ω is then

$$\alpha_{\beta\gamma}(\omega) = -e^2 \langle\langle\hat{R}_\beta|\hat{R}_\gamma\rangle\rangle_\omega, \quad (62)$$

and the sum rule of equation (60) takes the form

$$\int_0^\infty d\omega \text{Im} \alpha_{\beta\gamma}(\omega) = \frac{\pi e^2}{\hbar} (\langle\Psi_0|\hat{R}_\beta\hat{R}_\gamma|\Psi_0\rangle - \langle\Psi_0|\hat{R}_\beta|\Psi_0\rangle\langle\Psi_0|\hat{R}_\gamma|\Psi_0\rangle). \quad (63)$$

For any *finite* system with open boundary conditions, the two members of equation (63) are well defined and finite.

One may then argue that, for a metallic cluster in the large-size limit, the dipole polarizability diverges and the second cumulant moment in the right-hand side diverges as well; this divergence was indeed proposed long ago in the solid-state literature to identify the onset of conduction and the metal–insulator transition [53, 54]. However, caution is in order when addressing the thermodynamic limit of expressions like equation (63), based on unbound operators. The lesson drawn in recent years from the modern theory of polarization, and outlined in section 3, is relevant to the present issue as well. In short, the matrix elements in equations (57) and (63) may well happen to be dominated by the electron distribution *at the surface* of the cluster. Some further discussion, particularly of Kudinov’s work [54], is provided by SWM. Throughout this review, we safely proceed using only PBC where, among other features, no surface exists.

4.3. Current response

Unlike the dipole gauge, the current gauge *is* compatible with PBC. In fact within this gauge, the coupling of the system to an oscillating macroscopic field is, to first order in \mathbf{E} ,

$$\hat{H}'(\omega) = \sin(\omega t) \frac{e}{m\omega} \mathbf{E} \cdot \hat{\mathbf{P}} = \frac{ie}{2m\omega} (e^{-i\omega t} - e^{i\omega t}) \mathbf{E} \cdot \hat{\mathbf{P}}, \quad (64)$$

where $\hat{\mathbf{P}}$ is the momentum operator, equation (34).

Since the induced macroscopic current \mathbf{j} equals $e\delta\langle\hat{\mathbf{P}}\rangle/(mL^3)$, the definition of equation (56) leads to a linear response:

$$\mathbf{j} = \frac{ie^2}{2m\omega L^3} e^{-i\omega t} \sum_{\beta} \langle\langle\hat{\mathbf{P}}_{\alpha}|\hat{\mathbf{P}}_{\beta}\rangle\rangle_{\omega} E_{\beta} + \text{c.c.} \quad (65)$$

The frequency-dependent conductivity tensor is therefore

$$\sigma_{\alpha\beta}(\omega) = \frac{ie^2}{m^2\omega L^3} \langle\langle\hat{\mathbf{P}}_{\alpha}|\hat{\mathbf{P}}_{\beta}\rangle\rangle_{\omega}, \quad (66)$$

and the Kubo formula, equations (57) and (58), for the diagonal elements yields

$$\begin{aligned} \text{Re } \sigma_{\alpha\alpha}(\omega) &= \frac{\pi e^2}{\hbar m^2 L^3} \sum_n \frac{1}{\omega_{0n}} |\langle\Psi_0|\hat{\mathbf{P}}_{\alpha}|\Psi_n\rangle|^2 \delta(\omega - \omega_{0n}), & \omega > 0; \\ \text{Im } \sigma_{\alpha\alpha}(\omega) &= \frac{2e^2}{\hbar m^2 L^3} \sum_n |\langle\Psi_0|\hat{\mathbf{P}}_{\alpha}|\Psi_n\rangle|^2 \mathcal{P} \frac{1}{\omega^2 - \omega_{0n}^2}. \end{aligned} \quad (67)$$

Notice that $n = 0$ can be formally included in the sum, since the ground-state expectation value $\langle\Psi_0|\hat{\mathbf{P}}|\Psi_0\rangle$ vanishes. The conductivity given in equation (67) obeys the Kramers–Kronig relations in the form [49, 50]

$$\begin{aligned} \text{Re } \sigma_{\alpha\alpha}(\omega) &= \frac{2}{\pi} \mathcal{P} \int_0^{\infty} d\omega' \frac{\omega'}{\omega'^2 - \omega^2} \text{Im } \sigma_{\alpha\alpha}(\omega'); \\ \text{Im } \sigma_{\alpha\alpha}(\omega) &= -\frac{2}{\pi} \mathcal{P} \int_0^{\infty} d\omega' \frac{\omega'}{\omega'^2 - \omega^2} \text{Re } \sigma_{\alpha\alpha}(\omega'). \end{aligned} \quad (68)$$

4.4. Metals and insulators

The main quantity which discriminates between a conducting and nonconducting material at zero temperature is the low-frequency limit of the conductivity. As long as L remains finite,

the spectrum is discrete and the real part of the conductivity is either vanishing or diverging as a function of ω . Since the spectrum becomes dense in the thermodynamic limit, for any given ω the sum over the states in equation (67) converges to an integral, and the real part of the conductivity may eventually become a regular function of ω . I stress that, in general, the thermodynamic limit ($L \rightarrow \infty$) and the dc limit ($\omega \rightarrow 0$) *cannot* be interchanged.

Suppose that instead of being interested in the dc conductivity, we focus on an expression where some frequency integral of $\sigma_{\alpha\beta}(\omega)$ appears, such as the dispersion relations equation (68), or equation (89) below. Such integrated expressions hold at any finite L and also remain valid in the $L \rightarrow \infty$ limit: therefore the ω -integration and the thermodynamic limit *can* be interchanged.

At this point, it is worth mentioning that another method for regularizing the sum in equation (67) consists [43] in assuming the existence of a weak coupling between our N -electron system and the rest of the Universe, described phenomenologically by a broadening of each level by an amount Γ , much larger than the mean level spacing at the given L . Eventually, the limit $\Gamma \rightarrow 0$ is taken at fixed ω . Once more the two limits $\Gamma \rightarrow 0$ and $\omega \rightarrow 0$ limit do not commute, while instead the $\Gamma \rightarrow 0$ limit commutes with ω -integrations.

In a simple insulator—with a finite gap ε_g in the excitation spectrum—the real part of the conductivity strictly vanishes for $\omega < \varepsilon_g$. In other classes of more exotic insulators, the energy of the lowest-lying electronic excitation might converge to zero in the thermodynamic limit. However, in this case the momentum elements in equation (67) either vanish or go to zero fast enough with decreasing excitation energies to ensure the vanishing of the dc conductivity characterizing the insulating state of matter at zero temperature.

In a metal the dc conductivity does not vanish. More precisely, following the analysis of SWM, two cases must be considered:

- (i) ideal conductors where the electrons undergo free acceleration in a dc field; and
- (ii) nonideal conductors, such as disordered metal alloys, where scattering damps free acceleration.

In both cases, the real part of the conductivity includes a regular part which goes to a nonzero limit $\text{Re } \sigma_{\alpha\alpha}(0)$ for $\omega \rightarrow 0$. Additionally, in case (i) there is a singular term proportional to $\delta(\omega)$, accounting for the free-acceleration behaviour of the system electrons.

5. Electron localization

5.1. Fundamentals: one electron in one dimension

Let us start by considering the elementary problem of a one-dimensional potential well of unspecified shape. The Hamiltonian eigenfunctions $\psi(x)$ can be of two different kinds: either bound (localized, square integrable, exponentially decaying) or unbound (delocalized, non-normalizable, scattering states). This distinction is a sharp qualitative one, *provided that* we adopt the usual boundary conditions for $\psi(x)$ over $(-\infty, \infty)$; these are usually called ‘fixed’ or ‘open’ boundary conditions. A quantity of interest is the quadratic spread, or second cumulant moment, defined as $\langle x^2 \rangle_c = \langle \psi | x^2 | \psi \rangle - \langle \psi | x | \psi \rangle^2$; this quantity is finite for a bound state and diverges for an unbound one, thus providing a sharp localization criterion.

We focus in the following on a localized state, for instance the ground state of a given attractive well, and we call it $\psi^{(\infty)}(x)$ for reasons which will become clear in a moment. The corresponding density will be called $n_{\text{loc}}(x)$, for consistency with other published work [13, 16, 18]; it is furthermore expedient to choose the origin at the centre of $n_{\text{loc}}(x)$, in such a way that

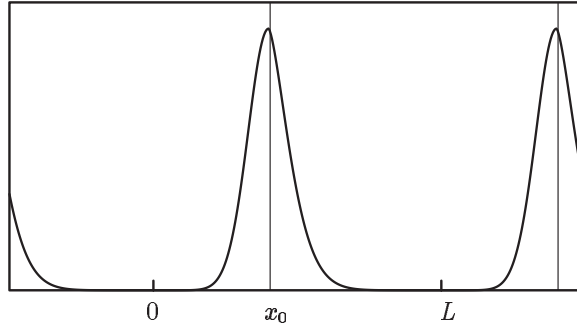


Figure 3. The distribution $|\psi(x)|^2$ of a single-particle orbital with periodic Born–von Kármán boundary conditions (PBC).

$$\int_{-\infty}^{\infty} dx x n_{\text{loc}}(x) = \int_{-\infty}^{\infty} dx x |\psi^{(\infty)}(x)|^2 = 0. \quad (69)$$

Next, we address the same attractive potential well using PBC over a segment of length L . In this case, apparently, there is no longer any sharp distinction: all eigenstates $\psi(x)$ are periodic, and so all look delocalized-like. Then we may ask: what is the relationship between $\psi(x)$ and $\psi^{(\infty)}(x)$? Clearly, with a suitable choice of the origin, the large- L limit of $\psi(x)$ must coincide with $\psi^{(\infty)}(x)$. A precise formulation of this statement is

$$\psi(x) \doteq \sum_{m=-\infty}^{\infty} \psi^{(\infty)}(x - x_0 - mL), \quad (70)$$

where the symbol \doteq , borrowed from Kohn [3, 4], means equality apart from terms which vanish exponentially with L . Notice that equation (70) can be spelled out by saying that the wavefunction ψ breaks up into a sum of functions which are localized in *disconnected* regions; the many-body generalization of a similar concept is central to Kohn’s theory of localization, section 5.2. Notice also that such disconnectedness results from the localized character of $\psi^{(\infty)}$, and does *not* apply to the scattering states of the given potential well. From equation (70) it also follows that

$$|\psi(x)|^2 \doteq \sum_{m=-\infty}^{\infty} n_{\text{loc}}(x - x_0 - mL). \quad (71)$$

Such a density is sketched in figure 3, where we have chosen L much larger than the exponential decay length of $\psi^{(\infty)}(x)$; this is the crucial feature ensuring that the two members in equations (70) and (71) are equal, to a good approximation.

We now wish to evaluate the quadratic spread of the charge distribution of the localized state, i.e.

$$\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 = \int_{-\infty}^{\infty} dx x^2 n_{\text{loc}}(x) = - \left. \frac{d^2 \tilde{n}_{\text{loc}}(k)}{dk^2} \right|_{k=0}. \quad (72)$$

This is a problem if we are working with PBC, and we therefore have access to $\psi(x)$ but *not* to $\psi^{(\infty)}(x)$: since the position operator x , as thoroughly discussed in section 3.1, is a forbidden operator for PBC, we cannot evaluate any moment in the usual way. Of course, $n_{\text{loc}}(x)$ *does* have well-defined moments, but inversion of equation (71) is not a good solution since it is highly nonunique. The way to cope with this problem has been shown by RS: following them, we define

$$z = \int_0^L dx e^{i(2\pi/L)x} |\psi(x)|^2. \quad (73)$$

The dimensionless complex number z , whose modulus is no larger than one, is just a Fourier coefficient of the periodic distribution $|\psi(x)|^2$; it can also be regarded as the single-particle one-dimensional limit of $z_N^{(\alpha)}$, equation (27). Using equation (71), this Fourier coefficient is expressed by means of the Fourier transform of n_{loc} as

$$z \doteq e^{i\frac{2\pi}{L}x_0} \tilde{n}_{\text{loc}}\left(-\frac{2\pi}{L}\right). \quad (74)$$

Owing to equations (69) and (72), the leading expansion of \tilde{n}_{loc} reads

$$\tilde{n}_{\text{loc}}(k) \simeq 1 - \frac{1}{2} \langle x^2 \rangle_c k^2. \quad (75)$$

Identifying then k with $-2\pi/L$ and using equation (74), we easily obtain

$$\langle x^2 \rangle_c \simeq 2 \left(\frac{L}{2\pi} \right)^2 (1 - |z|), \quad (76)$$

$$\langle x^2 \rangle_c \simeq - \left(\frac{L}{2\pi} \right)^2 \log |z|^2. \quad (77)$$

For a localized state, equations (76) and (77) go to the same finite limit for large L , but the latter expression is preferable for dealing with delocalized states as well. In fact for a scattering state one has $|\psi(x)|^2 \simeq 1/L$ throughout the periodicity cell, and hence a vanishing $|z|$ and an infinite quadratic spread.

In conclusion, equation (77) solves the problem of measuring the quadratic spread of a charge distribution with PBC, such as the (localized) one of figure 3. Furthermore, the same expression allows one, in the large- L limit, to sharply discriminate between localized and delocalized states: in the latter case, in fact, the quadratic spread defined by equation (77) diverges.

5.2. Localization according to Kohn

The milestone 1964 Kohn paper [3, 4], bearing the title ‘Theory of the insulating state’, vindicates within quantum mechanics the classical idea that electron localization is the reason that the electronic charge in insulators cannot flow freely under an applied dc field. The problem is: how does one *define* in a sharp manner localization/delocalization in the ground state of insulators/metals? Kohn’s solution is that the essential characteristic of insulators is *disconnectedness* of the ground wavefunction.

Let us consider our usual wavefunction Ψ_0 of the N -electron system with PBC over a cubic supercell of volume L^3 . We then define the three-dimensional vectors \mathbf{M} , whose components are integer multiples of L . We can then write the PBC as

$$\Psi_0(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i + \mathbf{M}, \dots, \mathbf{x}_N) = \Psi_0(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N), \quad (78)$$

where by $\mathbf{x}_i + \mathbf{M}$ we mean that the *spatial component* r_i of \mathbf{x}_i is translated by \mathbf{M} .

According to Kohn, the periodic wavefunction Ψ_0 is disconnected if it breaks down into the *sum* of terms $\Psi_{0,M}$ which are localized in nonoverlapping regions of the $3N$ -dimensional hyperspace:

$$|\Psi_0\rangle \doteq \sum_M |\Psi_{0,M}\rangle. \quad (79)$$

Here the sum runs over the infinite set of the \mathbf{M} -vectors, and all functions are defined in the whole hyperspace. Kohn’s criterion of localization in nonoverlapping regions is that for $\mathbf{M} \neq \mathbf{M}'$ the product function $\Psi_{0,M} \Psi_{0,M'}$ vanishes exponentially in the thermodynamic limit.

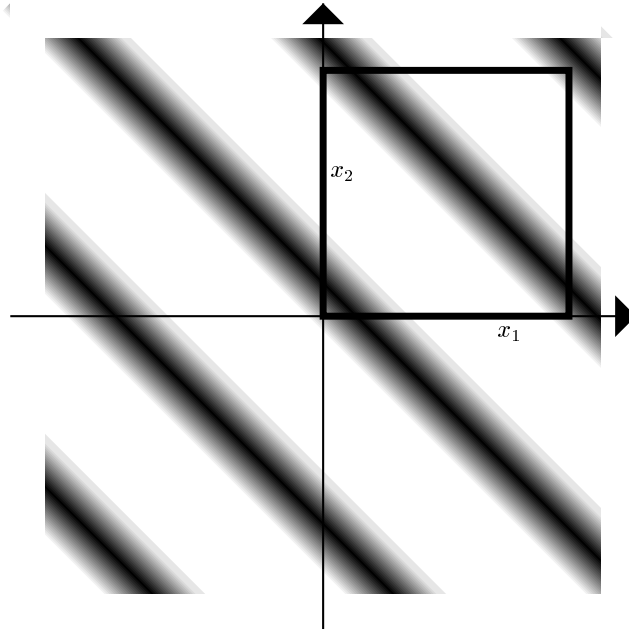


Figure 4. Visualization of Kohn's disconnectedness for a system of two spinless electrons in one dimension. The shaded areas are regions where $\Psi_0(x_1, x_2)$ is nonvanishing, and the square (of side L) is the periodicity supercell. For any given M (a one-dimensional vector in this simple case) the wavefunction $\Psi_{0,M}(x_1, x_2)$ in equation (79) is localized within a single stripe.

The most elementary example of Kohn's disconnectedness (for one particle in one dimension) is the one implied by equation (70) and figure 3. Another schematic picture of Kohn's disconnectedness is provided in figure 4, where we have $N = 2$ one-dimensional electrons (considered as spinless for the sake of simplicity). Their space coordinates are x_1 and x_2 , and the vector M is one dimensional. The nonoverlapping regions in the two-dimensional hyperspace are the diagonal stripes sketched in the figure. There are novel features with respect to the single-particle case: each of the $\Psi_{0,M}$ is *delocalized* in either the x_1 - or x_2 -variables, and the single-particle density $n(x)$ is delocalized as well. Nonetheless, localization in disconnected regions becomes evident if one considers the variable $x_1 - x_2$.

In the following we limit ourselves to illustrating how Kohn's localization works for the simple case of a crystalline system of independent electrons. As previously observed, the ground wavefunction Ψ_0 can be written in this case as a Slater determinant of WFs, equation (14), which obey PBC over the supercell of side L . In the thermodynamic limit each of the WFs goes to a well-defined function, which in the appropriate gauge is exponentially localized (section 2.6); we have used the (∞) superscript for these WFs. In full analogy to equation (70), the WFs can be written as

$$\langle \mathbf{r} | w_{nR_l} \rangle = w_n(\mathbf{r} - \mathbf{R}_l) \doteq \sum_M w_n^{(\infty)}(\mathbf{r} - \mathbf{R}_l - \mathbf{M}) = \sum_M \langle \mathbf{r} | w_{n, \mathbf{R}_l + \mathbf{M}}^{(\infty)} \rangle. \quad (80)$$

Using equation (80), it is then straightforward to prove [3, 4] that equation (14) becomes

$$|\Psi_0\rangle \doteq \sum_{M_1} \sum_{M_2} \cdots \sum_{M_{N_c}} |\Psi_{M_1 M_2 \dots M_{N_c}}^{(\infty)}\rangle, \quad (81)$$

where $N_c = N/(2n_b)$ is the number of cells in the supercell², and each $\Psi_{M_1 M_2 \dots M_{N_c}}^{(\infty)}$ is an N -electron determinant, obeying open boundary conditions:

$$|\Psi_{M_1 M_2 \dots M_{N_c}}^{(\infty)}\rangle = A \prod_n w_{n, R_1 + M_1}^{(\infty)} \dots w_{n, R_{N_c} + M_{N_c}}^{(\infty)} \overline{w}_{n, R_1 + M_1}^{(\infty)} \dots \overline{w}_{n, R_{N_c} + M_{N_c}}^{(\infty)}. \quad (82)$$

Kohn argues that the two wavefunctions $\Psi_{M_1 M_2 \dots M_{N_c}}^{(\infty)}$ and $\Psi_{M'_1 M'_2 \dots M'_{N_c}}^{(\infty)}$ have exponentially vanishing overlap unless $M_1 + M_2 + \dots + M_{N_c}$ equals $M'_1 + M'_2 + \dots + M'_{N_c}$.

Therefore the disconnectedness as above defined is verified, provided that we identify $\Psi_{0, M}$ in equation (79) with

$$|\Psi_{0, M}\rangle = \sum_{M_1 + M_2 + \dots + M_{N_c} = M} |\Psi_{M_1 M_2 \dots M_{N_c}}^{(\infty)}\rangle. \quad (83)$$

This disconnectedness is the key feature forbidding the free flow of charge in insulators under the application of a dc field, thus making the electrons polarizable but ‘bound’. Kohn conjectures that the same kind of disconnectedness is the cause of the insulating behaviour even in correlated and/or disordered systems, and quite generally in *any* insulator. A thorough discussion of Kohn’s localization and of its relationship to the modern theory of localization in insulators is provided by SWM.

5.3. Localization according to Resta and Sorella

The RS approach to electron localization stems from the modern theory of polarization [5–11]—owing particularly to the formulation of [12]—and is based on the same basic ingredients, namely the many-body phase operator of section 3.2 and its expectation values. According to RS, the electron localization in any condensed system can be measured by a Cartesian tensor, having the dimensions of a squared length. This localization tensor is the thermodynamic limit of expressions—such as equations (84) and (86) below—based on ground-state expectation values, where the latter are evaluated for N -electron wavefunctions with PBC. It is an intensive quantity, and has the key property of being finite in any insulator but divergent in metals: it provides therefore the sharp criterion needed to discriminate between the two classes of material. The localization tensor has the nature of a quadratic spread, or second cumulant moment; we borrow here notation due to SWM, and we call its elements $\langle r_\alpha r_\beta \rangle_c$, where the subscript *c* stands for ‘cumulant’.

The original RS paper only provides an expression for the diagonal elements of the tensor, which are sufficient to deal with high-symmetry materials. Inspired by equation (77), RS make the definition

$$\langle r_\alpha r_\alpha \rangle_c = -\frac{1}{N} \left(\frac{L}{2\pi} \right)^2 \log |z_N^{(\alpha)}|^2. \quad (84)$$

The equality obviously holds in the thermodynamic limit, whereas the symbol $=$ means \simeq (to leading order) if finite values of N and L are understood. In order to define the off-diagonal elements, it is expedient to introduce the notation

$$z_N^{(\alpha\beta)} = \langle \Phi_0(\kappa^{(\alpha)}) | \Phi_0(\kappa^{(\beta)}) \rangle = \langle \Psi_0 | \hat{U}(\kappa^{(\beta)} - \kappa^{(\alpha)}) | \Psi_0 \rangle, \quad (85)$$

using which the off-diagonal elements assume the form

$$\langle r_\alpha r_\beta \rangle_c = -\frac{1}{N} \left(\frac{L}{2\pi} \right)^2 \log \frac{|z_N^{(\alpha)}| |z_N^{(\beta)}|}{|z_N^{(\alpha\beta)}|}. \quad (86)$$

² N_c also equals M^3 , but we do not refer to M in the present section in order to avoid confusion with the vector \mathbf{M} , having a completely different meaning.

This expression was first published by Silvestrelli and Parrinello [20]; one can immediately verify that for $\alpha = \beta$ it reduces to the RS expression of equation (84). It will be clear in the following that the compact and elegant expressions of equations (84) and (86) enjoy the useful properties which have been anticipated, and more.

The main ingredient in equation (84) is the modulus of $z_N^{(\alpha)}$, the expectation value of the many-body phase operator. We notice that the phase of this same $z_N^{(\alpha)}$ was used in equation (41) to define the macroscopic polarization \mathbf{P} of the sample. This leads to the basic idea that in metals $z_N^{(\alpha)}$ goes to zero (fast enough with N), the localization tensor, equation (84), diverges, and \mathbf{P} is ill defined. Starting from the components of \mathbf{P} , it is expedient to define the first cumulant moment as

$$P_\alpha = \frac{eN}{L^3} \langle r_\alpha \rangle_c, \quad (87)$$

again an intensive quantity whose meaning becomes transparent upon realizing that L^3/N is the volume per electron. Equation (41) yields then

$$\langle r_\alpha \rangle_c = \frac{1}{N} \frac{L}{2\pi} \text{Im} \log z_N^{(\alpha)}. \quad (88)$$

A glance to equations (84) and (88) is enough for it to be obvious that polarization and localization are two aspects of the same phenomenon. This is indeed the main ‘philosophical’ claim of the RS paper.

Given the phenomenological link, outlined in the introduction, between macroscopic polarization and the insulating state of matter, RS postulated that a finite value of $\langle r_\alpha r_\beta \rangle_c$ is the essential characteristic of *any* insulator. Some anisotropic materials, such as graphite, can be conducting/insulating according to the Cartesian directions; RS suggest in such cases that the relevant tensor elements are divergent/finite.

The main RS postulate was corroborated in the original paper by two case studies: a crystalline system of independent electrons, where the statement is analytically proved, and a correlated one-dimensional model system, where the statement is numerically proved by explicit computation. Both these cases are discussed later, section 6.

Two important questions were not answered by RS. Given that in any insulator the localization tensor $\langle r_\alpha r_\beta \rangle_c$ is, at least in principle, a well-defined ground-state observable, the first question is whether this can be measured, and by which kind of experiments. The second question is whether $\langle r_\alpha r_\beta \rangle_c$ can be related in some way to dc conductivity, given that the vanishing of the latter characterizes—in addition to macroscopic polarization—the insulating state of matter. Both these questions were later answered by SWM, by the same unique token.

5.4. The work of Souza, Wilkens, and Martin

The tensor quantity which measures localization according to SWM coincides, in the thermodynamic limit, with $\langle r_\alpha r_\beta \rangle_c$ defined by RS, although it is arrived at by means of rather different algebra. The main result in the SWM paper is that $\langle r_\alpha r_\beta \rangle_c$ is indeed measurable: in fact it can be expressed as a frequency integral of the real part of the conductivity, equation (94) below, a quantity which is accessible to experiments. Furthermore, SWM argue that this frequency integral diverges in any metal, thus providing a general proof of the main RS postulate.

I have stressed before that $\langle r_\alpha r_\beta \rangle_c$ is a *ground-state* property, while conductivity is a property of the *excitations* of the many-electron system. In fact the SWM result is a *sum rule* of the fluctuation-dissipation kind, in the same class as those presented in section 4.1. Owing to the sum rule, one has access to a ground-state property by actually performing a measurement over the excited states.

We are not following SWM in the presentation of their main result; instead, we prefer to exploit the formalism, notation, and results developed throughout this review in order to prove their major point. The quantity of interest is the frequency integral

$$I_{\alpha\beta} = \int_0^\infty \frac{d\omega}{\omega} \operatorname{Re} \sigma_{\alpha\beta}(\omega), \quad (89)$$

and the issue is whether $I_{\alpha\beta}$ is a finite or a divergent tensor. Since the integral converges at its upper limit, the discussion focuses on the low-frequency behaviour of the integrand.

In a simple insulator—with a finite gap ε_g in the excitation spectrum—the real part of the conductivity, equation (67), vanishes for $\omega < \varepsilon_g$, and $I_{\alpha\beta}$ obviously converges. For other classes of more exotic insulators, with no gap in the excitation spectrum, the momentum elements in equation (67) either vanish or go to zero fast enough with decreasing excitation energies to ensure the vanishing of the dc conductivity characterizing the insulating state of matter at zero temperature. Therefore, according to SWM, the integral $I_{\alpha\beta}$ converges in *any* insulator.

Starting from equation (67), $I_{\alpha\beta}$ is cast as

$$I_{\alpha\beta} = \frac{\pi \hbar e^2}{m^2 L^3} \sum_{n \neq 0}' \frac{\langle \Psi_0 | \hat{P}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{P}_\beta | \Psi_0 \rangle}{(E_n - E_0)^2} = \frac{\pi \hbar e^2}{m^2 L^3} \sum_{n, n' \neq 0}' \frac{\langle \Psi_0 | \hat{P}_\alpha | \Psi_n \rangle}{E_n - E_0} \langle \Psi_n | \Psi_{n'} \rangle \frac{\langle \Psi_{n'} | \hat{P}_\beta | \Psi_0 \rangle}{E_{n'} - E_0}. \quad (90)$$

From equation (37), we make the replacement

$$\sum_{n \neq 0}' \langle \Psi_n | \frac{\langle \Psi_n | \hat{P}_\alpha | \Psi_0 \rangle}{E_n - E_0} \simeq \frac{mL}{2\pi \hbar} (e^{-i\gamma_N^{(\alpha)}} | \Phi_0(\boldsymbol{\kappa}^{(\alpha)}) \rangle - | \Psi_0 \rangle). \quad (91)$$

Notice that equation (37) has been obtained under the hypothesis of a finite-gap insulator; however, for a more general insulator where $1/(E_n - E_0)$ might diverge in the thermodynamic limit, such divergences are compensated by the matrix elements in the numerator, which must go to zero fast enough to ensure insulating behaviour of equation (67). We thus get for any insulator the finite tensor

$$I_{\alpha\beta} \simeq \frac{e^2}{4\pi \hbar L} (1 + |z_N^{(\alpha\beta)}| - |z_N^{(\alpha)}| - |z_N^{(\beta)}|) \simeq -\frac{e^2}{4\pi \hbar L} \log \frac{|z_N^{(\alpha)}| |z_N^{(\beta)}|}{|z_N^{(\alpha\beta)}|}, \quad (92)$$

where we have used equations (28) and (44), as well as

$$z_N^{(\alpha\beta)} = \langle \Psi_0 | \hat{U}(\boldsymbol{\kappa}^{(\beta)} - \boldsymbol{\kappa}^{(\alpha)}) | \Psi_0 \rangle \simeq e^{i(\gamma_N^{(\beta)} - \gamma_N^{(\alpha)})} |z_N^{(\alpha\beta)}|. \quad (93)$$

Comparing with the localization tensor, equation (86), one immediately obtains the main SWM result:

$$\langle r_\alpha r_\beta \rangle_c = \frac{\hbar L^3}{\pi e^2 N} \int_0^\infty \frac{d\omega}{\omega} \operatorname{Re} \sigma_{\alpha\beta}(\omega). \quad (94)$$

In a simple finite-gap insulator the diagonal elements fulfil the inequality

$$\langle r_\alpha r_\alpha \rangle_c < \frac{\hbar L^3}{\pi e^2 \varepsilon_g N} \int_{\varepsilon_g}^\infty \frac{d\omega}{\omega} \operatorname{Re} \sigma_{\alpha\alpha}(\omega) = \frac{\hbar^2}{2m\varepsilon_g}, \quad (95)$$

where the oscillator-strength sum rule has been used in the last equality. Equation (95) shows that the larger the gap, the smaller the upper bound to the diagonal element of the localization tensor; if the gap vanishes, there is no bound on it.

6. Localization in selected systems

6.1. Band insulators

We consider here a crystalline system of independent electrons, with n_b doubly occupied bands. The diagonal element of the localization tensor, equation (84), can be written using equation (50) as

$$\langle r_\alpha r_\alpha \rangle_c = -\frac{2}{N} \left(\frac{L}{2\pi} \right)^2 \log \prod_s \det S^\dagger(\mathbf{q}_s, \mathbf{q}_s + \boldsymbol{\kappa}^{(\alpha)}) S(\mathbf{q}_s, \mathbf{q}_s + \boldsymbol{\kappa}^{(\alpha)}). \quad (96)$$

This product, since it runs over *all* the M^3 mesh points, is nonvanishing (see also below, section 6.2).

In the thermodynamic limit, the expression in equation (96) becomes an integral over the reciprocal cell (or equivalently over the Brillouin zone):

$$\langle r_\alpha r_\alpha \rangle_c = \frac{a^3}{n_b (2\pi)^3} \int d\mathbf{q} \left(\sum_{n=1}^{n_b} \left\langle \frac{\partial}{\partial q_\alpha} u_{n\mathbf{q}} \middle| \frac{\partial}{\partial q_\alpha} u_{n\mathbf{q}} \right\rangle - \sum_{n,n'=1}^{n_b} \left\langle u_{n\mathbf{q}} \middle| \frac{\partial}{\partial q_\alpha} u_{n'\mathbf{q}} \right\rangle \left\langle \frac{\partial}{\partial q_\alpha} u_{n'\mathbf{q}} \middle| u_{n\mathbf{q}} \right\rangle \right). \quad (97)$$

The proof is relatively straightforward, starting from equation (97) and discretizing both the integral and the derivatives on the mesh defined in equation (10).

We have previously met in section 2.6 the same expression as in equation (97), and we have discussed its meaning according to MV: if we exploit the gauge freedom in order to write the ground-state wavefunction as the Slater determinant of optimally localized WFs, then in the thermodynamic limit the spherical second moment of such WFs is, on average, larger than the trace of our localization tensor, equations (18) and (20).

We emphasize, once more, that both equations (96) and (97) yield gauge-invariant expressions, which only depend on the N -electron wavefunction Ψ_0 , and *not* on the specific choice of the orbitals: the equivalent expression given in equation (21) makes this feature clear. It is worth noticing that in insulators the projector $P(\mathbf{r}, \mathbf{r}')$ decays exponentially with $|\mathbf{r} - \mathbf{r}'|$; hence the integral over all space in equation (21) converges despite the occurrence of the (unbound) position operator.

The nondiagonal elements $\langle r_\alpha r_\beta \rangle_c$, equation (86), also go, in the $M \rightarrow \infty$ limit, to the same expression as in equation (19), although the proof is less straightforward. Besides the trace, even the localization tensor itself bears some important meaning in the present case of an insulating crystal of independent electrons. Suppose that, unlike MV, we exploit the gauge freedom in order to write the ground-state wavefunction as the Slater determinant of orbitals that are optimally localized only *in a given direction*, meaning that the average quadratic spread in such direction is minimal. These orbitals have been called ‘hermaphrodite’ orbitals [23], since they are Wannier-like in the chosen direction, and Bloch-like in the orthogonal ones. It can be proved that the average quadratic spread of the optimally localized hermaphrodite orbitals in the given direction is *equal* to the appropriate projection of the localization tensor $\langle r_\alpha r_\beta \rangle_c$. Further discussion is provided in [23].

The localization tensor has recently been calculated from first principles—within density-functional theory in the local-density approximation [27]—for several tetrahedrally coordinated crystalline materials from the groups IV, III–V, and II–VI, having the diamond and zinc-blende structure [23]. As explained in section 3.3, what is indeed computed is the localization tensor of the noninteracting KS system; it is assumed that in this class of materials such a tensor is a good approximation to the corresponding quantity for the real interacting system. Since the materials are cubic, the tensor is indeed a scalar (multiple of the identity), and will be referred to as the squared localization length $\langle x^2 \rangle_c$.

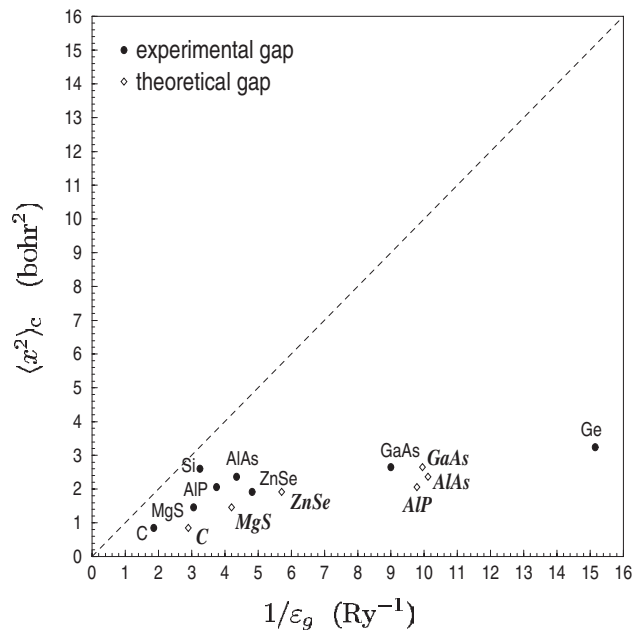


Figure 5. Squared localization length versus the inverse direct gap (theoretical and experimental), for several elemental and binary semiconductors. The inequality (95) is strongly verified. The points corresponding to Si and Ge with the theoretical gaps are off the scale; after [23].

This quantity is plotted in figure 5 versus the rhs of the inequality (95), where for the gap ε_g we have used both (i) the theoretical and (ii) the experimental values. In case (i) the inequality arises from an exact sum rule and *must* be satisfied; we are therefore checking the internal consistency of the calculations. Also, it may be noticed that the two members of the inequality differ in general by a rather large factor. In case (ii) there is no *a priori* guarantee that the inequality is verified, particularly given the fact that the experimental gap is systematically larger than the KS one. Nonetheless, the localization length is obtained here as a pure ground-state property, and it is well known that density-functional theory in the local-density approximation provides a good representation of the ground state, though not of the excitations. It is therefore interesting to verify that even for case (ii) the inequality (95) is strongly verified for all the materials studied.

The squared localization length ranges roughly between 1 and 3 bohr² for all the materials considered, diamond being the most localized and germanium the most delocalized. The trend is qualitatively expected, in agreement with SWM's statement that 'the larger the gap, the more localized the electrons are'. However, inequality (95) only provides a bound, and indeed a few materials show irregularities in their trend.

6.2. Band metals

We consider here a crystalline metal of independent electrons. The ground wavefunction can still be written as a Slater determinant of Bloch orbitals, equation (11). But then, given the presence of a Fermi surface, some \mathbf{q}_s of the mesh are *not* occupied. This has dramatic consequences for the expectation value $z_N^{(\alpha)}$, which indeed *vanishes*; hence the localization tensor $\langle r_\alpha r_\alpha \rangle_c$ diverges and, by the same token, the macroscopic polarization is ill defined.

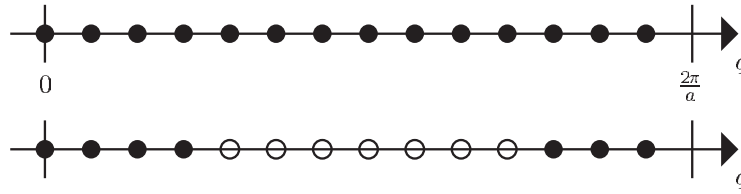


Figure 6. Discrete q -vectors, equation (10), in the reciprocal cell in one dimension, where $M = 14$ has been chosen. Black circles correspond to occupied q s, and empty circles to unoccupied ones. Top: insulator, with $N = 2M$ (filled band). Bottom: metal, with $N = M$ (half-filled band).

The algebra leading to the vanishing of $z_N^{(\alpha)}$ is best illustrated for the simple case of one band in one dimension, whose discrete Bloch vectors, equation (10), are shown in figure 6 for the case $M = 14$. The system is insulating if the number of electrons N equals $2M$ (filled band, top sketch), and a metal if $N = M$ (half-filled band, bottom sketch).

The *large* overlap matrix, equation (48), simplifies to

$$\mathcal{S}_{q_s, q_{s'}} = \frac{1}{L} \int_0^L dx \psi_{q_s}^*(x) \psi_{q_{s'}}(x) e^{i \frac{2\pi}{M} x}, \quad (98)$$

where the identity $\psi_{q_M}(x) \equiv \psi_{q_0}(x)$ is understood (the so-called periodic gauge). The elements of \mathcal{S} are nonzero whenever $s = s' + 1$. In the insulating case, owing to complete filling, both s and s' run over all the M -values; for any given s , there is always one (and only one) s' such that the matrix element in equation (98) is nonzero. That means that in any row of the \mathcal{S} -matrix—whose size is $M \times M$ —there is one, and only one, nonvanishing element. Under these circumstances, the determinant factors as a product of M nonzero numbers and is, therefore, nonvanishing.

The metallic case is very different. Since not all the q_s -vectors are occupied, the indices s and s' run over a subset of the M -values (figure 6): the matrix \mathcal{S} is of size $M/2 \times M/2$. There is one of the two q_s at the Fermi level for which the integrals in equation (98) are all vanishing, for any *occupied* s' . Therefore the matrix \mathcal{S} has a row of zeros, and its determinant vanishes, even at finite M : as anticipated, the localization length is infinite and macroscopic polarization is ill defined. The same proof can be obviously generalized to the three-dimensional multi-band case, at the price of a somewhat heavier algebra.

It is interesting to address the divergence of the localization tensor in a metal (of independent electrons) in an alternative way, focusing on the equivalent expression for $\langle r_\alpha r_\beta \rangle_c$ given in equation (21). This expression is easily demonstrated starting from equation (97) and using equation (15)—assuming therefore an *insulating* ground state. However, the same expression as in equation (21) holds its formal validity in the metallic case as well, where it diverges. To realize this, it is enough to imagine that the periodic potential is continuously varied until the insulating system undergoes a band crossing (or touching) and therefore becomes conducting. On the insulating side of the transition the localization tensor is a continuous function of the periodic crystalline potential and equals the rhs of equation (21); at the transition point and beyond—owing to the key SWM result—the lhs diverges and therefore the rhs must diverge as well. It is straightforward to verify this divergence for the simplest metal of all, namely, the free-electron gas, where the density matrix is known exactly in the thermodynamic limit [55]:

$$P(\mathbf{r}, \mathbf{r}') = \frac{1}{2} \rho(\mathbf{r}, \mathbf{r}') = \frac{3n_0}{2} \frac{j_1(k_F |\mathbf{r} - \mathbf{r}'|)}{k_F |\mathbf{r} - \mathbf{r}'|}, \quad (99)$$

where n_0 is the density, k_F is the Fermi wavevector, and j_1 is a Bessel function. Replacement of equation (99) into (21) results in a diverging integral, thus confirming that our localization tensor is formally infinite in this paradigmatic metal. Other, more realistic, metals feature this same divergence owing to the weak (polynomial) decay of $P(\mathbf{r}, \mathbf{r}')$ for large $|\mathbf{r} - \mathbf{r}'|$, which in turn arises from the sharpness of the Fermi surface in reciprocal space.

Finally, a warning is in order. It looks tempting to apply equation (21) even to the case of a correlated system—either insulating or metallic—using the most general density matrix in the form of equation (4), but this would be *incorrect*. Equation (21) only applies to a system of independent electrons, while we have stressed above that in a correlated system the expectation value of the many-body operator, equation (30), and hence our localization tensor, is *not* a function of the one-body reduced density matrix.

6.3. Correlated electrons: a model system

SWM have proved that the localization tensor is finite in any insulator and diverges in any metal (section 5.4). In the original RS paper, the issue is instead addressed by means of a correlated model system, whose ground wavefunction is computed by exact Lanczos diagonalizations. The same model was previously used in the literature [56, 57] to represent an insulator having a mixed ionic/covalent character. The model consists of a one-dimensional two-band Hubbard Hamiltonian at half-filling, whose ground wavefunction is explicitly correlated. The strength of the electron–electron interaction is measured by a single Hubbard parameter U . We simply call z_N the expectation value of the many-body operator, equation (27), in the one-dimensional case. For present purposes, it is enough to study the centrosymmetric geometry: the wavefunction is real, and whenever z_N does not vanish, its phase is either 0 or π .

According to some very recent work [24, 58], the same model Hamiltonian [56, 57] as was used by RS seems to have a very rich phase diagram, actually richer than suspected by RS. Since the focus here is on localization features, and not on the model itself, I will neglect other possible phases and I will only deal with the phase diagram as investigated by RS. As a function of U , the behaviour is as follows: at small U the system is a band insulator, while at a critical U_c it undergoes a transition to a Mott-like insulating phase. At the transition, the phase jumps by π ; it turns out that this occurrence is the main fact signalling the transition between the two, topologically distinct, insulating ground states. The phase of z_N , divided by π , is therefore a topological (integer) quantum number [59] which can be used as an order parameter to identify the two phases of the system: these correspond to qualitatively different organizations of the electrons in the ground-state wavefunction [57, 60].

In one dimension, equation (84) becomes

$$\langle x^2 \rangle_c = -\frac{1}{N} \left(\frac{L}{2\pi} \right)^2 \log |z_N|^2 = -\frac{N}{(2\pi n_0)^2} \log |z_N|^2, \quad (100)$$

where $n_0 = N/L$ is the average density. What happens to the localization length as a function of U ? The answer is provided by figure 7, where we show the dimensionless quantity $\mathcal{D} = -N \log |z_N|^2$ in the large- N limit. At $U = 0$ the system is noninteracting, and the squared localization length $\langle x^2 \rangle_c$ coincides with the second cumulant moment of the optimally localized WF of the occupied band. In the correlated case at $U \neq 0$, no Wannier analysis can be performed; yet $\langle x^2 \rangle_c$ maintains its role as a global measure of the localization of the electronic wavefunction as a whole. The localization length increases with U below the transition, diverges at the transition point U_c , and becomes localized again in the highly correlated regime, where in fact $\langle x^2 \rangle_c$ decreases with increasing U . Since the localization length remains finite at all values of U different from U_c , the system is always insulating except at a single point.

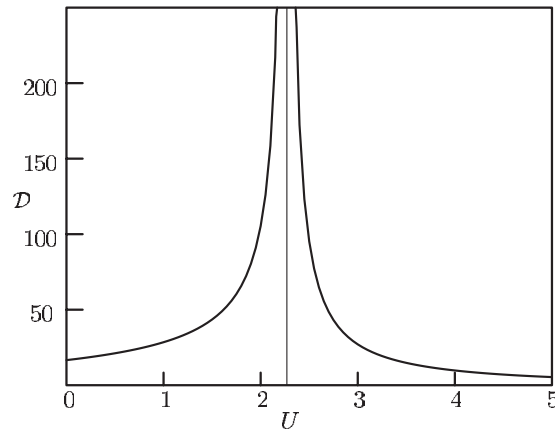


Figure 7. The dimensionless quantity $D = 4\pi n_0^2 \langle x^2 \rangle_c$ as a function of the electron–electron interaction parameter U for the model system. At low values of U the system is a band insulator, whereas at high values of U it is a Mott insulator. At the transition point the system is metallic and the localization length diverges; after RS [18].

What about the transition point? According to the previously stated viewpoint, the delocalized behaviour implies a metallic character of the many-electron system. Notice that $\langle x^2 \rangle_c$ is a pure ground-state property and apparently carries no information about the excitation spectrum of the system. Yet RS have explicitly verified—by exploiting the metastability of the Lanczos algorithm—that at the critical U -value there is indeed a level crossing. At the transition value U_c the ground state is twice degenerate and the lowest-lying excitation (at constant N) has vanishing energy, as is appropriate for a metal.

7. Conclusions

An insulator is distinguished from a metal (at zero temperature) by its vanishing dc conductivity and by its ability to sustain bulk macroscopic polarization. Both these features stem from a qualitatively different organization of the electrons in their ground state: electrons are localized in insulators and delocalized in metals. Such localization, however, is hidden in a rather subtle way into the many-body ground wavefunction. The present review provides a comprehensive survey of the modern approach to the problem: polarization, dc conductivity, and localization are addressed and explained by means of a unified, and remarkably simple, theoretical scheme.

The ‘theory of the insulating state’ starts from a milestone paper by Kohn, which appeared in 1964 [3]. Kohn argued that the electron localization is the cause of the vanishing of the dc conductivity in any kind of insulator, and that such localization can be sharply defined via the *disconnectedness* of the ground-state wavefunction (section 5.2). Substantial advances in the theory of the insulating state came only 35 years later, starting with the RS paper. At the root of the modern developments is the theory of macroscopic polarization, based on a Berry’s phase, which since 1993 has revolutionized our view of the phenomenon. According to RS, macroscopic polarization and electron localization are two aspects of the same phenomenon, and they can be addressed by the same formal tool, namely, the many-body phase operator (section 3.2), widely used throughout this review. The ground-state expectation value of this operator is a dimensionless complex number which vanishes in metals and is finite in insulators: in the latter case, the phase of this same number yields the macroscopic polarization. The

picture was later completed by SWM, who provided the missing link between RS localization and conductivity features. There is a direct relationship between the modern approach and Kohn's one: the many-body phase operator extracts in a very elegant and effective way the main connectedness feature of the many-body wavefunction. This is very clear when looking at the simple one-body case (section 5.1).

We have addressed here the most general condensed system: our compact formulation, in fact, deals on the same basis with crystalline and disordered systems, as well as with correlated- and independent-electron systems. The kind of localization discussed in this work is a global property of the many-electron system as a whole, *not* a feature of the individual orbitals.

We have also applied the general many-body formulation to the simple band structure case sketched in figure 2, i.e. specializing to a crystalline system of independent electrons. For such a system, the ground wavefunction is a Slater determinant; a relevant issue is whether it can be written as a determinant of *localized* one-particle orbitals. This issue, in fact, also discriminates between insulators and metals. We have shown that in a band insulator (section 6.1) the ground state *can* be written as a determinant of orbitals (WF) whose second moments are finite in the thermodynamic limit, while in a band metal (section 6.2) this is impossible.

I have shown that electron localization in different insulators can be quantitatively compared by means of a localization tensor, or second cumulant moment. This quantity is a pure ground-state property, yet it is experimentally accessible by means of conductivity experiments, i.e. actually measuring *excited-state* properties. This result, found by SWM, arises from a fluctuation-dissipation sum rule.

Metals are qualitatively different from insulators in that the localization tensor diverges therein. The theory, at its present stage, does not discriminate between normal metals and superconductors: all behave in the same way as long as the localization tensor is considered. An open issue is whether the present formalism can be modified or enhanced in order to even discriminate between these cases, given that Berry's phases have indeed been used on some occasions to deal with the superconducting transition [61].

The modern theory of localization reviewed in this work is very recent. Besides the analytical results, to date only a few published papers have provided actual computations. Numerical implementations of this theory in selected case studies may have two different aims: to demonstrate how the theory works or to exploit the theory for detecting the onset of the metal-insulator transition in a given system. In the literature published so far [20–24], there have been some studies of electron localization in correlated systems, while there are neither studies of disordered systems, nor investigations of any relationship to Anderson localization. However, the RS localization is expected to be applicable to such issues as well.

Acknowledgments

Invaluable discussions with R M Martin, N Marzari, F Mauri, G Ortíz, M Peressi, C Sgiarovello, S Sorella, I Souza, and D Vanderbilt are gratefully acknowledged. This review was written while the author was on sabbatical leave at the Max-Planck-Institut für Festkörperforschung in Stuttgart: I am grateful to O K Andersen and to several colleagues for creating a pleasant and stimulating environment there. This work was partially supported by the Office of Naval Research through grant N00014-96-1-0689.

References

- [1] Ashcroft N W and Mermin N D 1976 *Solid State Physics* (Philadelphia, PA: Saunders)
- [2] Kittel C 1996 *Introduction to Solid State Physics* 7th edn (New York: Wiley)

- [3] Kohn W 1964 *Phys. Rev. A* **133** 171
- [4] Kohn W 1968 *Many-Body Physics* ed C DeWitt and R Balian (New York: Gordon and Breach) p 351
- [5] Resta R 1992 *Ferroelectrics* **136** 51
- [6] King-Smith R D and Vanderbilt D 1993 *Phys. Rev. B* **47** 1651
- [7] Vanderbilt D and King-Smith R D 1993 *Phys. Rev. B* **48** 4442
- [8] Resta R 1993 *Europhys. Lett.* **22** 133
- [9] Ortíz G and Martin R M 1994 *Phys. Rev. B* **43** 14 202
- [10] Resta R 1994 *Rev. Mod. Phys.* **66** 899
- [11] Resta R 1997 *Europhys. News* **28** 18
- [12] Resta R 1998 *Phys. Rev. Lett.* **80** 1800
- [13] Resta R 1999 *Int. J. Quantum Chem.* **75** 599
- [14] Berry M V 1984 *Proc. R. Soc. A* **392** 45
- [15] Shapere A and Wilczek F (ed) 1989 *Geometric Phases in Physics* (Singapore: World Scientific)
- [16] Resta R 2000 *J. Phys.: Condens. Matter* **12** R107
- [17] See several papers in the most recent proceedings:
Proc. Workshop on First-Principles Calculations for Ferroelectrics 1995 *Ferroelectrics* **164**
Proc. Workshop on First-Principles Calculations for Ferroelectrics 1997 *Ferroelectrics* **194**
Cohan R E (ed) *Proc. Workshop on First-Principles Calculations for Ferroelectrics* 1998 *AIP Conf. Proc.* **436**
Cohan R E and Rabe K (eds) *Proc. Workshop on First-Principles Calculations for Ferroelectrics* 2000 *AIP Conf. Proc.* **535**
Krakauer H (ed) *Proc. Workshop on First-Principles Calculations for Ferroelectrics* 2001 *AIP Conf. Proc.* **582**
- [18] Resta R and Sorella S 1999 *Phys. Rev. Lett.* **82** 370
- [19] Souza I, Wilkens T and Martin R M 2000 *Phys. Rev. B* **62** 1666
- [20] Silvestrelli P L and Parrinello M 1999 *J. Chem. Phys.* **111** 3572
- [21] Aligia A A and Ortíz G 1999 *Phys. Rev. Lett.* **82** 2560
- [22] Aebischer C, Baeriswyl D and Noack R M 2001 *Phys. Rev. Lett.* **86** 468
- [23] Sgiarovello C, Peressi M and Resta R 2001 *Phys. Rev.* **64** 115202
- [24] Wilkens T and Martin R M 2001 *Phys. Rev. B* **63** 235108
- [25] McWeeny R 1992 *Methods of Molecular Quantum Mechanics* 2nd ed (London: Academic)
- [26] Pisani C, Dovesi R and Roetti C 1988 *Hartree-Fock Ab-Initio Treatment of Crystalline Systems (Springer Lecture Notes in Chemistry vol 48)* (Berlin: Springer)
- [27] Lundqvist S and March N H (ed) 1983 *Theory of the Inhomogeneous Electron Gas* (New York: Plenum)
- [28] Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
Car R and Parrinello M 1988 *Simple Molecular Systems at Very High Density (NATO ASI Series B, vol 186)* ed P Loubeyre, A Polian and N Boccara (New York: Plenum) p 455
Galli G and Pasquarello A 1993 *Computer Simulations in Chemical Physics* ed M Allan and D Tildesley (Dordrecht: Kluwer) p 261
- [29] Silvestrelli P L 1999 *Phys. Rev. B* **59** 9703
- [30] Wannier G H 1937 *Phys. Rev.* **52** 191
- [31] Blount E I 1962 *Solid State Physics* vol 13, ed H Ehrenreich, F Seitz and D Turnbull (New York: Academic) p 305
- [32] Kohn W 1959 *Phys. Rev. Lett.* **2** 393
- [33] des Cloizeaux J 1964 *Phys. Rev. A* **135** 685
des Cloizeaux J 1964 *Phys. Rev. A* **135** A697
- [34] Ismail-Beigi S and Arias T A 1999 *Phys. Rev. Lett.* **82** 2127
- [35] He L and Vanderbilt D 2001 *Phys. Rev. Lett.* **86** 5341
- [36] Kohn W 1996 *Phys. Rev. Lett.* **76** 3168
- [37] Marzari N and Vanderbilt D 1997 *Phys. Rev. B* **56** 12 847
- [38] Landau L D and Lifshitz E M 1984 *Electrodynamics of Continuous Media* (Oxford: Pergamon)
- [39] Gonze X, Ghosez P and Godby R W 1995 *Phys. Rev. Lett.* **74** 4035
- [40] Gonze X, Ghosez P and Godby R W 1997 *Phys. Rev. Lett.* **78** 294
Gonze X, Ghosez P and Godby R W 1997 *Phys. Rev. Lett.* **78** 2029
Martin R M and Ortíz G 1997 *Phys. Rev. B* **56** 1124
Vanderbilt D 1997 *Phys. Rev. Lett.* **79** 3966
Ortíz G, Souza I and Martin R M 1998 *Phys. Rev. Lett.* **80** 353
- [41] van Gisbergen S J A, Schipper P R T, Gritsenko O V, Baerends E J, Snijders J G, Champagne B and Kirtman B 1999 *Phys. Rev. Lett.* **83** 694
- [42] de Boeij P L, Kootstra F, Berger J A, van Leeuwen R and Snijders J G 2001 *J. Chem. Phys.* **115** 1995

- [43] Akkermans E 1997 *J. Math. Phys.* **38** 1781
- [44] Aharonov Y and Bohm D 1959 *Phys. Rev.* **115** 485
This is reprinted as
Aharonov Y and Bohm D 1989 *Geometric Phases in Physics* ed A Shapere and F Wilczek (Singapore: World Scientific) p 104
- [45] Feynman R P, Leighton R B and Sands M 1964 *The Feynman Lectures in Physics* vol 2 (Reading, MA: Addison-Wesley) section 15-4
- [46] Thouless D J 1983 *Phys. Rev. B* **27** 6083
Niu Q and Thouless D J 1984 *J. Phys. A: Math. Gen.* **17** 2453
- [47] Debernardi A, Bernasconi M, Cardona M and Parrinello M 1997 *Appl. Phys. Lett.* **71** 2692
Silvestrelli P L, Bernasconi M and Parrinello M 1997 *Chem. Phys. Lett.* **277** 478
Pasquarello A and Car R 1997 *Phys. Rev. Lett.* **79** 1766
Bernasconi M, Silvestrelli P L and Parrinello M 1998 *Phys. Rev. Lett.* **81** 1235
- [48] Zubarev D N 1960 *Sov. Phys.–Usp.* **3** 320
- [49] Zubarev D N 1974 *Non-Equilibrium Statistical Mechanics* (New York: Consultants Bureau)
- [50] Kubo R, Toda M and Hashitsume N 1985 *Statistical Physics II, Nonequilibrium Statistical Mechanics (Springer Series in Solid-State Sciences vol 31)* (Berlin: Springer)
- [51] Forster D 1975 *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions* (Reading, MA: Benjamin)
- [52] Feynman R P, Leighton R B and Sands M 1964 *The Feynman Lectures in Physics* vol 1 (Reading, MA: Addison-Wesley) section 41-1
- [53] Schönhammer K and Brenig W 1973 *Phys. Lett. A* **42** 447
- [54] Kudinov E K 1991 *Fiz. Tverd. Tela* **33** 2306 (Engl. transl. 1991 *Sov. Phys.–Solid State* **33** 1299)
- [55] Jones W and March N H 1985 *Theoretical Solid State Physics* (New York: Wiley) (reprinted (New York: Dover))
- [56] Egami T, Ishihara S and Tachiki M 1993 *Science* **261** 1307
Egami T, Ishihara S and Tachiki M 1994 *Phys. Rev. B* **49** 8944
Egami T, Ishihara S and Tachiki M 1994 *Phys. Rev. B* 16123
- [57] Resta R and Sorella S 1995 *Phys. Rev. Lett.* **87** 4738
- [58] Fabrizio M, Gogolin A O and Nersesyan A A 1999 *Phys. Rev. Lett.* **83** 2014
Takada Y and Kido M 2001 *J. Phys. Soc. Japan* **70** 21
Torio M E, Aligia A A and Ceccatto H A 2001 *Phys. Rev. B* **64** 121105(R)
- [59] Thouless D J 1998 *Topological Quantum Numbers in Nonrelativistic Physics* (Singapore: World Scientific)
- [60] Gidopoulos N, Sorella S and Tosatti E 2000 *Eur. Phys. J. B* **14** 217
- [61] Aligia A A and Gagliano E R 1998 *Physica C* **304** 29
Aligia A A 1999 *Europhys. Lett.* **45** 411
Aligia A A, Hallberg K, Batista C D and Ortíz G 2001 *Phys. Rev. B* **61** 7883