Electrical polarization and orbital magnetization: the modern theories

2010 J. Phys.: Condens. Matter 22 123201

(http://iopscience.iop.org/0953-8984/22/12/123201)

IOP Publishing Journal of Physics: Condensed Matter

J. Phys.: Condens. Matter **22** (2010) 123201 (19pp) doi:10.1088/0953-8984/22/12/123201

TOPICAL REVIEW

Electrical polarization and orbital magnetization: the modern theories

Raffaele Resta

Dipartimento di Fisica, Università di Trieste, Strada Costiera 11, I-34014 Trieste, Italy
and
CNR-INFM DEMOCRITOS National Simulation Center, Trieste, Italy

Received 7 December 2009, in final form 5 February 2010
Published 11 March 2010
Online at stacks.iop.org/JPhysCM/22/123201

Abstract

Macroscopic polarization **P** and magnetization **M** are the most fundamental concepts in any phenomenological description of condensed media. They are intensive vector quantities that intuitively carry the meaning of dipole per unit volume. But for many years both **P** and the orbital term in **M** evaded even a precise microscopic definition, and severely challenged quantum-mechanical calculations. If one reasons in terms of a finite sample, the electric (magnetic) dipole is affected in an extensive way by charges (currents) at the sample boundary, due to the presence of the unbounded position operator in the dipole definitions. Therefore **P** and the orbital term in **M**—phenomenologically known as bulk properties—apparently behave as surface properties; only spin magnetization is problemless. The field has undergone a genuine revolution since the early 1990s. Contrary to a widespread incorrect belief, **P** *has nothing to do* with the periodic charge distribution of the polarized crystal: the former is essentially a property of the *phase* of the electronic wavefunction, while the latter is a property of its *modulus*. Analogously, the orbital term in **M** has nothing to do with the periodic current distribution in the magnetized crystal. The modern theory of polarization, based on a Berry phase, started in the early 1990s and is now implemented in most first-principle electronic structure codes. The analogous theory for orbital magnetization started in 2005 and is partly work in progress. In the electrical case, calculations have concerned various phenomena (ferroelectricity, piezoelectricity, and lattice dynamics) in several materials, and are in spectacular agreement with experiments; they have provided thorough understanding of the behaviour of ferroelectric and piezoelectric materials. In the magnetic case the very first calculations are appearing at the time of writing (2010). Here I review both theories on a uniform ground in a density functional theory (DFT) framework, pointing out analogies and differences. Both theories are deeply rooted in geometrical concepts, elucidated in this work. The main formulae for crystalline systems express **P** and **M** in terms of Brillouin-zone integrals, discretized for numerical implementation. I also provide the corresponding formulae for disordered systems in a single **k**-point supercell framework. In the case of **P** the single-point formula has been widely used in the Car–Parrinello community to evaluate IR spectra.

Contents

1. Introduction

Polarization **P** and magnetization **M** are fundamental concepts that all undergraduates learn about in elementary courses [1, 2]. In view of this, it is truly extraordinary that until rather recently there was no generally accepted formula for both electrical polarization and orbital magnetization in condensed matter, even as a matter of principles. Computations of both **P** and **M** for real materials were therefore impossible. It is important to stress that we are addressing here 'polarization itself' and 'magnetization itself', while instead linear-response theory has satisfactorily provided **P** derivatives over the years and, more recently, even **M** derivatives.

In the case of **P**, a genuine change of paradigm was initiated by a couple of important papers [3, 4], after which the major development was introduced by King-Smith and Vanderbilt in 1992 (paper published in 1993 [5]). Other important advances continued during the 1990s [6, 7] and the so-called 'modern theory of polarization' has been at a mature stage for about a decade. Among other things, the modern theory shed new light on previous linear-response formulations. Several reviews have appeared in the literature: the very first one is [8] and the most recent ones are [9, 10].

In the case of **M** (or, more precisely, of the *orbital contribution* to **M**) a similar breakthrough only occurred in 2005 [11, 12], and the 'modern theory of magnetization' is partly work in progress.

It is worth emphasizing that most *ab initio* electronic structure codes on the market, for dealing with either crystalline or noncrystalline materials, implement the modern theory of polarization as a standard option. A nonexaustive list includes ABINIT [13], CPMD [14] CRYSTAL [15], QUANTUM-ESPRESSO [16], SIESTA [17], and VASP [18]. Implementations of the modern theory have been instrumental for more than a decade in the study of ferroelectric and piezoelectric materials [19–21]. The basic concepts of the modern theory of polarization have also started reaching a few textbooks [22], though very slowly; most of them are still plagued with erroneous concepts and statements.

At variance with the electrical case, the modern theory of magnetization is still in its infancy. Key developments are in progress [23–27], and first-principle calculations are just starting to appear [28, 29].

Macroscopic polarization may only occur in the absence of inversion symmetry, while macroscopic magnetization requires the absence of time-reversal symmetry. Another key difference is that polarization (as a bulk material property) only makes sense in insulating materials, while macroscopic magnetization exists both in insulators and metals. A material is insulating, in principle, only at $T = 0$, hence the modern theory of polarization is intrinsically a $T = 0$ theory. At variance with this, the modern theory of magnetization can be extended to $T \neq 0$ [24, 25].

Macroscopic polarization is the sum of two contributions: electronic and nuclear. Only the first term requires quantum-mechanical treatment, but it is mandatory to consider the two terms together, since overall charge neutrality is essential.

Macroscopic magnetization is a purely electronic phenomenon, but it is the sum of two contributions as well: spin magnetization and orbital magnetization. The latter occurs whenever time-reversal symmetry is broken in the spatial wavefunction. For instance, in a ferromagnet the spin–orbit interaction transmits the symmetry breaking from the spin degrees of freedom to the spatial (orbital) ones; the two contributions to the total magnetization can be resolved experimentally. Other examples include systems in *applied* magnetic fields. Whenever the unperturbed system is nonmagnetic and insulating, the induced magnetization is 100% of the orbital kind.

The modern theory of magnetization also allows the computation of NMR shielding tensors in condensed matter [28], in an alternative way with respect to the linear-response approach (in the long-wavelength limit) currently used for more than a decade by Mauri *et al* [30].

2. Macroscopics

2.1. Fundamentals

The basic microscopic quantities inside a material are the local microscopic fields $\mathbf{E}^{(\mathrm{micro})}(\mathbf{r})$ and $\mathbf{B}^{(\mathrm{micro})}(\mathbf{r})$, which fluctuate at the atomic scale. By definition, the macroscopic fields **E** and **B** are obtained by averaging them over a macroscopic length scale [2]. In a macroscopically homogeneous system the macroscopic fields **E** and **B** are constant, and in crystalline materials they coincide with the cell average of $\mathbf{E}^{(\mathrm{micro})}(\mathbf{r})$ and $\mathbf{B}^{(\mathrm{micro})}(\mathbf{r})$.

The constituent equations of electrostatics and magnetostatics in continuous media are, to linear order in the fields [1]

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}; \qquad \mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}, \tag{1}$$

where **P** and **M** are the macroscopic polarization and magnetization, respectively. All macroscopic quantities entering equation (1) may have a spatial dependence only in inhomogeneous regions, where a net electrical charge density

$\rho(\mathbf{r})$ and/or a dissipationless current density $\mathbf{j}(\mathbf{r})$ pile up according to

$$\nabla \cdot \mathbf{P}(\mathbf{r}) = -\rho(\mathbf{r}); \qquad \nabla \times \mathbf{M}(\mathbf{r}) = \frac{1}{c}\mathbf{j}(\mathbf{r}). \tag{2}$$

At an interface between two different homogeneous media $\mathbf{P}$ and $\mathbf{M}$ are, in general, discontinuous.

In the simple case of the surface of an homogeneously polarized and/or magnetized medium, $\mathbf{P}$ and $\mathbf{M}$ vanish on the vacuum side. Equations (2) implies the occurrence of a surface charge and/or a surface current

$$\sigma_{\text{surface}} = \mathbf{P} \cdot \mathbf{n}, \qquad \mathbf{K}_{\text{surface}} = c\,\mathbf{M} \times \mathbf{n}, \tag{3}$$

where $\mathbf{n}$ is the normal to the surface. Notice that $\mathbf{M}$ is a well defined quantity for either insulating or metallic materials; instead $\mathbf{P}$ is a nontrivial, material dependent, property only in insulating materials. In the metallic case σ_{surface} completely screens any electrical perturbation (Faraday-cage effect), hence $\mathbf{P}$ is trivial and universal.

We transform equations (1) using the dielectric and magnetic permeability tensors

$$\overleftrightarrow{\varepsilon} = \frac{\partial \mathbf{D}}{\partial \mathbf{E}}; \qquad \overleftrightarrow{\mu} = \frac{\partial \mathbf{B}}{\partial \mathbf{H}}, \tag{4}$$

$$\mathbf{P} = \mathbf{P}_0 + \frac{\overleftrightarrow{\varepsilon} - 1}{4\pi}\mathbf{E}; \qquad \mathbf{M} = \mathbf{M}_0 + \frac{\overleftrightarrow{\mu} - 1}{4\pi}\mathbf{H}. \tag{5}$$

Because of symmetry reasons, the polarization $\mathbf{P}_0$ in a null $\mathbf{E}$ field can be nonzero only if the unperturbed medium breaks inversion symmetry; analogously, the magnetization $\mathbf{M}_0$ in a null $\mathbf{B}$ field can be nonzero only if the unperturbed medium breaks time-reversal symmetry. For the sake of clarity, 'unperturbed' in the previous sentence includes cases where the solid has indeed a built-in perturbation *other* than a field (e.g. macroscopic strain, frozen phonon, and the like).

The modern theory of polarization, at least in its original form, only addresses $\mathbf{P}_0$, the polarization *in a null field*, also known (for reasons explained below) as the 'transverse' polarization. Quite analogously the modern theory of orbital magnetization, at the present stage of development, only addresses $\mathbf{M}_0$, the magnetization in a null field.

In condensed matter theory one addresses bulk quantities, with no reference to real finite samples with boundaries. The microscopic fields $\mathbf{E}^{(\text{micro})}(\mathbf{r})$ and $\mathbf{B}^{(\text{micro})}(\mathbf{r})$ are ideally measurable inside the material, with no reference to what happens outside a finite sample. Their macroscopic averages $\mathbf{E}$ and $\mathbf{B}$, i.e. the internal (or screened) macroscopic fields, are therefore the variables of choice for a first-principle description. It must be realized that, insofar as we address an infinite system with no boundaries, the macroscopic field (either $\mathbf{E}$ or $\mathbf{B}$) is just an arbitrary boundary condition. To realize this, it is enough to focus on the electrical case for a crystalline material. The microscopic charge density is neutral on average and lattice periodical; the value of $\mathbf{E}$ is just an arbitrary boundary condition for the integration of Poisson's equation. The usual choice (performed within all electronic structure codes) is to impose a lattice periodical Coulomb potential, i.e. $\mathbf{E} = 0$. Imposing a given nonzero value of $\mathbf{E}$ is equally legitimate (in insulators), although technically more difficult [31, 32].

In order to study the bulk material properties of a macroscopically homogeneous system it is quite convenient to address the infinite system with no boundaries. The above formulation of electrostatics and magnetostatics is sufficient and ideally suited for electronic structure theory: there is no need to address external (or unscreened) fields, as there is no need to address the auxiliary and unphysical fields $\mathbf{D}$ and $\mathbf{H}$.

For instance, the dielectric tensor $\overleftrightarrow{\varepsilon}$ defined by equation (4) is best addressed within electronic structure theory as

$$\overleftrightarrow{\varepsilon} = 1 + 4\pi \frac{\partial \mathbf{P}}{\partial \mathbf{E}}, \tag{6}$$

where only 'internal' quantities (well defined in the bulk of the sample), and no 'external' ones, appear. Obviously, for a homogeneous material, $\overleftrightarrow{\varepsilon}$ is a bulk material property, independent of the sample shape.

There are actually *two* different dielectric tensors: the genuinely static one, called $\overleftrightarrow{\varepsilon}_0$, and the so-called 'static high frequency', called $\overleftrightarrow{\varepsilon}_\infty$. The latter accounts for the electronic polarization only, and is also called the 'clamped-ion' dielectric tensor. Both are experimentally measurable [33].

2.2. Finite samples and shape issues

Even if there is no need to address finite samples and external versus internal fields from a theoretician's viewpoint, such a digression can be quite instructive given that experiments *are* performed over finite samples, often in external fields.

We start with the electrical case. Suppose a finite macroscopic sample is inserted in a constant external field $\mathbf{E}^{(\text{ext})}$: the microscopic field $\mathbf{E}^{(\text{micro})}(\mathbf{r})$ coincides with $\mathbf{E}^{(\text{ext})}$ far away from the sample, while it is different inside because of screening effects. If we choose an homogeneous sample of *ellipsoidal shape*, then the macroscopic average of $\mathbf{E}^{(\text{micro})}(\mathbf{r})$, i.e. the macroscopic screened field $\mathbf{E}$, is constant in the bulk of the sample.

The shape effects are embedded in the depolarization coefficients n_α, defined in [1], with $\sum_\alpha n_\alpha = 1$; Greek subscripts indicate Cartesian coordinates throughout. Special cases are the sphere ($n_x = n_y = n_z = 1/3$), the extremely prolate ellipsoid, i.e. a cylinder along z ($n_x = n_y = 1/2, n_z = 0$), and the extremely oblate one, i.e. a slab normal to z ($n_x = n_y = 0, n_z = 1$).

The main relationship between $\mathbf{E}$, $\mathbf{E}^{(\text{ext})}$, and $\mathbf{P}$ is [1]:

$$E_\alpha = E_\alpha^{(\text{ext})} - 4\pi n_\alpha P_\alpha. \tag{7}$$

When we consider a free-standing finite system, with *no* external field, equation (7) provides, by definition, the depolarizing field. In the simple case of a slab geometry the depolarizing field is $\mathbf{E} = -4\pi\mathbf{P}$ when $\mathbf{P}$ is normal to the slab, and $\mathbf{E} = 0$ when $\mathbf{P}$ is parallel to the slab: this is sketched in figure 1.

Quite generally, a vector field is called 'longitudinal' when it is curl-free, and 'transverse' when it is divergence-free: we analyse $\mathbf{P}(\mathbf{r})$ and $\mathbf{M}(\mathbf{r})$ as functions of a macroscopic

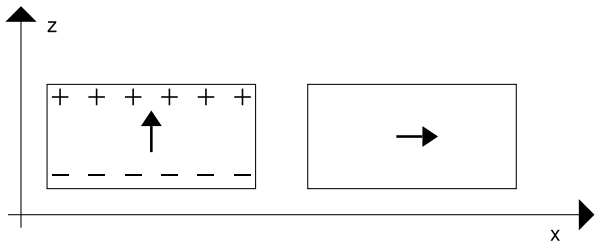

Figure 1. Electrical macroscopic polarization **P** in a slab normal to z, for a vanishing external field $\mathbf{E}^{(\mathrm{ext})}$. Left: when **P** is normal to the slab, a depolarizing field $\mathbf{E} = -4\pi\mathbf{P}$ is present inside the slab, and surface charges form, with areal density $\sigma_{\mathrm{surface}} = \mathbf{P}\cdot\mathbf{n}$. Right: when **P** is parallel to the slab, no depolarizing field and no surface charge is present.

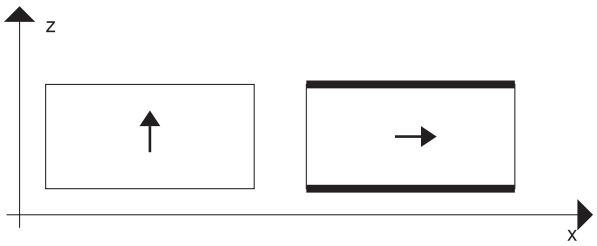

Figure 2. Macroscopic magnetization **M** in a slab normal to z, for a vanishing external field $\mathbf{B}^{(\mathrm{ext})}$. Left: when **M** is normal to the slab, no depolarizing field and no surface current is present. Right: when **M** is parallel to the slab, a demagnetizing field $\mathbf{B} = 4\pi\mathbf{M}$ is present inside the slab, and dissipationless currents $\mathbf{K}_{\mathrm{surface}} = c\,\mathbf{M}\times\mathbf{n}$ flow at the surfaces.

coordinate across the slab in this respect. The external fields are set to zero.

When $\mathbf{P}(\mathbf{r})$ it is normal to the slab we have $P_z = P_z(z)$ (independent of xy): hence at the slab boundary $\nabla\cdot\mathbf{P} \neq 0$, $\nabla\times\mathbf{P} = 0$: the normal polarization is longitudinal. When $\mathbf{P}(\mathbf{r})$ is parallel to the slab we have $P_x = P_x(z)$ (independent of xy): hence at the boundary $\nabla\cdot\mathbf{P} = 0$, $\nabla\times\mathbf{P} \neq 0$: the parallel polarization is transverse (see figure 1). Looking at equation (5), it is clear that the transverse polarization coincides with $\mathbf{P}_0$, while (for isotropic permittivity) the longitudinal one is $\mathbf{P} = \mathbf{P}_0/\varepsilon$. These are the extreme values; for an arbitrary ellipsoidal shape **P** will be intermediate between them.

A subtle issue is: which ε is to be used? ε_0 or ε_∞? The answer is always ε_0, with the only notable exception of lattice dynamics. If $\mathbf{P}_0$ is the polarization of a given 'frozen phonon' at zero field (i.e. transverse), the corresponding longitudinal polarization (for the same displacements pattern) is $\mathbf{P} = \mathbf{P}_0/\varepsilon_\infty$. This follows immediately e.g. from Huang's phenomenological theory [34, 35].

Next, we switch to the magnetic case. Again, by definition, the magnetization normal to the slab is longitudinal and the parallel one is transverse. According to [1], one has to replace **P** with **M**, **E** with **H**, and $\mathbf{E}^{(\mathrm{ext})}$ with $\mathbf{H}^{(\mathrm{ext})} = \mathbf{B}^{(\mathrm{ext})}$. The analogue of equation (7) is then

$$H_\alpha = B_\alpha^{(\mathrm{ext})} - 4\pi n_\alpha M_\alpha. \tag{8}$$

We eliminate **H** by means of equations (1); then for a uniformly magnetized ellipsoid in zero external field $\mathbf{B}^{(\mathrm{ext})}$ the demagnetizing field is

$$B_\alpha = 4\pi(1-n_\alpha)M_\alpha. \tag{9}$$

We consider once more the slab geometry, in which case $\mathbf{B} = 4\pi\mathbf{M}$ when **M** is parallel to the slab, and $\mathbf{B} = 0$ when **M** is normal to the slab: this is sketched in figure 2. In the case of isotropic permeability equations (4) and (5) lead to

$$\mathbf{M} = \mathbf{M}_0 + \frac{\mu-1}{4\pi\mu}\mathbf{B}. \tag{10}$$

It follows immediately that $\mathbf{M} = \mathbf{M}_0$ when the magnetization is normal to the slab (longitudinal), while it is easily verified that $\mathbf{M} = \mu\mathbf{M}_0$ when the magnetization is parallel (transverse). In analogy to the electrical case, these are the extreme values; for an arbitrary ellipsoidal shape **M** will be intermediate between them.

It is customary to write $\mu = 1 + 4\pi\chi$, where χ is the magnetic susceptibility. This can be positive or negative, but is fairly small, with the notable exception of ferromagnetic materials in a neighbourhood of the phase transition [1]. In most cases we can expand equation (10) as

$$M_\alpha \simeq M_{0,\alpha} + \chi B_\alpha = M_{0,\alpha} + 4\pi\chi(1-n_\alpha)M_\alpha. \tag{11}$$

For a spherical sample ($n_\alpha = 1/3$) the leading-order shape correction is

$$\mathbf{M} \simeq \left(1 + \frac{8\pi}{3}\chi\right)\mathbf{M}_0. \tag{12}$$

Finally, we summarize the slab results and we emphasize the key difference when the slab is in zero external fields. In the electrical case the *transverse* (i.e. parallel to the slab) polarization $\mathbf{P} = \mathbf{P}_0$ occurs in zero **E** (internal) field, while in the magnetic case it is the *longitudinal* (i.e. normal to the slab) magnetization $\mathbf{M} = \mathbf{M}_0$ which occurs in zero **B** (internal) field. This is confirmed by the absence of surface charges and surface currents in these geometries (see figures 1 and 2).

3. Microscopics

Intuitively, the macroscopic polarization **P** and magnetization **M** should be intensive vector properties carrying the meaning of electric/magnetic dipole per unit volume, but their *definition* in terms of microscopic quantities was an unsolved problem until the 1990s.

At a very elementary level, **P** is addressed by means of the time-honoured Clausius–Mossotti model [36], where the charge distribution of a polarized dielectric is regarded as the superposition of localized contributions, each providing an electric dipole. The model applies only to the extreme case of molecular crystals, where the polarizable units can be unambiguously identified; for any other material—including the alkali halides—such decomposition is severely nonunique [10, 37]. Incidentally, we anticipate that the modern theory indeed recovers the Clausius–Mossotti limit whenever it applies (see the end of section 6.5).

In a general crystalline insulator the electron distribution is periodic and delocalized over the whole unit cell, most notably in covalent materials. The popular textbooks typically attempt a microscopic definition of **P** in terms of the dipole moment per cell [33, 38], but such approaches are deeply flawed because there is no unique choice for the cell boundaries [39, 10]. Only a few undergraduate textbooks are free from such flaws (e.g. Marder [22]).

Macroscopic magnetization **M** is the sum of two terms, which are unambiguously defined in nonrelativistic (and semirelativistic) quantum mechanics: spin magnetization $\mathbf{M}^{(\mathrm{spin})}$ and orbital magnetization $\mathbf{M}^{(\mathrm{orb})}$. Experimentally, magneto-mechanical measurements, based on the Einstein–de Haas effect, provide the two terms separately. For instance, the values of $\mathbf{M}^{(\mathrm{spin})}$ and $\mathbf{M}^{(\mathrm{orb})}$ for the three ferromagnetic metals (Fe, Co, and Ni) have been accurately known for half a century [40].

From the viewpoint of the present review, $\mathbf{M}^{(\mathrm{spin})}$ is a dull quantity. Electronic structure codes routinely compute the spin density, which is a simple lattice periodical function. Its cell average (times a trivial factor) coincides with $\mathbf{M}^{(\mathrm{spin})}$. In other words, a 'dipolar density' is unambiguously identified, while the same does not happen in the orbital case.

For a finite sample, any surface effect contributes *nonextensively* to the total spin, and therefore cannot affect $\mathbf{M}^{(\mathrm{spin})}$ in the thermodynamic limit. The opposite happens for $\mathbf{M}^{(\mathrm{orb})}$, which suffers from the same problems as **P**. In the following, we will not address $\mathbf{M}^{(\mathrm{spin})}$ further, and we will use the symbol **M** to indicate $\mathbf{M}^{(\mathrm{orb})}$ only.

Given the intuitive meaning of **P** and **M**, it is tempting to define them as the dipole moment of a sample, divided by the volume V:

$$\begin{aligned} \mathbf{P} &= \frac{\mathbf{d}}{V} = \frac{1}{V}\int \mathrm{d}\mathbf{r}\,\mathbf{r}\rho^{(\mathrm{micro})}(\mathbf{r}), \\ \mathbf{M} &= \frac{\mathbf{m}}{V} = \frac{1}{2cV}\int \mathrm{d}\mathbf{r}\,\mathbf{r}\times\mathbf{j}^{(\mathrm{micro})}(\mathbf{r}), \end{aligned} \tag{13}$$

where $\rho^{(\mathrm{micro})}(\mathbf{r})$ and $\mathbf{j}^{(\mathrm{micro})}(\mathbf{r})$ are the microscopic charge and (orbital) current densities. Notice that there is no such thing as a 'dipolar density': the basic microscopic quantities are $\rho^{(\mathrm{micro})}(\mathbf{r})$ and $\mathbf{j}^{(\mathrm{micro})}(\mathbf{r})$. If the sample is *uniformly* polarized/magnetized, then the microscopic charge/current averages to zero in the bulk of the sample, while at the sample boundary a net charge piles up and/or a dissipationless current flows, in agreement with the macroscopic equation (3). Phenomenologically **P** and **M** are *bulk* material properties, while from the above considerations they apparently are *surface* properties. One may wonder, for instance, whether altering the surface (and only the surface) may result in a change of **P** and/or of **M**. This very fundamental problem was unsolved until the early 1990s for the electrical case, and until 2005 for the magnetic case.

Condensed matter theory universally adopts *periodic* (a.k.a. Born–von Kármán) boundary conditions (PBCs); in the special case of *crystalline* materials, PBCs lead to the Bloch theorem. One of the virtues of PBCs is that the system by construction has no surface. Therefore whatever one defines or computes within PBCs is by definition 'bulk': any surface effect is ruled out. But PBCs *do not* solve our problem, since the (unbounded) position operator **r** entering equations (13) is a 'forbidden' operator, *incompatible* with PBCs. The issue is then how to define and compute **P** and **M** within PBCs by means of formulae quite different from equations (13); therein, in fact, $\rho^{(\mathrm{micro})}(\mathbf{r})$ and $\mathbf{j}^{(\mathrm{micro})}(\mathbf{r})$ are assumed to vanish exponentially outside the finite sample. In the crystalline case, the basic ingredient of such formulae must be the Bloch orbitals of the occupied bands, while the forbidden **r** operator must *not* appear.

One important tenet of the modern theory is worth stressing: the macroscopic polarization (magnetization) of a uniformly polarized (magnetized) crystal has *nothing to do* with the lattice periodical charge (current) distribution—despite contrary statements in several textbooks, Actually, this tenet stems already from classical physics, as emphasized e.g. in the reference work of Hirst [41].

4. DFT, pseudopotentials, and more

The present work reviews the formulations which provide macroscopic electrical polarization and orbital magnetization in condensed matter in terms of single-particle orbitals, which assume the Bloch form in the crystalline case. The formulae are exact for noninteracting electrons, but the obvious aim is to implement them with Kohn–Sham (KS) orbitals, in a given DFT first-principle framework.

Since in this work we need to distinguish between insulators and metals, we stress that we mean 'KS insulator' and 'KS metal' throughout: that is, we discriminate whether the KS spectrum is gapped or gapless. In the class of 'simple' (i.e. computationally friendly) materials a genuine insulator (metal) is also a KS insulator (metal), although pathological cases (computationally unfriendly) do exist.

Having specified this, the key issue is then: does the KS polarization/magnetization coincide with the physical many-body one? The answer is subtle, and is different whether one chooses either 'open' boundary conditions (OBC), as appropriate for molecules and clusters, or PBCs (Born–von Kármán), as appropriate for condensed systems—either crystalline or disordered.

Within OBC the KS orbitals vanish at infinity. For a system of N electrons with $N/2$ doubly occupied orbitals $\varphi_j(\mathbf{r})$ the dipoles (electrical and magnetical) of the fictitious noninteracting KS system are then, in agreement with equations (13)

$$\begin{aligned} \mathbf{d} &= \mathbf{d}_{\mathrm{nuclear}} - 2e\sum_{j=1}^{N/2}\langle\varphi_j|\mathbf{r}|\varphi_j\rangle, \\ \mathbf{m} &= -\frac{e}{2c}\sum_{j=1}^{N/2}\langle\varphi_j|\mathbf{r}\times\mathbf{v}|\varphi_j\rangle, \end{aligned} \tag{14}$$

where $\mathbf{v} = \mathrm{i}[H, \mathbf{r}]$, and H is the KS Hamiltonian. Atomic Hartree units ($e = \hbar = m_e = 1$) are adopted in most of the following ($c \simeq 137$). The basic tenet of DFT is that the microscopic density of the fictitious noninteracting KS system coincides with the density of the interacting system: hence

equations (14) provides the exact many-body **d** for molecules and clusters. However, when considering a large system in the thermodynamic limit, the density in the surface region contributes *extensively* to the dipole. The magnetic case is different: the microscopic current in the noninteracting KS system *needs not* to be equal to the one in the interacting physical one. The drawback is in principle cured by the Vignale–Rasolt current DFT [42], although a simple, universal, and reliable functional to be applied in actual computations has still to appear [43–45].

The modern theory, as formulated below, provides formulae for **P** and **M** which are exact within PBCs for noninteracting electrons. However, within PBCs the macroscopic polarization **P** is *not* a function of the microscopic density, hence the value of **P** obtained from the KS orbitals, in general, is not the correct many-body **P**. This was first shown in 1995 by Gonze *et al* [46], and later discussed by several authors. A complete account of the issue can be found in [10]. Needless to say, the situation for **M** is no better.

Therefore, neither **P** nor **M** can be exactly expressed—even in principle—within standard DFT; but the exact DFT functional is obviously inaccessible, and even sometimes pathological. The practical issue is whether the current popular flavours of DFT provide an accurate approximation to the experimental values of **P** and **M** in a large class of materials.

In the electrical case a vast first-principle literature accumulated over the years—by either linear-response theory or the modern theory—typically shows errors of the order of 10–20% on permittivity, and much less on most other properties (infrared spectra, piezoelectricity, ferroelectricity) for many different materials. It is unclear which part of the error is to be attributed to DFT *per se*, and which part is to be attributed to the *approximations* to DFT. The above mentioned error refers to 3d systems (crystalline, amorphous, and liquid); the state of the art is much worse for quasi-1d systems (polymers), where the polarizabilities and hyperpolarizabilities can be off by orders of magnitude [47]. For such case studies the drawback is shown by computations within OBC, where DFT is in principle exact: hence the culprit is in the *approximate* functional.

In the magnetic case the experience is much more limited, and accumulated only by linear-response theory in the work of Mauri *et al* [48, 30, 49–51]. For the case studies addressed so far the error seems fairly small.

Next, we switch to discussing an issue related to the use of pseudopotentials, where a key difference between the electrical and magnetic case exists. In the former case, the pseudo-wavefunctions contain all of the information (to a very good approximation), and the formalism can be applied as it stands; in fact, it is implemented as such within the pseudopotential codes [13, 14, 16–18]. Quite to the contrary, in the latter case the orbital currents associated with the pseudo-wavefunctions miss very important physical contributions. While all-electron implementations have not yet appeared, the state-of-the art calculations [28, 29] combine the pseudopotential approach with a Blöchl-like PAW reconstruction for all elements beyond the first row, much in the same way as first shown by Pickard and Mauri in the framework of linear-response theory [49, 51].

5. Linear response

As stated in the very first paragraph of this work, we are mostly addressing the modern theory of polarization **P** and the modern theory of (orbital) magnetization **M**. Before the development of the modern theories, *derivatives* of **P** and **M** were accessible at the first-principle level via linear-response theory. Some (though not all) experimental observables related to **P** and **M** are by definition derivatives with respect to suitable perturbations. Several observables in this class have been computed over the years for many materials by means of specialized codes (see below).

The spontaneous polarization of a ferroelectric material and, analogously, the spontaneous orbital magnetization of a ferromagnetic material cannot be accessed via linear-response theory. Therefore such observables were ill-defined from an electronic structure viewpoint until the advent of the modern theories. Actually, the first computation ever of the spontaneous polarization of a ferroelectric was published in 1993 [52]. As for orbital magnetization, all computations on the market rely on the uncontrolled muffin-tin approximation; the first implementation of the modern theory (where such approximation is not needed) is appearing nowadays [29].

Even in the cases where the physical observable is by definition a derivative, it often proves convenient to evaluate such a derivative as a finite difference by means of the modern theory. This does not require a specialized code, in that it only needs a couple of ground-state calculations. The approach is particularly appealing when studying complex materials and/or using complex forms of exchange–correlation functionals. For instance, infrared spectra of liquids are routinely accessed via the modern theory of polarization [53–55].

5.1. Linear-response tensors

We indicate as $F(\mathbf{E}, \mathbf{H}, \lambda)$ the free energy per unit volume, where λ is the short-hand scalar notation for a macroscopic perturbation which is actually tensorial as well (e.g. zone-centre phonon, macroscopic strain, &C). We define λ such that $\lambda = 0$ is the equilibrium unperturbed value. We *exclude* from F the free energy of the free fields, which exists even in the absence of the material. At any λ value the polarization and the magnetization are the derivatives

$$\mathbf{P} = -\frac{\mathrm{d}F}{\partial \mathbf{E}}, \qquad \mathbf{M} = -\frac{\mathrm{d}F}{\partial \mathbf{H}}, \tag{15}$$

evaluated at $\mathbf{E} = 0$ and $\mathbf{H} = 0$. In this work we tacitly refer to the orbital term only in **M**.

The linear-response tensors are second derivatives of F. In particular $\partial\mathbf{P}/\partial\mathbf{E} = -\partial^2 F/\partial\mathbf{E}\partial\mathbf{E}$ is the electrical susceptibility and $\partial\mathbf{M}/\partial\mathbf{H} = -\partial^2 F/\partial\mathbf{H}\partial\mathbf{H}$ is the magnetic susceptibility. The common symbol $\overleftrightarrow{\chi}$ is customarily used for both tensors. It must be emphasized that the electrical $\overleftrightarrow{\chi}$ is definite positive and of the order one, while the magnetic $\overleftrightarrow{\chi}$ can have either sign and is fairly small, of the order $(1/137)^2$, except near a ferromagnetic transition or in superconducting materials. For this reason **H** can be safely replaced with **B** in many circumstances.

The evaluation of susceptibilities has been performed for several years by means of specialized linear-response codes, and is beyond the reach of the modern theories of polarization and magnetization, at least in their original version. An extension of the theory [31, 32], not discussed in this work, removes such a limitation in the electrical case. The magnetic case is universally dealt with by the long-wavelength linear-response approach of Mauri *et al* [30].

The mixed derivative $\overleftrightarrow{\alpha} = -\partial^2 F/\partial\mathbf{E}\partial\mathbf{H} = \partial\mathbf{M}/\partial\mathbf{E} = \partial\mathbf{P}/\partial\mathbf{H}$, named the magnetoelectric polarizability, is much in fashion nowadays given the current high interest in multiferroics [56]. It has been discovered very recently (2009) that the orbital magnetoelectric polarizability has some very nontrivial topological features [57]; preprints on this topic are appearing at the time of writing [58, 59].

The remaining linear-response tensors are the mixed second derivatives $\partial\mathbf{P}/\partial\lambda = -\partial^2 F/\partial\lambda\partial\mathbf{E}$ and $\partial\mathbf{M}/\partial\lambda = -\partial^2 F/\partial\lambda\partial\mathbf{H}$, evaluated at equilibrium ($\lambda = 0$). Specialized linear-response codes [16, 13] allow the computation of some of these tensors from first principles; by exploiting the symmetry of the mixed derivatives (Schwarz's theorem) there are usually two different paths, in principle equivalent but computationally very different in their implementation.

The use of a specialized code can be avoided (as said above) by evaluating the **P** and **M** derivatives as finite differences by means of the modern theories of polarization and magnetization.

5.2. Electrical case: pyroelectricity, piezoelectricity, and IR charges

The pyroelectric coefficient is defined as

$$\Pi_\alpha = \frac{\mathrm{d}P_\alpha}{\mathrm{d}T}, \tag{16}$$

the piezoelectric tensor as [60]

$$\gamma_{\alpha\beta\delta} = \frac{\partial P_\alpha}{\partial \epsilon_{\beta\delta}}, \tag{17}$$

and the dimensionless Born (or 'dynamical' or 'infrared') charge as

$$Z^*_{\mathrm{s},\alpha\beta} = \frac{V_\mathrm{c}}{e}\frac{\partial P_\alpha}{\partial u_{\mathrm{s},\beta}}, \tag{18}$$

that is as derivatives of **P** with respect to temperature T, strain $\epsilon_{\beta\delta}$, and displacement $\mathbf{u}_\mathrm{s}$ of sublattice s, respectively, where V_c is the primitive cell volume. In the above formulae, derivatives are to be taken at zero electric field and zero strain when these variables are not explicitly involved.

By interpreting the second mixed derivatives of F the other way around, we can define Π via the specific heat change linearly induced by a field at constant temperature, γ via the macroscopic stress linearly induced by a field at zero strain, and Z^* via the forces linearly induced by a field at the equilibrium geometry. This can be exploited in practice in linear-response calculations.

To the best of author's knowledge, pyroelectricity has never been investigated at the first-principle level in any material, although it is possibly within reach of finite-T Car–Parrinello simulations [61]. The other three tensor properties have been extensively studied in the literature, for many classes of materials, via linear-response theory. The first DFT computation ever of permittivity (for Si) appeared in 1986 [62] and of piezoelectric tensors (for the III–Vs) in 1989 [63]. Nowadays, most state-of-the-art linear-response calculations are based on the so-called 'density functional perturbation theory', as described e.g. in the comprehensive [64, 65], and implemented in the public-domain codes QUANTUM-ESPRESSO [16] and ABINIT [13].

Linear-response methods, also called—in quantum-chemistry jargon—'analytical derivative' methods, are not the unique tool to compute some of the above derivative properties: numerical differentiation in conjunction with the modern theory can be used as well. Since piezoelectric and infrared tensors are by definition zero-field properties, first-principle studies have widely and successfully used the modern theory within finite-difference schemes, particularly for complex materials, complex basis sets, and nonstandard functionals [66, 67].

Implementations of the modern theory have been instrumental in the study e.g. of piezoelectric and infrared properties of ferroelectric perovskites [21], as well as of the infrared spectra of liquid and amorphous materials [54, 55].

5.3. A closer look at IR charges (Born effective charge tensors)

The Born (or IR) effective charge tensor, equation (18), can equivalently be defined (as already observed) as the force $\mathbf{f}_\mathrm{s}$ linearly induced on a given nucleus s by a *macroscopic* **E** field of unit magnitude. This force can obviously be expressed as the *microscopic* $\mathbf{E}_\mathrm{s}$ field at site s, times the bare nuclear charge eZ_s:

$$f_{\mathrm{s},\alpha} = eZ^*_{\mathrm{s},\beta\alpha}E_\beta = eZ_\mathrm{s}E_{\mathrm{s},\alpha}. \tag{19}$$

Notice that the Cartesian tensor $\overleftrightarrow{Z}^*_\mathrm{s}$ is in general nonsymmetric. It follows that the local microscopic field at site s is related to the macroscopic one as

$$\frac{\partial E_{\mathrm{s},\alpha}}{\partial E_\beta} = Z^*_{\mathrm{s},\beta\alpha}/Z_\mathrm{s}. \tag{20}$$

In order to proceed further, we adopt in the following of this section an all-electron view (no pseudopotentials): therefore the perturbation induced by the displacement of nucleus s and its periodic replicas by an infinitesimal amount $\mathbf{u}_\mathrm{s}$ is identical to introducing in the unperturbed crystal an extra point dipole of magnitude $\mathbf{d}_\mathrm{s} = eZ_\mathrm{s}\mathbf{u}_\mathrm{s}$ (and its periodic replicas), where Z_s is the bare nuclear charge. The original definition of equation (18) can thus be recast as

$$Z^*_{\mathrm{s},\alpha\beta}/Z_\mathrm{s} = V_\mathrm{c}\frac{\partial P_\alpha}{\partial d_{\mathrm{s},\beta}}. \tag{21}$$

The above manipulations are useful to show that the NMR shielding tensor $\overleftrightarrow{\sigma}_\mathrm{s}$, introduced next, is the perfect magnetic analogue of $\overleftrightarrow{Z}^*_\mathrm{s}/Z_\mathrm{s}$.

5.4. Magnetic case: NMR shielding tensor

In nonmagnetic materials, the magnetic susceptibility is of purely orbital nature. Since the pioneering work of Mauri and Louie [30], this property has been successfully computed in many materials via linear response in the long-wavelength limit. Other properties, like the EPR g tensor for paramagnetic defects in solids, are also computed by suitable linear-response techniques which generalize the Mauri *et al* approach [68].

NMR spectroscopy [69] has been recognized since 1938 [70] to be a powerful experimental probe of local chemical environments, including structural and functional information on molecules, liquids, and increasingly, on solid-state systems.

The NMR nuclear shielding tensor $\overleftrightarrow{\sigma}_s$ by definition linearly relates the local microscopic magnetic field at a given nucleus $\mathbf{B}_s$ to the external macroscopic field $\mathbf{B}^{(\mathrm{ext})}$ applied to the finite sample:

$$\sigma_{s,\alpha\beta} = \delta_{\alpha\beta} - \frac{\partial B_{s,\alpha}}{\partial B^{(\mathrm{ext})}_{\beta}}. \tag{22}$$

It obviously depends on the sample shape (see section 2.2); it is expedient to start with a sample in the form of a slab, with $\mathbf{B}^{(\mathrm{ext})}$ normal to the slab, as in the left sketch of figure 2. For other shapes a correction is easily computed as a simple function of $\overleftrightarrow{\chi}$. The chosen shape has the virtue that the macroscopic screened field $\mathbf{B}$ inside the sample is equal to the external field $\mathbf{B}^{(\mathrm{ext})}$, hence

$$\sigma_{s,\alpha\beta} = \delta_{\alpha\beta} - \frac{\partial B_{s,\alpha}}{\partial B_{\beta}}, \tag{23}$$

whose electrical analogue is equation (20) with the obvious identification

$$Z^{*}_{s,\beta\alpha}/Z_s \longleftrightarrow \delta_{\alpha\beta} - \sigma_{s,\alpha\beta}. \tag{24}$$

The linear-response approach of Mauri *et al* [48]—called in the following the 'direct' approach–exploits equation (23) by computing the microscopic orbital currents linearly induced by a long-wavelength $\mathbf{B}$ field. Many improvements and applications have appeared in the literature for more than a decade since from the original paper [49, 51]. An alternative approach, based on Wannier functions in a supercell, was also proposed in 2001 by Sebastiani and Parrinello [71].

Very recently it has been demonstrated how to compute NMR shielding tensor $\overleftrightarrow{\sigma}_s$ via a 'converse' approach, by exploiting Schwarz's theorem and the modern theory of magnetization [28]. The logic can be easily explained having in mind the electrical analogue, section 5.3, and Schwarz's theorem. Equations (21) and (24) immediately yield

$$\delta_{\alpha\beta} - \sigma_{s,\alpha\beta} = V_c \frac{\partial M_{\beta}}{\partial m_{\alpha}}, \tag{25}$$

where it is understood that the derivative is taken at zero $\mathbf{B}$ field. In order to implement equation (25) in a first-principle calculation one applies an infinite array of point-like magnetic dipoles $\mathbf{m}_s$ to all equivalent sites and calculates the change in macroscopic orbital magnetization $\mathbf{M}$ by means of the modern theory. The vector potential corresponding to such perturbation is lattice periodical (since $\mathbf{B}$ is zero), and is easily inserted in the crystalline kinetic energy.

The very first test cases studied by this converse approach were some representative molecules in a supercell, crystalline diamond, and liquid water [28]. The induced $\mathbf{M}$ proves to be linear and stable over nine orders of magnitude, where $\mathbf{m}_s$ varies between 10^{-6} and 10^{3} Bohr magnetons. The results compare very favourably with previous results from the direct approach for the same systems [48, 72, 73].

In the converse approach one needs to perform three calculations for each site, but convergence of the perturbed Hamiltonian (starting from the unperturbed one) is quite fast and one can deal with a cell with hundreds of atoms. The main advantage, however, is that the converse method avoids a linear-response implementation (requiring substantial extra coding) and, furthermore, is implementable with any complex form of exchange–correlation functional, including DFT + U.

6. Modern theory of polarization

The modern theory of polarization is at a very mature stage. Several review papers have appeared in the literature. The very first one, [8], is still an highly cited classic (for crystalline systems); the most recent ones are [9, 10]. Here we summarize the basic concepts and the basic formulae, mostly aiming at comparing them with the modern theory of orbital magnetization, discussed below.

Most textbooks [33, 38] provide a flawed definition of $\mathbf{P}$, not implementable in practical computations [39]. A change of paradigm emerged in the early 1990s [3, 4]; the modern theory, based on a Berry phase, was founded by King-Smith and Vanderbilt soon afterwards [5]. At its foundation, the modern theory was limited to a crystalline system in an independent-electron—either KS or Hartree–Fock—framework. Later, the theory was extended to correlated and/or disordered systems [6, 7]. Here we are going to present some of the main formulae in reverse historical order.

The change of paradigm started with realizing that only *differences* of $\mathbf{P}$ are experimentally observable as bulk material properties. This is obvious for the derivative properties listed above; but even the 'spontaneous' $\mathbf{P}$ is not accessible as an equilibrium property.

The first calculation ever of spontaneous polarization was published in 1990 [3]. The case study was BeO: it has the simplest structure where inversion symmetry is absent (i.e. wurtzite), and furthermore its constituents are first-row atoms. The idea was to address the macroscopic polarization of a slab of finite thickness, with faces normal to the c axis, embedding it in an ad hoc medium which (1) has no bulk polarization for symmetry reasons, and (2) does not produce any geometrical or chemical perturbation at the interface. The optimal choice is a fictitious BeO in the zincblende structure. Because of obvious reasons, the system is periodically replicated in a supercell geometry (figure 3, top panel). The self-consistent calculation shows well localized interface charges, of opposite sign and equal magnitudes at the two nonequivalent interfaces (figure 3, bottom panel). The

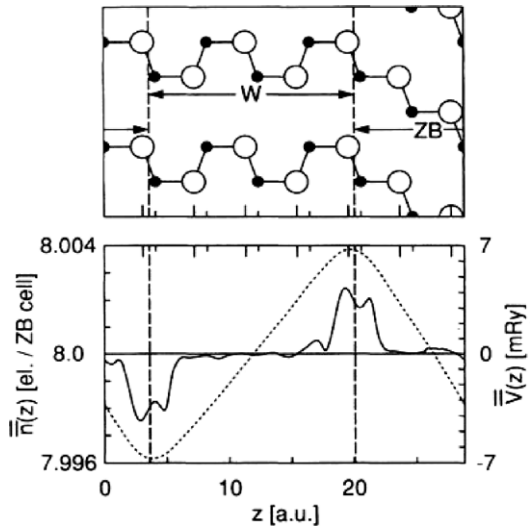

Figure 3. Top panel: the 14-atom BeO supercell in a vertical plane through the BeO bonds; the wurtzite (W) and zincblende (ZB) stackings are perspicuous. Bottom panel: macroscopic averages of the valence electron density (solid) and of the electrostatic potential (dotted).

interface charge is related, as in equation (3), to the difference in polarization between the two materials: $\sigma_{\text{interface}} = \Delta\mathbf{P} \cdot \mathbf{n}$. The computer experiment provides the value of $\sigma_{\text{interface}}$, and since $\mathbf{P}$ vanishes by symmetry in the zincblende slab, one thus obtains the bulk value of $\mathbf{P}$ in the wurtzite material.

It must be emphasized that the quantity really 'measured' in this computer experiment is $\Delta\mathbf{P}$, *not* the polarization $\mathbf{P}$ itself. After [3] was published, a study of the experimental literature showed that—contrary to an incorrect widespread belief—no experimental value of $\mathbf{P}$ in any wurtzite material exists: only estimates are available. Reference [3] marks, as said above, a change of paradigm: polarization must be *defined* by means of *differences*, and the concept of polarization 'itself' must be abandoned. After the modern theory of polarization appeared, the spontaneous polarization of BeO was computed as a Berry phase as well [74]. Not surprisingly, the result agrees quite well with [3], after taking into account that the former approach provides the *longitudinal* polarization (in a depolarizing field), and the latter the *transverse* polarization (in zero field): see the discussion in section 2.2.

At variance with BeO (which is pyroelectric but not ferroelectric), in ferroelectric materials the 'spontaneous' polarization has been measured, and tabulated in the literature, for several decades. Ferroelectrics are insulating solids characterized by a switchable macroscopic polarization $\mathbf{P}$. In such solids the value of $\mathbf{P}$ is generally nonzero at equilibrium, and the application of a large enough electric field switches the value of $\mathbf{P}$ among two (or more) different values. Ferroelectrics therefore undergo spontaneous symmetry breaking, and display a multistable equilibrium state. The most typical ferroelectric crystals are in the class of perovskite oxides: these have been much studied since the 1950s.

In most ferroelectrics the allowed values of $\mathbf{P}$ are equal in modulus and point along equivalent (enantiomorphous) symmetry directions. In a typical experiment the applied field switches the polarization from $\mathbf{P}$ to $-\mathbf{P}$, so that one speaks of polarization *reversal*: the quantity which is directly accessible to experiment is $\Delta P = 2|\mathbf{P}|$, *not* $\mathbf{P}$. An experimental determination of the spontaneous polarization is normally extracted from a measurement of the transient current flowing through the sample during an hysteresis cycle (figure 4).

The modern theory—in agreement with the experiment—avoids addressing the 'absolute' polarization of a given equilibrium state, quite in agreement with the experiments, which invariably measure polarization *differences*. Instead, it addresses differences in polarization between two states of the material that can be connected by an adiabatic switching process. The time-dependent Hamiltonian is assumed to remain insulating at all times, and the polarization difference is then equal to the time-integrated transient macroscopic current that flows through the insulating sample during the switching process:

$$\Delta\mathbf{P} = \mathbf{P}(\Delta t) - \mathbf{P}(0) = \int_0^{\Delta t} \mathrm{d}t\, \mathbf{j}(t). \tag{26}$$

In the adiabatic limit $\Delta t \to \infty$ and $\mathbf{j}(t) \to 0$, while $\Delta\mathbf{P}$ stays finite. Addressing currents (instead of charges) explains the occurrence of *phases* of the wavefunctions (instead of square moduli) in the modern theory. Eventually the time integration in equation (26) will be eliminated, leading to a two-point formula involving only the initial and final states.

6.1. *Single* **k***-point formula for supercell calculations*

For the sake of simplicity we deal with N electrons in a cubic supercell of size L. We choose the boundary condition that the microscopic field $\mathbf{E}^{(\text{micro})}(\mathbf{r})$ averages to zero over the supercell (see the discussion in section 2.1), hence the KS potential is supercell periodic. Notice that such choice corresponds to a vanishing macroscopic field $\mathbf{E}$ only insofar as the sample is homogeneous; otherwise (e.g. when simulating surfaces, interfaces, and polar molecules) the macroscopic field is in general nonzero in different supercell regions.

Suppose that $\varphi_j(\mathbf{r})$ are the occupied adiabatic eigenstates of the KS instantaneous Hamiltonian at time t, normalized to one in the supercell, and obeying PBCs therein; in other words they are obtained from diagonalizing the Hamiltonian at the Γ point at time t. We define the $N/2 \times N/2$ connection matrix

$$\mathsf{S}_{\alpha,jj'} = \langle\varphi_j|\mathrm{e}^{\mathrm{i}\frac{2\pi r_\alpha}{L}}|\varphi_{j'}\rangle, \tag{27}$$

which is an implicit function of the adiabatic time; notice that the operator in equation (27) is supercell periodic. The single-point Berry phase is defined as [7, 75, 76]

$$\gamma_\alpha = \mathrm{Im}\ln\det\mathsf{S}_\alpha; \tag{28}$$

this phase is gauge invariant, meaning that it is invariant for unitary transformations of the occupied orbitals between themselves.

Suppose that the nuclei are at sites $\mathbf{R}_m$ with charges Z_m; when they are adiabatically displaced the transient

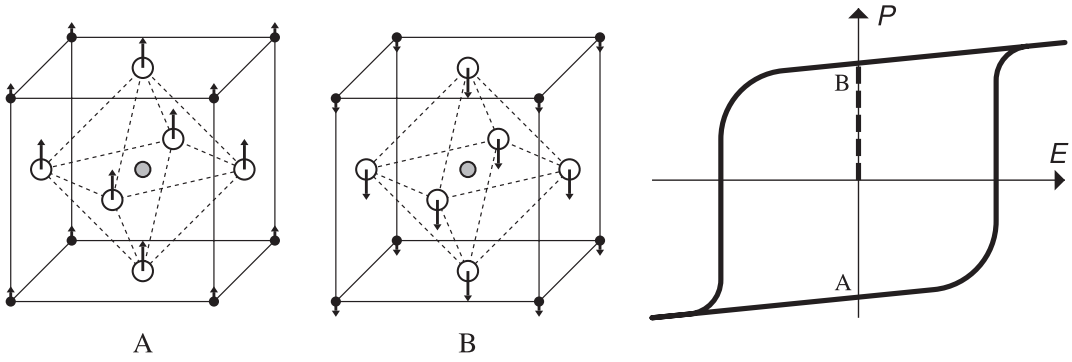

Figure 4. Left: the perovskite oxide $KNbO_3$ in the tetragonal structure. Solid, shaded, and empty circles represent K, Nb, and O atoms, respectively. The internal displacements (magnified by a factor 4) are indicated by arrows for two (A and B) enantiomorphous ferroelectric structures. An applied field switches between the two and reverses the polarization. Right: the polarization *difference* is typically measured via an hysteresis loop. The magnitude of the spontaneous polarization is also shown (vertical dashed segment); notice that spontaneous polarization is a zero-field property.

macroscopic electrical current (nuclear plus electronic) entering equation (26) is, in Hartree units,

$$\mathbf{j}(t) = \frac{1}{L^3}\sum_m Z_m \frac{\mathrm{d}\mathbf{R}_m}{\mathrm{d}t} + \mathbf{j}^{(\mathrm{el})}(t). \tag{29}$$

Notice that the overall charge neutrality of the system ($N = \sum_m Z_m$) is essential for dealing with dipolar properties. It can be shown, by means of linear-response theory, that the α-component of the electronic transient current is

$$j_\alpha^{(\mathrm{el})}(t) = -\frac{1}{\pi L^2}\frac{\mathrm{d}\gamma_\alpha(t)}{\mathrm{d}t}, \tag{30}$$

where γ_α is the instantaneous Berry phase of equation (28). This equation is correct to leading order in $1/L$ and for double occupancy [7, 75, 76]. Replacing it into equations (26) and (29) we get, in the large-L limit

$$\Delta P_\alpha = \frac{1}{L^3}\sum_m Z_m \Delta R_{m,\alpha} - \frac{1}{\pi L^2}[\gamma_\alpha(\Delta t) - \gamma_\alpha(0)]. \tag{31}$$

This is the two-point formula universally used, e.g. in Car–Parrinello [61] simulations, whenever polarization features are addressed [54, 55]; the generalization to noncubic supercells is trivial. Equation (31) is also routinely used to evaluate the dipole of an isolated molecule, whenever a supercell framework is desirable.

It is worth noticing that the nuclear and electronic terms contributing to $\Delta\mathbf{P}$ in equation (31) are *not* separately invariant for translation of the origin in the supercell. The key point is that their sum is indeed invariant, modulo the 'quantum' discussed below, section 6.4.

6.2. *Many* **k***-point formula for crystalline calculations*

Let us assume, for the sake of simplicity, a simple cubic lattice of lattice constant a. Then the Born–von-Kàrmàn period L is an integer multiple of the lattice constant: $L = Ma$, where $M \to \infty$ in the large-system limit. The most general crystal structure can be considered by means of a simple coordinate transformation [8]. The KS potential is lattice periodical, meaning that the macroscopic field $\mathbf{E}$ (i.e. the cell average of the microscopic one) vanishes.

The allowed Bloch vectors are discrete

$$\mathbf{k}_{s_1,s_2,s_3} = \frac{2\pi}{Ma}(s_1, s_2, s_3), \qquad s_\alpha = 0, 1, \ldots, M-1, \tag{32}$$

and the corresponding Bloch orbitals are $\psi_{n\mathbf{k}_{s_1,s_2,s_3}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{k}_{s_1,s_2,s_3}\cdot\mathbf{r}} u_{n\mathbf{k}_{s_1,s_2,s_3}}(\mathbf{r})$. The overlap matrix of equation (27) then becomes

$$\begin{aligned}\mathsf{S}_{\alpha,jj'} \to{}& \frac{1}{M^3}\int_0^L \mathrm{d}x \int_0^L \mathrm{d}y \int_0^L \mathrm{d}z\, \psi^*_{n\mathbf{k}}(\mathbf{r})\mathrm{e}^{\mathrm{i}\frac{2\pi r_\alpha}{Ma}}\psi_{n'\mathbf{k}'}(\mathbf{r}) \\ ={}& \int_0^a \mathrm{d}x \int_0^a \mathrm{d}y \int_0^a \mathrm{d}z\, u^*_{n\mathbf{k}}(\mathbf{r})\mathrm{e}^{\mathrm{i}(\mathbf{k}'-\mathbf{k}+\frac{2\pi r_\alpha}{Ma})}u_{n'\mathbf{k}'}(\mathbf{r}),\end{aligned} \tag{33}$$

where $\mathbf{k}$ and $\mathbf{k}'$ must be chosen within the discrete set. The $1/M^3$ factor owes to the fact that the $\varphi_j(\mathbf{r})$ orbitals entering equation (27) are normalized in the cube of volume L^3, while the Bloch orbitals ψ_n and u_n are normalized in the crystal cell of volume a^3. For given $\mathbf{k}$ and $\mathbf{k}'$ the size of the matrix on the rhs of equation (33) is n_b (the number of double-occupied bands), while the $\mathbf{k}$ and $\mathbf{k}'$ arguments run over M^3 discrete values. In fact, the total number of electrons in the Born–von-Kàrmàn box is $N = 2n_\mathrm{b}M^3$.

The key difference between the noncrystalline case and the crystalline one is that the connection matrix, equation (33), becomes very sparse in the latter case. Focusing, without loss of generality, on the x-component ($\alpha = 1$), and writing explicitly $\mathbf{k} = \mathbf{k}_{s_1,s_2,s_3}$ and $\mathbf{k}' = \mathbf{k}_{s'_1,s'_2,s'_3}$, its only nonzero elements are those with $s_1 = s'_1 + 1$, $s_2 = s'_2$, and $s_3 = s'_3$. With the usual definition of the scalar product between u_n orbitals

$$\langle u_{n\mathbf{k}} | u_{n'\mathbf{k}'}\rangle = \int_\mathrm{cell} \mathrm{d}\mathbf{r}\, u^*_{n\mathbf{k}}(\mathbf{r})u_{n'\mathbf{k}'}(\mathbf{r}), \tag{34}$$

these nonzero elements can be rewritten as

$$S_{nn'}(\mathbf{k}_{s_1+1,s_2,s_3}, \mathbf{k}_{s_1,s_2,s_3}) = \langle u_{n\mathbf{k}_{s_1+1,s_2,s_3}} | u_{n'\mathbf{k}_{s_1,s_2,s_3}}\rangle. \tag{35}$$

Owing to such sparseness, the determinant of the large matrix S_x (of size $N/2 = n_\mathrm{b}M^3$) in equation (27) factorizes into the product of M^3 determinants of the small matrices $S(\mathbf{k}, \mathbf{k}')$

(of size n_b each). The Berry phase defined in equation (28) then becomes

$$\gamma_x = \mathrm{Im}\ln \prod_{s_1,s_2,s_3=0}^{M-1} \det S(\mathbf{k}_{s_1+1,s_2,s_3}, \mathbf{k}_{s_1,s_2,s_3})$$
$$= -\sum_{s_2,s_3=0}^{M-1} \mathrm{Im}\ln \prod_{s_1=0}^{M-1} \det S(\mathbf{k}_{s_1,s_2,s_3}, \mathbf{k}_{s_1+1,s_2,s_3}). \quad (36)$$

If we use the symbol $\boldsymbol{\tau}_\ell$ for the nuclear positions in the unit cell, the main polarization formula, equation (31), becomes

$$\Delta P_x = \frac{1}{a^3}\sum_\ell Z_\ell \Delta\tau_{\ell,x} - \frac{1}{\pi M^2 a^2}[\gamma_x(\Delta t) - \gamma_x(0)], \quad (37)$$

and analogously for the other Cartesian components. This is the key formula implemented in most electronic structure codes for crystalline calculations [13, 15, 18, 16].

We notice that the u_{nk} orbitals entering S, equation (33), can be chosen with arbitrary phase factors (choice of the 'gauge'), but these factors cancel out in equation (36), leaving no arbitrariness. Furthermore, equation (36) is invariant by unitary transformations of the occupied orbitals at a given **k**. Therefore the discrete Berry phase in equation (36) is a global property of the occupied manifold as a whole; this is useful when the band numbering is nonunique (e.g. in the case of band crossings).

6.3. King-Smith and Vanderbilt formula

In order to make contact with the original continuum formulation by King-Smith and Vanderbilt it is expedient to define $\gamma_\alpha^{(\mathrm{crys})} = \gamma_\alpha/M^2$, and rewrite equation (37) as

$$\Delta P_x = \frac{1}{a^3}\sum_\ell Z_\ell \Delta\tau_{\ell,x} - \frac{1}{\pi a^2}[\gamma_x^{(\mathrm{crys})}(\Delta t) - \gamma_x^{(\mathrm{crys})}(0)]. \quad (38)$$

In the $M \to \infty$ limit the **k**-point mesh becomes dense. If the gauge is chosen in such a way that the overlap matrix $S_{nn'}(\mathbf{k},\mathbf{k}') = \langle u_{n\mathbf{k}}|u_{n'\mathbf{k}'}\rangle$ is a differentiable function of its arguments, the electronic term in equation (38) converges to a reciprocal-cell integral. In fact it can be shown that, in the $M \to \infty$ limit [5, 8, 76]

$$\gamma_x^{(\mathrm{crys})} = -\lim \frac{1}{M^2}\sum_{s_2,s_3=0}^{M-1} \mathrm{Im}\ln \prod_{s_1=0}^{M-1} \det S(\mathbf{k}_{s_1,s_2,s_3}, \mathbf{k}_{s_1+1,s_2,s_3}) \quad (39)$$
$$\to \frac{\mathrm{i}a^2}{(2\pi)^2}\int \mathrm{d}\mathbf{k} \frac{\partial}{\partial k'_x}\sum_{n=1}^{n_\mathrm{b}} S_{nn}(\mathbf{k},\mathbf{k}')\Big|_{\mathbf{k}'=\mathbf{k}}$$
$$= \frac{\mathrm{i}a^2}{(2\pi)^2}\int \mathrm{d}\mathbf{k}\sum_{n=1}^{n_\mathrm{b}}\left\langle u_{n\mathbf{k}}\left|\frac{\partial}{\partial k_x}u_{n\mathbf{k}}\right.\right\rangle. \quad (40)$$

Historically, equations (39) and (40) were derived first by King-Smith and Vanderbilt [5], and the single-point formula, equation (31), much later [7].

For the sake of completeness, we give also the formula for the most general crystalline lattice, with double band occupation. The electronic contribution to electronic polarization is the Brillouin-zone (BZ) integral

$$\mathbf{P}^{(\mathrm{el})} = -\frac{2\mathrm{i}}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\sum_{n=1}^{n_\mathrm{b}}\langle u_{n\mathbf{k}}|\nabla_\mathbf{k} u_{n\mathbf{k}}\rangle, \quad (41)$$

where it is understood that the expression must be used to evaluate polarization *differences* in a two-point formula, and the sum is over the occupied bands. The formula given here is in atomic Hartree units, for double occupancy, and for orbitals normalized to one over the crystal cell. The integral is over the BZ or, equivalently, over a reciprocal cell.

We recall that polarization (as a bulk material property) only makes sense in insulators, and that, in this work, we refer more precisely to 'KS insulators'. In fact, the integration is equations (40) and (41) is over the whole reciprocal cell or, equivalently, over the whole BZ. The spectrum has a gap and the number n_b of occupied orbitals is independent of **k**. Also, it is worth noticing that the integrand in equations (40) and (41) is *not* gauge invariant, in that it depends on the (arbitrary) choice of the phases of $|u_{n\mathbf{k}}\rangle$ at different **k**s; nonetheless the integral *is* gauge invariant (modulo the 'quantum' discussed below). More generally, the integral is invariant for any differentiable unitary mixing of the occupied $|u_{n\mathbf{k}}\rangle$ between themselves at a given **k**. A similar observation was made above about the discrete equation (36).

6.4. The polarization 'quantum'

Given that every phase is defined modulo 2π, all of the two-point formulae for $\Delta\mathbf{P}$ in terms of Berry phases are arbitrary modulo a polarization 'quantum'. This is the tradeoff one has to pay when switching from the adiabatic connection formula, equation (26)—where no such arbitrariness exists—to any of the two-point formulae given above.

In the single **k**-point case, equation (31), the 'quantum' is $2/L^2$: since we are interested in the large supercell limit, where the 'quantum' vanishes, the two-point formula is apparently useless. This is not the case, and in fact equation (31) is routinely used for evaluating polarization differences in noncrystalline materials. The key point is that the $L \to \infty$ limit is not actually needed; for an accurate description of a given material, it is enough to assume a *finite* L, actually larger than the relevant correlation lengths in the material. For any given length, the polarization 'quantum' $2/L^2$ sets an upper limit to the magnitude of a polarization difference accessible via the two-point formula, equation (31). The larger are the correlation lengths, the smaller is the accessible $\Delta\mathbf{P}$. This is no problem at all in practice, either when evaluating static derivatives by numerical differentiation, such as e.g. in [53, 77], or when performing Car–Parrinello simulations [54, 55]. In the latter case Δt is a Car–Parrinello time step (a few au), during which the polarization varies by a tiny amount, much smaller than the quantum $2/L^2$ (the typical size of a large simulation cell nowadays is $L \simeq 50$ au). Whenever needed, the drawback may be overcome by splitting Δt in equation (26) into several smaller time intervals, and by using the two-point formula for each of them.

It is worth emphasizing that—owing to supercell periodicity—even the classical nuclear term in equation (31) is affected by a similar indeterminacy, whenever a nuclear displacement $\Delta\mathbf{R}_m$ becomes of order L.

In the crystalline case translational invariance produces the much larger 'quantum' $2/a^2$. In fact, it is easily shown that equation (40) is gauge-invariant modulo 2π. The classical nuclear term has a similar indeterminacy.

Caution is in order in numerical work, when using equation (39) at finite M, since in general the $|u_{n\mathbf{k}}\rangle$ obtained from numerical diagonalization at the mesh points are *not* differentiable functions of $\mathbf{k}$. If each of the M^2 terms in the sum is chosen with arbitrary modulo 2π freedom, then $\gamma_\alpha^{(\mathrm{crys})}$ is unavoidably arbitrary modulo $2\pi/M^2$. A more clever choice is possible (and actually performed in practical implementations) as follows. One starts choosing arbitrarily one of the possible (modulo 2π) values for the first term in the sum ($s_2 = 0$ and $s_3 = 0$); for the remaining terms, it is possible to impose that nearest-neighbour phases differ by much less than 2π (if the mesh is dense enough). This choice is unique, and eliminates any residual arbitrariness, corresponding to the discrete average of a *continuous* function of $k_y k_z$, as indeed in equation (40). By this token the discrete Berry phase formula, equation (39), leads to the polarization 'quantum' $2/a^2$ (independent of M and L), indeed identical to the continuous one, and large enough to be harmless for most computations.

6.5. Wannier functions

The KS ground state is a Slater determinant of doubly occupied orbitals; any unitary transformation of the occupied states among themselves leaves the determinantal wavefunction invariant (apart for an irrelevant phase factor), and hence it leaves invariant any KS ground-state property.

For an insulating crystal, the KS orbitals are the Bloch states of completely occupied bands; these can be transformed to localized Wannier orbitals (or functions) WFs. This has been known since 1937 [78], but for many years the WFs have been mostly used as a formal tool; they became a popular topic in computational electronic structure only after the seminal work of Marzari and Vanderbilt [79]. A comprehensive review appeared as [80], and a public-domain implementation is in WANNIER90 [81]. If the crystal is metallic, the WFs can still be technically useful [51], but it must be emphasized that the ground state *cannot* be written as a Slater determinant of localized orbitals of any kind, as a matter of principle [82].

The transformation of the Berry phase formula equation (41) in terms of WFs provides an alternative, and perhaps more intuitive, viewpoint. The formal transformation has been known since the 1950s [83], although the physical meaning of the formalism was not understood until the advent of the modern theory of polarization.

The unitary transformation which defines the WF $w_{n\mathbf{R}}(\mathbf{r})$, labelled by band n and unit cell $\mathbf{R}$, within our normalization is

$$|w_{n\mathbf{R}}\rangle = \frac{V_\mathrm{c}}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}}\, |\psi_{n\mathbf{k}}\rangle. \tag{42}$$

If one then defines the 'Wannier centres' as $\mathbf{r}_{n\mathbf{R}} = \langle w_{n\mathbf{R}}|\mathbf{r}|w_{n\mathbf{R}}\rangle$, it is rather straightforward to prove that equation (41) is equivalent to

$$\mathbf{P}^{(\mathrm{el})} = -\frac{2}{V_\mathrm{c}}\sum_{n=1}^{n_\mathrm{b}} \mathbf{r}_{n\mathbf{0}}. \tag{43}$$

This means that the electronic term in the macroscopic polarization $\mathbf{P}$ is (twice) the dipole of the Wannier charge distributions in the central cell, divided by the cell volume. The nuclear term is obviously similar in form to equation (43).

WFs are severely gauge dependent, since the phases of the $|\psi_{n\mathbf{k}}\rangle$ appearing in equation (42) can be chosen arbitrarily. However, their centres are gauge-invariant modulo $\mathbf{R}$ (a lattice vector). Therefore $\mathbf{P}^{(\mathrm{el})}$ in equation (43) is affected by the same 'quantum' indeterminacy discussed above. We also stress, once more, that equation (43)—as well as equation (41)—is to be used in polarization *differences*, and does not define polarization itself.

The modern theory, when formulated in terms of WFs, becomes much more intuitive, and in a sense vindicates the venerable Clausius–Mossotti viewpoint [36]: in fact, the charge distribution is partitioned into localized contributions, each providing an electric dipole, and these dipoles yield the electronic term in $\mathbf{P}$. However, it is clear from equation (42) that the *phase* of the Bloch orbitals is essential to arrive at the right partitioning. Any decomposition based on charge only is severely nonunique and does not, in general, provide the right $\mathbf{P}$, with the notable exception of the extreme case of molecular crystals.

In the latter case, in fact, we may consider the set of WFs centred on a given molecule; their total charge distribution coincides—in the weakly interacting limit—with the electron density of the isolated molecule (possibly in a local field). This justifies the elementary Clausius–Mossotti viewpoint. It is worth mentioning that the dipole of a polar molecule is routinely computed in a supercell geometry via the single-point Berry phase [77]. The dipole value coincides with the one computed in a trivial way in the large supercell limit. Finite-size corrections, due to the local field (different in the two cases), can also be applied [84].

The case of alkali halides—where the model is often phenomenologically used—deserves a different comment [10]. The electron densities of isolated ions (with or without fields) are quite different from the corresponding WFs charge distributions, for instance because of orthogonality constraints: hence the model is *not* justified in its elementary form, despite contrary statements in the literature. For a detailed analysis, see [37].

7. Geometrical issues

7.1. Chern invariants and topological insulators

It has been observed that macroscopic polarization (as a bulk material property) only makes sense in insulating materials, while macroscopic orbital magnetization exists both in insulators and metals. Furthermore, magnetic insulators come in two classes: the 'nonexotic' and the 'exotic' ones,

called in the following 'normal insulators' and 'topological insulators', respectively.

Until recently the only known realization of a topological insulator was the quantum Hall effect (QHE): a 2d electron fluid in a perpendicular **B** field exhibits a new state of matter. The 'bulk' of the system is insulating, but there are circulating edge states which are robust ('topologically protected') in presence of disorder, and are responsible for the famous plateaus in the transverse conductivity. The same electron fluid can be described using toroidal boundary conditions, where no edge exists. In this case the signature of the quantum Hall state is a topological integer C_1 (Chern number of the first class) which geometrically characterizes the wavefunction. The Chern number is defined below, equation (45), only in the simple case of $\mathbf{B} = 0$; $C_1 = 0$ means a normal insulator. The Hall conductivity in the QHE regime is simply expressed in atomic Hartree units as

$$\sigma_{\mathrm{T}} = -C_1/2\pi, \tag{44}$$

or, in ordinary units, $\sigma_{\mathrm{T}} = -C_1 e^2/h$.

This result is due to Thouless and co-workers, both in the case of integer [85] and fractional [86] QHE. These two milestone papers mark the debut of geometrical concepts in electronic structure theory [87]. Notice that in the QHE regime, due to the presence of a macroscopic **B** field, the Hamiltonian *cannot* be lattice periodical.

A subsequent breakthrough on the theory side is the Haldane model Hamiltonian [88]: this can be considered as the archetype of topological insulators (see below). The model is comprised of a 2d honeycomb lattice with two tight-binding sites per primitive cell with site energies $\pm\Delta$, real first-neighbour hopping t_1, and complex second-neighbour hopping $t_2 \mathrm{e}^{\pm \mathrm{i}\varphi}$, as shown in figure 5. Within this two band model, one deals with insulators by taking the lowest band as occupied. The appeal of the model is that there is no macroscopic field, hence the vector potential and the Hamiltonian are lattice periodical and the single-particle orbitals always have the usual Bloch form. Essentially, the microscopic magnetic field can be thought as staggered (i.e. up and down in different regions of the cell), but its cell average vanishes. As a function of the flux parameter ϕ, this system undergoes a transition from zero Chern number (i.e. normal insulator) to $|C_1| = 1$ (i.e. topological insulator).

In general, the Chern number for *any* lattice periodical Hamiltonian in 2d is expressed in terms of the Bloch orbitals as

$$C_1 = \frac{\mathrm{i}}{2\pi} \sum_{n=1}^{n_{\mathrm{b}}} \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k} \, [\langle \partial u_{n\mathbf{k}}/\partial k_1 \mid \partial u_{n\mathbf{k}}/\partial k_2 \rangle - \langle \partial u_{n\mathbf{k}}/\partial k_2 \mid \partial u_{n\mathbf{k}}/\partial k_1 \rangle], \tag{45}$$

where the sum is over the occupied ns only, and the integral is over the 2d Brillouin zone (the formula here is given for single band occupancy). It is easily verified that C_1 is dimensionless, and in fact is quantized in integer units.

In 3d the Chern number, equation (45), is generalized to the (vector) Chern invariant

$$\mathbf{C} = \frac{\mathrm{i}}{2\pi} \sum_{n=1}^{n_{\mathrm{b}}} \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k} \, \langle \partial_{\mathbf{k}} u_{n\mathbf{k}} | \times | \partial_{\mathbf{k}} u_{n\mathbf{k}} \rangle, \tag{46}$$

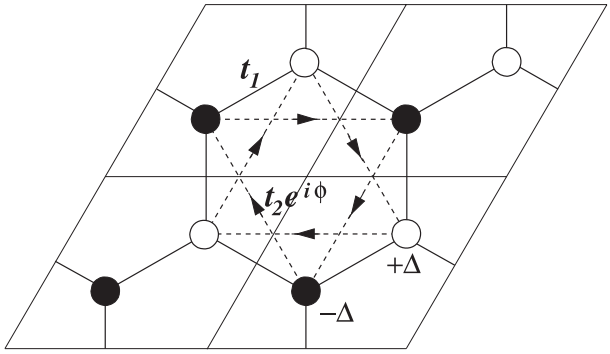

Figure 5. Four unit cells of the Haldane model [88]. Filled (open) circles denote sites with $E_0 = -\Delta$ ($+\Delta$). Solid lines connecting nearest neighbours indicate a real hopping amplitude t_1; dashed arrows pointing to a second-neighbour site indicate a complex hopping amplitude $t_2 \mathrm{e}^{\mathrm{i}\phi}$. Arrows indicate the sign of the phase ϕ for second-neighbour hopping.

with the usual meaning of the cross product between three-component bra and ket states. Here the integral is over the 3d Brillouin zone, and $\partial_{\mathbf{k}} = \partial/\partial\mathbf{k}$. The Chern invariant has the dimension of an inverse length, and in fact is quantized in units of reciprocal lattice vectors. Notice the analogy with—but also the key difference from—the Berry phase formula in the modern theory of polarization, equation (41).

Whenever the Chern invariant (number in 2d) is nonzero in a periodic Hamiltonian the bulk states are gapped, but there are topologically protected surface (edge in 2d) states which are conducting: we call 'Chern insulator' this kind of topological insulator. It is worth noticing that in a Chern insulator Wannier functions *cannot* exist [89] (despite the fact that the Hamiltonian eigenstates do have the Bloch form).

No microscopic realization of an insulator with nonzero Chern number (in 2d) or nonzero Chern invariant (in 3d), *in absence of a macroscopic* **B** *field* is known to date. A mesoscopic 2d Chern insulator, in the same spirit as the Haldane model Hamiltonian, was synthesized in 2008, and the quantization therein was demonstrated [90].

Since all known real materials were either conductors or normal insulators, the exotic insulators remained a curiosity of only academic interest for many years. The interest in topological insulators boomed after a 2005 paper by Kane and Mele, who proposed a novel invariant (called Z_2) to discriminate between normal and exotic insulators. Nontrivial values of this invariant occur in time-reversal symmetric systems; the fingerprint of a Z_2 topological insulator in 2d is the quantum spin Hall effect. Analogously to the ordinary quantum Hall effect, a finite sample displays topologically protected circulating edge states, but opposite spins are counterpropagating, and the total edge current vanishes. This exciting new field is clearly beyond the scope of the present review, given also that it concerns time-reversal symmetric systems. We only stress that a genuine revolution is underway in the field of topological insulators, and that the experimental realization of a topological insulator in 2d and 3d has been demonstrated. We provide a few references for orientation: [91–95].

7.2. Berry curvature and the anomalous Hall effect

The integrand in equations (45) and (46) is a key geometrical feature of the wavefunction within PBCs, and goes under the name of 'Berry curvature' (in both 2d and 3d), or, equivalently, of 'gauge field'. We define it per band, i.e.

$$\mathbf{\Omega}_n(\mathbf{k}) = \mathrm{i}\langle\partial_{\mathbf{k}} u_{n\mathbf{k}}| \times |\partial_{\mathbf{k}} u_{n\mathbf{k}}\rangle = -\,\mathrm{Im}\langle\partial_{\mathbf{k}} u_{n\mathbf{k}}| \times |\partial_{\mathbf{k}} u_{n\mathbf{k}}\rangle; \quad (47)$$

equivalently one can define the Berry curvature as the antisymmetric Cartesian tensor

$$\Omega_{n,\alpha\beta}(\mathbf{k}) = -2\,\mathrm{Im}\langle\partial u_{n\mathbf{k}}/\partial k_\alpha \mid \partial u_{n\mathbf{k}}/\partial k_\beta\rangle. \quad (48)$$

The Berry curvature is gauge invariant, hence in principle it leads to observable effects.

In the presence of time-reversal symmetry the Chern invariant vanishes, i.e. the Berry curvature integrates to zero over the BZ. However, it is identically zero only in centrosymmetric crystals. Instead, in crystals which are time-reversal symmetric but non-centrosymmetric, $\mathbf{\Omega}_n(\mathbf{k})$ contributes to the semiclassical equation of transport [96].

When time-reversal symmetry is broken and the system is metallic the integral of $\mathbf{\Omega}_n(\mathbf{k})$ over the occupied states provides a sizeable contribution to the anomalous Hall effect (AHE), discussed in the following.

In the absence of time-reversal symmetry (e.g. in a metallic ferromagnet) the transverse conductance is nonzero even in zero magnetic field. This is the AHE, discovered by E R Hall in 1881 (at about the same time as the normal Hall effect); it gathered a renewal of interest in the 2000s. The effect is due to several mechanisms, some of them extrinsic, and the relative role of the different mechanisms is still controversial; however, one important term in the AHE conductivity is intrinsic and purely geometrical. Using equation (48) this term is (in atomic Hartree units and for single band occupancy)

$$\sigma_{\alpha\beta} = -\frac{1}{(2\pi)^3}\sum_n \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\ \ f_{n\mathbf{k}}\Omega_{n,\alpha\beta}(\mathbf{k}), \quad (49)$$

where $f_{n\mathbf{k}} = \theta(\mu - \epsilon_{n\mathbf{k}})$ is the Fermi occupancy factor at $T = 0$, and μ is the Fermi energy. A formula similar, though not identical, to equation (49), was proposed as early as 1954 by Karplus and Luttinger [97]. The genuine Berry connection formula, equation (49), was established in 2002 [98], and implemented in first-principle calculations soon afterwards [99–101] for the three ferromagnetic metals Fe, Co, and Ni. It has been found that the integrand fluctuates wildly, and therefore the BZ integration is quite demanding.

It is worth pointing out that the 2d analogue of equation (49), when applied to a gapped crystal, coincides exactly with the QHE formula, equation (44). In both the QHE and AHE cases these topological formulae, derived within PBCs for a sample without boundaries, correspond to dissipationless boundary currents in finite samples.

8. Modern theory of magnetization

First of all we recall that both spin and orbital motion of the electrons contribute to the total magnetization. While spin magnetization can be calculated with high accuracy by standard state-of-the-art methods such as the spin-density functional theory (SDFT), orbital magnetization is the subject of investigations still in progress at the fundamental level. In this work, we refer to **M** as the macroscopic orbital magnetization in zero **B** field. This requires breaking of time-reversal symmetry in the spatial wavefunction, which can occur in several ways. An important paradigm is the 2d model Hamiltonian introduced by Haldane in 1988 [88] (figure 5); in real materials the time-reversal symmetry breaking can be due to spin–orbit interactions (as in ferromagnets), or to an explicit perturbation in nonmagnetic materials (as e.g. detailed in section 5.4).

8.1. Normal insulators

In a normal insulator (i.e. whenever the Chern invariant is zero) the Bloch orbitals can be chosen so as to obey $|\psi_{n\mathbf{k}+\mathbf{G}}\rangle = |\psi_{n\mathbf{k}}\rangle$ (the so-called periodic gauge), which in turn warrants the existence of WFs enjoying the usual properties. Under this condition, the formula yielding the macroscopic orbital magnetization in vanishing macroscopic **B** field is

$$\mathbf{M} = \frac{1}{c(2\pi)^3}\,\mathrm{Im}\sum_{n=1}^{n_\mathrm{b}}\int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\,\langle\partial_{\mathbf{k}} u_{n\mathbf{k}}| \times (H_{\mathbf{k}} + \epsilon_{n\mathbf{k}})\,|\partial_{\mathbf{k}} u_{n\mathbf{k}}\rangle. \quad (50)$$

The formula as given here is in atomic Hartree units ($c \simeq 137$) and for double band occupancy. As usual, $|u_{n\mathbf{k}}\rangle$ is the periodic part in a Bloch orbital, $\epsilon_{n\mathbf{k}}$ is the band energy, and $H_{\mathbf{k}}$ is the effective Hamiltonian acting on the u's, i.e. $H_{\mathbf{k}} = \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}} H \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}$. The orbitals are normalized to one over the crystal cell of volume V_c; the sum is over the occupied bands and the integral is over the whole BZ.

Equation (50) was first established for the single band case, independently in [11] (via the semiclassical method) and in [12] (addressing the ground state in term of WFs). In the latter case, computer simulations based on the 2d Haldane model Hamiltonian (figure 5) have been instrumental in order to arrive at the magnetization formula and to validate it. A precursor work, appeared in 2003 [102], provides the correct formula for the special case of the Hofstadter model Hamiltonian. The (nontrivial) extension to the many band case, as given in equations (50), was provided in 2006 by Ceresoli *et al* [23] (again, via WFs).

It is expedient to compare equation (50) with its electrical analogue, which is the King-Smith and Vanderbilt polarization formula, written in equation (41) (for double band occupancy and in atomic units as well). The main ingredient in both formulae are **k**-derivatives of the periodic $|u_{n\mathbf{k}}\rangle$ orbitals; additionally, the Hamiltonian and the band energies appear in **M**. A key difference is that in equation (41) the *integrand* is gauge dependent, and only the *integral* is gauge invariant; in the magnetic case, instead, both the integrand and the integral are gauge invariant. In the electrical case only the BZ integral—as in equation (41)—makes sense, and this is in agreement with the fact that bulk polarization **P** is well defined only in insulators (within our KS scheme, in 'KS insulators', to be more precise). In the magnetic case, instead,

the same integral appearing in equation (50), but limited to states below the Fermi level in metals, is gauge invariant and could make physical sense, since **M** is well defined even in metals. The actual formula for metals (see below, section 8.3) is similar, but somewhat different. A further key difference, worth emphasizing, is that there is no 'quantum' indeterminacy in the magnetic case.

An apparent paradox is that equation (50) does not appear at first sight to be invariant with respect to translation of the energy zero. However, the zero-Chern-invariant condition—compare equations (50)–(46)—enforces such invariance in any normal insulator.

The main magnetization formula, equation (50), for the orbital magnetization of a crystalline insulator can easily be implemented in existing first-principle electronic structure codes, making available the computation of the orbital magnetization in crystals and at surfaces. The **k**-derivatives therein must be discretized as finite differences; a gauge-invariant numerical algorithm to this aim is detailed in appendix A of [23].

8.2. *Single* **k***-point formula for supercell calculations*

We start by observing that equation (50) is invariant by cell doubling. In fact, starting with a cell (or supercell) of given size, we may regard the same physical system as having double periodicity (in all directions), in which case the integration domain in equation (50) gets 'folded' and shrinks by a factor 1/8, while the number of occupied eigenstates gets multiplied by a factor of 8. It is easy to realize that these are in fact *the same* eigenstates as in the unfolded case, apart possibly for a unitary transformation, irrelevant here. As for the discretized form of equation (50), it can be chosen to be numerically invariant by cell doubling within the computational tolerance chosen (by a suitable choice of the mesh).

The supercell is the obvious way of dealing with disordered systems, which can be regarded as a crystalline systems of large enough size. The actually required size depends on the relevant correlation lengths in the material addressed.

If we replace V_c in equation (50) with a (large) supercell volume V, the integral is approximated by the value of the integrand at $\mathbf{k} = 0$ times the reciprocal volume $(2\pi)^3/V$. If there are N electrons in the supercell the formula is

$$\mathbf{M} = \frac{1}{cV}\,\mathrm{Im}\sum_{n=1}^{N/2}\langle\partial_{\mathbf{k}}u_{n0}|\times(H_0+\epsilon_{n0})\,|\partial_{\mathbf{k}}u_{n0}\rangle. \tag{51}$$

This formula has been proposed in [27] and validated, once more, via simulations based on the 2d Haldane model Hamiltonian. One key virtue of equation (51) is its gauge invariance in a generalized sense, that is for arbitrary unitary mixing of the occupied $|u_{n\mathbf{k}}\rangle$ among themselves.

The $\mathbf{k} = 0$ derivatives $|\partial_{\mathbf{k}}u_{n0}\rangle$ appearing in equation (51) deserve further discussion, since here we no longer have any mesh, only *one* reciprocal point (the Γ point). One possible approach is to evaluate such derivatives via perturbation theory, i.e.

$$|\partial_{\mathbf{k}}u_{n0}\rangle = \sum_{m\neq n}|u_{m0}\rangle\frac{\langle u_{m0}|\mathbf{v}|u_{n0}\rangle}{\epsilon_{m0}-\epsilon_{n0}}, \tag{52}$$

where **v** is the velocity operator

$$\mathbf{v} = \mathrm{i}[H,\mathbf{r}] = \nabla_{\mathbf{k}}H_{\mathbf{k}}|_{\mathbf{k}=0}. \tag{53}$$

Equation (52) is convenient for tight-binding implementations, where the sum is over a small number of terms. We also notice that the matrix representation of **r**, for use in equation (53), is usually taken to be diagonal on the tight-binding basis.

However, equation (52) is not convenient for a first-principle implementation, since it would require the evaluation of slowly convergent perturbation sums. This can be avoided taking a different approach. If $\mathbf{b}_j$ are the shortest reciprocal vectors of the supercell, and ∂_j indicates the partial **k**-derivative in the direction of $\mathbf{b}_j$, then by definition

$$|\partial_j u_{n0}\rangle = \lim_{\lambda\to 0}\frac{1}{\lambda|\mathbf{b}_j|}(|u_{n\,\lambda\mathbf{b}_j}\rangle - |u_{n0}\rangle). \tag{54}$$

For a large enough supercell, the limit is approximated by taking $\lambda = 1$. Next we wish to evaluate $|u_{n\mathbf{b}_j}\rangle$ without actually diagonalizing the Hamiltonian at $\mathbf{k}\neq 0$. To this aim, we notice that the state $\mathrm{e}^{-\mathrm{i}\mathbf{b}_j\cdot\mathbf{r}}|u_{n0}\rangle$ obeys periodic boundary conditions and is an eigenstate of $H_{n\mathbf{b}_j}$ corresponding, possibly, to a different occupied eigenvalue and to a different phase choice. In other words, no further diagonalization is needed to identify the manifold spanned by the occupied eigenstates $|u_{n\mathbf{b}_j}\rangle$ appearing in equation (54). It is then easy to evaluate equation (51), provided a specific gauge is enforced. In fact, equation (51) is gauge invariant by unitary mixing of the occupied eigenstates. We further observe that the eigenstates $|u_{n\mathbf{k}}\rangle$ obtained from numerical diagonalization are *not* analytical functions of **k**—as instead is implicitly assumed in equation (54). But this feature causes no harm after the gauge transformation is performed. The algorithm performing the required gauge transformation is detailed in [27], and is inspired by [103, 104].

The single-point formula, equation (51), has been implemented in a first-principle framework to evaluate the NMR shielding tensor in liquid water, via the converse approach discussed in section 5.4 [28].

The formula is ideally suited for implementation in time-dependent Car–Parrinello simulations, as in the corresponding electrical case (section 6.1), but an important caveat is in order: whenever time-reversal symmetry is absent, the *classical* nuclear equation of motion needs to be modified.

The nuclear kinetic energy includes, in general, *two* vector potentials: the obvious one of magnetic origin, and an additional one of geometric origin (or 'gauge potential'). The latter, in the absence of time-reversal symmetry, is *not* curl-free in the region visited by the classical system. Therefore the two vector potentials give rise to two different—and often competing—forces of the Lorentz kind in the Newton equation of motion: one is the familiar Lorentz force of classical electromagnetism [2], and the other has a quantum origin. For more details, see section 5.2 in [76], and for a toy-model implementation see [105].

8.3. Chern insulators and metals

We switch here to single band occupancy, having in mind e.g. a ferromagnet (where the orbitals for up and down spins are different). The macroscopic magnetization per spin channel is

$$\mathbf{M} = \frac{1}{2c(2\pi)^3}\,\mathrm{Im}\sum_n \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\; f_{n\mathbf{k}} \langle \partial_{\mathbf{k}} u_{n\mathbf{k}}| \times (H_{\mathbf{k}} + \epsilon_{n\mathbf{k}} - 2\mu)\,|\partial_{\mathbf{k}} u_{n\mathbf{k}}\rangle, \tag{55}$$

where $f_{n\mathbf{k}}$ is, as above, the Fermi occupancy factor at $T = 0$ and μ is the Fermi energy. The formula applies to both Chern insulators (defined as insulators with nonzero Chern invariant) and metals. Equation (55) is obviously invariant by translation of the energy zero and coincides with (one half of) equation (50) in the case of a normal insulator. In fact, the role of μ is irrelevant if the Chern invariant, equation (46), vanishes. In both Chern insulators and metals the magnetization depends on μ, as it must be (see section 8.4).

Equation (55) was first derived from semiclassical arguments for the single band case in [11]. Subsequently, equation (55) was heuristically assumed in [23] and validated via computer experiments. Needless to say, the transformation to WFs leading to the proof of equation (50) could not be used for Chern insulators and for metals.

The numerical validation of equation (55) was based once more on the 2d Haldane model Hamiltonian (figure 5) at various fillings, and required two sets of simulations based on (i) OBCs, corresponding to a finite sample with a boundary, and (ii) PBCs. In case (i) the magnetization **M** was computed in the trivial way by means of equations (14), while in case (ii) **M** was computed by means of equation (55), discretized on a numerical grid and exploiting a smearing technique (Fermi–Dirac occupancy at finite T). The two sets of computations indeed converged to the *same* **M** *value* in the large-system limit [23].

More recently, a quantum derivation of equation (55) beyond the semiclassical regime, and based on perturbation theory, was published by Shi *et al* [25]. Only one first-principle implementation exists at the time of writing [29]: this concerns the orbital contribution to the spontaneous magnetization of Fe, Co, and Ni.

In order to proceed further, it is expedient to write equation (55) identically as the sum of two terms, each of them separately gauge invariant, hence in principle separately measurable

$$\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2$$

$$\mathbf{M}_1 = \sum_n \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\; f_{n\mathbf{k}}\; \mathbf{m}_n(\mathbf{k}),$$

$$\mathbf{m}_n(\mathbf{k}) = \frac{1}{2c(2\pi)^3}\,\mathrm{Im}\langle \partial_{\mathbf{k}} u_{n\mathbf{k}}| \times (H_{\mathbf{k}} - \epsilon_{n\mathbf{k}})\,|\partial_{\mathbf{k}} u_{n\mathbf{k}}\rangle; \tag{56}$$

$$\mathbf{M}_2 = \frac{1}{c(2\pi)^3}\sum_n \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\; f_{n\mathbf{k}}\,(\mu - \epsilon_{n\mathbf{k}})\,\mathbf{\Omega}_n(\mathbf{k}). \tag{57}$$

Here $\mathbf{\Omega}_n(\mathbf{k})$ is the Berry curvature, equation (47), and $\mathbf{m}_n(\mathbf{k})$ coincides with the semiclassical formula for the magnetization of a wavepacket in the nth band [96].

8.4. Finite temperature formula

It has already been stated that—at variance with the electrical analogue **P**—orbital magnetization **M** is a well defined physical property even at nonzero temperature. The Fermi occupancy factor as a function of μ (chemical potential) and β (inverse temperature) becomes

$$f_{n\mathbf{k}} = \frac{1}{\mathrm{e}^{\beta(\epsilon_{n\mathbf{k}}-\mu)} + 1}. \tag{58}$$

According to Shi *et al* [25] the finite temperature orbital magnetization can be written as $\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2$, where $\mathbf{M}_1$ is identical in form to equation (56), whereas $\mathbf{M}_2$ instead becomes

$$\mathbf{M}_2 = \frac{1}{c(2\pi)^3}\sum_n \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\; \frac{1}{\beta}\ln[1 + \mathrm{e}^{-\beta(\epsilon_{n\mathbf{k}}-\mu)}]\,\mathbf{\Omega}_n(\mathbf{k}), \tag{59}$$

whose $T \to 0$ limit coincides with equation (57). The formulae given here hold for any crystalline system: normal insulators, Chern insulators, and metals.

Next we take the μ derivatives of the two terms in the magnetization formula:

$$\frac{\partial \mathbf{M}_1}{\partial \mu} = \sum_n \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\; \frac{\partial f_{n\mathbf{k}}}{\partial \mu}\; \mathbf{m}_n(\mathbf{k}) \tag{60}$$

$$\frac{\partial \mathbf{M}_2}{\partial \mu} = \frac{1}{c(2\pi)^3}\sum_n \int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\;\; f_{n\mathbf{k}}\;\mathbf{\Omega}_n(\mathbf{k}). \tag{61}$$

We notice that at low temperature $\partial f_{n\mathbf{k}}/\partial\mu$ is essentially a δ at the Fermi surface, and we analyse the three cases in the $T \to 0$ limit.

(i) For a normal insulator μ falls in an energy gap and the Berry connection integrates to zero over the BZ. Ergo the magnetization **M** is μ-independent.

(ii) For a Chern insulator μ falls in a bulk gap, ergo $\partial\mathbf{M}_1/\partial\mu$ vanishes, while $\partial\mathbf{M}_2/\partial\mu$ is quantized and proportional to the Chern invariant, equation (46):

$$\frac{\partial \mathbf{M}}{\partial \mu} = -\frac{1}{c(2\pi)^2}\mathbf{C}. \tag{62}$$

The physical interpretation of this equation is best understood in 2d, where the analogue of equation (62) reads

$$\frac{\partial M}{\partial \mu} = -\frac{C_1}{2\pi c}, \tag{63}$$

and C_1 is the Chern number. We address a finite sample cut from a Chern insulator. Owing to equations (2) a macroscopic current of intensity $I = cM$ circulates at the edge of any two-dimensional uniformly magnetized sample, hence equation (63) yields

$$\frac{\mathrm{d}I}{\mathrm{d}\mu} = -\frac{C_1}{2\pi}. \tag{64}$$

The role of chiral edge states is elucidated, for example [106, 107], by considering a vertical strip of width ℓ, where the currents at the right and left boundaries are $\pm I$. The net current vanishes insofar as μ is constant throughout the sample.

When an electric field $\mathbf{E}$ is applied across the sample, the right and left chemical potentials differ by $\Delta\mu = E\ell$ and the two edge currents no longer cancel. Our equation (64) is consistent with the known quantum Hall results. In fact, according to equation (64), the net current is $\Delta I \simeq -C_1\Delta\mu/2\pi$, while the transverse conductivity is defined by $\Delta I = \sigma_{\mathrm{T}}E\ell$. We thus arrive at equation (49).

Remarkably, the above equations state that the contribution of edge states is indeed a bulk quantity, and can be evaluated in the thermodynamic limit by adopting periodic boundary conditions where the system has no edges. As already observed, this feature may look counterintuitive, but this fascinating behaviour has been known for more than 20 years in QHE theory [85, 106].

(iii) In a metal both μ-derivatives contribute. The first term $\partial\mathbf{M}_1/\partial\mu$ is nontopological and has bulk nature. It involves only the states at the Fermi level, and simply measures the magnetization change due to the change in occupation of these bulk states. At variance with this, $\partial\mathbf{M}_2/\partial\mu$ is topological and is related to the contribution of chiral boundary states in a finite sample, very similarly to the Chern insulator case discussed above.

8.5. Transport

In this section we allow for space inhomogeneity at a macroscopic scale. If $\mathbf{r}$ is a macroscopic coordinate, we may define a local chemical potential $\mu(\mathbf{r})$ and a local inverse temperature $\beta(\mathbf{r})$. Replacing them in equations (56) and (59) we get the expression for the local magnetization $\mathbf{M}(\mathbf{r}) = \mathbf{M}_1(\mathbf{r}) + \mathbf{M}_2(\mathbf{r})$. According to equations (2), a macroscopic dissipationless current flows in the bulk of the material. However, this 'magnetization current' is divergence-free, hence *unobservable* in a transport experiment (for a thorough discussion see [108]).

Let $g(\mathbf{k}, \mathbf{r})$ be the out-of-equilibrium phase space distribution of the carriers. In the semiclassical theory of transport [96] the carriers are actually wavepackets in a neighbourhood of the Fermi level. In the simple case where a single band crosses the Fermi level, and in the absence of a mechanical force, the wavepacket velocity is simply $\dot{\mathbf{r}} = \partial\epsilon_{\mathbf{k}}/\partial\mathbf{k}$, hence the current density seems to be

$$\mathbf{j}(\mathbf{r}) = -\frac{1}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\, g(\mathbf{r}, \mathbf{k})\, \dot{\mathbf{r}}. \tag{65}$$

This is not quite correct: since the wavepacket rotates around its centre, there is an extra contribution to the current. According to [109] the complete expression is

$$\begin{aligned}\mathbf{j}(\mathbf{r}) &= -\frac{1}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\, g(\mathbf{r}, \mathbf{k})\, \dot{\mathbf{r}} \\ &\quad + \nabla \times \frac{c}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\, f(\mathbf{r}, \mathbf{k})\mathbf{m}(\mathbf{k}) \\ &= -\frac{1}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\, g(\mathbf{r}, \mathbf{k})\, \dot{\mathbf{r}} + c\,\nabla \times \mathbf{M}_1(\mathbf{r}),\end{aligned} \tag{66}$$

where $g(\mathbf{r}, \mathbf{k})$ is approximated with the equilibrium $f(\mathbf{r}, \mathbf{k})$ in the extra term for a linear order expression.

Equation (66) is the expression for the physical current, but this is not yet the transport current. As shown by Xiao *et al* [24] the contribution from the unobservable magnetization current (discussed above) must be discounted. The correct expression for the *transport* current is

$$\begin{aligned}\mathbf{j}_{\mathrm{tr}}(\mathbf{r}) &= \mathbf{j}(\mathbf{r}) - c\nabla \times \mathbf{M}(\mathbf{r}) \\ &= -\frac{1}{(2\pi)^3}\int_{\mathrm{BZ}} \mathrm{d}\mathbf{k}\, g(\mathbf{r}, \mathbf{k})\, \dot{\mathbf{r}} - c\,\nabla \times \mathbf{M}_2(\mathbf{r}).\end{aligned} \tag{67}$$

The first term therein explicitly depends on the out-of-equilibrium $g(\mathbf{r}, \mathbf{k})$, hence on the details of relaxation processes. The second term, instead, is independent of relaxation processes and is therefore an intrinsic linear response of the system, having geometrical origin; we focus on this second term only in the following, calling it $\mathbf{j}_2(\mathbf{r})$. If we assume the system in thermal equilibrium, hence $\beta(\mathbf{r})$ constant, then $\partial/\partial r_\beta = \partial\mu/\partial r_\beta\ \partial/\partial\mu$, and equation (67) yields

$$\frac{1}{c}j_{2,\alpha} = -\varepsilon_{\alpha\beta\gamma}\frac{\partial M_{2,\gamma}}{\partial r_\beta} = -\varepsilon_{\alpha\beta\gamma}\frac{\partial\mu}{\partial r_\beta}\frac{\partial M_{2,\gamma}}{\partial\mu}, \tag{68}$$

where $\varepsilon_{\alpha\beta\gamma}$ is the antisymmetric tensor, and now equation (61) can be used.

If we transform the Berry curvature from vector to tensor form—equations (47) and (48)—we get $\Omega_{n,\alpha\beta} = \varepsilon_{\alpha\beta\gamma}\Omega_{n,\gamma}$, where $\varepsilon_{\alpha\beta\gamma}$ is the antisymmetric tensor. Equation (61) then yields

$$\varepsilon_{\alpha\beta\gamma}\frac{\partial M_{2,\gamma}}{\partial\mu} = -\frac{1}{c}\sigma_{\alpha\beta}, \tag{69}$$

where $\sigma_{\alpha\beta}$ is the topological AHE conductivity, equation (49). Finally equation (68) is rewritten for the transport current as

$$j_\alpha = \sigma_{\alpha\beta}\frac{\partial\mu}{\partial r_\beta}. \tag{70}$$

If we replace $\nabla\mu$ with the electric field, equation (70) becomes identical to the familiar formula for the transverse conductivity. The results of [24] prove thus that the Einstein relation continues to hold for transverse conductivity in absence of time-reversal symmetry. We note that the formula given here addresses intrinsic effects only.

8.6. Dichroic f-sum rule

The differential absorption of left and right circularly polarized light by magnetic materials is known as magnetic circular dichroism. In the past 15 years or so a sum rule for x-ray magnetic circular dichroism (XMCD) has been extensively used at synchrotron facilities to obtain information about orbital magnetism in solids. The relationship between the $\mathbf{M}$ measured via the XCMD sum rule and the $\mathbf{M}$ provided by the modern theory has been addressed in 2008 by Souza and Vanderbilt [26], and will be reviewed here.

A very important caveat is in order at the very beginning. We are going to assume in the following that the KS energies and orbitals—both occupied and empty—provide a faithful excitation spectrum of the crystalline system. Clearly, this is a severe approximation, whose accuracy may be doubtful in

many materials. We recall, nonetheless, that even the ground-state magnetization is—strictly speaking—beyond the reach of standard DFT: see the discussion in section 4. The sum rule discussed here is of course exact for noninteracting electrons.

If we define as $\langle\boldsymbol{\sigma}\rangle''_{\mathrm{A}}$ the frequency-integrated XCMD spectrum in vector notation, the main result of [26] is, in atomic Hartree units,

$$\langle\boldsymbol{\sigma}''_{\mathrm{A}}\rangle = \pi c\mathbf{M}_1 \tag{71}$$

where $\mathbf{M}_1$ is given by equation (56). In other words, only one of the two (gauge-invariant) terms into which we have partitioned $\mathbf{M}$ is measured by the XMCD spectrum. The term $\mathbf{M}_2$, equation (57), is missing; its quantitative importance is unknown.

Equation (71) is proved for normal insulators only—once more, via a transformation to WFs—although it is possibly valid for Chern insulators and metals as well. In the interpretation of [26] the XCMD sum rule probes the gauge-invariant part of the self-rotation of the occupied WFs.

9. Conclusions

We have reviewed here, on a common ground, both the modern theory of polarization $\mathbf{P}$ and the modern theory of orbital magnetization $\mathbf{M}$. The former theory (or its existence at least) is well known in the electronic structure community. It is implemented as a standard option in most codes [13–15, 17, 16, 18], has revolutionized the theory of ferroelectric and piezoelectric materials [19–21], and started reaching—although very slowly—the elementary textbooks. The theory of magnetization, instead, is still in its infancy. The very first *ab initio* implementations [28, 29] are appearing at the time of writing (2010). Previous computations of orbital magnetization in solids have invariantly relied on the uncontrolled muffin-tin approximation.

The modern theories address $\mathbf{P}$ and $\mathbf{M}$ in zero macroscopic fields $\mathbf{E}$ and $\mathbf{B}$. The meaning of this apparently counterintuitive situation is thoroughly discussed (section 2).

The presentation given here is strictly within a KS scheme, whose limitations are however discussed (section 4). We provide formulae both for crystalline solids, where $\mathbf{P}$ and $\mathbf{M}$ are Brillouin-zone integrals (discretized for numerical work), and for noncrystalline condensed systems in a single $\mathbf{k}$-point supercell framework.

At the KS level both theories are in (I dare saying) a definitive shape. Instead, when dealing with explicitly correlated wavefunctions (such as within quantum Monte Carlo), a successful formula exists for $\mathbf{P}$ [6, 7]—not discussed here—but not yet for $\mathbf{M}$.

Acknowledgments

The developments reported in this review cover a period of many years, during which I have profited enormously from discussions with many co-workers and colleagues, too many to be mentioned individually. Here I want to acknowledge explicitly Qian Niu and David Vanderbilt, whose contributions dominate this review. The work in this area has been continuously supported for many years by the Office of Naval Research, thanks to Wally Smith's vision. My current ONR grant is N00014-07-1-1095.

References

[1] Landau L D and Lifshitz E M 1984 *Electrodynamics of Continuous Media* (Oxford: Pergamon)
[2] Jackson J D 1975 *Classical Electrodynamics* (New York: Wiley)
[3] Posternak M, Baldereschi A, Catellani A and Resta R 1990 *Phys. Rev. Lett.* **64** 1777
[4] Resta R 1992 *Ferroelectrics* **136** 51
[5] King-Smith R D and Vanderbilt D 1993 *Phys. Rev.* B **47** 1651
[6] Ortíz G and Martin R M 1994 *Phys. Rev.* B **43** 14202
[7] Resta R 1998 *Phys. Rev. Lett.* **80** 1800
[8] Resta R 1994 *Rev. Mod. Phys.* **66** 899
[9] Vanderbilt D and Resta R 2006 *Conceptual Foundations of Materials: A Standard Model for Ground- and Excited-State Properties* ed S G Louie and M L Cohen (Amsterdam: Elsevier) p 139
[10] Resta R and Vanderbilt D 2007 *Physics of Ferroelectrics: A Modern Perspective* (*Topics in Applied Physics* vol 105) ed Ch H Ahn, K M Rabe and J-M Triscone (Berlin: Springer) p 31
[11] Xiao D, Shi J and Niu Q 2005 *Phys. Rev. Lett.* **95** 137204
[12] Thonhauser T, Ceresoli D, Vanderbilt D and Resta R 2005 *Phys. Rev. Lett.* **95** 137205
[13] http://www.abinit.org/
[14] http://www.cpmd.org/
[15] http://www.crystal.unito.it/
[16] http://www.quantum-espresso.org
[17] http://www.uam.es/departamentos/ciencias/fismateriac/siesta/
[18] http://cms.mpi.univie.ac.at/vasp/
[19] Resta R 2003 *Modelling Simul. Mater. Sci. Eng.* **11** R69
[20] Duan W H and Liu Z R 2006 *Curr. Opin. Solid State Mater. Sci.* **10** 40
[21] Rabe K M, Triscone J-M and Ahn Ch H (ed) 2007 *Physics of Ferroelectrics: A Modern Perspective* (*Topics in Applied Physics* vol 105) (Berlin: Springer)
[22] Marder M P 2000 *Condensed Matter Physics* (New York: Wiley)
[23] Ceresoli D, Thonhauser T, Vanderbilt D and Resta R 2006 *Phys. Rev.* B **74** 024408
[24] Xiao D, Yao Y, Fang Z and Niu Q 2006 *Phys. Rev. Lett.* **97** 026603
[25] Shi J, Vignale G, Xiao D and Niu Q 2007 *Phys. Rev. Lett.* **99** 197202
[26] Souza I and Vanderbilt D 2008 *Phys. Rev.* B **77** 054438
[27] Ceresoli D and Resta R 2007 *Phys. Rev.* B **76** 012405
[28] Thonhauser T, Ceresoli D, Mostofi A A, Marzari N, Resta R and Vanderbilt D 2009 *J. Chem. Phys.* **131** 101101
[29] Ceresoli D, Gerstmann U, Seitsonen A P and Mauri F 2010 *Phys. Rev.* B **81** 060409
[30] Mauri F, Pfrommer B G and Louie S G 1996 *Phys. Rev. Lett.* **77** 5300
[31] Souza I, Ìñìguez J and Vanderbilt D 2002 *Phys. Rev. Lett.* **89** 117602
[32] Umari P and Pasquarello A 2002 *Phys. Rev. Lett.* **89** 157602
[33] Kittel C 1996 *Introduction to Solid State Physics* 7th edn (New York: Wiley)
[34] Huang K 1950 *Proc. R. Soc.* A **203** 178
[35] Born M and Huang K 1954 *Dynamical Theory of Crystal Lattices* (Oxford: Oxford University Press)
[36] Mossotti O F 1850 *Mem. Mat. Fis. Soc. Ital. Sci. Res. Mod.* **24** 49
Clausius R 1879 *Die Mechanische Behandlung der Electrica* (Berlin: Vieweg)
[37] Umari P, Dal Corso A and Resta R 2001 *Fundamental Physics of Ferroelectrics: 2001 Williamsburg Workshop* ed H Krakauer (Woodbury, NY: AIP) p 107

[38] Ashcroft N W and Mermin N D 1976 *Solid State Physics* (Philadelphia, PA: Saunders)
[39] Martin R M 1974 *Phys. Rev.* B **9** 1998
[40] Meyer A J P and Asch G 1961 *J. Appl. Phys.* **32** S330
[41] Hirst L L 1997 *Rev. Mod. Phys.* **69** 607
[42] Vignale G and Rasolt M 1988 *Phys. Rev.* B **37** 10685
[43] Ebert H, Battocletti M and Gross E K U 1997 *Europhys. Lett.* **40** 545
[44] Pittalis S, Kurth S, Helbig N and Gross E K U 2006 *Phys. Rev.* A **74** 062511
[45] Sharma S, Pittalis S, Kurth S, Shallcross S, Dewhurst J K and Gross E K U 2006 *Phys. Rev.* B **76** 100401
[46] Gonze X, Ghosez Ph and Godby R W 1995 *Phys. Rev. Lett.* **74** 4035
[47] van Gisbergen S J A, Koostra F, Schipper P R T, Gritsenko O V, Snijders J G and Baerends E J 1998 *Phys. Rev.* A **57** 2556
[48] Mauri F and Louie S G 1996 *Phys. Rev. Lett.* **76** 4246
[49] Pickard C J and Mauri F 2001 *Phys. Rev.* B **63** 245101
Pickard C J and Mauri F 2003 *Phys. Rev. Lett.* **91** 196401
[50] Marques M A L, d'Avezac M and Mauri F 2006 *Phys. Rev.* B **73** 125433
[51] Yates J R, Pickard C J and Mauri F 2007 *Phys. Rev.* B **76** 024401
[52] Resta R, Posternak M and Baldereschi A 1993 *Phys. Rev. Lett.* **70** 1010
[53] Pasquarello A and Car R 1997 *Phys. Rev. Lett.* **79** 1766
[54] Silvestrelli P L, Bernasconi M and Parrinello M 1997 *Chem. Phys. Lett.* **277** 478
[55] Sharma M, Resta R and Car R 2005 *Phys. Rev. Lett.* **95** 187401
[56] Picozzi S and Ederer C 2009 *J. Phys.: Condens. Matter* **21** 303201
[57] Essin A M, Moore J E and Vanderbilt D 2009 *Phys. Rev. Lett.* **102** 146805
[58] Essin A M, Turner A M, Moore J E and Vanderbilt D 2010 arXiv:1002.0290v1
[59] Malashevich A, Souza I, Coh S and Vanderbilt D 2010 arXiv:1002.0300v1
[60] Martin R M 1972 *Phys. Rev.* B **5** 1607
[61] Car R and Parrinello M 1985 *Phys. Rev. Lett.* **55** 2471
[62] Baroni S and Resta R 1986 *Phys. Rev.* B **33** 7017
[63] de Gironcoli S, Baroni S and Resta R 1989 *Phys. Rev. Lett.* **62** 2853
[64] Baroni S, de Gironcoli S, Dal Corso A and Giannozzi P 2001 *Rev. Mod. Phys.* **73** 515
[65] Wu X, Vanderbilt D and Hamann D 2005 *Phys. Rev.* B **72** 035105
[66] Massidda S, Resta R, Posternak M and Baldereschi A 1995 *Phys. Rev.* B **52** 16977
[67] Massidda S, Posternak M, Baldereschi A and Resta R 1999 *Phys. Rev. Lett.* **82** 430
[68] Pickard C J and Mauri F 2002 *Phys. Rev. Lett.* **88** 086403
[69] Grant D M and Harris R K (ed) 1996 *Encyclopedia of NMR* (London: Wiley)
[70] Rabi I I, Zacharias J R, Millman S and Kusch P 1938 *Phys. Rev.* **53** 318
[71] Sebastiani D and Parrinello M 2001 *J. Phys. Chem.* A **105** 1951
[72] Sebastiani D and Parrinello M 2002 *ChemPhysChem* **3** 675
[73] Pfrommer B G, Mauri F and Louie S G 2000 *J. Am. Chem. Soc.* **122** 123
[74] Resta R, Posternak M and Baldereschi A 1993 *Mat. Res. Soc. Symp. Proc.* **291** 647
[75] Resta R 1999 *Int. J. Quantum Chem.* **75** 599
[76] Resta R 2000 *J. Phys.: Condens. Matter* **12** R107
[77] Pasquarello A and Resta R 2003 *Phys. Rev.* B **68** 174302
[78] Wannier G H 1937 *Phys. Rev.* **52** 191
[79] Marzari N and Vanderbilt D 1997 *Phys. Rev.* B **56** 12847
[80] Marzari N, Souza I and Vanderbilt D http://www.psi-k.org/newsletters/News_57/Highlight_57.pdf
[81] Mostofi A A, Lee Y-S, Souza I, Vanderbilt D and Marzari N 2008 *Comput. Phys. Commun.* **178** 685
[82] Resta R 2006 *J. Chem. Phys.* **124** 104104
[83] Blount E I 1962 *Solid State Physics* vol 13, ed H Ehrenreich, F Seitz and D Turnbull (New York: Academic) p 305
[84] Dabo I, Kozinsky B, Singh-Miller N E and Marzari N 2008 *Phys. Rev.* B **77** 115139
[85] Thouless D J, Kohmoto M, Nightingale M P and den Nijs M 1982 *Phys. Rev. Lett.* **49** 405
[86] Niu Q, Thouless D J and Wu Y S 1985 *Phys. Rev.* B **31** 3372
[87] Thouless D J 1998 *Topological Quantum Numbers in Nonrelativistic Physics* (Singapore: World Scientific)
[88] Haldane F D M 1988 *Phys. Rev. Lett.* **61** 2015
[89] Thonhauser T and Vanderbilt D 2006 *Phys. Rev.* B **74** 235111
[90] Taillefumier M, Dugaev V K, Canals B, Lacroix C and Bruno P 2008 *Phys. Rev.* B **78** 155330
[91] Sheng D N, Weng Z Y, Sheng L and Haldane F D M 2006 *Phys. Rev. Lett.* **97** 036808
[92] Zhang S-C 2008 *Physics* **1** 6
[93] Chen Y L *et al* 2009 *Science* **325** 178
[94] Moore J E 2009 *Physics* **2** 82
[95] Qi X L and Zhang S C 2010 *Phys. Today* **63** (1) 38
[96] Sundaram G and Niu Q 1999 *Phys. Rev.* B **59** 14915
[97] Karplus R and Luttinger J M 1954 *Phys. Rev.* **95** 1154
[98] Jungwirth T, Niu Q and MacDonald A H 2002 *Phys. Rev. Lett.* **88** 207208
[99] Fang Z *et al* 2003 *Science* **301** 92
[100] Yao Y, Kleinman L, MacDonald A H, Sinova J, Jungwirth T, Wang D-S, Wang E and Niu Q 2004 *Phys. Rev. Lett.* **92** 037204
[101] Wang X, Vanderbilt D, Yates J R and Souza I 2007 *Phys. Rev.* B **76** 195109
[102] Gat O and Avron J E 2003 *New J. Phys.* **5** 44
[103] Sai N, Rabe K M and Vanderbilt D 2002 *Phys. Rev.* B **66** 104108
[104] Souza I, Íñiguez J and Vanderbilt D 2004 *Phys. Rev.* B **69** 085106
[105] Ceresoli D and Tosatti E 2007 *Phys. Rev.* B **75** 161101
[106] Halperin B I 1982 *Phys. Rev.* B **25** 2185
[107] Yoshioka D 2002 *The Quantum Hall Effect* (Berlin: Springer)
[108] Cooper N R, Halperin B I and Ruzin I M 1997 *Phys. Rev.* B **55** 2344
[109] Culcer D, Sinova J, Sinitsyn N A, Jungwirth T, MacDonald A H and Niu Q 2004 *Phys. Rev. Lett.* **93** 046602