Theory of orbital magnetoelectric response

Andrei Malashevich[1,3]**, Ivo Souza**[1]**, Sinisa Coh**[2]
and David Vanderbilt[2]

[1] Department of Physics, University of California, Berkeley,
CA 94720-7300, USA
[2] Department of Physics and Astronomy, Rutgers University, Piscataway,
NJ 08854-8019, USA
E-mail: andreim@berkeley.edu

Abstract. We extend the recently developed theory of bulk orbital magnetization to finite electric fields, and use it to calculate the orbital magnetoelectric (ME) response of periodic insulators. Working in the independent-particle framework, we find that the finite-field orbital magnetization can be written as a sum of three gauge-invariant contributions, one of which has no counterpart at zero field. The extra contribution is collinear with and explicitly dependent on the electric field. The expression for the orbital magnetization is suitable for first-principles implementations, allowing one to calculate the ME response coefficients by numerical differentiation. Alternatively, perturbation-theory techniques may be used, and for that purpose we derive an expression directly for the linear ME tensor by taking the first field-derivative analytically. Two types of terms are obtained. One, the 'Chern–Simons' term, depends only on the unperturbed occupied orbitals and is purely isotropic. The other, 'Kubo' terms, involve the first-order change in the orbitals and give isotropic as well as anisotropic contributions to the response. In ordinary ME insulators all terms are generally present, while in strong Z_2 topological insulators only the Chern–Simons term is allowed, and is quantized. In order to validate the theory, we have calculated under periodic boundary conditions the linear ME susceptibility for a 3D tight-binding model of an ordinary ME insulator, using both the finite-field and perturbation-theory expressions. The results are in excellent agreement with calculations on bounded samples.

[3] Author to whom any correspondence should be addressed.

Contents

1. Introduction

In insulating materials in which both spatial inversion and time-reversal symmetries are broken, a magnetic field $\boldsymbol{B}$ can induce a first-order electric polarization $\boldsymbol{P}$, and conversely an electric field $\mathcal{E}$ can induce a first-order magnetization $\boldsymbol{M}$ [1, 2]. This linear magnetoelectric (ME) effect is described by the susceptibility tensor

$$\alpha_{da} = \left.\frac{\partial P_d}{\partial B_a}\right|_{\boldsymbol{B}=0} = \left.\frac{\partial M_a}{\partial \mathcal{E}_d}\right|_{\mathcal{E}=0}, \tag{1}$$

where indices label spatial directions. This tensor can be divided into a 'frozen-ion' contribution that occurs even when the ionic coordinates are fixed, and a 'lattice-mediated' contribution corresponding to the remainder. Each of these two contributions can be decomposed further according to whether the magnetic interaction is associated with spins or orbital currents, giving four contributions to α in total.

All of these contributions, except the frozen-ion orbital one, are relatively straightforward to evaluate, at least in principle, and *ab initio* calculations have started to appear. For example, the lattice-mediated spin-magnetization response was calculated in [3] for Cr_2O_3 and in [4] for $BiFeO_3$ (including the strain deformation effects that are present in the latter), and calculations based on the converse approach (polarization response to a Zeeman field) were recently reported [5]. One generally expects the lattice-mediated couplings to be larger than the frozen-ion ones, and insofar as the spin–orbit interaction can be treated perturbatively, interactions involving spin magnetization are typically larger than the orbital ones. However, we shall see that there are situations in which the spin–orbit interaction cannot be treated perturbatively, and in which the frozen-ion orbital contribution is expected to be dominant. Therefore, it is desirable to have a complete description that accounts for all four contributions.

The frozen-ion orbital contribution is, in fact, the one part of ME susceptibility for which there is at present no satisfactory theoretical or computational framework, although some progress towards that goal was made in two recent works [6, 7]. Following Essin *et al* [7], we refer to it as the 'orbital magnetoelectric polarizability' (OMP). For the remainder of

this paper, we will focus exclusively on this contribution to (1), and shall denote it simply by α. Accordingly, the symbol $\boldsymbol{M}$ will be used henceforth for the orbital component of the magnetization.

The question we pose to ourselves is the following: what is the quantum-mechanical expression for the tensor α of a generic three-dimensional (3D) band insulator? We note that the conventional perturbation-theory expression for α [8, 9] does not apply to Bloch electrons, as it involves matrix elements of unbounded operators. The proper expressions for $\boldsymbol{P}$ [10] and $\boldsymbol{M}$ [11]–[14] in periodic crystals have been derived, but so far only at $\boldsymbol{B} = 0$ and $\boldsymbol{\mathcal{E}} = 0$, respectively. The evaluation of equation (1) therefore remains an open problem.

Phenomenologically, the most general form of α is a 3×3 matrix where all nine components are independent. Dividing it into traceless and isotropic parts, the latter is conveniently expressed in terms of a single dimensionless parameter θ as

$$\alpha_{da}^{\theta} = \frac{\theta e^2}{2\pi hc} \delta_{da}. \tag{2}$$

The presence of an isotropic ME coupling is equivalent to the addition of a term proportional to $\theta \boldsymbol{\mathcal{E}} \cdot \boldsymbol{B}$ to the electromagnetic Lagrangian. Such a term describes 'axion electrodynamics' [15] and (2) may therefore also be referred to as the 'axion OMP'. The electrodynamic effects of the axion field are elusive (in fact, the very existence of α^{θ} was debated until recently: see [16, 17] and references therein). For example, in a finite, static sample cut from a uniform ME medium, those effects are only felt at the surface [15, 18]. In particular, α^{θ} gives rise to a surface Hall effect [19].

An essential feature of the axion theory is that a change of θ by 2π leaves the electrodynamics invariant [15]. The profound implications for the ME response of materials were recognized by Qi *et al* [6], and discussed further by Essin *et al* [7]. These authors showed that there is a part of the isotropic OMP that remains ambiguous up to integer multiples of 2π in the corresponding θ until the surface termination of the sample is specified. For example, a change by $2\pi n$ occurs if the surface is modified by adsorbing a quantum anomalous Hall layer. Hence this particular contribution to θ can be formulated as a bulk quantity only modulo a quantum of indeterminacy, in much the same way as the electric polarization $\boldsymbol{P}$ [10, 20]. A microscopic expression for it was derived in the framework of single-particle band theory by the above authors. It is given by the Brillouin-zone integral of the Chern–Simons form [21] in k-space, which is a multivalued global geometric invariant reminiscent of the Berry-phase expression for $\boldsymbol{P}$ [10]. We denote henceforth this 'geometric' contribution to the OMP as the Chern–Simons OMP (CSOMP).

A remarkable outcome of this analysis is the prediction [6] of a purely isotropic 'topological ME effect', associated with the CSOMP, in a newly discovered class of time-reversal invariant insulators known as Z_2 topological insulators [22]–[24]. As a result of the multivaluedness of θ, the presence of time-reversal symmetry in the bulk, which takes θ into $-\theta$, is consistent with two solutions: $\theta = 0$, corresponding to ordinary insulators, and $\theta = \pi$, corresponding to strong Z_2 topological insulators[4]. The latter case is non-perturbative in the spin–orbit interaction, and $\theta = \pi$ amounts to a rather large ME susceptibility (in Gaussian units

---

[4] An analogous situation occurs in the theory of polarization: inversion symmetry, which takes $\boldsymbol{P}$ into $-\boldsymbol{P}$, allows for a nontrivial solution, which does not include $\boldsymbol{P} = 0$ in the 'lattice' of values [20]. An important difference is that while θ is a directly measurable response, only *changes* in $\boldsymbol{P}$ are detectable, so that the experimental implications of the nontrivial solution are less clear in this case.

it is $1/4\pi$ times the fine structure constant, or $\sim 6 \times 10^{-4}$, to be compared with $\sim 1 \times 10^{-4}$ for the total ME response of Cr_2O_3 at low temperature [25]).

It is not clear from these recent works, however, whether the isotropic CSOMP constitutes the full OMP response of a generic insulator. It does appear to do so for the tight-binding model studied in [7], whose ME response was correctly reproduced by the Chern–Simons expression even when the parameters were tuned to break time-reversal and inversion symmetries (i.e. for generic θ not equal to 0 or π). On the other hand, other considerations seem to demand additional contributions. For example, it is not difficult to construct tight-binding models of molecular crystals in which it is clear that the OMP cannot be purely isotropic.

In this work we derive, using rigorous quantum-mechanical arguments, an expression for the OMP tensor α of band insulators, written solely in terms of bulk quantities (the periodic Hamiltonian and ground state Bloch wavefunctions, and their first-order change in an electric field). We restrict our derivation to non-interacting Hamiltonians, as the essential physics we wish to describe occurs already at the single-particle level. We find that in crystals with broken time-reversal and inversion symmetries there are, in addition to the CSOMP term discussed in [6, 7], extra terms that generally contribute to both the trace and the traceless parts of α.

Our theoretical approach closely mimics one type of ME response experiment: a finite electric field $\mathcal{E}$ is applied to a bounded sample, and the (orbital) magnetization is calculated in the presence of the field. Then the thermodynamic limit is taken at fixed field. This key step in the derivation must be done carefully, so that crucial surface contributions are not lost in the process, and here we follow the Wannier-based approach of references [12, 13], adapted to $\mathcal{E} \neq 0$. Finally, the linear response coefficient $\alpha_{da} = \partial M_a / \partial \mathcal{E}_d$ is extracted in the limit that $\mathcal{E}$ goes to zero.

In a concurrent work by Essin *et al* and one of us [26], an alternative approach was taken, which is closer in spirit to the calculation in [10] of the change in polarization as an integrated current: the adiabatic current induced in an infinite crystal by a change in its Hamiltonian in the presence of a magnetic field is computed, and then expressed as a total time derivative. The two approaches are complementary and lead to the same expression for α, illuminating it from different angles.

The paper is organized as follows. In section 2, we derive the bulk expression for $\boldsymbol{M}(\mathcal{E})$, and reorganize it into three gauge-invariant contributions, one of which yields directly the CSOMP response. The gauge-invariant decomposition of $\boldsymbol{M}(\mathcal{E})$ is done at first in k-space for periodic crystals, and then also for bounded samples working in real space. In section 3, we derive a k-space formula for the OMP tensor α by taking analytically the field derivative of $\boldsymbol{M}(\mathcal{E})$. Numerical tests on a tight-binding model of a ME insulator are presented at appropriate places throughout the paper in order to validate the bulk expressions for $\boldsymbol{M}(\mathcal{E})$ and α. In appendix A, we describe the tight-binding model, as well as technical details on how the various formulas are implemented on a k-point grid. Appendices B and C contain derivations of certain results given in the main text.

2. Orbital magnetization in finite electric field

2.1. Preliminaries

The orbital magnetization $\boldsymbol{M}$ is defined as the orbital moment per unit volume,

$$\boldsymbol{M} = -\frac{e}{2cV} \sum_i \langle \psi_i | \boldsymbol{r} \times \boldsymbol{v} | \psi_i \rangle. \tag{3}$$

Here $e > 0$ is the magnitude of the electron charge, V is the sample volume and $|\psi_i\rangle$ are the occupied eigenstates. While this expression can be directly implemented when using open boundary conditions, the electronic structure of crystals is more conveniently calculated and interpreted using periodic boundary conditions, in order to take advantage of Bloch's theorem. This poses, however, serious difficulties in dealing with the circulation operator $\boldsymbol{r} \times \boldsymbol{v}$, because of the unbounded and nonperiodic nature of the position operator $\boldsymbol{r}$. These subtle issues were fully resolved only recently, with the derivation of a bulk expression for $\boldsymbol{M}$ directly in terms of the extended Bloch states [11]–[14].

In previous derivations, the crystal was taken to be under shorted electrical boundary conditions. We shall extend the derivation given in [12, 13] to the case where a static homogeneous electric field $\boldsymbol{\mathcal{E}}$ is present, so that the full Hamiltonian reads

$$\mathcal{H} = \mathcal{H}^0 + e\boldsymbol{\mathcal{E}} \cdot \boldsymbol{r}. \tag{4}$$

The derivation, carried out for an insulator with N valence bands within the independent-particle approximation, involves transforming the set of occupied eigenstates $|\psi_i\rangle$ of $\mathcal{H}$ into a set of Wannier-type (i.e. localized and orthonormal) orbitals $|w_i\rangle$ and expressing $\boldsymbol{M}(\boldsymbol{\mathcal{E}})$ in the Wannier representation. This is done at first for a finite sample cut from a periodic crystal, and eventually the thermodynamic limit is taken at fixed field.

Before continuing, two remarks are in order. Firstly, the assumption that it is possible to construct well-localized Wannier functions (WFs) spanning the valence bands is only valid if the Chern invariants of the valence bandstructure vanish identically [27]. This requirement is satisfied by normal band insulators as well as by Z_2 topological insulators, but not by quantum anomalous Hall insulators [28], which thus far remain hypothetical. Secondly, because of Zener tunnelling, an insulating crystal does not have a well-defined ground state in a finite electric field. Nonetheless, upon slowly ramping up the field to the desired value, the electron system remains in a quasistationary state that is, for all practical purposes, indistinguishable from a truly stationary state. This is the state we shall consider in the ensuing derivation. As discussed in [29, 30], it is Wannier- and Bloch-representable, even though the Hamiltonian (4) is not lattice-periodic.

2.2. k-space expression

Our derivation of a k-space (bulk) expression for $\boldsymbol{M}(\boldsymbol{\mathcal{E}})$ is carried out mostly in real space, using a Wannier representation. It is only in the last step that we switch to reciprocal space, by expressing the crystalline WFs $|\boldsymbol{R}n\rangle$ in terms of the cell-periodic Bloch functions $|u_{n\boldsymbol{k}}\rangle$ via [31]

$$|\boldsymbol{R}n\rangle = V_c \int [\mathrm{d}k] \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{R})} |u_{n\boldsymbol{k}}\rangle, \tag{5}$$

where $\boldsymbol{R}$ is a lattice vector, V_c is the unit-cell volume, $[\mathrm{d}k] \equiv \mathrm{d}^3k/(2\pi)^3$ and the integral is over the first Brillouin zone.

We begin with a finite sample immersed in a field $\boldsymbol{\mathcal{E}}$, divide it up into an interior region and a surface region, and assign each WF to either one. The boundary between the two regions is chosen in such a way that the fractional volume of the surface region goes to zero as $V \to \infty$, but deep enough that WFs near the boundary are bulk-like. Following [12, 13], equation (3) for the orbital magnetization can then be rewritten as an interior contribution plus a surface contribution, denoted respectively as the 'local circulation' (LC) and the 'itinerant circulation'

(IC). Remarkably, in the thermodynamic limit *both* can be expressed solely in terms of the interior-region crystalline WFs or, equivalently, in terms of the bulk Bloch functions, as shown in the above references at $\mathcal{E} = 0$ and below for $\mathcal{E} \neq 0$. Specifically, we shall show that

$$M = M^{\mathrm{LC}} + M^{\mathrm{IC},0} + M^{\mathrm{IC},\mathcal{E}}, \tag{6a}$$

where

$$M_a^{\mathrm{LC}} = -\gamma \epsilon_{abc} \,\mathrm{Im} \sum_n^N \int [\mathrm{d}k] \langle \partial_b u_{n\boldsymbol{k}} | H_{\boldsymbol{k}}^0 | \partial_c u_{n\boldsymbol{k}} \rangle \tag{6b}$$

is the contribution from the interior WFs,

$$M_a^{\mathrm{IC},0} = -\gamma \epsilon_{abc} \,\mathrm{Im} \sum_{nm}^N \int [\mathrm{d}k] \langle \partial_b u_{n\boldsymbol{k}} | \partial_c u_{m\boldsymbol{k}} \rangle H_{mn\boldsymbol{k}}^0 \tag{6c}$$

is the part of the surface contribution coming from the zero-field Hamiltonian, and

$$M_a^{\mathrm{IC},\mathcal{E}} = -\gamma \epsilon_{abc} \,\mathrm{Im} \sum_{nm}^N \int [\mathrm{d}k] \langle \partial_b u_{n\boldsymbol{k}} | \partial_c u_{m\boldsymbol{k}} \rangle e\mathcal{E} \cdot \boldsymbol{A}_{mn\boldsymbol{k}} \tag{6d}$$

is the part of the surface contribution coming from the electric field term in the Hamiltonian (4). In the above expressions $\gamma = -e/(2\hbar c)$,

$$H_{\boldsymbol{k}}^0 = \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \mathcal{H}^0 \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}, \tag{7}$$

$$H_{mn\boldsymbol{k}}^0 = \langle u_{m\boldsymbol{k}} | H_{\boldsymbol{k}}^0 | u_{n\boldsymbol{k}} \rangle \tag{8}$$

and $\boldsymbol{A}_{mn\boldsymbol{k}}$ is the Berry connection matrix defined in equation (14) below.

Having stated the result, we now present the derivation, starting with the interior contribution M^{LC}. Using $[r_i, r_j] = 0$, the velocity operator $\boldsymbol{v} = (\mathrm{i}/\hbar)[\mathcal{H}, \boldsymbol{r}]$ becomes $(\mathrm{i}/\hbar)[\mathcal{H}^0, \boldsymbol{r}]$, so that the circulation operator $\boldsymbol{r} \times \boldsymbol{v}$ is unaffected by the electric field. It immediately follows that the local circulation part M^{LC} is given in terms of the field-polarized states $|u_{n\boldsymbol{k}}\rangle$ by the same expression, equation (6b), as was derived in [13] for the zero-field case.

Consider now the contribution $M^{\mathrm{IC}} = M^{\mathrm{IC},0} + M^{\mathrm{IC},\mathcal{E}}$ from the surface WFs $|w_s\rangle$. For large samples it takes the form [13]

$$M^{\mathrm{IC}} = -\frac{e}{2cN_{\mathrm{c}}V_{\mathrm{c}}} \sum_s^{\mathrm{surf}} \boldsymbol{r}_s \times \boldsymbol{v}_s, \tag{9}$$

where N_{c} is the number of crystal cells of volume V_{c}, $\boldsymbol{r}_s = \langle w_s | \boldsymbol{r} | w_s \rangle$ and

$$\boldsymbol{v}_s = \langle w_s | \boldsymbol{v} | w_s \rangle = \frac{2}{\hbar} \,\mathrm{Im} \langle w_s | \boldsymbol{r} \mathcal{H} | w_s \rangle. \tag{10}$$

Note that $\mathcal{H} | w_s \rangle$ already belongs to the occupied manifold spanned by $P = \sum_j^{\mathrm{occ}} |w_j\rangle\langle w_j|$, since we assume a (quasi)stationary state. Thus, we can insert a P between $\boldsymbol{r}$ and $\mathcal{H}$ above, and using (4) we obtain

$$\boldsymbol{v}_s = \sum_j^{\mathrm{occ}} \left(\boldsymbol{v}_{\langle js \rangle}^0 + \boldsymbol{v}_{\langle js \rangle}^{\mathcal{E}} \right), \tag{11}$$

where $\boldsymbol{v}_{\langle js \rangle}^0 = (2/\hbar) \,\mathrm{Im}[\boldsymbol{r}_{sj} \mathcal{H}_{js}^0]$ is the same as in [12, 13] and $\boldsymbol{v}_{\langle js \rangle}^{\mathcal{E}} = (2e/\hbar) \,\mathrm{Im}[\boldsymbol{r}_{sj}(\boldsymbol{r}_{js} \cdot \mathcal{E})]$ is a new term.

The reasoning [12, 13] by which M^{IC} can be recast in terms of the bulk WFs $|\boldsymbol{R}n\rangle$ relies on the exponential localization of the WFs and on certain properties of $\boldsymbol{v}^0_{\langle js\rangle}$ (antisymmetry under $j \leftrightarrow s$ and invariance under lattice translations deep inside the crystallite) which are shared by $\boldsymbol{v}^{\mathcal{E}}_{\langle js\rangle}$. Hence we can follow similar steps as in those works, arriving at

$$M_a^{\mathrm{IC},\mathcal{E}} = \frac{e}{4cV_{\mathrm{c}}}\epsilon_{abc}\sum_{\boldsymbol{R}}\sum_{mn}^{N} v^{\mathcal{E}}_{\langle 0m,\boldsymbol{R}n\rangle,b}R_c, \tag{12}$$

and similarly for $M_a^{\mathrm{IC},0}$ with v^0 substituting for $v^{\mathcal{E}}$. The latter is identical to the expression for M_a^{IC} valid at $\mathcal{E} = 0$ [12, 13], and upon converting to k-space becomes (6c).

Let us now turn to $M_a^{\mathrm{IC},\mathcal{E}}$ and write (12) as $(e^2/2c\hbar V_{\mathrm{c}})\epsilon_{abc}\mathcal{E}_d$ Im $W_{bd,c}$ where

$$W_{bd,c} = \sum_{\boldsymbol{R}}\sum_{mn}^{N} \langle \boldsymbol{R}n|r_b|\boldsymbol{0}m\rangle\langle\boldsymbol{0}m|r_d|\boldsymbol{R}n\rangle R_c. \tag{13}$$

In order to recast this expression as a k-space integral, it is useful to introduce the $N \times N$ Berry connection matrix

$$A_{mn\boldsymbol{k},b} = \mathrm{i}\langle u_{m\boldsymbol{k}}|\partial_b u_{n\boldsymbol{k}}\rangle = A^*_{nm\boldsymbol{k},b}, \tag{14}$$

where $\partial_b \equiv \partial/\partial k_b$. It satisfies the relation [31, 32]

$$\langle \boldsymbol{R}n|r_b|\boldsymbol{0}m\rangle = V_c \int [\mathrm{d}k]A_{nm\boldsymbol{k},b}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}}. \tag{15}$$

We also need

$$R_c\langle \boldsymbol{R}n|r_d|\boldsymbol{0}m\rangle = \mathrm{i}V_c \int [\mathrm{d}k](\partial_c A_{nm\boldsymbol{k},d})\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}}, \tag{16}$$

which follows from (15). Using these two relations, (13) becomes

$$W_{bd,c} = \mathrm{i}V_{\mathrm{c}}\sum_{mn}^{N}\int [\mathrm{d}k]A_{mn\boldsymbol{k},d}\partial_c A_{nm\boldsymbol{k},b}, \tag{17}$$

and we arrive at (6d).

The sum of equations (6b)–(6d) gives the desired k-space expression for $\boldsymbol{M}(\mathcal{E})$. In the limit $\mathcal{E} \to 0$, the term $\boldsymbol{M}^{\mathrm{IC},\mathcal{E}}$ vanishes, and equation (31) of [13] is recovered.

We have implemented (6b)–(6d) for the tight-binding model of appendix A. Since for small electric fields $\boldsymbol{M}(\mathcal{E})$ differs only slightly from $\boldsymbol{M}(0)$, in order to observe the effect of the electric field we consider differences in magnetization rather than the absolute magnetization. Therefore, in all our numerical tests we evaluated the OMP tensor α_{da}. With the help of (6b)–(6d) we calculated it as $\Delta M_a/\Delta\mathcal{E}_d$, using small fields $\mathcal{E}_d = \pm 0.01$. We then repeated the calculation on finite samples cut from the bulk crystal, using (3) in place of (6b)–(6d). Figure 1 shows the value of the zz and zy components of α plotted as a function of the parameter φ, the phase of one of the complex hopping amplitudes (see appendix A for details). The very precise agreement between the solid and dashed lines confirms the correctness of the k-space formula. The same level of agreement was found for the other components of α.
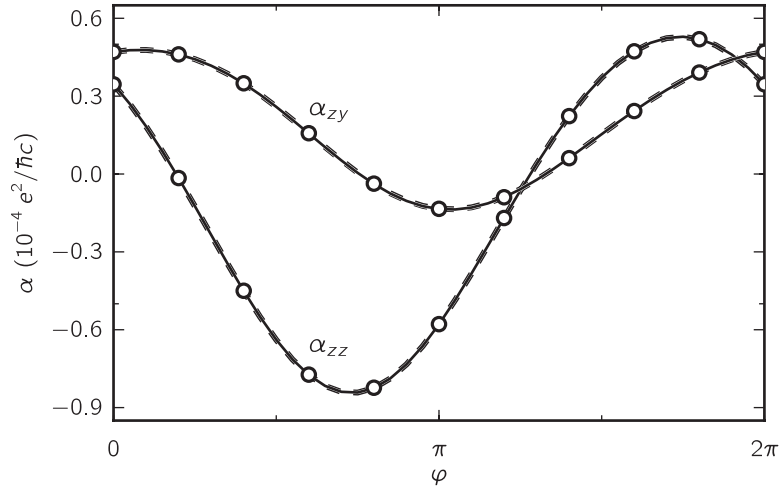
Figure 1. The zz and zy components of the OMP tensor α of the tight-binding model described in appendix A, as a function of the parameter φ. The two lower bands are treated as occupied. Solid line: extrapolation from finite-size samples using numerical differentiation of the finite-field magnetization calculated from (3). Dashed line: numerical differentiation of the finite-field magnetization calculated using (6b)–(6d) discretized on a k-space grid. Open circles: linear-response calculation in k-space using discretized versions of (47a)–(47c).

2.3. Gauge-invariant decomposition

2.3.1. Periodic crystals. Equation (6a) for $\boldsymbol{M}(\mathcal{E})$ is valid in an arbitrary gauge, that is, the sum of its three terms given by (6b)–(6d)—but not each term individually—remains invariant under a unitary transformation

$$|u_{n\boldsymbol{k}}\rangle \rightarrow \sum_{m}^{N} |u_{m\boldsymbol{k}}\rangle U_{mn\boldsymbol{k}} \tag{18}$$

among the valence-band states at each $\boldsymbol{k}$. In order to make the gauge invariance of (6a) manifest, it is convenient to first manipulate it into a different form, given in terms of certain canonical objects, which we now define. We begin by introducing the covariant k-derivative of a valence state [30],

$$|\tilde{\partial}_b u_{n\boldsymbol{k}}\rangle = Q_{\boldsymbol{k}}\, |\partial_b u_{n\boldsymbol{k}}\rangle, \tag{19}$$

where $Q_{\boldsymbol{k}} = 1 - P_{\boldsymbol{k}}$ and

$$P_{\boldsymbol{k}} = \sum_{j=1}^{N} |u_{j\boldsymbol{k}}\rangle\langle u_{j\boldsymbol{k}}|. \tag{20}$$

The covariant and ordinary derivatives are related by

$$|\partial_b u_{n\boldsymbol{k}}\rangle = |\tilde{\partial}_b u_{n\boldsymbol{k}}\rangle - \mathrm{i}\sum_{m}^{N} A_{mn\boldsymbol{k},b}|u_{m\boldsymbol{k}}\rangle. \tag{21}$$

IOP Institute of Physics Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

The generalized metric-curvature tensor is [31]

$$F_{nm\mathbf{k},bc} = \langle \tilde{\partial}_b u_{n\mathbf{k}} | \tilde{\partial}_c u_{m\mathbf{k}} \rangle = F^*_{mn\mathbf{k},cb}. \tag{22}$$

Viewed as an $N \times N$ matrix over the band indices, F is gauge-covariant, changing as

$$F_{nm\mathbf{k},bc} \rightarrow (U^\dagger_{\mathbf{k}} F_{\mathbf{k},bc} U_{\mathbf{k}})_{nm} \tag{23}$$

under the transformation (18). We also note the relation

$$\langle \partial_b u_{n\mathbf{k}} | \partial_c u_{m\mathbf{k}} \rangle = F_{nm\mathbf{k},bc} + (A_{\mathbf{k},b} A_{\mathbf{k},c})_{nm}. \tag{24}$$

We shall make use of two more gauge-covariant objects,

$$H^0_{nm\mathbf{k},b} = \mathrm{i}\langle u_{n\mathbf{k}} | H^0_{\mathbf{k}} | \tilde{\partial}_b u_{m\mathbf{k}} \rangle \tag{25}$$

and

$$H^0_{nm\mathbf{k},bc} = \langle \tilde{\partial}_b u_{n\mathbf{k}} | H^0_{\mathbf{k}} | \tilde{\partial}_c u_{m\mathbf{k}} \rangle, \tag{26}$$

which enter the relation

$$\langle \partial_b u_{n\mathbf{k}} | H^0_{\mathbf{k}} \partial_c u_{m\mathbf{k}} \rangle = H^0_{nm\mathbf{k},bc} + \left[A_{\mathbf{k},b} H^0_{\mathbf{k},c} + \left(H^0_{\mathbf{k},b} \right)^\dagger A_{\mathbf{k},c} + A_{\mathbf{k},b} H^0_{\mathbf{k}} A_{\mathbf{k},c} \right]_{nm}. \tag{27}$$

Coming back to equations (6a)–(6d), for M^{LC}_a we use (27) and for M^{IC}_a we use (24), leading to

$$M_a = -\gamma \epsilon_{abc} \int [\mathrm{d}k] \, \mathrm{Im} \, \mathrm{tr}\left[H^0_{bc} + 2A_b H^0_c + H^0 F_{bc} + e\mathcal{E}_d A_d F_{bc} + e\mathcal{E}_d A_d A_b A_c \right], \tag{28}$$

where 'tr' denotes the electronic trace over the occupied valence bands and we have dropped the subscript $\mathbf{k}$. The second term can be rewritten using

$$H^0_{nm,c} = -e\mathcal{E}_d F_{nm,dc}. \tag{29}$$

(To obtain this relation start from the generalized Schrödinger equation satisfied by the valence states at $\mathcal{E} \neq 0$ [33],

$$H^0 | u_n \rangle = \sum_{m}^{N} (H^0_{mn} + e\mathcal{E} \cdot A_{mn}) | u_m \rangle - \mathrm{i}e\mathcal{E}_d | \partial_d u_n \rangle, \tag{30}$$

and multiply through by $\langle \tilde{\partial}_c u_m |$.)

Let us define the quantities

$$\widetilde{M}^{\mathrm{LC}}_a = -\gamma \epsilon_{abc} \int [\mathrm{d}k] \, \mathrm{Im} \, \mathrm{tr}\left[H^0_{bc} \right], \tag{31}$$

$$\widetilde{M}^{\mathrm{IC}}_a = -\gamma \epsilon_{abc} \int [\mathrm{d}k] \, \mathrm{Im} \, \mathrm{tr}\left[H^0 F_{bc} \right] \tag{32}$$

and

$$M^{\mathrm{CS}}_a = -e\gamma \epsilon_{abc}\mathcal{E}_d \int [\mathrm{d}k] \, \mathrm{Im} \, \mathrm{tr}\left[2A_b F_{cd} + F_{bc} A_d + A_b A_c A_d \right]. \tag{33}$$

The total magnetization is given by their sum

$$M_a = \widetilde{M}_a^{\mathrm{LC}} + \widetilde{M}_a^{\mathrm{IC}} + M_a^{\mathrm{CS}}. \tag{34a}$$

Referring to (22) and (26) the first two terms read, in a more conventional notation,

$$\widetilde{M}_a^{\mathrm{LC}} = -\gamma\,\epsilon_{abc} \int [\mathrm{d}k] \sum_n^N \mathrm{Im}\langle\tilde{\partial}_b u_{nk}|H_{\boldsymbol{k}}^0|\tilde{\partial}_c u_{nk}\rangle \tag{34b}$$

and

$$\widetilde{M}_a^{\mathrm{IC}} = -\gamma\,\epsilon_{abc} \int [\mathrm{d}k] \sum_{nm}^N \mathrm{Im}\left(H_{nmk}^0\langle\tilde{\partial}_b u_{mk}|\tilde{\partial}_c u_{nk}\rangle\right). \tag{34c}$$

These are the only terms that remain in the limit $\mathcal{E} \to 0$, in agreement with equation (43) of [13]. At finite field they depend on $\mathcal{E}$ implicitly via the wavefunctions.

We now show that the term $\boldsymbol{M}^{\mathrm{CS}}$, which gathers all the contributions with an explicit dependence on $\mathcal{E}$, can be recast as

$$M_a^{\mathrm{CS}} = e\gamma\,\mathcal{E}_a \int [\mathrm{d}k]\epsilon_{ijk}\,\mathrm{tr}\left[A_i\partial_j A_k - \frac{2\mathrm{i}}{3}A_i A_j A_k\right]. \tag{34d}$$

To do so it is convenient to introduce the Berry curvature tensor

$$\Omega_{nm,ab} = \mathrm{i}F_{nm,ab} - \mathrm{i}F_{nm,ba} = -\Omega_{nm,ba}, \tag{35}$$

where $F_{nm,ab}$ was defined in (22). A few lines of algebra show that

$$\Omega_{nm,ab} = \partial_a A_{nm,b} - \partial_b A_{nm,a} - \mathrm{i}[A_a, A_b]_{nm}. \tag{36}$$

In order to go from (33) to (34d), use (35) to write $\mathrm{Im}\,\mathrm{tr}[F_{bc}A_d]$ as $-\frac{1}{2}\,\mathrm{tr}[A_d\Omega_{bc}]$ and $-2\,\mathrm{Im}\,\mathrm{tr}[A_bF_{dc}]$ as $\mathrm{tr}[A_b\Omega_{dc}]$, and then replace $\Omega_{nm,bc}$ in these expressions with $\epsilon_{abc}\Omega_{nm,a}$, where $\Omega_{nm,a} = \frac{1}{2}\epsilon_{abc}\Omega_{nm,bc}$ is the Berry curvature tensor written in axial-vector form. This leads to

$$M_a^{\mathrm{CS}} = e\gamma \int [\mathrm{d}k](\mathcal{E}_a\,\mathrm{tr}[\boldsymbol{\Omega}\cdot\boldsymbol{A}] - \mathcal{E}_d\epsilon_{abc}\,\mathrm{Im}\,\mathrm{tr}[A_bA_cA_d]). \tag{37}$$

The first term is parallel to the field, and can be rewritten with the help of (36):

$$\mathrm{tr}[\boldsymbol{\Omega}\cdot\boldsymbol{A}] = \epsilon_{ijk}\,\mathrm{tr}[A_i\partial_j A_k - \mathrm{i}A_i A_j A_k]. \tag{38}$$

While not immediately apparent, the second term in (37) also points along the field. To see this, write

$$\sum_{bcd} \mathcal{E}_d\epsilon_{abc}\,\mathrm{Im}\,\mathrm{tr}[A_bA_cA_d] = \mathcal{E}_a \sum_{bc} \epsilon_{abc}\,\mathrm{Im}\,\mathrm{tr}[A_aA_bA_c] + \sum_{d\neq a}\sum_{bc} \epsilon_{abc}\,\mathrm{Im}\,\mathrm{tr}[A_bA_cA_d], \tag{39}$$

where we suspended momentarily the implied summation convention. The last term vanishes because the factor ϵ_{abc} forces $d \neq a$ to equal either b or c, producing terms such as $\mathrm{Im}\,\mathrm{tr}[A_bA_bA_c]$ which vanish identically as A_b is Hermitian. Rewriting $\mathcal{E}_a \sum_{bc} \epsilon_{abc}\,\mathrm{Im}\,\mathrm{tr}[A_aA_bA_c]$ as $(\mathcal{E}_a/3)\sum_{ijk} \epsilon_{ijk}\,\mathrm{Im}\,\mathrm{tr}[A_iA_jA_k]$ and restoring the summation convention, we arrive at (34d).

Equations (34b)–(34d), which constitute the main result of this section, are separately gauge invariant. For $\widetilde{\boldsymbol{M}}^{\mathrm{LC}}$ and $\widetilde{\boldsymbol{M}}^{\mathrm{IC}}$ this is apparent already from (31) and (32), whose *integrands* are gauge-invariant, being traces over gauge-covariant matrices. In contrast, equation (34d) for
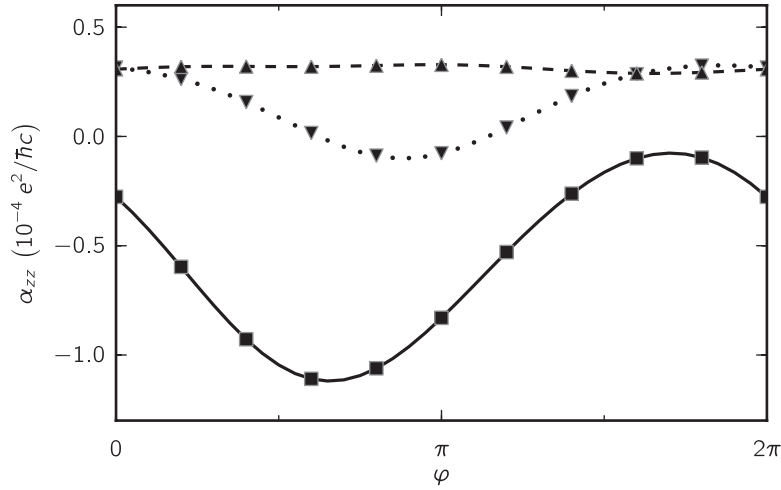
Figure 2. Decomposition of the α_{zz} curve in figure 1 into the gauge-invariant contributions $\tilde{\alpha}_{zz}^{\mathrm{LC}}$ (solid lines), $\tilde{\alpha}_{zz}^{\mathrm{IC}}$ (dashed line) and $\alpha_{zz}^{\mathrm{CS}}$ (dotted line) calculated in k-space using finite differences in $\mathcal{E}$. Symbols denote the same contributions evaluated for bounded samples, also using finite differences.

$\boldsymbol{M}^{\mathrm{CS}}$ only becomes invariant after taking the integral on the right-hand side over the entire Brillouin zone (the integrand being familiar from differential geometry as the Chern–Simons 3-form [21, 34]).

The Chern–Simons contribution (34d) has several remarkable features: (i) as already noted, it is perfectly isotropic, remaining parallel to $\mathcal{E}$ for arbitrary orientations of $\mathcal{E}$ relative to the crystal axes; (ii) being isotropic, it vanishes in less than three dimensions, which intuitively can be understood because already in two dimensions polarization must be in the plane of the system and magnetization must be out of the plane; (iii) for $N > 1$ valence bands it is a multivalued bulk quantity with a quantum of arbitrariness $(e^2/hc)\mathcal{E}_a$, a fact that is connected with the possibility of a cyclic adiabatic evolution that would change (47a) below for θ by 2π [6].

We have repeated the calculation of the OMP presented in figure 1 using (34b)–(34d) instead of (6b)–(6d), finding excellent agreement between them. The electric field derivative of the decomposition (34a) gives the corresponding decomposition of the OMP tensor (1),

$$\alpha = \tilde{\alpha}^{\mathrm{LC}} + \tilde{\alpha}^{\mathrm{IC}} + \alpha^{\mathrm{CS}}, \tag{40}$$

where each term is also gauge-invariant. The zz components of these terms are plotted separately in figure 2.

2.3.2. Finite samples. It is natural to ask whether the gauge-invariant decomposition of the orbital magnetization given in equation (34a) can be made already for finite samples, before taking the thermodynamic limit and switching to periodic boundary conditions. This has previously been done in the case $\mathcal{E} = 0$, where $\boldsymbol{M}^{\mathrm{CS}} = 0$ and $\widetilde{\boldsymbol{M}}^{\mathrm{LC}}$ and $\widetilde{\boldsymbol{M}}^{\mathrm{IC}}$ take the form [35]

$$\widetilde{M}_a^{\mathrm{LC}} = \frac{e}{2\hbar cV}\epsilon_{abc} \operatorname{Im}\operatorname{Tr}[Pr_b Q\mathcal{H}^0 Qr_c] \tag{41a}$$

and

$$\widetilde{M}_a^{\mathrm{IC}} = \frac{e}{2\hbar cV}\epsilon_{abc} \operatorname{Im}\operatorname{Tr}[P\mathcal{H}^0 Pr_b Qr_c]. \tag{41b}$$

Here P and $Q = 1 - P$ are the projection operators onto the occupied and empty subspaces, respectively, and 'Tr' denotes the electronic trace over the entire Hilbert space. These two expressions, which are manifestly gauge-invariant, remain valid at finite field, reducing to (34b) and (34c) in the thermodynamic limit.

We now complete this picture for $\mathcal{E} \neq 0$ by showing that the remaining contribution $\boldsymbol{M}^{\mathrm{CS}} = \boldsymbol{M} - \widetilde{\boldsymbol{M}}^{\mathrm{LC}} - \widetilde{\boldsymbol{M}}^{\mathrm{IC}}$ can also be written in trace form, as

$$M_a^{\mathrm{CS}} = -\frac{e^2}{3\hbar cV}\mathcal{E}_a \epsilon_{ijk}\, \mathrm{Im}\, \mathrm{Tr}\,[Pr_i Pr_j Pr_k]. \tag{41c}$$

We first recast the orbital magnetization (3) as

$$M_a = -\frac{e}{2cV}\epsilon_{abc}\mathrm{Tr}\,[Pr_b v_c] = \frac{e}{2\hbar cV}\epsilon_{abc}\, \mathrm{Im}\, \mathrm{Tr}\,\left[Pr_b \mathcal{H}^0 r_c\right] \tag{42}$$

and then subtract (41a) and (41b) from it to find, after some manipulations,

$$M_a^{\mathrm{CS}} = -\frac{e}{\hbar cV}\epsilon_{abc}\, \mathrm{Im}\, \mathrm{Tr}\,[Q\mathcal{H}^0 Pr_b Pr_c]. \tag{43}$$

Replacing $\mathcal{H}^0$ with $\mathcal{H} - e\mathcal{E}_d r_d$ and using $Q\mathcal{H}P = 0$,

$$M_a^{\mathrm{CS}} = -\frac{e^2}{\hbar cV}\epsilon_{abc}\mathcal{E}_d\, \mathrm{Im}\, \mathrm{Tr}\,[Pr_d Pr_b Pr_c]. \tag{44}$$

The imaginary part of the trace vanishes if any two of the indices b, c or d are the same, and therefore d must be equal to a. Using the cyclic property we conclude that all non-vanishing terms in the sum over b and c are identical, leading to (41c). This part of the field-induced magnetization is clearly isotropic, with a coupling strength (see equation (2)) given by

$$\theta^{\mathrm{CS}} = -\frac{4\pi^2}{3V}\epsilon_{ijk}\, \mathrm{Im}\, \mathrm{Tr}[Pr_i Pr_j Pr_k]. \tag{45}$$

This expression can assume nonzero values because the Cartesian components of the projected position operator $P\boldsymbol{r}P$ do not commute [31].

We have used (41a)–(41c) to evaluate the OMP contributions $\tilde{\alpha}^{\mathrm{LC}}$, $\tilde{\alpha}^{\mathrm{IC}}$ and α^{CS} for finite samples, finding excellent agreement with the k-space calculations using (34b)–(34d). As an example, the finite-sample results for the zz component are plotted as the symbols in figure 2.

3. Linear-response expression for the OMP tensor

In sections 2.2 and 2.3.1, expressions were given for evaluating $\boldsymbol{M}(\mathcal{E})$ under periodic boundary conditions. Used in conjunction with finite-field *ab initio* methods for periodic insulators [29, 36], they allow one to calculate the OMP tensor by finite differences. Alternatively, the electric field may be treated perturbatively [33]. With this approach in mind, we shall now take the $\mathcal{E}$-field derivative in (1) analytically and obtain an expression for the OMP tensor that is amenable to density-functional perturbation-theory implementation [37]. It should be kept in mind that in the context of self-consistent-field (SCF) calculations the 'zero-field' part of the Hamiltonian (4),

$$\mathcal{H}^0 = -\frac{\hbar^2}{2m}\nabla^2 + V_{\mathrm{SCF}}(\boldsymbol{r}) \tag{46}$$

does depend on $\mathcal{E}$ implicitly, through the charge density. As we will see, this dependence gives rise to additional local-field screening terms in the expression for the OMP.

We shall only consider the case where the OMP is calculated for a reference state at zero field, which we indicate by a superscript '0'. Upon inserting (34a) into (1), we obtain the three gauge-invariant OMP terms in (40). The term α^{CS} is clearly of the isotropic form (2), with

$$\theta^{\mathrm{CS}} = -\frac{1}{4\pi} \int \mathrm{d}^3 k \, \epsilon_{ijk} \, \mathrm{tr}\left[A_i^0 \partial_j A_k^0 - \frac{2\mathrm{i}}{3} A_i^0 A_j^0 A_k^0 \right]. \tag{47a}$$

This is the same expression as obtained previously by heuristic methods [6, 7][5]. The other two terms were not considered in the previous works. They are

$$\tilde{\alpha}_{da}^{\mathrm{LC}} = \gamma \epsilon_{abc} \int [\mathrm{d}k] \sum_n^N \mathrm{Im}(2\langle \tilde{\partial}_b u_{nk}^0 | (\partial_c H_k^0) | \tilde{\partial}_D u_{nk}^0 \rangle - \langle \tilde{\partial}_b u_{nk}^0 | (\partial_D H_k^0) | \tilde{\partial}_c u_{nk}^0 \rangle) \tag{47b}$$

and

$$\tilde{\alpha}_{da}^{\mathrm{IC}} = \gamma \epsilon_{abc} \int [\mathrm{d}k] \sum_{mn}^N \mathrm{Im}(2\langle \tilde{\partial}_b u_{nk}^0 | \tilde{\partial}_D u_{mk}^0 \rangle \langle u_{mk}^0 | (\partial_c H_k^0) | u_{nk}^0 \rangle - \langle \tilde{\partial}_b u_n^0 | \tilde{\partial}_c u_m^0 \rangle \langle u_m^0 | (\partial_D H_k^0) | u_n^0 \rangle), \tag{47c}$$

where ∂_D denotes the field-derivative $\partial/\partial\mathcal{E}_d$ and

$$|\tilde{\partial}_D u_{nk}^0\rangle \equiv Q_k \, |\partial_D u_{nk}\rangle|_{\mathcal{E}=0} \tag{48}$$

are the first-order field-polarized states projected onto the unoccupied manifold. The terms containing $\partial_D H_k^0$ describe the screening by local fields. They vanish for tight-binding models such as the one in this work, but should be included in self-consistent calculations, in the way described in [37]. We shall sometimes refer to $\tilde{\alpha}^{\mathrm{LC}}$ and $\tilde{\alpha}^{\mathrm{IC}}$ as 'Kubo' contributions because, unlike the Chern–Simons term, they involve first-order changes in the occupied orbitals and Hamiltonian, in a manner reminiscent of conventional linear-response theory[6].

Equations (47a)–(47c) are the main result of this section. The derivation of (47b) and (47c) is somewhat laborious and is sketched in appendix B. We emphasize that the Kubo-like terms, besides endowing the tensor α with off-diagonal elements, also generally contribute to its trace, which therefore is not purely geometric. Writing the isotropic part of the OMP response in the form (2), we then have

$$\theta = \theta^{\mathrm{CS}} + \theta^{\mathrm{Kubo}}. \tag{49}$$

The two contributions are plotted for our model in figure 3. Moreover, the open circles in figure 1 show the zz and zy components of the OMP tensor computed from (47a)–(47c), confirming that the analytic field derivative of the magnetization was taken correctly.

In the case of an insulator with a single valence band, the partition (40) of the OMP tensor acquires some interesting features. The terms $\tilde{\alpha}^{\mathrm{IC}}$ and α^{CS} become purely itinerant, i.e. they vanish in the limit of a crystal composed of non-overlapping molecular units, with one electron per molecule. Also, the first term in expression (47c) for $\tilde{\alpha}^{\mathrm{IC}}$—the only term for tight-binding models—becomes traceless, as can be readily verified in a Hamiltonian gauge (where $H_k^0 |u_{nk}^0\rangle = E_{nk}^0 |u_{nk}^0\rangle$) with the help of the perturbation theory formula [33]

$$|\tilde{\partial}_D u_{nk}^0\rangle = \mathrm{i}e \sum_{m>N} \frac{|u_{mk}^0\rangle \langle u_{mk}^0|}{E_n^0 - E_m^0} |\partial_d u_{nk}^0\rangle. \tag{50}$$

[5] An inconsistency in the published literature regarding the numerical prefactor in (47a) has been resolved; see [38].

[6] The terminology 'Kubo terms' for $\tilde{\alpha}^{\mathrm{LC}}$ and $\tilde{\alpha}^{\mathrm{IC}}$ is only meant to be suggestive. A Kubo-type linear-response calculation of the OMP should produce all three terms, including α^{CS}.
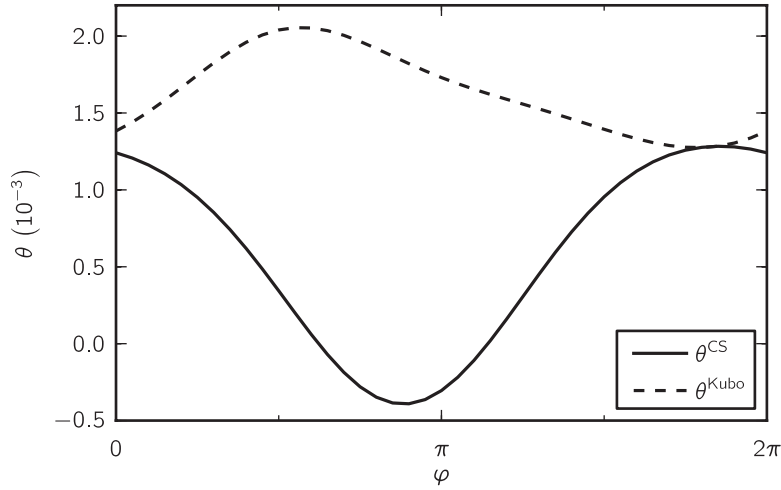
Figure 3. Contributions to the isotropic OMP from the Chern–Simons term α^{CS} and from the Kubo-like terms $\tilde{\alpha}^{LC}$ and $\tilde{\alpha}^{IC}$, expressed in terms of the dimensionless coupling strength θ in (2). Model parameters are the same as for figure 1.

In order to verify these features numerically, we calculated the various contributions treating only the lowest band of our tight-binding model as occupied. The molecular limit was taken by setting to zero the hoppings between neighbouring eight-site cubic 'molecules.'

It could have been anticipated from the outset that the Chern–Simons term (47a) could not be the entire expression for the OMP, based on the following argument [26]. Consider an insulator with $N > 1$ valence bands, all of which are isolated from one another. By looking at α_{da} as $\partial P_d / \partial B_a$, one can argue that since each band carries a certain amount of polarization $\boldsymbol{P}^{(n)}$, the total OMP should satisfy the relation

$$\alpha = \sum_n^N \alpha^{(n)}, \tag{51}$$

where $\alpha^{(n)}$ is the OMP one would obtain by filling band n while keeping all other bands empty. We shall refer to this property as the 'band-sum-consistency' of the OMP. It only holds exactly for models without charge self-consistency (see the analytic proof in appendix C), but that suffices for the purpose of the argument. We note that the Chern–Simons contribution (47a) alone is *not* band-sum-consistent, because the second term therein vanishes for a single occupied band. Hence an additional contribution, also band-sum-inconsistent, must necessarily exist. Indeed, both $\tilde{\alpha}^{LC}$ and $\tilde{\alpha}^{IC}$ are band-sum-inconsistent, in such a way that the total OMP satisfies (51). This is illustrated in figure 4 for our tight-binding model.

4. Summary and outlook

In summary, we have developed a theoretical framework for calculating the frozen-ion orbital-magnetization response (OMP) to a static electric field. This development fills an important gap in the microscopic theory of the magnetoelectric effect, paving the way to first-principles calculations of the full response. While the OMP is often assumed to be small compared to the
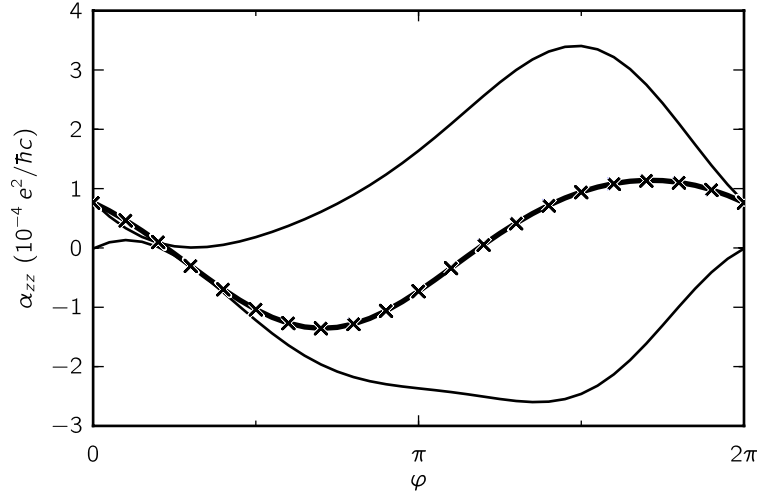
Figure 4. Comparison between α_{zz} calculated treating the two lowest bands as occupied (crosses) and the sum $\alpha_{zz}^{(1)} + \alpha_{zz}^{(2)}$ (thick solid line), where $\alpha_{zz}^{(n)}$ (thin solid lines) correspond to treating only the lowest band ($n = 1$, upper line) or the second-lowest band ($n = 2$, lower line) as occupied. Model parameters are the same as for figure 1, except that the second-lowest on-site energy in table A.1 is raised from -6.0 to -5.0 in order to keep the two lowest bands well separated.

lattice-mediated and spin-magnetization parts of the ME response, there is no *a priori* reason why it should always be so. In fact, in strong Z_2 topological insulators it is the only contribution that survives, and the predicted value is large compared to that of prototypical magnetoelectrics. Although the measurement of the $\theta = \pi$ ME effect in topological insulators is challenging, as time-reversal symmetry must be broken to gap the surfaces [6, 7, 23], there may be other related materials where those symmetries are broken already in the bulk. The present formalism should be helpful in the ongoing computational search for such materials with a large and robust OMP.

A key result of this work is a k-space expression for the orbital magnetization of a periodic insulator under a finite electric field $\mathcal{E}$ (equations (6a)–(6d) or, equivalently, (34a)–(34d)). In addition to the terms (34b) and (34c) already present at zero field [13], in three dimensions the field-dependent magnetization comprises an additional purely isotropic 'Chern–Simons' term, given by (34d). This new term depends explicitly on $\mathcal{E}$ and only implicitly on H_k^0, while the converse is true for the other terms. Moreover, it is a multivalued quantity, with a quantum of arbitrariness $\boldsymbol{M}_0 = \mathcal{E}e^2/hc$ along $\mathcal{E}$. Thus, the analogy with the Berry-phase theory of electric polarization [10, 20], where a similar quantum arises, becomes even more profound at finite electric field.

The Chern–Simons term $\boldsymbol{M}^{\mathrm{CS}}$ is responsible for the geometric part of the OMP response discussed in [6, 7] in connection with topological insulators. We have clarified that in materials with broken time-reversal and inversion symmetries in the bulk, the CSOMP does not generally constitute the full response, as the remaining orbital magnetization terms, $\widetilde{\boldsymbol{M}}^{\mathrm{LC}}$ and $\widetilde{\boldsymbol{M}}^{\mathrm{IC}}$, can also depend linearly on $\mathcal{E}$. Their contribution to the OMP, given by (47b) and (47c), is twofold: (i) to modify the isotropic coupling strength θ and (ii) to introduce an anisotropic component of the response.

Another noteworthy result is equation (45) for the Chern–Simons OMP of finite systems. One appealing feature of this expression is that it allows one to calculate the CSOMP without

the need to choose a particular gauge. Instead, its k-space counterpart, equation (47a), requires for its numerical evaluation a smoothly varying gauge for the Bloch states across the Brillouin zone. Equation (45) is also the more general of the two, as it can be applied to noncrystalline or otherwise disordered systems.

We conclude by enumerating a few questions that are raised by the present work. Do the individual gauge-invariant OMP terms identified here in a one-electron picture remain meaningful for interacting systems, and can they be separated experimentally? (This appears to be the case for $\widetilde{M}^{\text{LC}}$ and $\widetilde{M}^{\text{IC}}$ at $\mathcal{E} = 0$ [35].) How does the expression (47a)–(47c) for the linear OMP response change when the reference state is under a finite electric field $\mathcal{E}$? Finally, we note that equation (41c) for the CSOMP of finite systems has a striking resemblance to a formula given by Kitaev [39] for the 2D Chern invariant characterizing the integer quantum Hall effect. Can this connection be made more precise, in view of the fact that the quantum of indeterminacy in θ^{CS} is associated with the possibility of changing the Chern invariant of the surface layers? These questions are left for future studies.

Acknowledgments

Appendix A. Tight-binding model and technical details

A.1. Tight-binding model

We have chosen for our tests a model of an ordinary (that is, non-topological) insulator. The prerequisites were the following. It should break both time-reversal and inversion symmetries, as the OMP tensor otherwise vanishes identically. It should be 3D, as the geometric part of the response vanishes otherwise. Its symmetry should be sufficiently low to render all nine components of the OMP tensor nonzero. Finally, it should have multiple valence bands, for generality.

We opted for a spinless model on a cubic lattice. It can be obtained starting with a one-site simple cubic model, doubling the cell in each direction, and assigning random on-site energies E_i and complex first-neighbour hoppings $t_{j \rightarrow i} = t e^{i\phi_{j \rightarrow i}}$ of fixed magnitude $t = 1$. The Hamiltonian reads
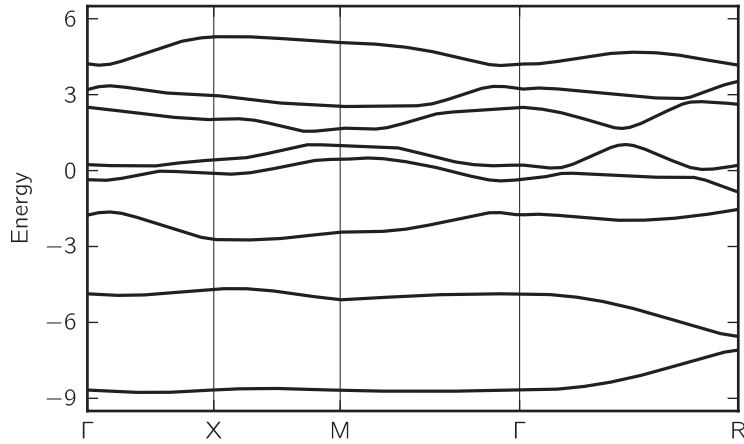
$$\mathcal{H}^0 = \sum_i E_i c_i^\dagger c_i + \sum_{\langle i j \rangle} e^{i\phi_{j \rightarrow i}} c_i^\dagger c_j, \tag{A.1}$$

where $i = (x, y, z)$ labels the sites and $\langle i j \rangle$ denotes pairs of nearest-neighbour sites. The values of E_i on two of the eight sites were adjusted to ensure a finite gap everywhere in the Brillouin zone between the two lowest bands (chosen as the valence bands) and the remaining six. We also made sure that nonzero phases $\phi_{j \rightarrow i}$ were not restricted to 2D square-lattice planes; otherwise those are mirror symmetry planes, whose existence is sufficient to make the diagonal elements of the OMP tensor vanish. In our calculations, all the model parameters were kept fixed except for one phase, which was scanned over the range $[0, 2\pi]$, and the results are plotted as a function

Table A.1. Parameters of the tight-binding model. Columns I–III give the site coordinates $i = (x, y, z)$, in units of the lattice constant $a = 1$ of the $2 \times 2 \times 2$ primitive cubic cell. Column IV contains the on-site energies E_i, and the last three columns contain the phases of the complex nearest-neighbour hopping amplitudes along bonds in the negative $\hat{x}$, $\hat{y}$ and $\hat{z}$ directions.

x	y	z	E_i	$\phi_{(i+\hat{x}/2)\rightarrow i}$	$\phi_{(i+\hat{y}/2)\rightarrow i}$	$\phi_{(i+\hat{z}/2)\rightarrow i}$
0.0	0.0	0.0	−6.5	$\varphi \in [0, 2\pi]$	0.5π	1.7π
0.5	0.0	0.0	0.9	1.3π	0.2π	0.5π
0.5	0.5	0.0	1.4	0.8π	1.4π	0.6π
0.0	0.5	0.0	1.2	0.3π	1.9π	1.0π
0.0	0.0	0.5	−6.0[a]	1.4π	0.8π	0.3π
0.5	0.0	0.5	1.5	0.6π	1.7π	0.7π
0.5	0.5	0.5	0.8	0.8π	0.6π	1.2π
0.0	0.5	0.5	1.2	1.9π	0.3π	1.4π

[a] In figure 4 the value −5.0 was used instead.

Figure A.1. Band structure of the cubic-lattice tight-binding model given by (A.1), for the choice of parameters in table A.1 and $\varphi = 0$.

of this phase φ. For reference, the on-site energies and the phases of the hopping amplitudes are listed in table A.1. The energy bands are shown in figure A.1 for $\varphi = 0$.

In order to couple the system to the electric field and to be able to define its orbital magnetization, the position operator r must be specified along with $\mathcal{H}^0$. We have chosen the simplest representation where r is diagonal in the tight-binding basis.

A.2. Technical details

The calculations employing periodic boundary conditions were carried out on an $80 \times 80 \times 80$ k-point mesh, and the k-space implementation of finite electric fields was done using the method discussed in section V of [30]. The open boundary condition calculations used cubic samples containing $L \times L \times L$ eight-site unit cells, that is, $2L + 1$ sites along each edge. For large L, we

expect the magnetization to scale as

$$M(L) = M + \frac{a}{L} + \frac{b}{L^2} + \frac{c}{L^3}, \tag{A.2}$$

where a, b and c account for face, edge and corner corrections, respectively [13]. Calculations of $M(L)$ under small fields were done using $L = 4, 5, 6, 7$, and then fitted to (A.2) in order to extract the value M of the magnetization in the $L \to \infty$ limit. The differences between OMP values calculated in various ways as shown in figures 1 and 2 were of the order of $10^{-7} e^2/\hbar c$ or less.

Before evaluating the k-space expressions for $M(\mathcal{E})$ ((6a)–(6d) and (34b)–(34d)) and α ((47a)–(47c)) on a grid, they need to be properly discretized. The presence of the gauge-dependent Berry connection in (6d) demands the use of a 'smooth gauge' for its evaluation, where the valence Bloch states given by (18) are smoothly varying functions of k. This is achieved by projecting a set of trial orbitals onto the set of occupied Bloch eigenstates according to the prescription in equations (62)–(64) of [31]. (For the tight-binding model discussed below, when treating the two lowest bands as occupied, the two trial orbitals are chosen as delta functions located at the two sites with lowest on-site energy.) If needed, this one-shot projection procedure can be improved upon by finding an optimally smooth gauge using methods based on minimizing the real-space spread of the WFs [31], but we found our results to change negligibly when performing this extra step. In a smooth gauge the needed k-derivatives of the Bloch states and of the Berry connection matrix are then evaluated by straightforward numerical differentiation. Note that (6b) and (6c) should be evaluated in the same smooth gauge as (6d), as these three equations are not separately gauge-invariant. A smooth gauge must also be used for (34d) and (47a), because, as discussed in section 2.3.1, the Chern–Simons 3-form is locally gauge-dependent.

The same strategy can be used to discretize (34b) and (34c). However, since the k-derivatives appearing in those equations are covariant, the discretized form of the covariant derivative (19) given in [13, 30] may be used instead, circumventing the need to work in a smooth gauge. We have implemented both approaches, finding excellent agreement between them.

Finally, we come to equations (47b) and (47c). In addition to the k-derivative of the valence Bloch states, we need their (covariant) field-derivative (48), as well as the k-derivative of H_k^0. The latter quantity is easily calculated within the tight-binding method, and for the former we used the linear-response expression (50). Note that this requires choosing the unperturbed states to be in the Hamiltonian gauge. This choice precludes calculating the k-derivative on the right-hand side of (50) by straightforward finite differences, which can only be done in a smooth gauge. But because $\langle u_{mk}^0 | \partial_d u_{nk}^0 \rangle$ equals $\langle u_{mk}^0 | \tilde{\partial}_d u_{nk}^0 \rangle$ for $m > N$, the discretized covariant derivative approach may be used instead. Alternatively, one can evaluate the ordinary k-derivative by summation over states as

$$|\partial_d u_{nk}^0\rangle = \sum_{m \neq n} |u_{mk}^0\rangle \frac{\langle u_{mk}^0 | (\partial_d H_k^0) | u_{nk}^0 \rangle}{E_{nk}^0 - E_{mk}^0}. \tag{A.3}$$

We note that this formula may not be used to calculate the geometric term (47a), because it induces locally a parallel transport gauge ($A_{nn}^0 = 0$), which cannot be enforced globally since the Brillouin zone is a closed space.

Appendix B. Derivation of equations (47*b*) and (47*c*)

For notational simplicity we drop the crystal momentum index $\boldsymbol{k}$. So, for example, $|u_{nk}\rangle$ shall be denoted by $|u_n\rangle$. In order to calculate the OMP terms

$$\tilde{\alpha}_{da}^{\text{LC}} = \partial_D \widetilde{M}_a^{(\text{LC})}\big|_{\mathcal{E}=0} \tag{B.1}$$

and

$$\tilde{\alpha}_{da}^{\text{IC}} = \partial_D \widetilde{M}_a^{(\text{IC})}\big|_{\mathcal{E}=0} \tag{B.2}$$

starting from (31) and (32), we shall first examine the field- and k-derivatives of certain basic quantities.

We begin by noting that the field-derivative $\partial_D P = -\partial_D Q$ of the projection operator (20) can be written as

$$\partial_D P = \sum_n^N \left(|\tilde{\partial}_D u_n\rangle\langle u_n| + |u_n\rangle\langle\tilde{\partial}_D u_n| \right) \equiv \tilde{\partial}_D P, \tag{B.3}$$

in terms of the covariant field-derivative (48) (a similar expression holds for the k-derivative). This follows from a relation analogous to (21):

$$|\partial_D u_n\rangle = |\tilde{\partial}_D u_n\rangle - \mathrm{i}\sum_l^N A_{ln,D}|u_l\rangle, \tag{B.4}$$

where

$$\boldsymbol{A}_{ln,D} = \mathrm{i}\langle u_l|\partial_D u_n\rangle = A_{nl,D}^* \tag{B.5}$$

is the Berry connection matrix along the parametric direction $\mathcal{E}_d$. With the help of (B.3) the field derivative of (22) becomes

$$\partial_D F_{nm,bc} = \langle\partial_{Db}^2 u_n|Q|\partial_c u_m\rangle + \langle\partial_b u_n|Q|\partial_{Dc}^2 u_m\rangle + \mathrm{i}(F_{bD}A_c)_{nm} - \mathrm{i}(A_b F_{Dc})_{nm}, \tag{B.6}$$

where F_{bD} is obtained from F_{bd} by replacing ∂_d with ∂_D. We shall also need the field- and k-derivatives of the matrix H_{nm}^0 defined by (8):

$$\partial_D(H_{nm}^0)\big|_{\mathcal{E}=0} = \mathrm{i}\left[A_D^0, H^0\right]_{nm} + (\partial_D H_{\text{op}}^0)_{nm}\big|_{\mathcal{E}=0}, \tag{B.7}$$

$$\partial_c(H_{nm}^0)\big|_{\mathcal{E}=0} = \mathrm{i}\left[A_c^0, H^0\right]_{nm} + (\partial_c H_{\text{op}}^0)_{nm}\big|_{\mathcal{E}=0}, \tag{B.8}$$

where we introduced the notation $(\partial_{D,c} H_{\text{op}}^0)_{nm} \equiv \langle u_n|\partial_{D,c} H^0|u_m\rangle$, where 'op' indicates that the derivative is taken on the operator itself, not its matrix representation. These two relations follow directly from (B.4) and (21). We will also make use of identities such as

$$\operatorname{Re}\operatorname{tr}[X F_{bc}] = \operatorname{Re}\operatorname{tr}\left[X^\dagger F_{cb}\right]. \tag{B.9}$$

In particular, if X and Y are Hermitian,

$$\operatorname{Re}\operatorname{tr}[X Y F_{bc}] = \operatorname{Re}\operatorname{tr}[Y X F_{cb}]. \tag{B.10}$$

We are now ready to evaluate (B.2):

$$\tilde{\alpha}_{da}^{\text{IC}} = -\gamma\epsilon_{abc}\int[\mathrm{d}k]\operatorname{Im}\operatorname{tr}\left[F_{bc}\partial_D H^0 + H^0\partial_D F_{bc}\right]\big|_{\mathcal{E}=0}. \tag{B.11}$$

IOP Institute of Physics Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Inserting (B.6) and (B.7) on the right-hand side generates a number of terms. Some can be combined upon interchanging dummy indices $b \leftrightarrow c$ and invoking (B.10), leading to

$$\tilde{\alpha}_{da}^{\text{IC}} = -\gamma\epsilon_{abc}\int[dk](2\text{Re}\,\text{tr}\left[A_D H^0 F_{bc} + H^0 F_{bD} A_c\right] + \text{Im}\,\text{tr}\left[F_{bc}\partial_D H_{\text{op}}^0\right]$$

$$+2\text{Im}\sum_{mn}^{N} H_{mn}^0 \langle\partial_{Db}^2 u_n|Q|\partial_c u_m\rangle)\Big|_{\mathcal{E}=0}. \tag{B.12}$$

Integrating the last term by parts in k_b and using (B.3) and (B.8) again produces a number of terms, most of which cancel out. The end result reads

$$\tilde{\alpha}_{da}^{\text{IC}} = \gamma\epsilon_{abc}\int[dk]\,\text{Im}\,\text{tr}\left[2F_{bD}\partial_c H_{\text{op}}^0 - F_{bc}\partial_D H_{\text{op}}^0\right]\Big|_{\mathcal{E}=0}. \tag{B.13}$$

Similarly, (B.1) can be evaluated by repeatedly using (B.3) and integrating by parts the terms with mixed field- and k-derivatives, yielding

$$\tilde{\alpha}_{da}^{\text{LC}} = \gamma\epsilon_{abc}\int[dk]\,\text{Im}\,\text{tr}\left[2(\partial_c H^0)_{bD} - (\partial_D H^0)_{bc}\right]\Big|_{\mathcal{E}=0}, \tag{B.14}$$

where $(\partial_c H^0)_{bD}$ and $(\partial_D H^0)_{bc}$ are defined in analogy with (26), e.g.

$$(\partial_c H^0)_{nmk,bD} = \langle\tilde{\partial}_b u_{nk}|(\partial_c H_k^0)|\tilde{\partial}_D u_{mk}\rangle. \tag{B.15}$$

Equations (B.14) and (B.13) are respectively equivalent to (47b) and (47c) in the main text. The gauge invariance of these equations follows from the fact that they are written as traces over gauge-covariant objects. (We also note that the covariant derivative transforms according to (18) regardless of the parameter with respect to which the differentiation is carried out.)

Appendix C. Band-sum consistency of the OMP

Here we show analytically that the OMP tensor α satisfies the band-additivity relation (51) in models without charge self-consistency. In order to isolate the contribution $\alpha^{(n)}$ coming from valence band n (assumed to be well separated in energy from all other bands), we choose the Hamiltonian matrix to be diagonal at zero field, i.e. $H_{mnk}^0(\mathcal{E}=0) = E_{nk}^0 \delta_{mn}$. If in addition we use a parallel-transport gauge for the linear electric field perturbation [33] (this is achieved by setting to zero the matrix A_D defined in (B.5)), we find, using (B.7), $\partial_D H_{mnk}^0|_{\mathcal{E}=0} = 0$. With the help of these two relations, the field-derivative $\partial_D M_a|_{\mathcal{E}=0}$ of (6a) is easily taken. From the first two terms therein, we obtain (dropping index k)

$$2\gamma\epsilon_{abc}\int[dk]\sum_n^N \text{Im}\langle\partial_b u_n|\partial_c(H^0 + E_n)|\partial_D u_n\rangle\Bigg|_{\mathcal{E}=0}. \tag{C.1}$$

In the parallel-transport gauge, $|\partial_D u_n\rangle$ is given by (50), and combining the resulting expression with the field derivative of the third term in (6a) yields

$$\alpha_{da} = 2e\gamma\epsilon_{abc}\int[dk]\sum_n^N \text{Re}\langle\partial_b u_n^0|\partial_c(H^0 + E_n^0)\left(\sum_{l>N}\frac{|u_l^0\rangle\langle u_l^0|}{E_n^0 - E_l^0}\right)|\partial_d u_n^0\rangle$$

$$-e\gamma\epsilon_{abc}\int[dk]\sum_{mn}^N \text{Re}(\langle u_m^0|\partial_d u_n^0\rangle\langle\partial_b u_n^0|\partial_c u_m^0\rangle). \tag{C.2}$$

To find $\alpha_{da}^{(n)}$ we replace $\sum_{l>N}$ with $\sum_{l\neq n}$ and reduce sums $\sum_{mn}^{N}$ and $\sum_{n}^{N}$ to single terms. Inserting these expressions into (51) and splitting $\sum_{l\neq n}$ into $\sum_{l>N}$ and $\sum_{l\neq n}^{N}$, some terms cancel and others can be combined, leading to

$$\epsilon_{abc} \int [\mathrm{d}k] \sum_n^N \sum_{m\neq n}^N \mathrm{Re}\left[\langle u_m^0|\partial_d u_n^0\rangle \left(\frac{\langle \partial_b u_n^0|\partial_c(H^0+E_n^0)|u_m^0\rangle}{E_n^0-E_m^0} + \frac{1}{2}\langle \partial_b u_n^0|\partial_c u_m^0\rangle\right)\right] = 0. \qquad (\text{C.3})$$

The lhs is proportional to the difference between α_{da} and $\sum_n^N \alpha_{da}^{(n)}$, and vanishes as a result of an exact cancellation between the terms (n,m) and (m,n) in the double sum. The integrand of the (n,m) term is

$$\epsilon_{abc}\,\mathrm{Re}\left[\langle u_m^0|\partial_d u_n^0\rangle \left(\frac{\langle \partial_b u_n^0|\partial_c(H^0+E_n^0)|u_m^0\rangle}{E_n^0-E_m^0} + \frac{1}{2}\langle \partial_b u_n^0|\partial_c u_m^0\rangle\right)\right], \qquad (\text{C.4})$$

and after some manipulations the integrand of the (m,n) term becomes

$$\epsilon_{abc}\,\mathrm{Re}\left[\langle u_m^0|\partial_d u_n^0\rangle \left(\frac{\langle u_n^0|\partial_c(H^0+E_m^0)|\partial_b u_m^0\rangle}{E_n^0-E_m^0} + \frac{1}{2}\langle \partial_b u_n^0|\partial_c u_m^0\rangle\right)\right]. \qquad (\text{C.5})$$

The final step is to use the identity

$$\frac{\langle \partial_b u_n^0|\partial_c(H^0+E_n^0)|u_m^0\rangle}{E_n^0-E_m^0} = \frac{\langle \partial_b u_n^0|E_m^0-H^0|\partial_c u_m^0\rangle}{E_n^0-E_m^0} - \partial_c(E_n^0+E_m^0)\frac{\langle u_n^0|\partial_b u_m^0\rangle}{E_n^0-E_m^0}. \qquad (\text{C.6})$$

(This identity follows from the relation

$$(H^0-E_m^0)|\partial_c u_m^0\rangle = -(\partial_c H^0 - \partial_c E_m^0)|u_m^0\rangle, \qquad (\text{C.7})$$

which in turn can be obtained by expanding $H^0|u_m^0\rangle = E_m^0|u_m^0\rangle$ to first order in the change in wavevector $\boldsymbol{k}$.) The quantity (C.4) + (C.5) then becomes

$$\epsilon_{abc}\,\mathrm{Re}\left[\langle u_m^0|\partial_d u_n^0\rangle \left(\frac{\langle \partial_b u_n^0|E_m^0-H^0|\partial_c u_m^0\rangle}{E_n^0-E_m^0} + \frac{\langle \partial_c u_n^0|E_n^0-H^0|\partial_b u_m^0\rangle}{E_n^0-E_m^0} + \langle \partial_b u_n^0|\partial_c u_m^0\rangle\right)\right]. \qquad (\text{C.8})$$

Interchanging $b \leftrightarrow c$ in the second term and combining with the first yields minus the third term, which concludes the proof.

References

[1] O'Dell T 1970 *The Electrodynamics of Magneto-Electric Media* (Amsterdam: North-Holland)
[2] Fiebig M 2005 *J. Phys. D: Appl. Phys.* **38** R123
[3] Íñiguez J 2008 *Phys. Rev. Lett.* **101** 117201
[4] Wojdel J and Íñiguez J 2009 *Phys. Rev. Lett.* **103** 267205
[5] Delaney K T and Spaldin N A 2009 arXiv:0912.1335
[6] Qi X, Hughes T L and Zhang S C 2008 *Phys. Rev.* B **78** 195424
[7] Essin A M, Moore J E and Vanderbilt D 2009 *Phys. Rev. Lett.* **102** 146805
[8] Raab R E and De Lange O L 2005 *Multipole Theory in Electromagnetism* (Oxford: Clarendon)
[9] Barron L D 2004 *Molecular Light Scattering and Optical Activity* (Cambridge: Cambridge University Press)
[10] King-Smith R D and Vanderbilt D 1993 *Phys. Rev.* B **47** 1651

[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] Ceresoli D, Thonhauser T, Vanderbilt D and Resta R 2006 *Phys. Rev.* B **74** 024408

[14] Shi J, Vignale G, Xiao D and Niu Q 2007 *Phys. Rev. Lett.* **99** 197202

[15] Wilczek F 1987 *Phys. Rev. Lett.* **58** 1799

[16] Raab R E and Sihvola A H 1997 *J. Phys. A: Math. Gen.* **30** 1335

[17] Hehl F W, Obukhov Y N, Rivera J P and Schmid H 2008 *Phys. Rev.* A **77** 022106

[18] Obukhov Y N and Hehl F W 2005 *Phys. Lett.* A **341** 357

[19] Widom A, Friedman M H and Srivastava Y 1986 *J. Phys. A: Math. Gen.* **19** L175

[20] Resta R and Vanderbilt D 2007 *Physics of Ferroelectrics: A Modern Perspective* ed K M Rabe, C H Ahn and J M Triscone (Berlin: Springer) pp 31–68

[21] Chern S S and Simons J 1974 *Ann. Math.* **99** 48

[22] Qi X L and Zhang S C 2010 *Phys. Today* **63** 33

[23] Hasan M Z and Kane C L 2010 arXiv:1002.3895

[24] Moore J E 2010 *Nature* **464** 194

[25] Wiegelmann H, Jansen A G M, Wyder P, Rivera J P and Schmid H 1994 *Ferroelectrics* **162** 141

[26] Essin A, Turner A, Moore J and Vanderbilt D 2010 arXiv:1002.0290

[27] Thonhauser T and Vanderbilt D 2006 *Phys. Rev.* B **74** 235111

[28] Haldane F D M 1988 *Phys. Rev. Lett.* **61** 2015

[29] Souza I, Íñiguez J and Vanderbilt D 2002 *Phys. Rev. Lett.* **89** 117602

[30] Souza I, Íñiguez J and Vanderbilt D 2004 *Phys. Rev.* B **69** 085106

[31] Marzari N and Vanderbilt D 1997 *Phys. Rev.* B **56** 12847

[32] Blount E I 1962 *Solid State Physics* vol 13 ed F Seitz and D Turnbull (New York: Academic)

[33] Nunes R W and Gonze X 2001 *Phys. Rev.* B **63** 155107

[34] Nakahara M 1998 *Geometry, Topology and Physics* (Bristol: Institute of Physics Publishing)

[35] Souza I and Vanderbilt D 2008 *Phys. Rev.* B **77** 054438

[36] Umari P and Pasquarello A 2002 *Phys. Rev. Lett.* **89** 157602

[37] Baroni S, de Gironcoli S, Corso A D and Giannozzi P 2001 *Rev. Mod. Phys.* **73** 515

[38] Essin A, Moore J E and Vanderbilt D 2009 *Phys. Rev. Lett.* **103** 259902

[39] Kitaev A 2006 *Ann. Phys., NY* **321** 2