Dichroic f-sum rule and the orbital magnetization of crystals

Ivo Souza[1] and David Vanderbilt[2]

[1]*Department of Physics, University of California, Berkeley, CA 94720, USA*
[2]*Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854-8019, USA*

(Dated: September 15, 2007)

We consider the magnetic circular dichroism spectrum of a crystal with broken time-reversal symmetry in the electric-dipole approximation. Using the Kubo formula for the absorptive part of the antisymmetric optical conductivity, the frequency integral is recast as a ground-state property. We show that in insulators this quantity is proportional to the circulation of the Wannier orbitals around their centers (more precisely, to the gauge-invariant part thereof). This differs from the net circulation, or ground state orbital magnetization, which has two additional contributions: (i) the remaining Wannier self-rotation, and (ii) the "itinerant" circulation arising from the center-of-mass motion of the Wannier orbitals, both on the surface and in the interior of the sample. Contributions (i) and (ii) are not separately meaningful, since their individual values depend on the particular choice of Wannier functions. Their sum is however gauge-invariant, and can be inferred from a combination of two experiments: a measurement of the magneto-optical spectrum over a sufficiently wide range to evaluate the sum rule, and a gyromagnetic determination of the total orbital magnetization.

PACS numbers: 78.20.Ls, 75.10.Lp, 73.43.-f

I. INTRODUCTION

Optical sum rules provide a link between excitation spectra and ground-state properties. The best-known example is the f-sum rule of atomic physics.[1] It relates the frequency-integrated absorption of linearly polarized light to the number of valence electrons. In this work we consider the analogous result for circularly polarized light. For non-magnetic systems the circular f-sum rule is simply the average of the f-sum rules for the two linearly-polarized components of the beam, again yielding the total number of electrons. If, however, the system is magnetized, either spontaneously or by an applied field, this is no longer the case; there is a small correction that flips sign when either the magnetization of the sample or the helicity of the incident light is reversed. We are interested in what information this correction to the circular f-sum rule provides about the magnetization. Special emphasis will be placed on insulating systems, for which an intuitive picture in terms of localized Wannier orbitals can be given.

The differential absorption of left- and right-circularly-polarized light by magnetic materials is known as magnetic circular dichroism (MCD). The object of interest in this work can thus be viewed as a "dichroic" f-sum rule for the integrated MCD spectrum. Such a sum rule was first derived by Hasegawa and Howard for the special case of a hydrogen atom in a magnetic field.[2] They showed that it is proportional to the quantum-mechanical expectation value of the orbital angular momentum operator, i.e., to the orbital moment. It has been assumed that this conclusion generalizes trivially to many-electron systems such as solids.[3,4] This is not so, as shown by Oppeneer, who obtained the correct sum rule for that case.[5] He observed that it yields a quantity that is subtly different from the orbital magnetization, and should instead be viewed as one of two terms adding up to the orbital magnetization $\mathbf{M}_{\rm orb}$.

In a separate development, a rigorous theory of orbital magnetization in crystals was recently formulated.[6,7,8,9] Interestingly, it also identifies two separate contributions to $\mathbf{M}_{\rm orb}$. One key result of the present work is to recast the dichroic f-sum rule in the language of this modern theory, elucidating its physical content. Conversely, the sum rule solves an open problem in the theory of Refs. 6,7,8,9 as raised explicitly in Ref. 8: whether the two gauge-invariant contributions to $\mathbf{M}_{\rm orb}$ identified therein are separately measurable in principle. The present work answers this question in the affirmative.

Although we will mostly focus on crystalline solids, we find it useful to start in Sec. II by discussing the sum rule in the more general context of bounded samples under open boundary conditions. The detailed treatment of periodic crystals is deferred until Sec. III, after which we conclude in Sec. IV with a summary and outlook. In Appendices A, B, and C we derive and elaborate on some results quoted in the main text. In particular, Appendix A discusses the relation between the dichroic f-sum rule and three other sum rules from the literature.

II. BOUNDED SAMPLES

A. Preliminaries

In this work we are interested in systems displaying broken time-reversal symmetry in the spatial wavefunctions. A typical example would be a ferromagnet such as iron in which the exchange interaction breaks time-reversal symmetry in the spin channel and this symmetry breaking is then transmitted to the orbital degrees of freedom by the spin-orbit interaction. Other exam-

ples include systems in applied magnetic fields, and also certain spinless model Hamiltonians such as the Haldane model.[10]

We work in the independent-particle approximation. The interaction with light will be treated in the electric-dipole approximation, valid at not-too-high frequencies. This should be adequate provided that the sum rule saturates before higher-order contributions, such as electric quadrupole and magnetic dipole terms, become significant. The oscillator strength for the transition between one-electron states n and m is $f^{(\hat{\boldsymbol{\epsilon}})}_{n\to m} = (2m_e/\hbar\omega_{mn})|\hat{\boldsymbol{\epsilon}}\cdot\mathbf{v}_{nm}|^2$, where $\hat{\boldsymbol{\epsilon}}$ is the polarization. For light propagating along $\hat{\boldsymbol{\alpha}}\times\hat{\boldsymbol{\beta}}$ with circular polarization $\hat{\boldsymbol{\epsilon}} = (\hat{\boldsymbol{\alpha}} \pm i\hat{\boldsymbol{\beta}})/\sqrt{2}$,

$$f^{(\pm)}_{n\to m} = \frac{1}{2}f'_{nm,\alpha\alpha} + \frac{1}{2}f'_{nm,\beta\beta} \mp f''_{nm,\alpha\beta}, \tag{1}$$

where we have introduced the complex quantity

$$f_{nm,\alpha\beta} = (2m_e/\hbar\omega_{mn})v_{nm,\alpha}v_{mn,\beta}. \tag{2}$$

Here α,β label Cartesian directions, $\hbar\omega_{mn} = \mathcal{E}_m - \mathcal{E}_n$, $v_{nm,\alpha}$ are velocity matrix elements, and m_e is the electron mass. Note that the object $f_{nm,\alpha\beta}$ is Hermitian in the Cartesian indices. Thus its real and imaginary parts, f' and f'', are symmetric and antisymmetric respectively. The dichroism of the transition is

$$f^{(+)}_{n\to m} - f^{(-)}_{n\to m} = -2f''_{nm,\alpha\beta}. \tag{3}$$

Consider now a macroscopic system (e.g., a sample of volume V cut from a bulk crystal) and decompose its optical conductivity $\sigma(\omega)$ in three different ways: (i) real and imaginary parts, σ' and σ''; (ii) symmetric and antisymmetric parts, $\sigma_{\rm S}$ and $\sigma_{\rm A}$; (iii) Hermitian and anti-Hermitian parts, $\sigma_{\rm H}$ and $\sigma_{\rm AH}$. Then

$$\sigma_{\rm H} = \sigma'_{\rm S} + i\sigma''_{\rm A} \tag{4}$$

and

$$\sigma_{\rm AH} = \sigma'_{\rm A} + i\sigma''_{\rm S}, \tag{5}$$

where the Cartesian indices have been omitted. The properties of $\sigma(\omega)$ can be summarized by noting that the Hermitian part is dissipative while the anti-Hermitian part is reactive, and the symmetric part is "ordinary" while the antisymmetric part is "dichroic." At $T = 0$ the dissipative (or absorptive) part is

$$\sigma_{\rm H}(\omega) = \frac{\pi e^2}{2m_e V}\sum_n^{\rm occ}\sum_m^{\rm empty} f_{nm}\delta(\omega-\omega_{mn}), \tag{6}$$

where $-e$ is the electron charge. For circularly polarized light propagating along $\hat{\mathbf{z}}$,

$$\sigma^{(\pm)}_{\rm abs}(\omega) = \frac{1}{2}\sigma'_{{\rm S},xx}(\omega) + \frac{1}{2}\sigma'_{{\rm S},yy}(\omega) \mp \sigma''_{{\rm A},xy}(\omega). \tag{7}$$

Thus the dichroism is given by the imaginary part of the antisymmetric optical conductivity,[11]

$$\sigma^{(+)}_{\rm abs}(\omega) - \sigma^{(-)}_{\rm abs}(\omega) = -2\sigma''_{{\rm A},xy}(\omega). \tag{8}$$

This vanishes for time-reversal-invariant systems, as can be seen from the Onsager relation $\sigma_{\alpha\beta}(\mathbf{H}) = \sigma_{\beta\alpha}(-\mathbf{H})$.

B. Dichroic f-sum rule

With the notation

$$\langle f\rangle \equiv \int_0^\infty f(\omega)d\omega, \tag{9}$$

the dichroic f-sum rule relates the integrated MCD spectrum $\langle\sigma''_{{\rm A},\alpha\beta}\rangle$ to a certain ground-state property of the system. To see how, we begin by expressing $\sigma''_{{\rm A},\alpha\beta}(\omega)$ as the imaginary part of the Kubo formula (6). Combining with Eq. (2) and taking the integral,

$$\langle\sigma''_{{\rm A},\alpha\beta}\rangle = \frac{\pi e^2}{\hbar V}\sum_n^{\rm occ}\sum_m^{\rm empty}{\rm Im}\left(\frac{\langle n|\hat{v}_\alpha|m\rangle\langle m|\hat{v}_\beta|n\rangle}{\omega_{mn}}\right). \tag{10}$$

Using the identity

$$\frac{\langle n|\hat{v}_\alpha|m\rangle}{\omega_{mn}} = -i\langle n|\hat{r}_\alpha|m\rangle \tag{11}$$

and defining the projector onto the empty states $\hat{Q} = \sum_m^{\rm empty}|m\rangle\langle m|$,

$$\langle\sigma''_{{\rm A},\alpha\beta}\rangle = -\frac{\pi e^2}{2\hbar V}\sum_n^{\rm occ}\langle n|\hat{r}_\alpha\hat{Q}\hat{v}_\beta|n\rangle - (\alpha\leftrightarrow\beta). \tag{12}$$

Introducing the pseudo-vector $\boldsymbol{\sigma}''_{\rm A} = (1/2)\epsilon_{\alpha\beta\gamma}\sigma''_{{\rm A},xy}$ and $\hat{P} = \sum_m^{\rm occ}|n\rangle\langle n|$, this can be written more concisely as

$$\langle\boldsymbol{\sigma}''_{\rm A}\rangle = -\frac{\pi e^2}{2\hbar V}{\rm Tr}[\hat{P}\hat{\mathbf{r}}\times\hat{Q}\hat{\mathbf{v}}]. \tag{13}$$

Eq. (13) is the dichroic f-sum rule, also obtained in Ref. 5. Using the closure relation $\hat{Q} = \hat{1} - \hat{P}$, it becomes apparent that the right-hand-side depends exclusively on the occupied states, and is closely related to the ground-state orbital magnetization $\mathbf{M}_{\rm orb} = \gamma{\rm Tr}[\hat{P}\hat{\mathbf{r}}\times\hat{\mathbf{v}}]$, where $\gamma = -(e/2cV)$ in electrostatic units (esu). Writing

$$\mathbf{M}_{\rm orb} = \mathbf{M}^{\rm (I)}_{\rm SR} + \boldsymbol{\Delta}\mathbf{M}, \tag{14}$$

with

$$\mathbf{M}^{\rm (I)}_{\rm SR} = \gamma{\rm Tr}[\hat{P}\hat{\mathbf{r}}\times\hat{Q}\hat{\mathbf{v}}] \tag{15}$$

and

$$\boldsymbol{\Delta}\mathbf{M} = \gamma{\rm Tr}[\hat{P}\hat{\mathbf{r}}\times\hat{P}\hat{\mathbf{v}}] \tag{16}$$

(the notation will be explained shortly), Eq. (13) becomes

$$\langle\boldsymbol{\sigma}''_{\rm A}\rangle = \frac{\pi ec}{\hbar}\mathbf{M}^{\rm (I)}_{\rm SR}. \tag{17}$$

Hence the sum rule yields an orbital quantity $\mathbf{M}^{\rm (I)}_{\rm SR}$ with units of magnetization, but differing from the actual orbital magnetization by the remainder $\boldsymbol{\Delta}\mathbf{M}$.

Two of the three quantities in Eq. (14) are independently measurable. The left-hand side can be determined from gyromagnetic experiments,[12] while $\mathbf{M}_{\rm SR}^{\rm (I)}$ on the right-hand side is obtainable from magneto-optical experiments via the sum rule. Thus, their difference $\mathbf{\Delta M}$ can also be determined in principle. However, measuring $\mathbf{M}_{\rm SR}^{\rm (I)}$ and $\mathbf{\Delta M}$ independently will be of only limited interest unless some physical meaning can be attached to each of them separately. With this goal in mind we shall now make contact with the recent theory of macroscopic orbital magnetization.

C. Relation to the orbital magnetization

The results obtained so far are fairly general. To proceed further we specialize to insulating samples. For the present purposes "insulating" means that the ground state wavefunction can be written as a Slater determinant of well-localized orthonormal molecular orbitals $|w_i\rangle$, which we will generically refer to as Wannier functions (WFs) even when the sample does not have a crystalline interior.[13] This definition encompasses a broad range of systems, both macroscopic and microscopic, but it excludes metals and Chern insulators,[14] which are not Wannier-representable in the above sense.

By invariance of the trace, the orbital magnetization can be expressed in the Wannier representation as

$$\mathbf{M}_{\rm orb} = \gamma \sum_i^{\rm occ} \langle w_i|\hat{\mathbf{r}} \times \hat{\mathbf{v}}|w_i\rangle. \tag{18}$$

In Ref. 6 this was decomposed as[15]

$$\mathbf{M}_{\rm orb} = \mathbf{M}_{\rm SR} + \mathbf{M}_{\rm IC} \tag{19}$$

where

$$\mathbf{M}_{\rm SR} = \gamma \sum_i^{\rm occ} \langle w_i|(\hat{\mathbf{r}} - \overline{\mathbf{r}}_i) \times \hat{\mathbf{v}}|w_i\rangle \tag{20}$$

arises from the circulation of the occupied WFs around their centers $\overline{\mathbf{r}}_i = \langle w_i|\hat{\mathbf{r}}|w_i\rangle = \mathbf{r}_{ii}$ ("self-rotation"), while

$$\mathbf{M}_{\rm IC} = \overline{\mathbf{r}}_i \times \langle w_i|\hat{\mathbf{v}}|w_i\rangle \tag{21}$$

is the circulation arising from the motion of the centers of mass of the WFs.

It is well known that the WFs of a given system are not uniquely defined; unitary mixing among the WFs is allowed, giving rise to a "gauge freedom." In practice one deals with this issue by choosing, among the infinitely many possible gauges, a particular one that has certain desirable properties. A common strategy is to work in the gauge that minimizes the quadratic spread of the WFs, producing so-called maximally-localized WFs.[13] Naturally, any physical observable (e.g., $\mathbf{M}_{\rm orb}$) is necessarily invariant under a change of gauge. This is unfortunately not the case for the individual terms $\mathbf{M}_{\rm SR}$ and $\mathbf{M}_{\rm IC}$ in Eqs. (20)–(21), which turn out to be gauge-dependent. This is to be expected since these quantities do not take the form of traces, unlike those in the decomposition introduced earlier via Eqs. (14)–(16).

The two decompositions (14)–(16) and (19)–(21) are not unrelated, however. To see this, we insert the identity $\hat{1} = \hat{Q} + \hat{P}$ at the location of the cross product in Eq. (20) to obtain

$$\mathbf{M}_{\rm SR} = \mathbf{M}_{\rm SR}^{\rm (I)} + \mathbf{M}_{\rm SR}^{\rm (II)}, \tag{22}$$

where $\mathbf{M}_{\rm SR}^{\rm (I)}$ is the quantity defined in Eq. (15) (since $\langle w_i|\hat{Q} = 0$), and

$$\begin{aligned} \mathbf{M}_{\rm SR}^{\rm (II)} &= \gamma\Big({\rm Tr}[\hat{P}\hat{\mathbf{r}} \times \hat{P}\hat{\mathbf{v}}] - \sum_i^{\rm occ} \overline{\mathbf{r}}_i \times \overline{\mathbf{v}}_i\Big) \\ &= \gamma \sum_{i,j\neq i}^{\rm occ} \mathbf{r}_{ij} \times \mathbf{v}_{ji}. \end{aligned} \tag{23}$$

In this way we have segregated the gauge-dependence of $\mathbf{M}_{\rm SR}$ to the term $\mathbf{M}_{\rm SR}^{\rm (II)}$, isolating a gauge-invariant part $\mathbf{M}_{\rm SR}^{\rm (I)}$ which turns out to be precisely the quantity defined in Eq. (15) and appearing in the sum rule (17). When the gauge-dependent self-rotation $\mathbf{M}_{\rm SR}^{\rm (II)}$ is combined with the gauge-dependent itinerant circulation $\mathbf{M}_{\rm IC}$, it forms the gauge-invariant quantity $\mathbf{\Delta M}$ of Eq. (16). The relation between the decompositions (14)–(16) and (19)–(21) can be summarized by writing

$$\mathbf{M}_{\rm orb} = \mathbf{M}_{\rm SR}^{\rm (I)} + \underbrace{\mathbf{M}_{\rm SR}^{\rm (II)} + \mathbf{M}_{\rm IC}}_{\mathbf{\Delta M}}. \tag{24}$$

There is a remarkable parallelism between the decomposition (22) of the Wannier self-rotation (20) and the decomposition[13]

$$\Omega = \Omega_{\rm I} + \widetilde{\Omega} \tag{25}$$

of the Wannier spread

$$\Omega = \sum_n^{\rm occ} \langle w_i|(\hat{\mathbf{r}} - \overline{\mathbf{r}}_i)^2|w_i\rangle \tag{26}$$

into a gauge-invariant part

$$\Omega_{\rm I} = \sum_\alpha {\rm Tr}\,[\hat{P}\hat{r}_\alpha\hat{Q}\hat{r}_\alpha] \tag{27}$$

and a gauge-dependent part

$$\widetilde{\Omega} = \sum_{i,j\neq i}^{\rm occ} |\mathbf{r}_{ij}|^2. \tag{28}$$

The similarities between Eqs. (15) and (27), and between Eqs. (23) and (28), are striking. (Interestingly,

the gauge-invariant spread $\Omega_{\rm I}$ is related to the "ordinary" absorption spectrum by a second sum rule, as discussed in Ref. 16 and Appendix A. In addition, the interpretation of $\Omega_{\rm I}$ as a measure of the quadratic quantum fluctuations, or "quantum spread," of the many-electron center of mass[16] is mirrored by $\mathbf{M}_{\rm SR}^{\rm (I)}$ having the meaning of a center-of-mass circulation, as discussed in Appendix B.) First-principles calculations show that for maximally-localized WFs, $\widetilde{\Omega}$ is typically much smaller than $\Omega_{\rm I}$.[13] Indeed, the minimization of the spread acts precisely to reduce $\widetilde{\Omega}$ as much as possible. In general $\widetilde{\Omega}$ cannot be made to vanish exactly in two or higher dimensions, since the non-commutivity of $\hat{P}\hat{x}\hat{P}$, $\hat{P}\hat{y}\hat{P}$, and $\hat{P}\hat{z}\hat{P}$ implies that the off-diagonal $\mathbf{r}_{ij}$ cannot all be zero. In practice, however, they can become quite small. According to Eq. (23), $\mathbf{M}_{\rm SR}^{\rm (II)}$ would also vanish if all off-diagonal $\mathbf{r}_{ij}$ were precisely zero. Hence we expect the self-rotation of maximally-localized WFs to be dominated by the gauge-invariant part as well.[17]

The fact that $\mathbf{\Delta M}$ is composed of self-rotation and itinerant-circulation parts which are not separately gauge-invariant means that angular momentum can be converted back and forth between $\mathbf{M}_{\rm SR}$ and $\mathbf{M}_{\rm IC}$ via gauge transformations. This will be discussed in more detail in Sec. III D; here we simply note that the two parts are similar in that both originate from the spatial overlap between neighboring WFs. This is evident from the definition of $\mathbf{M}_{\rm SR}^{\rm (II)}$, and for $\mathbf{M}_{\rm IC}$ it follows from writing $\overline{\mathbf{v}}_i$ in terms of the "current donated from one Wannier orbital to its neighbors" as in Ref. 6. $\mathbf{\Delta M}$ can therefore be interpreted as an *interorbital* contribution to $\mathbf{M}_{\rm orb}$, even though it includes part of the self-rotation, while $\mathbf{M}_{\rm SR}^{\rm (I)}$ is the purely *intraorbital* portion. (Similarly, $\Omega_{\rm I}$ and $\widetilde{\Omega}$ are the intraorbital and interorbital parts of the Wannier spread, respectively.)

III. CRYSTALLINE SOLIDS

In this Section we apply the general formalism of Sec. II to crystalline solids, recasting the relevant quantities in the form of Brillouin zone integrals. We start in Sec. III A by rederiving the dichroic f-sum rule for Bloch electrons. In the remaining subsections we explore the connections between this bulk reformulation and the theory of orbital magnetization in crystals.[6,7,8,9]

A somewhat unsatisfying aspect of that theory as developed in Ref. 8 is the lack of consistency in the way the orbital magnetization was decomposed, in the following sense. One partition ($\mathbf{M}_{\rm orb} = \mathbf{M}_{\rm LC} + \mathbf{M}_{\rm IC}$ in their notation[15]) was made for bounded samples, after which the thermodynamic limit was taken for each term separately. The resulting k-space expressions were then combined to form the total $\mathbf{M}_{\rm orb}$. Finally, working in k-space, a *different* partition ($\mathbf{M}_{\rm orb} = \widetilde{\mathbf{M}}_{\rm LC} + \widetilde{\mathbf{M}}_{\rm IC}$) was identified whose individual terms were gauge-invariant, unlike those of the original decomposition. In the process, however, the intuitive real-space interpretation of the original decomposition was lost, and the separate meanings of the two terms in the gauge-invariant decomposition was left unclear.

Here, instead, we shall work from the very beginning with the two gauge-invariant terms $\mathbf{M}_{\rm SR}^{\rm (I)}$ and $\mathbf{\Delta M}$, which afford a simple real-space interpretation in terms of WFs. They are first identified for fragments with a crystalline interior (crystallites) in Sec. III B. The thermodynamic limit of each term is then taken, producing the reciprocal-space expressions of Eqs. (44)–(45) (the details of the derivation can be found in Appendix C). Interestingly, we find that our gauge-invariant terms $\mathbf{M}_{\rm SR}^{\rm (I)}$ and $\mathbf{\Delta M}$ differ from – but are simply related to – those of the gauge-invariant decomposition of Ref. 8. In the particular case of an insulator with a single valence band, on the other hand, they reduce exactly to the terms identified in Ref. 7, as will be discussed in Sec. III C. Because the work of Ref. 7 is based on a semiclassical picture of wavepacket dynamics, however, it is not easily generalized to a multiband gauge-invariant framework as is done here.

In Eq. (24) of Sec. II the decomposition $\mathbf{\Delta M} = \mathbf{M}_{\rm SR}^{\rm (II)} + \mathbf{M}_{\rm IC}$ for insulating systems was obtained by working in the Wannier representation. For insulating crystallites $\mathbf{M}_{\rm IC}$ can be divided further into a "surface" part $\mathbf{M}_{\rm IC}^{\rm (surf)}$ and an "interior" part $\mathbf{M}_{\rm IC}^{\rm (int)}$. The interplay between the resulting three contributions to $\mathbf{\Delta M}$ will be the focus of the final two subsections. Single-band insulators are discussed in Sec. III C. The general case of multiband insulators is considered in Sec. III D, where the gauge-transformation properties of those terms is analyzed.

A. Dichroic f-sum rule

The first step is to rewrite the Kubo formula (6) in a form appropriate for periodic crystals, where dipole transitions connect valence and conduction Bloch states with the same crystal momentum $\mathbf{k}$. Eq. (2) becomes, dropping the index $\mathbf{k}$ for conciseness,

$$f_{nm,\alpha\beta} = -(2m_e\omega_{mn}/\hbar)\langle u_n|\partial_\alpha u_m\rangle\langle u_m|\partial_\beta u_n\rangle, \quad (29)$$

where $\partial_\alpha \equiv \partial/\partial k_\alpha$ and we have used the relation[18] $v_{nm,\alpha} = \omega_{mn}\langle u_n|\partial_\alpha u_m\rangle$ for $m \neq n$, with $|u_n\rangle$ a cell-periodic Bloch state. Eq. (6) now reads

$$\sigma_{\rm H}(\omega) = \frac{\pi e^2}{2m_e}\int \frac{d\mathbf{k}}{(2\pi)^3}\sum_n^{\rm occ}\sum_m^{\rm empty} f_{nm}\delta(\omega-\omega_{mn}). \quad (30)$$

Consider the frequency integral of $\sigma_{\rm H}(\omega)$,

$$\langle\sigma_{\rm H}\rangle = \frac{\pi e^2}{2m_e}\int \frac{d\mathbf{k}}{(2\pi)^3}\sum_n^{\rm occ}\sum_m^{\rm empty} f_{nm}. \quad (31)$$

The dichroic f-sum rule will be obtained from the imaginary part of this complex quantity, while the real part yields the ordinary f-sum rule (see Appendix A).

Using Eq. (29) to expand the summation,

$$\begin{aligned}
&\sum_n^{\rm occ}\sum_m^{\rm empty} f_{nm,\alpha\beta} = \\
&= -\frac{2m_e}{\hbar^2}\sum_n^{\rm occ}\sum_m^{\rm empty}\langle u_n|\partial_\alpha u_m\rangle(E_m-E_n)\langle u_m|\partial_\beta u_n\rangle \\
&= -\frac{2m_e}{\hbar^2}\sum_n^{\rm occ}\sum_m^{\rm empty}\langle \partial_\alpha u_n|u_m\rangle(E_n-E_m)\langle u_m|\partial_\beta u_n\rangle \\
&= \frac{2m_e}{\hbar^2}(g_{\mathbf{k},\alpha\beta}-h_{\mathbf{k},\alpha\beta}),
\end{aligned} \tag{32}$$

where we have introduced a set of notations as follows:

$$b_{\mathbf{k},\alpha\beta} = \sum_n^{\rm occ}\langle\widetilde{\partial}_\alpha u_n|\widetilde{\partial}_\beta u_n\rangle, \tag{33}$$

$$g_{\mathbf{k},\alpha\beta} = \sum_n^{\rm occ}\langle\widetilde{\partial}_\alpha u_n|\hat{H}|\widetilde{\partial}_\beta u_n\rangle, \tag{34}$$

and

$$h_{\mathbf{k},\alpha\beta} = \sum_n^{\rm occ} E_n\,\langle\widetilde{\partial}_\alpha u_n|\widetilde{\partial}_\beta u_n\rangle. \tag{35}$$

The symbol $\widetilde{\partial}$ denotes the covariant derivative,[8,19] defined as $|\widetilde{\partial}_\alpha u_{n\mathbf{k}}\rangle = \hat{Q}_{\mathbf{k}}|\partial_\alpha u_{n\mathbf{k}}\rangle$, where $\hat{Q}_{\mathbf{k}} = \sum_m^{\rm empty}|u_{m\mathbf{k}}\rangle\langle u_{m\mathbf{k}}|$. The imaginary part of $b_{\mathbf{k},\alpha\beta}$ is essentially the Berry curvature while its real part is related to the quantum metric (Ref. 13, Appendix C; we discuss the physical content of $b_{\mathbf{k},\alpha\beta}$ in Appendix A). Quantities $g_{\mathbf{k},\alpha\beta}$ and $h_{\mathbf{k},\alpha\beta}$ are similar to $b_{\mathbf{k},\alpha\beta}$ except that they carry an extra factor of Hamiltonian or energy. Note that $b_{\mathbf{k},\alpha\beta}$ corresponds to the quantity $f_{\mathbf{k},\alpha\beta}$ in Ref. 8, while $g_{\mathbf{k},\alpha\beta}$ and $h_{\mathbf{k},\alpha\beta}$ are the same as in that work.

With these definitions Eq. (31) becomes

$$\langle\sigma_{{\rm H},\alpha\beta}\rangle = \frac{\pi e^2}{\hbar^2}\int\frac{d\mathbf{k}}{(2\pi)^3}\,(g_{\mathbf{k},\alpha\beta}-h_{\mathbf{k},\alpha\beta}). \tag{36}$$

The imaginary part reads, in vector form,

$$\langle\boldsymbol{\sigma}''_{\rm A}\rangle = \frac{\pi e^2}{\hbar^2}\int\frac{d\mathbf{k}}{(2\pi)^3}\,(\mathbf{g}''_{\mathbf{k}}-\mathbf{h}''_{\mathbf{k}}). \tag{37}$$

This is the dichroic f-sum rule in the Bloch representation.

We can now compare this result with the decomposition obtained in Ref. 8, where the ground-state orbital magnetization was partitioned into two gauge-invariant terms as

$$\mathbf{M}_{\rm orb} = \widetilde{\mathbf{M}}_{\rm LC} + \widetilde{\mathbf{M}}_{\rm IC} \tag{38}$$

where

$$\widetilde{\mathbf{M}}_{\rm LC} = \frac{e}{\hbar c}\int\frac{d\mathbf{k}}{(2\pi)^3}\,\mathbf{g}''_{\mathbf{k}}, \tag{39}$$

$$\widetilde{\mathbf{M}}_{\rm IC} = \frac{e}{\hbar c}\int\frac{d\mathbf{k}}{(2\pi)^3}\,\mathbf{h}''_{\mathbf{k}}. \tag{40}$$

We thus arrive at our main result

$$\langle\boldsymbol{\sigma}''_{\rm A}\rangle = \frac{\pi ec}{\hbar}\big(\widetilde{\mathbf{M}}_{\rm LC}-\widetilde{\mathbf{M}}_{\rm IC}\big) \tag{41}$$

relating the integrated MCD spectrum to the components of the orbital magnetization. Note that the sum rule is proportional to the *difference* between the gauge-invariant contributions of Ref. 8. By independently measuring the sum of $\widetilde{\mathbf{M}}_{\rm LC}$ and $\widetilde{\mathbf{M}}_{\rm IC}$ via gyromagnetic experiments[12] and the difference via the magneto-optical sum rule, the value of each individual term can indeed be measured in principle, resolving an open problem posed in Ref. 8.

Strictly speaking, Eqs. (38)–(40) as written are valid for conventional insulators only. The generalization to metals and Chern insulators is subtle, but the understanding emerging from Refs. 7,8,9 is that it the appropriate generalization is obtained by making the replacements $H\to H-\mu$ and $E_n\to E_n-\mu$ in Eqs. (34) and (35), where μ is the electron chemical potential. Clearly $g_{\mathbf{k}}-h_{\mathbf{k}}$, and with it the sum rule (37), are insensitive to these substitutions.[20]

Comparing Eqs. (38) and (41) for extended crystals with Eqs. (14) and (17) for bounded samples, it appears plausible that the two partitions (14) and (38) of $\mathbf{M}_{\rm orb}$ ought to be related by

$$\mathbf{M}^{\rm (I)}_{\rm SR} = \widetilde{\mathbf{M}}_{\rm LC} - \widetilde{\mathbf{M}}_{\rm IC}, \tag{42}$$

$$\boldsymbol{\Delta}\mathbf{M} = 2\widetilde{\mathbf{M}}_{\rm IC}, \tag{43}$$

or explicitly,

$$\mathbf{M}^{\rm (I)}_{\rm SR} = \frac{e}{\hbar c}\int\frac{d\mathbf{k}}{(2\pi)^3}\,(\mathbf{g}''_{\mathbf{k}}-\mathbf{h}''_{\mathbf{k}}), \tag{44}$$

$$\boldsymbol{\Delta}\mathbf{M} = \frac{2e}{\hbar c}\int\frac{d\mathbf{k}}{(2\pi)^3}\,\mathbf{h}''_{\mathbf{k}}. \tag{45}$$

The correctness of these identities is demonstrated in Appendix C by taking the thermodynamic limit of results derived in the next subsection.

B. The magnetization of an insulating crystallite

To gain a deeper insight into the expressions derived from the reciprocal-space Kubo formula in the previous section, we now specialize the results obtained for bounded samples in Sec. II C to the case that the sample has a crystalline interior. Working in the Wannier representation, we are then able to establish connections between the k-space and Wannier viewpoints and associate a local physical picture with the various terms appearing in the bulk orbital magnetization.

Following Refs. 6 and 8, we divide our crystallite into "surface" and "interior" regions. This division is largely arbitrary, and it only needs to satisfy two requirements: (i) the border between the two regions should be placed sufficiently deep inside the sample where the local environment is already crystalline, and (ii) the surface region should occupy a non-extensive volume in the thermodynamic limit. The Wannier orbitals spanning the ground state are assigned to each region. Those in the interior converge exponentially to the bulk WFs $|\mathbf{R}n\rangle$ ($\mathbf{R}$ is a lattice vector), and those on the surface will be denoted by $|w_s\rangle$.

We first divide the orbital magnetization into self-rotation (SR) and itinerant-circulation (IC) contributions according to Eqs. (19)–(21). In the thermodynamic limit the SR part, which only involves the relative coordinate $\hat{\mathbf{r}}-\overline{\mathbf{r}}_i$, is dominated by the interior region. Invoking translational invariance,

$$\mathbf{M}_{\rm SR} = \gamma_c \sum_n \left[\langle \mathbf{0}n|\hat{\mathbf{r}}\times\hat{\mathbf{v}}|\mathbf{0}n\rangle - \overline{\mathbf{r}}_n\times\overline{\mathbf{v}}_n\right], \tag{46}$$

where $\gamma_c = -e/(2cV_c)$, with V_c the cell volume, $\overline{\mathbf{r}}_n = \langle \mathbf{0}n|\hat{\mathbf{r}}|\mathbf{0}n\rangle$, and $\overline{\mathbf{v}}_n = \langle \mathbf{0}n|\hat{\mathbf{v}}|\mathbf{0}n\rangle$. Henceforth summations over band-like indices span the valence-band states.

Next we break down the self-rotation as in Eq. (22), setting $\hat{P} = \sum_{\mathbf{R}}\sum_n |\mathbf{R}n\rangle\langle\mathbf{R}n|$:

$$\mathbf{M}_{\rm SR}^{\rm (I)} = \gamma_c {\rm Re}\,{\rm tr}_c[\hat{P}\hat{\mathbf{r}}\times\hat{Q}\hat{\mathbf{v}}] \tag{47}$$

$$\begin{aligned}\mathbf{M}_{\rm SR}^{\rm (II)} &= \gamma_c\Big({\rm Re}\,{\rm tr}_c[\hat{P}\hat{\mathbf{r}}\times\hat{P}\hat{\mathbf{v}}] - \sum_n \overline{\mathbf{r}}_n\times\overline{\mathbf{v}}_n\Big)\\ &= \gamma_c\sum_n\sum_{\mathbf{R}m\neq\mathbf{0}n} {\rm Re}\left\{\langle\mathbf{0}n|\hat{\mathbf{r}}|\mathbf{R}m\rangle\times\langle\mathbf{R}m|\hat{\mathbf{v}}|\mathbf{0}n\rangle\right\}.\end{aligned} \tag{48}$$

The symbol ${\rm tr}_c$ denotes the trace per unit cell. Note that we have taken the real part of the traces explicitly; this was not needed in Eqs. (15) and (23) for bounded samples, where the traces were automatically real.

Now we turn to the IC term (21) in Eq. (19). Unlike $\mathbf{M}_{\rm SR}$, in the thermodynamic limit it generally has contributions from both interior *and* surface regions:[6,8]

$$\mathbf{M}_{\rm IC} = \mathbf{M}_{\rm IC}^{\rm (int)} + \mathbf{M}_{\rm IC}^{\rm (surf)}. \tag{49}$$

The interior part becomes

$$\mathbf{M}_{\rm IC}^{\rm (int)} = \gamma_c\sum_n \overline{\mathbf{r}}_n\times\overline{\mathbf{v}}_n \tag{50}$$

where it was necessary to use

$$\sum_n \overline{\mathbf{v}}_n = 0 \tag{51}$$

when exploiting the translational invariance. Eq. (51) expresses the fact that no macroscopic current, or dynamic polarization,[19] flows through the bulk in a stationary state. Because of this constraint, the quantity (50) necessarily vanishes for insulators with a single valence band. In multiband insulators it takes the form of an *intracell* itinerant circulation: the WF centers in each cell can have a net circulation while their collective center-of-mass remains at rest.

Finally, the surface contribution is

$$\mathbf{M}_{\rm IC}^{\rm (surf)} = \gamma\sum_{s=1}^{N_s}\overline{\mathbf{r}}_s\times\overline{\mathbf{v}}_s. \tag{52}$$

It was shown in Refs. 6 and 8 that in the thermodynamic limit this can be recast as

$$\mathbf{M}_{\rm IC}^{\rm (surf)} = -\overline{\gamma}_c\,{\rm Im}\sum_{mn\mathbf{R}}\mathbf{R}\times\langle m\mathbf{0}|\hat{\mathbf{r}}|n\mathbf{R}\rangle\langle n\mathbf{R}|\hat{H}|m\mathbf{0}\rangle, \tag{53}$$

where $\overline{\gamma}_c = \gamma_c/\hbar$. This result is remarkable in that it expresses a circulation in the surface region solely in terms of matrix elements between the interior WFs, in a way that does not depend on the precise location of the boundary between the two regions (provided that the boundary satisfies the two criteria mentioned earlier). We emphasize that it holds for *crystalline* insulators only.

Whereas $\mathbf{M}_{\rm IC}^{\rm (int)}$ is an intracell-like term, in the bulk form (53) $\mathbf{M}_{\rm IC}^{\rm (surf)}$ is seen to have an *intercell* character, vanishing in the "Clausius-Mossotti" limit of zero overlap between WFs belonging to different cells. The assignment of the bulk WFs to specific cells is however not unique, and by making a different choice it is possible to convert between "intracell" $\mathbf{M}_{\rm IC}^{\rm (int)}$ and "intercell" $\mathbf{M}_{\rm IC}^{\rm (surf)}$. For this and other reasons to be detailed in Sec. III D, the interior and surface parts of $\mathbf{M}_{\rm orb}$ are in general not physically well-defined, even in crystalline insulators. Collecting terms, the full orbital magnetization reads

$$\mathbf{M}_{\rm orb} = \mathbf{M}_{\rm SR}^{\rm (I)} + \underbrace{\mathbf{M}_{\rm SR}^{\rm (II)} + \mathbf{M}_{\rm IC}^{\rm (int)} + \mathbf{M}_{\rm IC}^{\rm (surf)}}_{\boldsymbol{\Delta}\mathbf{M}}, \tag{54}$$

which is similar to Eq. (24) except that the IC term has been separated into interior and surface parts.

This Wannier-based decomposition of the magnetization of a crystallite follows closely that of Ref. 8. Two differences are worth noting. First, we have emphasized the distinction between Wannier self-rotation and itinerant circulation. In Ref. 8 the emphasis was more on the separation between the surface contribution $\mathbf{M}_{\rm IC}^{\rm (surf)}$ (denoted by $\mathbf{M}_{\rm IC}$ in that work) and the interior contribution $\mathbf{M}_{\rm LC} = \gamma_c{\rm tr}_c[\hat{P}\hat{\mathbf{r}}\times\hat{\mathbf{v}}]$ containing the net magnetic dipole density of the WFs in a crystalline cell. This "local circulation" includes all of the self-rotation as well as the intercell part of the itinerant circulation. In the present notation the decomposition of Ref. 8 reads

$$\mathbf{M}_{\rm orb} = \underbrace{\mathbf{M}_{\rm SR}^{\rm (I)} + \mathbf{M}_{\rm SR}^{\rm (II)} + \mathbf{M}_{\rm IC}^{\rm (int)}}_{\mathbf{M}_{\rm LC}} + \mathbf{M}_{\rm IC}^{\rm (surf)}. \tag{55}$$

Note that for one-band insulators $\mathbf{M}_{\rm IC}^{\rm (int)} = 0$, in which case the interior contribution coincides with the

self-rotation, and the surface part with the itinerant circulation.[6] Second, by identifying a gauge-invariant part of the self-rotation, we have been able to organize the four resulting terms into the two gauge-invariant groups indicated in Eq. (54).

The present viewpoint appears to be more useful for arriving at a simple physical picture for the sum rule. It has the additional advantage of being applicable to disordered and microscopic systems, for which the distinction between interior and surface contributions loses meaning.

C. One-band insulators

We begin our discussion of $\mathbf{M}_{\rm orb}$ in insulators with a single valence band by considering the remainder $\mathbf{\Delta M}$. We saw in Sec. III B that, of the three terms into which it is naturally decomposed in the Wannier representation, one of them vanishes if there is only one WF per cell,

$$\mathbf{M}_{\rm IC}^{\rm (int)} = 0. \tag{56}$$

Remarkably, the two surviving terms become identical,

$$\mathbf{M}_{\rm SR}^{\rm (II)} = \mathbf{M}_{\rm IC}^{\rm (surf)} = \frac{\mathbf{\Delta M}}{2}, \tag{57}$$

and thus *individually* gauge-invariant. This follows from Eqs. (42)–(43) in the one-band limit. Indeed, the quantities $\widetilde{\mathbf{M}}_{\rm LC}$ and $\widetilde{\mathbf{M}}_{\rm IC}$ therein were defined in Ref. 8 in such a way that for one-band insulators they reduce to the quantities $\mathbf{M}_{\rm LC} = \mathbf{M}_{\rm SR}$ and $\mathbf{M}_{\rm IC}^{\rm (surf)}$ in Eq. (55). It is then seen that Eqs. (42) and (43) correspond to the first and second equalities in Eq. (57) respectively. We emphasize that Eqs. (56)–(57) only hold for crystalline WFs which respect the full translational symmetry of the crystal. If, for instance, a larger unit cell is used (effectively folding the Brillouin zone and turning the system into a multiband insulator), the additional gauge freedom can be used to construct WFs for which Eqs. (56)–(57) no longer hold.

Consider now the full orbital magnetization. For one-band insulators the reciprocal-space expressions (44)–(45) reduce to

$$\mathbf{M}_{\rm SR}^{\rm (I)} = \frac{e}{2\hbar c}\,{\rm Im}\int \frac{d\mathbf{k}}{(2\pi)^3}\,\langle\partial_{\mathbf{k}}u_{\mathbf{k}}|\times(\hat{H}_{\mathbf{k}}-E_{\mathbf{k}})|\partial_{\mathbf{k}}u_{\mathbf{k}}\rangle \tag{58}$$

and

$$\mathbf{\Delta M} = \frac{e}{\hbar c}\int \frac{d\mathbf{k}}{(2\pi)^3}\,E_{\mathbf{k}}\,{\rm Im}\,\langle\partial_{\mathbf{k}}u_{\mathbf{k}}|\times|\partial_{\mathbf{k}}u_{\mathbf{k}}\rangle. \tag{59}$$

Their sum $\mathbf{M}_{\rm orb}$ is given by the right-hand-side of Eq. (58) with $-E_{\mathbf{k}}$ replaced by $+E_{\mathbf{k}}$, which is the expression originally obtained in Refs. 6 and 7. Moreover, the individual contributions $\mathbf{M}_{\rm SR}^{\rm (I)}$ and $\mathbf{\Delta M}$ coincide with those identified in Ref. 7. Instead, the derivation of Ref. 6 leads to the alternative – but, for one-band insulators, also gauge-invariant – partition into the "interior" and "surface" parts $\mathbf{M}_{\rm LC} = \mathbf{M}_{\rm SR} = \mathbf{M}_{\rm SR}^{\rm (I)} + \mathbf{\Delta M}/2$ and $\mathbf{M}_{\rm IC}^{\rm (surf)} = \mathbf{\Delta M}/2$.

While the individual terms $\mathbf{M}_{\rm SR}^{\rm (I)}$ and $\mathbf{\Delta M}$ agree, for single-band insulators, with those of Ref. 7, we interpret them somewhat differently here. Eq. (58) of Ref. 7 had the meaning of an intrinsic magnetic moment associated with the self-rotation of the carrier wavepackets. According to the present derivation, that term is only *part* of the Wannier self-rotation. As for Eq. (59), in the derivation of Ref. 7 it was seen to arise from a Berry-phase correction to the electronic density of states, and was subsequently claimed to be associated with a boundary current circulation.[21] Instead, according to the present viewpoint only *half* of it originates in the itinerant circulation $\mathbf{M}_{\rm IC}^{\rm (surf)}$ of the surface WFs, while the other half is ascribed to the remaining self-rotation $\mathbf{M}_{\rm SR}^{\rm (II)}$ of the WFs in the bulk.

D. Gauge transformations for multiband insulators

In multiband insulators all three terms $\mathbf{M}_{\rm SR}^{\rm (II)}$, $\mathbf{M}_{\rm IC}^{\rm (int)}$, and $\mathbf{M}_{\rm IC}^{\rm (surf)}$ can be nonzero. However, their individual values are not physically meaningful, since a gauge transformation can redistribute the total $\mathbf{\Delta M}$ among them. In particular, it is interesting to consider gauge transformations that shift the location of a WF by a lattice vector.

A general gauge transformation takes the form[13]

$$|u_{n\mathbf{k}}\rangle \to \sum_m |u_{m\mathbf{k}}\rangle U_{mn}(\mathbf{k}) \tag{60}$$

where $U_{mn}(\mathbf{k})$ is an $N_b \times N_b$ unitary matrix in the band indices. We assume that a transformation of this kind has already been applied to transform from the Hamiltonian eigenstates at each $\mathbf{k}$ to a set of states that are smooth in $\mathbf{k}$ from which the WFs are to be constructed. We can then interpose an additional diagonal gauge transformation

$$|u_{n\mathbf{k}}\rangle \to e^{-i\mathbf{k}\cdot\mathbf{R}_n}|u_{n\mathbf{k}}\rangle, \tag{61}$$

where $\mathbf{R}_n$ is a real-space lattice vector; this has the effect of shifting the location of WF n by $\mathbf{R}_n$. For a one-band insulator, or if $\mathbf{R}_n$ is the same for all bands, this amounts to shifting the choice of the "home" unit cell. However, in the multiband case different WFs can be shifted differently, corresponding to the freedom in choosing which WFs "belong" to the home unit cell.

For example, Fig. 1 shows four cells of a model two-dimensional crystal consisting of "molecular magnets" disposed on a square lattice with lattice constant a. Before the transformation (61), the home unit cell contains the four WFs shown in bold in Panel (a). Applying the transformation with $\mathbf{R}_2 = a\hat{x}$ and $\mathbf{R}_m = 0$ for all other WFs changes the selection of the "basis" of WFs belonging to the home cell to be that shown in Panel (b).

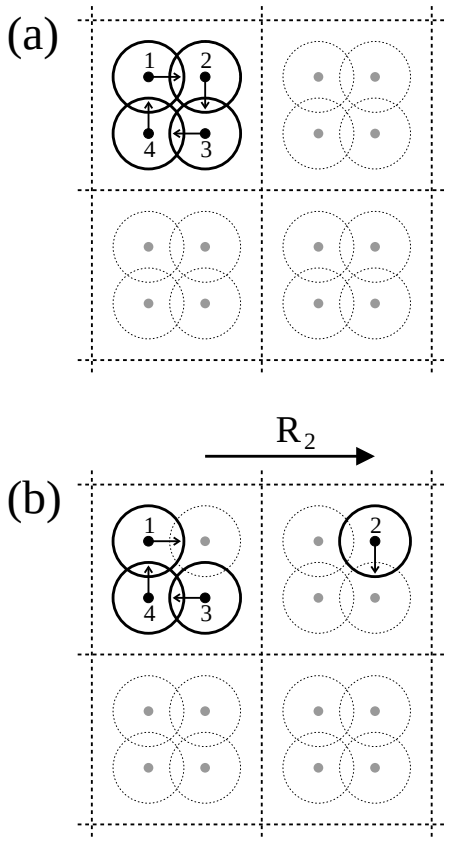

FIG. 1: Schematic model of a molecular crystal with one molecule per cell and four WFs per molecule. The range of the orbitals is indicated by the overlapping circles, and their center-of-mass velocities $\overline{\mathbf{v}}_n$ are denoted by arrows. The two panels show in bold two possible choices of "Wannier basis."

How does this affect the individual terms composing $\mathbf{\Delta M}$? Clearly the self-rotation (46) is not affected. According to Eq. (50), $\mathbf{M}_{\rm IC}^{\rm (int)}$ changes by $\gamma_c \mathbf{R}_2 \times \overline{\mathbf{v}}_2$. To preserve the overall invariance of $\mathbf{\Delta M}$ the remaining term $\mathbf{M}_{\rm IC}^{\rm (surf)}$ must change by an equal and opposite amount. Let us see in more detail how this comes about.

We begin with a general formal derivation. The k-space expression for $\mathbf{M}_{\rm IC}^{\rm (surf)}$ is given by[8]

$$\mathbf{M}_{\rm IC}^{\rm (surf)} = \frac{e}{2\hbar c}{\rm Im}\sum_{mn}\int\frac{d\mathbf{k}}{(2\pi)^3}E_{mn,\mathbf{k}}\langle\partial_{\mathbf{k}}u_{n\mathbf{k}}|\times|\partial_{\mathbf{k}}u_{m\mathbf{k}}\rangle. \tag{62}$$

A few steps of algebra show that under the transformation (61) $\mathbf{M}_{\rm IC}^{\rm (surf)}$ changes by

$$\frac{e}{\hbar c}\mathbf{R}_2\times\sum_m\int\frac{d\mathbf{k}}{(2\pi)^3}{\rm Re}\left\{\langle u_2|\partial_{\mathbf{k}}u_m\rangle\langle u_m|\hat{H}_{\mathbf{k}}|u_2\rangle\right\}. \tag{63}$$

Replacing $\langle u_2|\partial_{\mathbf{k}}u_m\rangle$ by $-\langle\partial_{\mathbf{k}}u_2|u_m\rangle$ allows to identify a term $\hat{P}_{\mathbf{k}}\hat{H}_{\mathbf{k}}=\hat{H}_{\mathbf{k}}$ in the above expression, which becomes

$$-\frac{e}{\hbar c}\mathbf{R}_2\times\int\frac{d\mathbf{k}}{(2\pi)^3}{\rm Re}\left\{\langle u_2|\hat{H}_{\mathbf{k}}|\partial_{\mathbf{k}}u_2\rangle\right\}. \tag{64}$$

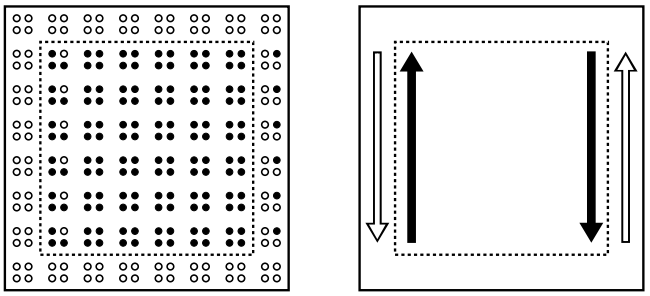

FIG. 2: A finite sample cut from the bulk crystal of Fig. 1. With the choice of Wannier basis of Fig. 1(a), "interior" and "surface" WFs lie inside and outside the dashed line respectively; with that of Fig. 1(b), they are denoted by solid and open circles respectively. Right panel: open and solid arrows show the extra "itinerant currents" ($\mathbf{M}_{\rm IC}^{\rm (surf)}$ and $\mathbf{M}_{\rm IC}^{\rm (int)}$ respectively) associated with the latter choice.

Comparison with the Wannier velocity[19]

$$\overline{\mathbf{v}}_n = -\frac{2V_c}{\hbar}\int\frac{d\mathbf{k}}{(2\pi)^3}{\rm Re}\left\{\langle u_n|\hat{H}_{\mathbf{k}}|\partial_{\mathbf{k}}u_n\rangle\right\} \tag{65}$$

then produces the desired result $-\gamma_c\mathbf{R}_2\times\overline{\mathbf{v}}_2$.

Coming back to the example in Fig. 1, the intramolecular orbital overlap gives rise to the nonzero velocities $\overline{\mathbf{v}}_n$ indicated by the arrows. With the choice of Wannier basis of Panel (a), the collective circulation of the Wannier centers in each cell results in a finite $\mathbf{M}_{\rm IC}^{\rm (int)}$, while from Eq. (53) $\mathbf{M}_{\rm IC}^{\rm (surf)}$ vanishes, since there is negligible intercell overlap. When the configuration of Panel (b) is chosen, $\mathbf{M}_{\rm IC}^{\rm (surf)}$ becomes $-\gamma_c\mathbf{R}_2\times\overline{\mathbf{v}}_2$. From the present viewpoint this nonzero value is made possible by the intramolecular (but now *inter*cell) overlap between the second WF of each cell with WFs one, three, and four from the cell shifted by $\mathbf{R}_2$.

To view $\mathbf{M}_{\rm IC}^{\rm (surf)}$ as a surface contribution rather than a bulk intercell term, we consider now a finite sample of the same crystal (Fig. 2), which has been divided into surface and interior regions. In deciding which WFs are "interior-like" and which are "surface-like" we shall require that all WFs assigned to the same cell must belong to the same region. If the Wannier basis of Fig. 1(a) is used, the surface region can be chosen to comprise the outermost layer of molecules, so that the border between the two regions is given by the dashed line. The four WFs on each molecule form a unit with some internal IC circulation but zero center-of-mass velocity. The total sample magnetization is the sum of all such internal circulations, which in the large-sample limit is interior-dominated, so that $\mathbf{M}_{\rm IC}^{\rm (surf)}\to 0$.

If the Wannier basis of Fig. 1(b) is chosen instead, the upper and lower surface regions are still composed of the outermost layer of molecules. However, the left surface now contains, in addition, one WF from each molecule in the second layer. Those lone surface WFs carry a down-

ward particle "IC current" which extends along the left surface and is indicated by an open arrow on the right panel. A corresponding IC current appears on the right surface, and together they yield $\mathbf{M}_{\rm IC}^{\rm (surf)} = -(e/2ca)|\overline{\mathbf{v}}_2|\hat{\mathbf{z}}$, which agrees with the result $\mathbf{M}_{\rm IC}^{\rm (surf)} = -\gamma_c \mathbf{R}_2 \times \overline{\mathbf{v}}_2$ found earlier using a purely bulk argument [in this example $\gamma_c = -e/(2ca^2)$]. A change of gauge should not change any physical quantity, such as the actual current flowing on the left surface. Since it appears to change by adding the open arrow, there must be another equal and opposite contribution (the adjacent solid arrow). This contribution is the "interior" IC current carried by the remaining three WFs (filled circles) on the molecules of the second layer.

The situation just described is reminiscent of the "quantum of polarization" in the theory of dielectric polarization,[22] where a change of Wannier basis like that leading from Fig. 1(a) to Fig. 1(b) shifts the polarization by a quantum and also changes the surface charge by one electron per surface cell area. This might suggest that the full gauge invariance of the interior and surface parts of $\mathbf{M}_{\rm orb}$ discussed in Sec. III C for single-band insulators becomes, in multiband insulators, a gauge-invariance modulo $\gamma_c \mathbf{R}_n \times \overline{\mathbf{v}}_n$. While true for this particular example, this is generally not so.[23] Even for this model it will cease to be true as soon as the molecules start overlapping significantly. When this happens, the value of $\mathbf{M}_{\rm IC}^{\rm (surf)}$ can be tuned continuously using other types of gauge transformations, e.g., the continuous diagonal transformation

$$|u_{n\mathbf{k}}\rangle \to e^{i\theta_{n\mathbf{k}}}|u_{n\mathbf{k}}\rangle \tag{66}$$

with $\theta_{n,\mathbf{k}+\mathbf{G}} = \theta_{n\mathbf{k}}$. When applied to the $n = 2$ states this produces a change in $\mathbf{M}_{\rm IC}^{\rm (surf)}$ given by Eq. (64) with $\mathbf{R}_2$ therein replaced by a factor of $\partial_{\mathbf{k}}\theta_{2\mathbf{k}}$ in the integrand. Since both $\overline{\mathbf{r}}_2$ and $\overline{\mathbf{v}}_2$ remain invariant (the former was shown in Ref. 18 and latter follows from Eq. (51) together with the fact that all other $\overline{\mathbf{v}}_m$ are unaffected), so does $\mathbf{M}_{\rm IC}^{\rm (int)}$. The change in $\mathbf{M}_{\rm IC}^{\rm (surf)}$ must therefore be absorbed by $\mathbf{M}_{\rm SR}^{\rm (II)}$.

To summarize, the transformation (61) transfers discrete amounts of itinerant circulation between the interior and surface regions, while the transformation (66) converts continuously between interior self-rotation and surface itinerant circulation. Finally, under the most general transformation (60) all three gauge-dependent terms in Eq. (54) can be affected simultaneously, so that only their sum $\mathbf{\Delta M}$ is unique and physically meaningful.

IV. SUMMARY AND OUTLOOK

We have presented an exact sum rule for the MCD spectrum, elucidated its physical interpretation, and related it to the recent rigorous formulation of orbital magnetization in crystals. In insulating systems the sum rule probes the gauge-invariant part $\mathbf{M}_{\rm SR}^{\rm (I)}$ of the self-rotation of the occupied Wannier orbitals. The total orbital magnetization has a second, less obvious contribution $\mathbf{\Delta M}$, arising from the overlap between neighboring WFs. It comprises both self-rotation (SR) and itinerant-circulation (IC) parts in proportions which depend on the precise choice of WFs, while $\mathbf{\Delta M}$ itself has a unique value. Although the intuitive interpretation in terms of the occupied WFs is restricted to Wannier-representable systems such as conventional insulators, the terms $\mathbf{M}_{\rm SR}^{\rm (I)}$ and $\mathbf{\Delta M}$ are in fact well-defined for all electron systems, including metals and Chern insulators.

The practical importance of the sum rule is that it provides a decomposition of $\mathbf{M}_{\rm orb}$ into its gauge-invariant constituents, using a combination of gyromagnetic and magneto-optical measurements. This should provide valuable information on the intraorbital (or localized) versus interorbital (or itinerant) character of orbital magnetism. For example, it has been suggested (Ref. 24, Appendix B) that the anomalously large g-factors of Bi might be caused by itinerant circulations very much like the ones discussed here. On the basis of the present work it should now be possible to test this conjecture.

In the last decade and a half an approximate sum rule for the X-ray MCD spectrum[4] has been used extensively to obtain site-specific information about orbital magnetism in solids. The resulting orbital moments have been compared with gyromagnetic measurements of $\mathbf{M}_{\rm orb}$.[25] Such a comparison may not be appropriate in systems where the remainder $\mathbf{\Delta M}$ is significant. It would therefore be of great interest to find such systems defying the conventional wisdom about the connection between the MCD spectrum and the orbital moments.

The ideas discussed in this work should be most relevant for materials displaying appreciable orbital magnetism and, in particular, appreciable *interorbital* effects which might enhance the ratio $|\mathbf{\Delta M}/\mathbf{M}_{\rm SR}^{\rm (I)}|$. These criteria do not favor band ferromagnets. First, their orbital magnetization tends to be relatively small. In the transition metal ferromagnets Fe, Co, and Ni, for example, it accounts for less than 10% of the spontaneous magnetization.[12] (For comparison, the field-induced orbital magnetization of the d paramagnetic metals can be as large as the induced spin magnetization.[26]) Secondly, ferromagnetism is favored by narrow bands and localized orbitals, for which interorbital effects are expected to be relatively minor. Paramagnets and diamagnets with relatively wide bands (e.g., the s-p metals and semiconductors) therefore appear to be more promising candidates. An important direction for future work is to perform first-principles calculations of $\mathbf{M}_{\rm SR}^{\rm (I)}$ and $\mathbf{\Delta M}$ for real materials.

In Appendix A we place the dichroic f-sum rule in the broader context of other known sum rules. We note in particular that by taking different inverse-frequency moments, the interband MCD spectrum can be related to two other phenomena resulting from broken time-reversal symmetry, namely the ground state orbital magnetization and the intrinsic anomalous Hall effect. These are

generally expected to coexist, and this is indeed the case for ferromagnets, where all three occur spontaneously. In the case of Pauli paramagnets, however, the intrinsic Hall mechanism of Karplus and Luttinger has received little if any attention. On the other hand, it is known that Pauli paramagnets can display an appreciable field-induced MCD spectrum.[27,28] This raises the question as to what role the Berry curvature may play in their "ordinary" (field-induced) Hall effect. Such a "dissipationless" contribution is undoubtedly present in principle by virtue of the sum rule (A7). Quantitative estimates and a discussion of the underlying mechanisms will be presented in a future communication.[29]

To conclude, we have shown how to relate the localized ($\mathbf{M}_{\rm SR}^{\rm (I)}$) and itinerant ($\mathbf{\Delta M}$) parts of the orbital magnetization to magneto-optical and gyromagnetic observables. This should allow one to probe more deeply into the nature of magnetism in solids and other many-electron systems than previously possible.

Acknowledgments

This work was supported by NSF Grant DMR-0549198.

APPENDIX A: OTHER SUM RULES

In this Appendix we derive three additional sum rules for Bloch electrons and discuss their relation to the dichroic f-sum rule. All four involve inverse-frequency moments $\langle\omega^{-p}\sigma_{\rm H}\rangle$ [in the notation of Eq. (9)] of the absorption spectrum (4). They are given by $p = 0$ and $p = 1$, and in each case two sum rules are obtained by taking the real and imaginary parts: one ordinary and the other dichroic, respectively.

We first consider $p = 0$. From the imaginary part of Eq. (36) we obtained the dichroic f-sum rule (41). To discuss the real part we revert from (36) to the form (31),

$$\langle\sigma'_{{\rm S},\alpha\beta}\rangle = \frac{\pi e^2}{2m_e}\int\frac{d\mathbf{k}}{(2\pi)^3}\sum_n^{\rm occ}\sum_m^{\rm empty} f'_{nm,\alpha\beta}. \tag{A1}$$

Since $f_{nm,\alpha\beta} = -[f_{mn,\alpha\beta}]^*$,

$$\begin{aligned}\sum_n^{\rm occ}\sum_m^{\rm empty} f'_{nm,\alpha\beta} &= \sum_n^{\rm occ}\sum_{m\neq n} f'_{nm,\alpha\beta}\\ &= \sum_n^{\rm occ}\left[\delta_{\alpha\beta} - \left(\frac{m_e}{m_e^*}\right)_{n,\alpha\beta}\right],\end{aligned} \tag{A2}$$

where the second equality is the effective-mass theorem. Hence we find

$$\langle\sigma'_{{\rm S},\alpha\beta}\rangle = \frac{\pi e^2}{2m_e}\int\frac{d\mathbf{k}}{(2\pi)^3}\sum_n^{\rm occ}\left[\delta_{\alpha\beta} - \left(\frac{m_e}{m_e^*}\right)_{n,\alpha\beta}\right], \tag{A3}$$

the *modified f-sum rule*[30] for the ordinary spectrum.

To obtain the two sum rules for $p = 1$ we again start from Eq. (30), but now replace Eq. (32) by

$$\sum_n^{\rm occ}\sum_m^{\rm empty}\frac{f_{nm,\alpha\beta}}{\omega_{mn}} = \frac{2m_e}{\hbar}\,b_{\mathbf{k},\alpha\beta}, \tag{A4}$$

where $b_{n,\alpha\beta}$ was defined in Eq. (33). Thus

$$\langle\omega^{-1}\sigma_{\rm H}\rangle = \frac{\pi e^2}{\hbar}\int\frac{d\mathbf{k}}{(2\pi)^3}\,b_{\mathbf{k}}. \tag{A5}$$

For the dichroic part, noting that $\mathbf{\Omega_k} = -2\mathbf{b}''_{\mathbf{k}}$ is the Berry curvature summed over the occupied states at $\mathbf{k}$, and comparing with the Karplus-Luttinger Hall conductivity[31]

$$\boldsymbol{\sigma}'_{\rm A}(\omega=0) = -\frac{e^2}{\hbar}\int\frac{d\mathbf{k}}{(2\pi)^3}\,\mathbf{\Omega_k}, \tag{A6}$$

one finds the *Hall sum rule*,

$$\langle\omega^{-1}\boldsymbol{\sigma}''_{\rm A}\rangle = \frac{\pi}{2}\boldsymbol{\sigma}'_{\rm A}(\omega=0). \tag{A7}$$

This is the $\omega\to 0$ limit of the Kramers-Krönig relation for the antisymmetric conductivity.[11] Since only the *interband* part of the optical conductivity was included on the left-hand-side, the *intrinsic* dc Hall conductivity was obtained on the right-hand-side. Extrinsic contributions to the latter (e.g., skew scattering) arise from intraband terms in the former.

Finally consider the ordinary (real) part of Eq. (A5). The quantity $b'_{n,\alpha\beta}$ is the quantum metric.[13] It is related to the localization tensor $\Lambda_{\alpha\beta}$ of insulators by[32]

$$\Lambda_{\alpha\beta} = \frac{V}{N}\int\frac{d\mathbf{k}}{(2\pi)^3}\,b'_{\mathbf{k},\alpha\beta}, \tag{A8}$$

where N/V is the electron density. Hence we recover the *electron localization sum rule*[16]

$$\langle\omega^{-1}\sigma'_{{\rm S},\alpha\beta}\rangle = \frac{\pi e^2 N}{\hbar V}\Lambda_{\alpha\beta}. \tag{A9}$$

In summary, we have in Eqs. (36) and (A5) two general sum rules for the zero-th and first inverse frequency moments of the optical absorption, respectively. Taking imaginary and real parts of (36) gives the dichroic f-sum rule (41) and the modified ordinary f-sum rule (A3), while taking imaginary and real parts of (A5) gives the Hall sum rule (A7) and the electron localization sum rule (A9).

Besides emerging from a unified formalism, the four sum rules display certain similarities. For instance, it will be shown in Appendix B that the dichroic f-sum rule yields the circulation of the electronic center of mass, while the trace of the localization tensor yields its quantum spread. Moreover, in a one-particle picture each quantity can be viewed as the gauge-invariant part of the

corresponding property (self-rotation or spread) of the Wannier orbitals, as discussed in Sec. II C for bounded systems. There is however one important difference between the behavior of the two quantities in the thermodynamic limit. While the center-of-mass circulation remains well-defined for metals, the trace of the localization tensor is only meaningful for insulators, diverging in metals.[16,32] Interestingly, the delocalization of electrons in metals is also responsible for a correction to the f-sum rule. Contrary to the canonical f-sum rule for atoms,[1] the modified f-sum rule (A3) does not yield the number density of valence electrons in a metal, due to the presence of the last term on the right-hand-side. This term appears because the Bloch states are extended and do not vanish at infinity.[30] The fact that the correction term nevertheless vanishes for insulators is a consequence of the localized nature of insulating many-body wavefunctions in *configuration* space.[33]

We conclude by noting that Eq. (A7) provides an extreme example of how sum rules from atomic physics can change qualitatively when applied to extended systems. Indeed, the corresponding sum rule for bound systems produces a vanishing result,[34]

$$\langle \omega^{-1}\sigma''_{\mathrm{A},\alpha\beta}\rangle = \frac{\pi e^2}{\hbar V}\mathrm{Im\,Tr}[\hat{P}\hat{r}_\alpha\hat{Q}\hat{r}_\beta] = 0. \tag{A10}$$

In contrast, the bulk formula (A7) produces for Chern insulators a quantized Hall conductivity, and it also describes the intrinsic anomalous Hall conductivity of ferromagnetic metals.[31] This apparent contradiction highlights the subtleties associated with the process of taking the thermodynamic limit and switching from open to periodic boundary conditions for non-Wannier-representable systems. Such issues are still not fully resolved in the theory of orbital magnetization. While a general derivation of the bulk formula for $\mathbf{M}_{\mathrm{orb}}$ has been given working from the outset with a periodic crystal,[9] derivations which start from finite crystallites and take them to the thermodynamic limit (Refs. 6 and 8 and Appendix C) are presently restricted to conventional insulators.

APPENDIX B: DICHROIC f-SUM RULE AND THE MANY-BODY WAVEFUNCTION

In the main text we interpreted the dichroic f-sum rule, and the associated decomposition (14) of $\mathbf{M}_{\mathrm{orb}}$, in an independent-particle picture based on WFs. It is also possible to relate these quantities directly to properties of the many-electron wavefunction, without invoking any particular single-particle representation. In preparation for that, let us first discuss a one-electron system (e.g., a hydrogen atom in a magnetic field). Its absorption spectrum is composed of sharp lines, and is more conveniently described in terms of an oscillator strength rather than an optical conductivity. Inserting Eq. (2) into Eq. (3) and using the relation (11) to replace one of the velocity matrix elements,

$$f''_{nm,\alpha\beta} = -\frac{m_e}{\hbar}\big[\langle n|\hat{r}_\alpha|m\rangle\langle m|\hat{v}_\beta|n\rangle - (\alpha\leftrightarrow\beta)\big]. \tag{B1}$$

Summing over $m\neq n$ and using the closure relation together with $\langle n|\hat{\mathbf{v}}|n\rangle = 0$ one finds, in vector notation,

$$\sum_{m\neq n}\mathbf{f}''_{nm} = -\frac{m_e}{\hbar}\langle n|\hat{\mathbf{r}}\times\hat{\mathbf{v}}|n\rangle. \tag{B2}$$

This is the original dichroic f-sum rule of Hasegawa and Howard,[2] with the orbital angular momentum appearing on the right-hand-side; in the notation of Sec. II B it reads $\langle\boldsymbol{\sigma}''_{\mathrm{A}}\rangle = (\pi ec/\hbar)\,\mathbf{M}_{\mathrm{orb}}$ (since here $\boldsymbol{\Delta}\mathbf{M}=0$).

We now generalize the discussion to N-electron systems. In this context $\hat{\mathbf{r}} = \sum_{i=1}^N \hat{\mathbf{r}}_i$ and $\hat{\mathbf{v}} = \sum_{i=1}^N \hat{\mathbf{v}}_i$, and it is crucial to make a distinction between the one-particle operator $\hat{\boldsymbol{\Lambda}}^{(1)} = \sum_{i=1}^N \hat{\mathbf{r}}_i\times\hat{\mathbf{v}}_i$ and the two-particle operator $\hat{\boldsymbol{\Lambda}}^{(2)} = \hat{\mathbf{r}}\times\hat{\mathbf{v}}$ appearing in Eq. (B2), as emphasized in Ref. 35. The former is related to the electronic angular momentum and orbital magnetization, while the latter is related to the contribution to the angular momentum coming from the circulation of the electronic center of mass. (In the classical context, for example, a pair of electrons orbiting 180° out of phase in the same circular orbit would have $\hat{\boldsymbol{\Lambda}}^{(1)}\neq 0$ but $\hat{\boldsymbol{\Lambda}}^{(2)}=0$.)

The derivation of the dichroic sum rule for the N-electron case proceeds as before, except that the velocity matrix elements in Eq. (2) become $\mathbf{v}_{nm} = \langle\Psi_n|\hat{\mathbf{v}}|\Psi_m\rangle$, where $|\Psi_m\rangle$ are now many-body eigenstates. The result is still given by Eq. (B2), with $|n\rangle$ replaced by $|\Psi_n\rangle$. Indeed, it is natural to define the many-body generalization of Eq. (15) as

$$\mathbf{M}^{(\mathrm{I})}_{\mathrm{SR}} = \gamma\langle\Psi_n|\hat{\boldsymbol{\Lambda}}^{(2)}|\Psi_m\rangle \tag{B3}$$

so that Eq. (17) continues to hold. From this many-body perspective the difference $\boldsymbol{\Delta}\mathbf{M}$ with respect to the full $\mathbf{M}_{\mathrm{orb}}$ is seen to arise from the cross terms $\sum_{i,j\neq i}^N \hat{\mathbf{r}}_i\times\hat{\mathbf{v}}_j$ in $\hat{\boldsymbol{\Lambda}}^{(2)} - \hat{\boldsymbol{\Lambda}}^{(1)}$.

To recover from (B3) the independent-particle expression (15) we specialize to the case where $|\Psi_m\rangle$ is a single Slater determinant. In second-quantized notation $\hat{\mathbf{r}} = \sum_{ij}\mathbf{r}_{ij}c_i^\dagger c_j$, $\hat{\mathbf{v}} = \sum_{ij}\mathbf{v}_{ij}c_i^\dagger c_j$, and $|\Psi_0\rangle = c_1^\dagger\ldots c_N^\dagger|0\rangle$, where i and j label orthogonal one-particle states. Then Eq. (B3) becomes

$$\mathbf{M}^{(\mathrm{I})}_{\mathrm{SR}} = \gamma\big(\langle 0|c_N\ldots c_1\big)\Big(\sum_{ij}\mathbf{r}_{ij}c_i^\dagger c_j\Big)\times \tag{B4}$$

$$\times\ \Big(\sum_{kl}\mathbf{v}_{kl}c_k^\dagger c_l\Big)\big(c_1^\dagger\ldots c_N^\dagger|0\rangle\big). \tag{B5}$$

Terms in which the indices do not pair can immediately be eliminated. Furthermore, pairings of the form ($k=l$, $i=j$) give no contribution, since this leads to $(\sum_i^{\mathrm{occ}}\mathbf{r}_{ii})\times(\sum_k^{\mathrm{occ}}\mathbf{v}_{kk})$ which vanishes because

$\langle\Psi_0|\hat{\mathbf{v}}|\Psi_0\rangle = 0$. The only surviving terms are those with $(j=k,\, i=l)$, yielding

$$\begin{aligned}\mathbf{M}_{\rm SR}^{\rm (I)} &= \gamma\sum_{ij}\mathbf{r}_{ij}\times\mathbf{v}_{ji}\big(\langle\Psi_0|c_jc_j^\dagger\big)\big(c_i^\dagger c_i|\Psi_0\rangle\big)\\ &= \gamma\sum_{ij}\mathbf{r}_{ij}\times\mathbf{v}_{ji}(1-n_j)n_i = \gamma\sum_{i}^{\rm occ}\sum_{j}^{\rm empty}\mathbf{r}_{ij}\times\mathbf{v}_{ji},\end{aligned} \tag{B6}$$

where n_i is the state occupancy. Clearly the expression on the right-hand-side is equivalent to that in Eq. (14).

A similar analysis can be made for the other sum rules presented in Appendix A. For example, the counterpart of the Hall sum rule for a bounded many-electron system reads

$$\sum_{m\neq 0}\frac{\mathbf{f}''_{0m}}{\omega_{m0}} = -\frac{im_e}{\hbar}\langle\Psi_0|\hat{\mathbf{r}}\times\hat{\mathbf{r}}|\Psi_0\rangle = 0, \tag{B7}$$

which was termed in Ref. 3 the *Kuhn sum rule*. The independent-particle form (A10) can be recovered from (B7) along the lines of Eqs. (B4)–(B6). As for the electron localization sum rule, it yields the second cumulant-moment of the quantum distribution of the many-electron center-of-mass.[16] In the independent-particle limit this reduces to $\Lambda_{\alpha\beta} = (1/N){\rm Tr}[\hat{P}\hat{r}_\alpha\hat{Q}\hat{r}_\beta]$, whose trace yields the gauge-invariant WF spread (27). The bulk formula (A8) for insulating crystals can be recovered in the thermodynamic limit following the strategy described below for the orbital magnetization.

APPENDIX C: THERMODYNAMIC LIMIT

In this Appendix we start from the expressions (15) and (16) for $\mathbf{M}_{\rm SR}^{\rm (I)}$ and $\mathbf{\Delta M}$ of insulating crystallites and, by taking the thermodynamic limit in the Wannier representation, turn them into the reciprocal-space expressions (44) and (45).

Before proceeding, recall that the quantities $\mathbf{g_k}$ and $\mathbf{h_k}$ entering Eqs. (44)–(45) were defined in Eqs. (34)–(35) in the context of the "Hamiltonian gauge" in which n labels a Bloch energy eigenstate. Here, we work with a generalized Wannier representation as in Sec. III B, where n labels a Wannier function and $|u_{n\mathbf{k}}\rangle$ is the state of Bloch symmetry (generally not an energy eigenstate) constructed from that Wannier function.[13] The two representations are related by a k-dependent unitary rotation as in Eq. (60). Then Eq. (34) remains valid in the present context, since it already takes the form of a trace, while Eq. (35) is now replaced by

$$h_{\mathbf{k},\alpha\beta} = \sum_{nm}E_{nm}(\mathbf{k})\langle\widetilde{\partial}_\alpha u_{m\mathbf{k}}|\widetilde{\partial}_\beta u_{n\mathbf{k}}\rangle, \tag{C1}$$

where $E_{nm}(\mathbf{k}) = \langle u_{n\mathbf{k}}|\hat{H}_k|u_{m\mathbf{k}}\rangle$. With $\mathbf{g_k}$ and $\mathbf{h_k}$ written as traces in this way, it is evident that each is a gauge-invariant quantity.[8]

1. Gauge-invariant self-rotation $\mathbf{M}_{\rm SR}^{\rm (I)}$

For insulating crystallites in the thermodynamic limit, Eq. (15) can be replaced by Eq. (47). Thus we need to establish the equivalence between Eqs. (47) and (44). Using

$$\mathbf{v} = -\frac{i}{\hbar}[\hat{\mathbf{r}},\hat{H}] \tag{C2}$$

and specializing to the z-component of Eq. (47),

$$M_{{\rm SR},z}^{\rm (I)} = \frac{e}{\hbar c V_c}{\rm Im\,tr}_c[\hat{P}\hat{x}\hat{Q}\hat{H}\hat{Q}\hat{y} - \hat{P}\hat{H}\hat{P}\hat{x}\hat{Q}\hat{y}]. \tag{C3}$$

The second term above may be expanded as a trace in the Wannier representation as

$${\rm tr}_c[\hat{P}\hat{H}\hat{P}\hat{x}\hat{Q}\hat{y}] = \sum_{\mathbf{R}}\sum_{mn}\langle\mathbf{0}m|\hat{H}|\mathbf{R}n\rangle\langle\mathbf{R}n|\hat{x}\hat{Q}\hat{y}|\mathbf{0}m\rangle. \tag{C4}$$

Then using the identities

$$\langle\mathbf{0}m|\hat{H}|\mathbf{R}n\rangle = V_c\int\frac{d\mathbf{k}}{(2\pi)^3}\,e^{-i\mathbf{k}\cdot\mathbf{R}}E_{mn,\mathbf{k}}, \tag{C5}$$

$$\langle\mathbf{R}n|\hat{x}\hat{Q}\hat{y}|\mathbf{0}m\rangle = V_c\int\frac{d\mathbf{k}}{(2\pi)^3}\,e^{i\mathbf{k}\cdot\mathbf{R}}\langle\widetilde{\partial}_x u_{n\mathbf{k}}|\widetilde{\partial}_y u_{m\mathbf{k}}\rangle, \tag{C6}$$

we obtain

$$\frac{1}{V_c}{\rm tr}_c[\hat{P}\hat{H}\hat{P}\hat{x}\hat{Q}\hat{y}] = \int\frac{d\mathbf{k}}{(2\pi)^3}\,h_{\mathbf{k},\alpha\beta}. \tag{C7}$$

Using a similar argument, it follows that

$$\frac{1}{V_c}{\rm tr}_c[\hat{P}\hat{x}\hat{Q}\hat{H}\hat{Q}\hat{y}] = \int\frac{d\mathbf{k}}{(2\pi)^3}\,g_{\mathbf{k},\alpha\beta}. \tag{C8}$$

Combining the above two equations with Eq. (C3) then yields Eq. (44).

2. Gauge-invariant remainder $\mathbf{\Delta M}$

To take the thermodynamic limit of $\mathbf{\Delta M}$ we start from Eq. (16) and apply it to a large crystallite to arrive at Eq. (45). Focusing on the z-component,

$$(\Delta M)_z = \gamma{\rm Tr}[\hat{P}\hat{x}\hat{Q}\hat{v}_y] - \gamma{\rm Tr}[\hat{P}\hat{y}\hat{Q}\hat{v}_x]. \tag{C9}$$

Now use Eq. (C2) to obtain

$$(\Delta M)_z = -2i\overline{\gamma}{\rm Tr}[\hat{x}\hat{P}\hat{y}\hat{P}\hat{H}\hat{P} - \hat{y}\hat{P}\hat{x}\hat{P}\hat{H}\hat{P}], \tag{C10}$$

where we defined $\overline{\gamma} = \gamma/\hbar$ and replaced $\hat{P}\hat{H}$ by the more symmetrical form $\hat{P}\hat{H}\hat{P}$. Using $-i{\rm Tr}[\hat{\mathcal{O}} - \mathcal{O}^\dagger] = 2{\rm Im\,Tr}[\hat{\mathcal{O}}]$, this becomes

$$(\Delta M)_z = 4\overline{\gamma}{\rm Im\,Tr}[\hat{x}\hat{P}\hat{y}(\hat{P}\hat{H}\hat{P})]. \tag{C11}$$

At this point we are still considering a bounded sample. To obtain a bulk expression we first need to manipulate Eq. (C11) into a form where the unbounded operators $\hat{x}$ and $\hat{y}$ are sandwiched between $\hat{P}$ and $\hat{Q}$, as in Eq. (C3). That ensures that ill-defined diagonal position matrix elements between the extended Bloch states do not occur. We will make use of the following rules for finite-dimensional Hermitian matrices A, B, C, and D:

$$\text{(i) } \mathrm{Im\,Tr}[ABCD] = \mathrm{Im\,Tr}[DABC] \tag{C12}$$

$$\text{(ii) } \mathrm{Im\,Tr}[ABCD] = -\mathrm{Im\,Tr}[DCBA] \tag{C13}$$

$$\text{(iii) } \mathrm{Im\,Tr}[AB] = 0 \tag{C14}$$

and, if any two of the matrices A, B, and C commute,

$$\text{(iv) } \mathrm{Im\,Tr}[ABC] = 0. \tag{C15}$$

Rule (i) is the cyclic permutation, rule (ii) is reversal, rule (iii) follows from (i) and (ii), and rule (iv) follows from (iii). Replacing the first $\hat{P}$ in Eq. (C11) by $\hat{1} - \hat{Q}$ and applying rule (iv) to the term containing $\hat{1}$ ($[\hat{x}, \hat{y}] = 0$ and $\hat{P}\hat{H}\hat{P}$ is Hermitian), we obtain

$$(\Delta M)_z = -4\overline{\gamma}\mathrm{Im\,Tr}[\hat{P}\hat{H}\hat{P}\hat{x}\hat{Q}\hat{y}], \tag{C16}$$

which has the desired form.

Now we invoke Wannier-representability to write

$$\mathrm{Tr}[\hat{P}\hat{H}\hat{P}\hat{x}\hat{Q}\hat{y}] = \sum_j \langle w_j|\hat{H}\hat{P}(\hat{x}-\overline{x}_j)\hat{Q}(\hat{y}-\overline{y}_j)|w_j\rangle \tag{C17}$$

(note that $\hat{P}\overline{\mathbf{r}}\hat{Q} = 0$). Since only the relative coordinate appears, the contribution from the surface orbitals is non-extensive, vanishing in the thermodynamic limit. We are then left with a bulk-like expression:

$$\mathrm{Tr}[\hat{P}\hat{H}\hat{P}\hat{x}\hat{Q}\hat{y}] \to \sum_{\mathbf{R}m}\sum_{\mathbf{R}'n} \langle \mathbf{R}m|\hat{H}|\mathbf{R}'n\rangle\langle \mathbf{R}'n|\hat{x}\hat{Q}\hat{y}|\mathbf{R}m\rangle. \tag{C18}$$

Both matrix elements on the right-hand-side depend on $\mathbf{R}$ and $\mathbf{R}'$ only through $\mathbf{R}'-\mathbf{R}$, and therefore, comparing with Eq. (C4),

$$\frac{1}{N_c}\mathrm{Tr}[\hat{P}\hat{H}\hat{P}\hat{x}\hat{Q}\hat{y}] \to \mathrm{tr}_c[\hat{P}\hat{H}\hat{P}\hat{x}\hat{Q}\hat{y}], \tag{C19}$$

where N_c is the number of crystalline cells in the sample. Combining Eqs. (C7), (C16) and (C19) one obtains Eq. (45), which concludes the proof.

[1] H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic Press, 1957).
[2] H. Hasegawa and R. E. Howard, J. Phys. Chem. Solids **21**, 179 (1961).
[3] D. Y. Smith, Phys. Rev. B **13**, 5303 (1976).
[4] B. T. Thole, P. Carra, F. Sette, and G. van der Laan, Phys. Rev. Lett. **68**, 1943 (1992).
[5] P. M. Oppeneer, J. Magn. Magn. Mat. **188**, 275 (1998).
[6] T. Thonhauser, D. Ceresoli, D. Vanderbilt, and R. Resta, Phys. Rev. Lett. **95**, 137205 (2005).
[7] D. Xiao, J. Shi, and Q. Niu, Phys. Rev. Lett. **95**, 137204 (2005).
[8] D. Ceresoli, T. Thonhauser, D. Vanderbilt, and R. Resta, Phys. Rev. B **74**, 024408 (2006).
[9] J. Shi, G. Vignale, D. Xiao, and Q. Niu, arXiv:0704.3824 (2007).
[10] F. D. M. Haldane, Phys. Rev. Lett. **61**, 2015 (1988).
[11] H. S. Bennett and E. A. Stern, Phys. Rev. **137**, A448 (1965).
[12] C. Kittel, *Introduction to Solid State Physics* (Wiley, 1953), 1st ed.
[13] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).
[14] T. Thonhauser and D. Vanderbilt, Phys. Rev. B **74**, 235111 (2006).
[15] The quantity $\mathbf{M}_{\mathrm{SR}}$ was called $\mathbf{M}_{\mathrm{LC}}$ in Ref. 6. One reason for using a different notation here is that in Ref. 8 $\mathbf{M}_{\mathrm{LC}}$ was defined (for crystals) in a different way, by replacing the Wannier center $\overline{\mathbf{r}}_i$ in Eq. (20) by the lattice vector $\mathbf{R}$ labeling the cell to which the WF "belongs." The two definitions turn out to be equivalent for single-band insulators (the case of interest in Ref. 6). In Sec. III B we discuss the precise relation between our $\mathbf{M}_{\mathrm{SR}}$ and the $\mathbf{M}_{\mathrm{LC}}$ of Ref. 8 in the multiband case.
[16] I. Souza, T. Wilkens, and R. M. Martin, Phys. Rev. B **62**, 1666 (2000).
[17] However, since the off-diagonal $\mathbf{r}_{ij}$ enter the self-rotation as a first power and the spread as a second power, the non-invariant part may play a larger role in the former quantity than in the latter.
[18] R. D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993).
[19] I. Souza, J. Íñiguez, and D. Vanderbilt, Phys. Rev. B **69**, 085106 (2004).
[20] For metals the low-frequency part of the absorption spectrum contains intraband (Drude) contributions in addition to the interband absorption. Only the latter was taken into account in the derivation of the dichroic f-sum rule.
[21] D. Xiao, Y. Yao, Z. Fang, and Q. Niu, Phys. Rev. Lett. **97**, 026603 (2006).
[22] D. Vanderbilt and R. D. King-Smith, Phys. Rev. B **48**, 4442 (1993).
[23] The quantum of polarization can ultimately be traced to the quantization of the charge in a Wannier function. On the contrary, neither the self-rotation nor the center-of-mass velocity of a Wannier function is quantized; these can change continuously under a multiband gauge transformation.
[24] Y. Yafet, Solid State Phys. **14**, 1 (1963).

[25] C. T. Chen, Y. U. Idzerda, H.-J. Lin, N. V. Smith, G. Meigs, E. Chaban, G. H. Ho, E. Pellegrin, and F. Sette, Phys. Rev. Lett. **75**, 152 (1995).
[26] A. Hjelm, J. Trygg, O. Eriksson, B. Johansson, and J. M. Wills, Int. J. Mod. Phys. B **9**, 2735 (1995).
[27] A. N. Yaresko, L. Uba, S. Uba, A. Y. Perlov, R. Gontarz, and V. N. Antonov, Phys. Rev. B **58**, 7648 (1998).
[28] H. Ebert and S. Man'kovsky, Phys. Rev. Lett. **90**, 077404 (2003).
[29] J. R. Yates, D. Vanderbilt, and I. Souza, (in preparation).
[30] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys* (Clarendon Press, 1936).
[31] Y. Yao, L. Kleinman, A. H. MacDonald, J. Sinova, T. Jungwirth, D.-S. Wang, E. Wang, and Q. Niu, Phys. Rev. Lett. **92**, 037204 (2004).
[32] R. Resta, J. Phys.: Condens. Matter **14**, R625 (2002).
[33] W. Kohn, in *Many-Body Physics*, edited by C. DeWitt and R. Balian (Gordon and Breach, New York, 1968), p. 351.
[34] To see that the middle expression in Eq. (A10) vanishes, write $\mathrm{Im\,Tr}[\hat{P}\hat{r}_\alpha\hat{Q}\hat{r}_\beta]$ as $\mathrm{Im\,Tr}[(\hat{P})(\hat{r}_\alpha\hat{r}_\beta)] - \mathrm{Im\,Tr}[(\hat{P}\hat{r}_\alpha\hat{P})(\hat{P}\hat{r}_\beta\hat{P})]$, note that $\hat{P}$, $\hat{r}_\alpha\hat{r}_\beta$, $\hat{P}\hat{r}_\alpha\hat{P}$ and $\hat{P}\hat{r}_\beta\hat{P}$ are all Hermitian, and use Eq. (C14).
[35] J. Kunes and P. M. Oppeneer, Phys. Rev. B **61**, 15774 (2000).