Ab initio calculation of the anomalous Hall conductivity by Wannier interpolation

Xinjie Wang,[1] Jonathan R. Yates,[2,3] Ivo Souza,[2,3] and David Vanderbilt[1]

[1]*Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854-8019, USA*
[2]*Department of Physics, University of California, Berkeley, California 94720, USA*
[3]*Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

The intrinsic anomalous Hall conductivity in ferromagnets depends on subtle spin-orbit-induced effects in the electronic structure, and recent *ab initio* studies found that it was necessary to sample the Brillouin zone at millions of k-points to converge the calculation. We present an efficient first-principles approach for computing this quantity. We start out by performing a conventional electronic-structure calculation including spin-orbit coupling on a uniform and relatively coarse k-point mesh. From the resulting Bloch states, maximally localized Wannier functions are constructed which reproduce the *ab initio* states up to the Fermi level. The Hamiltonian and position-operator matrix elements, needed to represent the energy bands and Berry curvatures, are then set up between the Wannier orbitals. This completes the first stage of the calculation, whereby the low-energy *ab initio* problem is transformed into an effective tight-binding form. The second stage only involves Fourier transforms and unitary transformations of the small matrices setup in the first stage. With these inexpensive operations, the quantities of interest are interpolated onto a dense k-point mesh and used to evaluate the anomalous Hall conductivity as a Brillouin zone integral. The present scheme, which also avoids the cumbersome summation over all unoccupied states in the Kubo formula, is applied to bcc Fe, giving excellent agreement with conventional, less efficient first-principles calculations. Remarkably, we find that about 99% of the effect can be recovered by keeping a set of terms depending only on the Hamiltonian matrix elements, not on matrix elements of the position operator.

PACS number(s): 71.15.Dx, 71.70.Ej, 71.18.+y, 75.50.Bb

I. INTRODUCTION

The Hall resistivity of a ferromagnet depends not only on the magnetic induction, but also on the magnetization; the latter dependence is known as the anomalous Hall effect (AHE).[1] The AHE is used for investigating surface magnetism, and its potential for investigating nanoscale magnetism, as well as for magnetic sensors and memory devices applications, is being considered.[2] Theoretical investigations of the AHE have undergone a revival in recent years, and have also lead to the proposal for a spin counterpart, the spin Hall effect, which has subsequently been realized experimentally.

The first theoretical model of the AHE was put forth by Karplus and Luttinger,[3] who showed that it can arise in a perfect crystal as a result of the spin-orbit interaction of polarized conduction electrons. Later, two alternative mechanisms, skew scattering[4] and side jump scattering,[5] were proposed by Smit and Berger, respectively. In skew scattering the spin-orbit interaction gives rise to an asymmetric scattering cross section even if the defect potential is symmetric, and in side-jump scattering it causes the scattered electron to acquire an extra transverse translation after the scattering event. These two mechanisms involve scattering from impurities or phonons, while the Karplus-Luttinger contribution is a scattering-free band-structure effect. The different contributions to the AHE are critically reviewed in Ref. 6. Perhaps because an intuitive physical picture was lacking, the Karplus-Luttinger theory was strongly disputed in the early literature. Attempts at estimating its magnitude on the basis of realistic band-structure calculations were also rare.[7]

In recent years, new insights into the Karplus-Luttinger contribution have been obtained by several authors,[8–12] who

reexamined it in the modern language of Berry's phases. The term $\mathbf{\Omega}_n(\mathbf{k})$ in the equations below was recognized as the Berry curvature of the Bloch states in reciprocal space, a quantity which had previously appeared in the theory of the integer quantum Hall effect,[13] and also closely related to the Berry-phase theory of polarization.[14] The dc anomalous Hall conductivity (AHC) is simply given as the Brillouin zone (BZ) integral of the Berry curvature weighted by the occupation factor of each state,

$$\sigma_{xy} = -\frac{e^2}{\hbar} \sum_n \int_{BZ} \frac{d\mathbf{k}}{(2\pi)^3} f_n(\mathbf{k}) \Omega_{n,z}(\mathbf{k}), \quad (1)$$

where $\sigma_{xy} = -\sigma_{yx}$ is the antisymmetric part of the conductivity. While this can be derived in several ways, it is perhaps most intuitively understood from the semiclassical point of view, in which the group velocity of an electron wave packet in band n is[9,15]

$$\dot{\mathbf{r}} = \frac{1}{\hbar} \frac{\partial \mathcal{E}_{n\mathbf{k}}}{\partial \mathbf{k}} - \dot{\mathbf{k}} \times \mathbf{\Omega}_n(\mathbf{k}). \quad (2)$$

The second term, often overlooked in elementary textbook derivations, is known as the "anomalous velocity." The expression for the current density then acquires a new term $ef_n(\mathbf{k})\dot{\mathbf{k}} \times \mathbf{\Omega}_n(\mathbf{k})$ which, with $\dot{\mathbf{k}} = -e\mathbf{E}/\hbar$, leads to Eq. (1).

Recently, first-principles calculations of Eq. (1) were carried out for the ferromagnetic perovskite SrRuO$_3$ by Fang *et al.*,[16] and for a transition metal, bcc Fe, by Yao *et al.*[17] In both cases the calculated values compared well with experimental data, lending credibility to the intrinsic mechanism. The most striking feature of these calculations is the strong

and rapid variation of the Berry curvature in k-space. In particular, there are sharp peaks and valleys at places where two energy bands are split by the spin-orbit coupling across the Fermi level. In order to converge the integral, the Berry curvature has to be evaluated over millions of k-points in the Brillouin zone. In the previous work this was done via a Kubo formula involving a large number of unoccupied states; the computational cost was very high, even for bcc Fe, with only one atom in the unit cell.

In this paper, we present an efficient method for computing the intrinsic AHC. Unlike the conventional approach, it does not require carrying out a full *ab initio* calculation for every k-point where the Berry curvature needs to be evaluated. The actual *ab initio* calculation is performed on a much coarser k-point grid. By a postprocessing step, the resulting Bloch states below and immediately above the Fermi level are then mapped onto well-localized Wannier functions. In this representation it is then possible to interpolate the Berry curvature onto any desired k-point with very little computational effort and essentially no loss of accuracy.

The paper is organized as follows. In Sec. II we introduce the basic definitions and describe the Kubo-formula approach used in previous calculations of the intrinsic AHC. In Sec. III our Wannier-based approach is described. The details of the band-structure calculation and Wannier-function construction are described in Sec. IV, followed by an application of the method to bcc Fe in Sec V. Finally, Sec. VI contains a brief summary and discussion.

II. DEFINITIONS AND BACKGROUND

The key ingredient in the theory of the intrinsic anomalous Hall effect is the Berry curvature $\mathbf{\Omega}_n(\mathbf{k})$, defined as

$$\mathbf{\Omega}_n(\mathbf{k}) = \nabla \times \mathbf{A}_n(\mathbf{k}), \tag{3}$$

where $\mathbf{A}_n$ is the Berry connection,

$$\mathbf{A}_n(\mathbf{k}) = i\langle u_{n\mathbf{k}}|\nabla_\mathbf{k}|u_{n\mathbf{k}}\rangle. \tag{4}$$

The integral of the Berry curvature over a surface bounded by a closed path in k-space is the Berry phase of that path.[18] In what follows it will be useful to write the Berry curvature as a second-rank antisymmetric tensor:

$$\Omega_{n,\gamma}(\mathbf{k}) = \epsilon_{\alpha\beta\gamma}\Omega_{n,\alpha\beta}(\mathbf{k}), \tag{5}$$

$$\Omega_{n,\alpha\beta}(\mathbf{k}) = -2\,\mathrm{Im}\left\langle\frac{\partial u_{n\mathbf{k}}}{\partial k_\alpha}\bigg|\frac{\partial u_{n\mathbf{k}}}{\partial k_\beta}\right\rangle, \tag{6}$$

where the Greek letters indicate Cartesian coordinates, $\epsilon_{\alpha\beta\gamma}$ is the Levi-Civita tensor, and $u_{n\mathbf{k}}$ are the cell-periodic Bloch functions.

With this notation we rewrite the quantity we wish to evaluate, Eq. (1), as

$$\sigma_{\alpha\beta} = -\frac{e^2}{\hbar}\int_{BZ}\frac{d\mathbf{k}}{(2\pi)^3}\Omega_{\alpha\beta}(\mathbf{k}), \tag{7}$$

where we have introduced the *total* Berry curvature

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_n f_n(\mathbf{k})\Omega_{n,\alpha\beta}(\mathbf{k}). \tag{8}$$

Direct evaluation of Eq. (6) poses a number of practical difficulties related to the presence of k-derivatives of Bloch states, as will be discussed in the next section. In previous work[16,17] these were circumvented by recasting Eq. (6) as a Kubo formula,[7,13] where the k-derivatives are replaced by sums over states:

$$\Omega_{n,\alpha\beta}(\mathbf{k}) = -2\,\mathrm{Im}\sum_{m\neq n}\frac{v_{nm,\alpha}(\mathbf{k})v_{mn,\beta}(\mathbf{k})}{[\omega_m(\mathbf{k})-\omega_n(\mathbf{k})]^2}, \tag{9}$$

where $\omega_n(\mathbf{k})=\mathcal{E}_{n\mathbf{k}}/\hbar$ and the matrix elements of the Cartesian velocity operators $\hat{v}_\alpha=(i/\hbar)[\hat{H},\hat{r}_\alpha]$ are given by[19]

$$v_{nm,\alpha}(\mathbf{k}) = \langle\psi_{n\mathbf{k}}|\hat{v}_\alpha|\psi_{m\mathbf{k}}\rangle = \frac{1}{\hbar}\left\langle u_{n\mathbf{k}}\bigg|\frac{\partial\hat{H}(\mathbf{k})}{\partial k_\alpha}\bigg|u_{m\mathbf{k}}\right\rangle, \tag{10}$$

where $\hat{H}(\mathbf{k})=e^{-i\mathbf{k}\cdot\hat{\mathbf{r}}}\hat{H}e^{i\mathbf{k}\cdot\hat{\mathbf{r}}}$. The merit of Eq. (9) lies in its practical implementation on a finite k-grid using only the wave functions at a single k-point. As is usually the case for such linear-response formulas, sums over pairs of occupied states can be avoided in the $T=0$ version of Eqs. (8) and (9) for the total Berry curvature,

$$\Omega_{\alpha\beta}(\mathbf{k}) = -2\,\mathrm{Im}\sum_v\sum_c\frac{v_{vc,\alpha}(\mathbf{k})v_{cv,\beta}(\mathbf{k})}{[\omega_c(\mathbf{k})-\omega_v(\mathbf{k})]^2}, \tag{11}$$

where v and c subscripts denote valence (occupied) and conduction (unoccupied) bands, respectively. However, the evaluation of this formula requires the cumbersome summation over unoccupied states. Even if practical calculations truncate the summation to some extent, the computation could be time-consuming. Moreover, the time required to calculate the matrix elements of the velocity operator in Eq. (9) or Eq. (11) is not negligible.

III. EVALUATION OF THE BERRY CURVATURE BY WANNIER INTERPOLATION

In view of the above-mentioned drawbacks of the Kubo formula for practical calculations, it would be highly desirable to have a numerical scheme based on the "geometric formula" (6), in terms of the occupied states only. The difficulties in implementing that formula arise from the k-derivatives therein. Since in practice one always replaces the Brillouin zone integration by a discrete summation, an obvious approach would be to use a finite-difference representation of the derivatives on the k-point grid. However, this requires some care: a straightforward discretization will yield results which depend on the choice of phases of the Bloch states, even though Eq. (6) is in principle invariant under such "diagonal gauge transformations." The problem becomes more acute in the presence of band crossings and avoided crossings, because then it is not clear which two states at neighboring grid points should be taken as "partners" in a finite-differences expression. (Moreover, since the system is a metal, at $T=0$ the occupation can be different at neighboring k-points.) Successful numerical strategies for
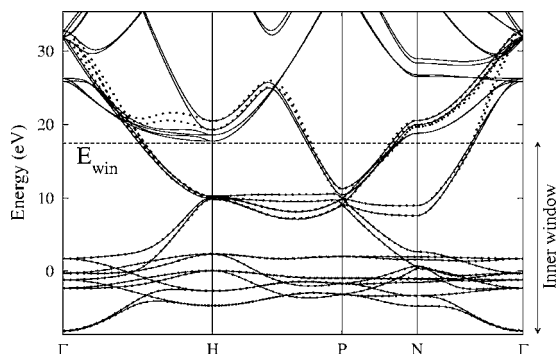
FIG. 1. Band structure of bcc Fe with spin-orbit coupling included. Solid lines: original band structure from a conventional first-principles calculation. Dotted lines: Wannier-interpolated band structure. The zero of energy is the Fermi level.

dealing with problems of this nature have been developed in the context of the Berry-phase theory of polarization of insulators, and a workable finite-difference scheme which combines those ideas with Wannier interpolation is sketched in Appendix B.

We present here a different, more powerful strategy that also relies on a Wannier representation of the low-energy electronic structure. We will show that it is possible to express the needed derivatives analytically in terms of the Wannier functions, so that no finite-difference evaluation of a derivative is needed in principle. The use of Wannier functions allows us to achieve this while still avoiding the summation over all empty states which appears in the Kubo formula as a result of applying conventional $\mathbf{k} \cdot \mathbf{p}$ perturbation theory.

A. Wannier representation

We begin by using the approach of Souza, Marzari, and Vanderbilt[20] to construct a set of Wannier functions (WFs) for the metallic system of interest. For insulators, one normally considers a set of WFs that span precisely the space of occupied Bloch states. Here, since we have a metallic system and we want to have well-localized WFs, we choose a number of WFs larger than the number $N_\mathbf{k}$ of occupied states at any $\mathbf{k}$, and only insist that the space spanned by the WFs should include, as a subset, the space of the occupied states, plus the first few empty states. Thus these partially occupied WFs will serve here as a kind of "exact tight-binding basis" that can be used as a compact representation of the low-energy electronic structure of the metal.

This is illustrated in Fig. 1, where the band structure of bcc Fe is shown. The details of the calculations will be presented later in Sec. IV. The solid lines show the full *ab initio* band structure, while the dashed lines show the bands obtained within the Wannier representation using $M = 18$ WFs per cell (nine of each spin; see Sec. IV B). In the method of Ref. 20, one specifies an energy E_{win} lying somewhat above the Fermi energy E_f, and insists on finding a set of WFs spanning all the *ab initio* states in an energy window up to E_{win}. In the calculation of Fig. 1 we chose $E_{\text{win}} \simeq 18$ eV, and it is evident that there is an essentially perfect match between the fully *ab initio* and the Wannier-represented bands up to, but not above, E_{win}. Clearly, a Wannier-based calculation of any property of the occupied manifold, such as the intrinsic AHC, should be in excellent agreement with a direct *ab initio* evaluation, provided that E_{win} is set above E_f.

We shall assume that we have M WFs per unit cell denoted as $|\mathbf{R}n\rangle$, where $n = 1, \dots, M$ and $\mathbf{R}$ labels the unit cell. We shall also assume that the Bloch-like functions given by the phased sum of WFs

$$|u_{n\mathbf{k}}^{(\text{W})}\rangle = \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot(\hat{\mathbf{r}}-\mathbf{R})}|\mathbf{R}n\rangle \tag{12}$$

span the actual Bloch eigenstates $|u_{n\mathbf{k}}\rangle$ of interest ($n = 1, \dots, N_\mathbf{k}$) at each $\mathbf{k}$ (clearly M must be $\geq N_\mathbf{k}$ everywhere in the BZ). It follows that, if we construct the $M \times M$ Hamiltonian matrix

$$H_{nm}^{(\text{W})}(\mathbf{k}) = \langle u_{n\mathbf{k}}^{(\text{W})}|\hat{H}(\mathbf{k})|u_{m\mathbf{k}}^{(\text{W})}\rangle \tag{13}$$

and diagonalize it by finding an $M \times M$ unitary rotation matrix $U(\mathbf{k})$ such that

$$U^\dagger(\mathbf{k})H^{(\text{W})}(\mathbf{k})U(\mathbf{k}) = H^{(\text{H})}(\mathbf{k}), \tag{14}$$

where $H_{nm}^{(\text{H})}(\mathbf{k}) = \mathcal{E}_{n\mathbf{k}}^{(\text{H})}\delta_{nm}$, then $\mathcal{E}_{n\mathbf{k}}^{(\text{H})}$ will be identical to the true $\mathcal{E}_{n\mathbf{k}}$ for all occupied bands. The corresponding Bloch states,

$$|u_{n\mathbf{k}}^{(\text{H})}\rangle = \sum_m |u_{m\mathbf{k}}^{(\text{W})}\rangle U_{mn}(\mathbf{k}), \tag{15}$$

will also be identical to the true eigenstates $|u_{n\mathbf{k}}\rangle$ for $\mathcal{E} \leq E_\text{f}$. (In the scheme of Ref. 20, these properties will actually hold for energies up to E_{win}.) However, the band energies and Bloch states will *not* generally match the true ones at the energies higher than E_{win}, as shown in Fig. 1. We thus use the superscript "H" to distinguish the projected band energies $\mathcal{E}_{n\mathbf{k}}^{(\text{H})}$ and eigenvectors $|u_{n\mathbf{k}}^{(\text{H})}\rangle$ from the true ones $\mathcal{E}_{n\mathbf{k}}$ and $|u_{n\mathbf{k}}\rangle$, keeping in mind that this distinction is only significant in the higher-energy unoccupied region ($\mathcal{E} > E_{\text{win}}$) of the projected band structure.

The unitary rotation of states expressed by the matrix $U(\mathbf{k})$ is often referred to as a "gauge transformation," and we shall adopt this terminology here. We shall refer to the Wannier-derived Bloch-like states $|u_{n\mathbf{k}}^{(\text{W})}\rangle$ as belonging to the Wannier (W) gauge, while the eigenstates $|u_{n\mathbf{k}}^{(\text{H})}\rangle$ of the projected band structure are said to belong to the Hamiltonian (H) gauge.

Quantities such as the Berry connection $\mathbf{A}_n(\mathbf{k})$ of Eq. (4) and the Berry curvature $\Omega_{n,\alpha\beta}(\mathbf{k})$ of Eq. (6) clearly depend upon the gauge in which they are expressed. [The curvature is actually invariant under the subset of gauge transformations of the diagonal form $U_{nm}(\mathbf{k}) = e^{i\phi_{n\mathbf{k}}}\delta_{nm}$, which is also the remaining gauge freedom within the Hamiltonian gauge.] The quantity that we wish to calculate, Eq. (8), is most naturally expressed in the Hamiltonian gauge, where it takes the form

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_{n=1}^{M} f_n(\mathbf{k})\Omega_{n,\alpha\beta}^{(H)}(\mathbf{k}). \qquad (16)$$

Here $\Omega_{n,\alpha\beta}^{(H)}(\mathbf{k})$ is given by Eq. (6) with $|u_{n\mathbf{k}}\rangle \rightarrow |u_{n\mathbf{k}}^{(H)}\rangle$. It is permissible to make this substitution because the projected band structure matches the true one for all occupied states. In practice one may take for the occupation factor $f_n(\mathbf{k}) = \theta(E_f - \mathcal{E}_{n\mathbf{k}})$ (as done in the present work), or introduce a small thermal smearing.

Our strategy now is to see how the right-hand side of Eq. (16) can be obtained by starting with quantities that are defined and computed first in the Wannier gauge and then transformed into the Hamiltonian gauge. The resulting scheme can be viewed as a generalized Slater-Koster interpolation, which takes advantage of the smoothness in k-space of the Wannier-gauge objects, a direct consequence of the short range of the Wannier orbitals in real space.

B. Gauge transformations

Because the gauge transformation of Eq. (15) involves a unitary rotation among several bands, it is useful to introduce generalizations of the quantities in Eqs. (4) and (6) having two band indices instead of one. Thus we define

$$A_{nm,\alpha}(\mathbf{k}) = i\langle u_n | \partial_\alpha u_m \rangle \qquad (17)$$

and

$$\Omega_{nm,\alpha\beta}(\mathbf{k}) = \partial_\alpha A_{nm,\beta} - \partial_\beta A_{nm,\alpha} = i\langle \partial_\alpha u_n | \partial_\beta u_m \rangle - i\langle \partial_\beta u_n | \partial_\alpha u_m \rangle, \qquad (18)$$

where every object in each of these equations should consistently carry either a (W) or (H) label. (We have now suppressed the $\mathbf{k}$ subscripts and introduced the notation $\partial_\alpha = \partial/\partial k_\alpha$ for conciseness.) In this notation, Eq. (16) becomes

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_{n=1}^{M} f_n(\mathbf{k})\Omega_{nn,\alpha\beta}^{(H)}(\mathbf{k}). \qquad (19)$$

Note that when $\Omega_{\alpha\beta}$ appears without a (W) or (H) superscript, as on the left-hand side of this equation, it denotes the total Berry curvature on the left-hand side of Eq. (16).

The matrix representation of an ordinary operator such as the Hamiltonian or the velocity can be transformed from the Wannier to the Hamiltonian gauge, or vice versa, just by operating on the left and right by $U^\dagger(\mathbf{k})$ and $U(\mathbf{k})$, as in Eq. (14); such a matrix is called "gauge-covariant." Unfortunately, the matrix objects in Eqs. (17) and (18) are not gauge-covariant because they involve k-derivatives acting on the Bloch states. For example, a straightforward calculation shows that

$$A_\alpha^{(H)} = U^\dagger A_\alpha^{(W)} U + iU^\dagger \partial_\alpha U, \qquad (20)$$

where each object is an $M \times M$ matrix and matrix products are implied throughout. For every matrix object O, we define

$$\bar{O}^{(H)} = U^\dagger O^{(W)} U \qquad (21)$$

so that, by definition, $\bar{O}^{(H)} = O^{(H)}$ only for gauge-covariant objects.

The derivative $\partial_\alpha U$ may be obtained from ordinary perturbation theory. We adopt a notation in which $\|\phi_m\rangle\rangle$ is the mth M-component column vector of matrix U, so that $\langle\langle\phi_n\|H^{(W)}\|\phi_m\rangle\rangle = \mathcal{E}_n \delta_{nm}$; the stylized bra-ket notation is used to emphasize that objects like $H^{(W)}$ and $\|\phi_n\rangle\rangle$ are $M \times M$ matrices and M-component vectors, i.e., operators and state vectors in the "tight-binding space" defined by the WFs, not in the original Hilbert space. Perturbation theory with respect to the parameter $\mathbf{k}$ takes the form

$$\|\partial_\alpha \phi_n\rangle\rangle = \sum_{l \neq n} \frac{\langle\langle\phi_l\|H_\alpha^{(W)}\|\phi_n\rangle\rangle}{\mathcal{E}_n^{(H)} - \mathcal{E}_l^{(H)}} \|\phi_l\rangle\rangle, \qquad (22)$$

where $H_\alpha^{(W)} \equiv \partial_\alpha H^{(W)}$. In matrix notation this can be written

$$\partial_\alpha U_{mn} = \sum_l U_{ml} D_{ln,\alpha}^{(H)} = (U D_\alpha^{(H)})_{mn}, \qquad (23)$$

where

$$D_{nm,\alpha}^{(H)} \equiv (U^\dagger \partial_\alpha U)_{nm} = \begin{cases} \dfrac{\bar{H}_{nm,\alpha}^{(H)}}{\mathcal{E}_m^{(H)} - \mathcal{E}_n^{(H)}} & \text{if } n \neq m \\ 0 & \text{if } n = m \end{cases} \qquad (24)$$

and $\bar{H}_{nm,\alpha}^{(H)} = (U^\dagger H_\alpha^{(W)} U)_{nm}$ according to Eq. (21). Note that while $\Omega_{\alpha\beta}$ and A_α are Hermitian in the band indices, $D_\alpha^{(H)}$ is instead anti-Hermitian. The gauge choice implicit in Eqs. (22) and (24) is $\langle\langle\phi_n\|\partial_\alpha\phi_n\rangle\rangle = (U^\dagger \partial_\alpha U)_{nn} = 0$ (this is the so-called "parallel transport" gauge).

Using Eq. (23), Eq. (20) becomes

$$A_\alpha^{(H)} = \bar{A}_\alpha^{(H)} + iD_\alpha^{(H)} \qquad (25)$$

and the derivative of Eq. (15) becomes

$$|\partial_\alpha u_n^{(H)}\rangle = \sum_m |\partial_\alpha u_m^{(W)}\rangle U_{mn} + \sum_m |u_m^{(H)}\rangle D_{mn,\alpha}^{(H)}. \qquad (26)$$

Plugging the latter into Eq. (18), we finally obtain, after a few manipulations, the matrix equation

$$\Omega_{\alpha\beta}^{(H)} = \bar{\Omega}_{\alpha\beta}^{(H)} - [D_\alpha^{(H)}, \bar{A}_\beta^{(H)}] + [D_\beta^{(H)}, \bar{A}_\alpha^{(H)}] - i[D_\alpha^{(H)}, D_\beta^{(H)}]. \qquad (27)$$

The band-diagonal elements $\Omega_{nn,\alpha\beta}^{(H)}(\mathbf{k})$ then need to be inserted into Eq. (19).

Equation (27) can also be derived from Eq. (25), by combining it with the first line of Eq. (18):

$$\Omega_{\alpha\beta}^{(H)} = \partial_\alpha \bar{A}_\beta^{(H)} - \partial_\beta \bar{A}_\alpha^{(H)} - i[D_\alpha^{(H)}, D_\beta^{(H)}], \qquad (28)$$

where we have used $i(\partial_\alpha U)^\dagger \partial_\beta U = -iD_\alpha^{(H)} D_\beta^{(H)}$. Invoking Eq. (21) we find

$$\partial_\alpha \bar{A}_\beta^{(H)} - \partial_\beta \bar{A}_\alpha^{(H)} = -[D_\alpha^{(H)}, \bar{A}_\beta^{(H)}]$$
$$+ [D_\beta^{(H)}, \bar{A}_\alpha^{(H)}] + U^\dagger(\partial_\alpha A_\beta^{(W)} - \partial_\beta A_\alpha^{(W)})U. \qquad (29)$$

The last term on the right-hand side is $\bar{\Omega}_{\alpha\beta}^{(H)}$, and thus we recover Eq. (27).

C. Discussion

We expect, based on Eq. (9), that the largest contributions to the AHC will come from regions of k-space where there are small energy splittings between bands (for example, near spin-orbit-split avoided crossings).[16] In the present formulation, this will give rise to small energy denominators in Eq. (24), leading to very large $D_\alpha^{(\mathrm{H})}$ values in those regions. These large and spiky contributions will then propagate into $A_\alpha^{(\mathrm{H})}$ and $\Omega_{\alpha\beta}^{(\mathrm{H})}$, whereas $A_\alpha^{(\mathrm{W})}$ and $\Omega_{\alpha\beta}^{(\mathrm{W})}$, and also $\bar{A}_\alpha^{(\mathrm{H})}$ and $\bar{\Omega}_{\alpha\beta}^{(\mathrm{H})}$, will remain with their typically smaller values. Thus, these spiky contributions will be present in the second and third terms, and especially in the fourth term, of Eq. (27). The contributions of these various terms are illustrated for the case of bcc Fe in Sec. V A, and we show there that the last term typically makes by far the dominant contribution, followed by the second and third terms, and then by the first term.

The dominant fourth term can be recast in the form of a Kubo formula as

$$\Omega_{n,\alpha\beta}^{DD} = -2\,\mathrm{Im}\sum_{m\neq n}\frac{\langle\langle\phi_n\|H_\alpha^{(\mathrm{W})}\|\phi_m\rangle\rangle\langle\langle\phi_m\|H_\beta^{(\mathrm{W})}\|\phi_n\rangle\rangle}{(\mathcal{E}_m^{(\mathrm{H})}-\mathcal{E}_n^{(\mathrm{H})})^2}.$$
(30)

The following differences between this equation and the true Kubo formula, Eq. (9), should, however, be kept in mind. First, the summation in Eq. (30) is restricted to the M-band projected band structure. Second, above E_{win} the projected band structure deviates from the original *ab initio* one. Third, even below E_{win}, where they do match exactly, the "effective tight-binding velocity matrix elements" appearing in Eq. (30) differ from the true ones, given by Eq. (10). [The relation between them is particularly simple for energies below E_{win},

$$v_{nm,\alpha}^{(\mathrm{H})} = \frac{1}{\hbar}\bar{H}_{nm,\alpha}^{(\mathrm{H})} - \frac{i}{\hbar}(\mathcal{E}_m^{(\mathrm{H})}-\mathcal{E}_n^{(\mathrm{H})})\bar{A}_{nm,\alpha}^{(\mathrm{H})},$$
(31)

and follows from combining the identity[19] $A_{nm,\alpha} = i\langle\psi_n|\hat{v}_\alpha|\psi_m\rangle/(\omega_m-\omega_n)$, valid for $m\neq n$, with Eqs. (24) and (25)]. All these differences are, however, exactly compensated by the previous three terms in Eq. (27). We emphasize that all terms in that equation are defined strictly within the projected space spanned by the Wannier functions.

We note in passing that it is possible to rewrite Eq. (27) in such a way that the large spiky contributions are isolated into a single term. This alternative formulation, which turns out to be related to a gauge-covariant curvature tensor, will be described in Appendix A.

D. Sum over occupied bands

In the above, we have proposed to evaluate $\Omega_{nn,\alpha\beta}^{(\mathrm{H})}$ from Eq. (27) and then insert it into the band sum, Eq. (19), in order to compute the AHC. However, this approach has the shortcoming that small splittings (avoided crossings) between a pair of *occupied* bands n and m lead to large values of $D_{nm,\alpha}^{(\mathrm{H})}$, and thus to large but canceling contributions to the AHC coming from $\Omega_{nn,\alpha\beta}^{(\mathrm{H})}$ and $\Omega_{mm,\alpha\beta}^{(\mathrm{H})}$. Here, we rewrite the total Berry curvature (19) in such a way that the cancellation is explicit.

Inserting Eq. (27) into Eq. (19) and interchanging dummy labels $n\leftrightarrow m$ in certain terms, we obtain

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_n f_n\bar{\Omega}_{nn,\alpha\beta}^{(\mathrm{H})} + \sum_{nm}(f_m-f_n)(D_{nm,\alpha}^{(\mathrm{H})}\bar{A}_{mn,\beta}^{(\mathrm{H})}$$
$$- D_{nm,\beta}^{(\mathrm{H})}\bar{A}_{mn,\alpha}^{(\mathrm{H})} + iD_{nm,\alpha}^{(\mathrm{H})}D_{mn,\beta}^{(\mathrm{H})}).$$
(32)

The factors of (f_m-f_n) insure that terms arising from pairs of fully occupied states give no contribution. Thus the result of this reformulation is that individual terms in Eq. (32) have large spiky contributions only when avoided crossings or near-degeneracies occur across the Fermi energy. This approach is therefore preferable from the point of view of numerical stability, and it is the one that we have implemented in the current work.

As expected from the discussion in Sec. III C and shown later in Sec. V B, the dominant term in Eq. (32) is the last one,

$$\Omega_{\alpha\beta}^{DD} = i\sum_{nm}(f_m-f_n)D_{nm,\alpha}^{(\mathrm{H})}D_{mn,\beta}^{(\mathrm{H})}$$
(33)

or, in a more explicitly Kubo-like form,

$$\Omega_{\alpha\beta}^{DD} = i\sum_{nm}(f_m-f_n)\frac{\bar{H}_{nm,\alpha}^{(\mathrm{H})}\bar{H}_{mn,\beta}^{(\mathrm{H})}}{(\mathcal{E}_m^{(\mathrm{H})}-\mathcal{E}_n^{(\mathrm{H})})^2}.$$
(34)

In the zero-temperature limit, the latter can easily be cast into a form like Eq. (30), but with a double sum running over occupied bands n and unoccupied bands m, very reminiscent of the original Kubo formula in Eq. (11).

We remark that $(1/\hbar)\bar{H}_{nm,\alpha}^{(\mathrm{H})}$ coincides with the "effective tight-binding velocity operator" of Ref. 21. This is an approximate tight-binding velocity operator. Comparison with Eqs. (31) and (39) below shows that it is lacking the contributions which involve matrix elements of the position operator between the WFs.[22] We now recognize in Eq. (22) the standard result from $\mathbf{k}\cdot\mathbf{p}$ perturbation theory, but in terms of the approximate momentum operator. Using that equation, Eq. (30) can be cast as the tight-binding-space analog of Eq. (6),

$$\Omega_{n,\alpha\beta}^{DD} = -2\,\mathrm{Im}\langle\langle\partial_\alpha\phi_{n\mathbf{k}}\|\partial_\beta\phi_{n\mathbf{k}}\rangle\rangle.$$
(35)

This allows us to rewrite Eq. (34) in a form that closely resembles the total Berry curvature, Eq. (16):

$$\Omega_{\alpha\beta}^{DD} = \sum_{n=1}^M f_n\Omega_{n,\alpha\beta}^{DD}.$$
(36)

E. Evaluation of the Wannier-gauge matrices

Equation (32) is our primary result. To review, recall that this is a condensed notation expressing the $M\times M$ matrix $\Omega_{nm,\alpha\beta}^{(\mathrm{H})}(\mathbf{k})$ in terms of the matrices $\bar{\Omega}_{nm,\alpha\beta}^{(\mathrm{H})}(\mathbf{k})$, etc. The basic ingredients needed are the four matrices $H^{(\mathrm{W})}$, $H_\alpha^{(\mathrm{W})}$, $A_\alpha^{(\mathrm{W})}$, and $\Omega_{\alpha\beta}^{(\mathrm{W})}$ at a given $\mathbf{k}$. Diagonalization of the first of them yields the energy eigenvalues needed to find the occupation factors f_n. It also provides the gauge transformation U which

is then used to construct $\bar{H}_\alpha^{(\mathrm{H})}$, $\bar{A}_\alpha^{(\mathrm{H})}$, and $\bar{\Omega}_{\alpha\beta}^{(\mathrm{H})}$ from the other three objects via Eq. (21). Finally, $\bar{H}_\alpha^{(\mathrm{H})}$ is inserted into Eq. (24) to obtain $D_\alpha^{(\mathrm{H})}$, and all terms in Eq. (32) are evaluated.

In this section we explain how to obtain the matrices $H^{(\mathrm{W})}(\mathbf{k})$, $H_\alpha^{(\mathrm{W})}(\mathbf{k})$, $A_\alpha^{(\mathrm{W})}(\mathbf{k})$, and $\Omega_{\alpha\beta}^{(\mathrm{W})}(\mathbf{k})$ at an arbitrary point $\mathbf{k}$ for use in the subsequent calculations described above.

1. Fourier transform expressions

The four needed quantities can be expressed as follows:

$$H_{nm}^{(\mathrm{W})}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle, \qquad (37)$$

$$H_{nm,\alpha}^{(\mathrm{W})}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} iR_\alpha \langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle, \qquad (38)$$

$$A_{nm,\alpha}^{(\mathrm{W})}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle, \qquad (39)$$

$$\Omega_{nm,\alpha\beta}^{(\mathrm{W})}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} (iR_\alpha \langle \mathbf{0}n|\hat{r}_\beta|\mathbf{R}m\rangle - iR_\beta \langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle). \qquad (40)$$

(The notation $|\mathbf{0}n\rangle$ refers to the nth WF in the home unit cell $\mathbf{R}=\mathbf{0}$.) Equation (37) follows by combining Eqs. (12) and (13), while Eq. (39) follows by combining Eqs. (12) and (17). Equations (38) and (40) are then obtained from Eqs. (37) and (39) using $H_{nm,\alpha} = \partial_\alpha H_{nm}$ and Eq. (18), respectively.

It is remarkable that the only real-space matrix elements that are required between WFs are those of the four operators $\hat{H}$ and $\hat{r}_\alpha$ ($\alpha = x$, y, and z). Because the WFs are strongly localized, these matrix elements are expected to decay rapidly as a function of lattice vector $\mathbf{R}$, so that only a modest number of them need to be computed and stored once and for all. Collectively, they define our "exact tight-binding model" and suffice to allow subsequent calculation of all needed quantities. Furthermore, the short range of these matrix elements in real space insures that the Wannier-gauge quantities on the left-hand sides of Eqs. (37)–(40) will be smooth functions of $\mathbf{k}$, thus justifying the earlier discussion in which it was argued that these objects should have no rapid variation or enhancement in k-space regions where avoided crossings occur. (Recall that such large, rapidly varying contributions only appear in the $D^{(\mathrm{H})}$ matrices and in quantities that depend upon them.) It should, however, be kept in mind that Eq. (32) is not written directly in terms of the smooth quantities (37)–(40), but rather in terms of those quantities transformed according to Eq. (21). The resulting objects are not smooth, since the matrices U change rapidly with $\mathbf{k}$. However, even while not smooth, they remain small.

2. Evaluation of real-space matrix elements

We conclude this section by discussing the calculation of the fundamental matrix elements $\langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle$ and $\langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle$. There are several ways in which these could be computed, and the choice could well vary from one implementation to another. One possibility would be to construct the WFs in

real space, say on a real-space grid, and then to compute the Hamiltonian and position-operator matrix elements directly on that grid. In the context of a code that uses a real-space basis (e.g., localized orbitals or grids), this might be the best choice. However, in the context of plane-wave methods it is usually more convenient to work in reciprocal space if possible. This is in the spirit of the Wannier-function construction scheme,[20,23] which is formulated as a postprocessing step after a conventional *ab initio* calculation carried out on a uniform k-point grid. (In the following we will use the symbol $\mathbf{q}$ to denote the points of this *ab initio* mesh, to distinguish them from arbitrary or interpolation-grid points denoted by $\mathbf{k}$.)

The end result of the Wannier-construction step are M Bloch-like functions $|u_{n\mathbf{q}}^{(\mathrm{W})}\rangle$ at each $\mathbf{q}$. The WFs are obtained from them via a discrete Fourier transform:

$$|\mathbf{R}n\rangle = \frac{1}{N_q^3} \sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot(\mathbf{R}-\hat{\mathbf{r}})} |u_{n\mathbf{q}}^{(\mathrm{W})}\rangle. \qquad (41)$$

This expression follows from inverting Eq. (12). If the *ab initio* mesh contains $N_q \times N_q \times N_q$ points, the resulting WFs are really periodic functions over a supercell of dimensions $L \times L \times L$, where $L = N_q a$ and a is the lattice constant of the unit cell. The idea then is to choose L large enough that the rapid decay of the localized WFs occurs on a scale much smaller than L. This ensures that the matrix elements $\langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle$ and $\langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle$ between a pair of WFs separated by more than $L/2$ are negligible, so that further refinement of the *ab initio* mesh will have a negligible impact on the accuracy of Wannier-interpolated quantities. (In particular, the interpolated band structure, Fig. 1, is able to reproduce tiny features of the full band structure, such as spin-orbit-induced avoided crossings, even if they occur on a length scale much smaller than the *ab initio* mesh spacing.) While the choice of a reciprocal-space cell spanned by the vectors $\mathbf{q}$ is immaterial, because of the periodicity of reciprocal space, this is not so for the vectors $\mathbf{R}$. In practice we choose the $N_q \times N_q \times N_q$ vectors $\mathbf{R}$ to be evenly distributed on the Wigner-Seitz supercell of volume $N_q^3 a^3$ centered around $\mathbf{R}=\mathbf{0}$.[20] This is the most isotropic choice possible, ensuring that the strong decay of the matrix elements for $|\mathbf{R}| \sim L/2$ is achieved irrespective of direction.

The matrix elements of the Hamiltonian are obtained from Eq. (41) as

$$\langle \mathbf{0}n|\hat{H}|\mathbf{R}m\rangle = \frac{1}{N_q^3} \sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot\mathbf{R}} H_{nm}^{(\mathrm{W})}(\mathbf{q}), \qquad (42)$$

which is the reciprocal of Eq. (37), with the sum running over the coarse *ab initio* mesh points. The position matrix is obtained similarly by inverting Eq. (39):

$$\langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m\rangle = \frac{1}{N_q^3} \sum_{\mathbf{q}} e^{-i\mathbf{q}\cdot\mathbf{R}} A_{nm,\alpha}^{(\mathrm{W})}(\mathbf{q}). \qquad (43)$$

The matrix $A_{nm,\alpha}^{(\mathrm{W})}(\mathbf{q})$ is then evaluated by approximating the k-derivatives in Eq. (17) by finite-differences on the *ab initio* mesh using the expression[23]

$$A_{nm,\alpha}^{(W)}(\mathbf{q}) \simeq i\sum_{\mathbf{b}} w_b b_\alpha(\langle u_{n\mathbf{q}}^{(W)}|u_{m,\mathbf{q+b}}^{(W)}\rangle - \delta_{nm}), \qquad (44)$$

where $\mathbf{b}$ are the vectors connecting $\mathbf{q}$ to its nearest neighbors on the *ab initio* mesh. This approximation is valid because in the Wannier gauge the Bloch states vary smoothly with $\mathbf{k}$. We note that the overlap matrices appearing on the right-hand side are available "for free" as they have already been computed and stored during the WF construction procedure. This is also the case for the matrices $H^{(W)}(\mathbf{q})$ needed in Eq. (42).

It should be kept in mind that the k-space finite-difference procedure outlined above entails an error of order $O(\Delta q^2)$ in the values of the position operator matrix elements, where Δq is the *ab initio* mesh spacing. The importance of such an error is easily assessed by trying denser q-point meshes; in our case, we find that it is not a numerically significant source of error for the $8\times8\times8$ mesh that we employ in our calculations. [In large measure this is simply because less than 2% of the total AHC comes from terms that depend on these position-operator matrix elements, as will be discussed in Sec. V. Indeed, we find that the $O(\Delta q^2)$ convergence of this small contribution hardly shows in the convergence of the total AHC, which empirically appears to be approximately exponential in the *ab initio* mesh density.] However, if the $O(\Delta q^2)$ convergence is a source of concern, one could adopt the direct real-space mesh integration method mentioned at the beginning of this section, which should be free of such errors.

IV. COMPUTATIONAL DETAILS

In this section we present some of the detailed steps of the calculations as they apply to our test system of bcc Fe. First, we describe the first-principles band-structure calculations that are carried out initially. Second, we discuss the procedure for constructing maximally localized Wannier functions for the bands of interest following the method of Souza, Marzari, and Vanderbilt.[20] Third, we discuss the variable treatment of the spin-orbit interaction within these first-principles calculations, which is useful for testing the dependence of the AHC on the spin-orbit coupling strength.

A. Band-structure calculation

Fully relativistic band-structure calculations for bcc Fe in its ferromagnetic ground state at the experimental lattice constant $a=5.42$ Bohr are carried out using the PWSCF code.[24] A kinetic-energy cutoff of 60 Hartree is used for the plane-wave expansion of the valence wave functions (400 Hartree for the charge densities). Exchange and correlation effects are treated with the Perdew, Burke, and Ernzerhof generalized-gradient approximation.[25]

The core-valence interaction is described here by means of norm-conserving pseudopotentials which include spin-orbit effects[26,27] in separable Kleinman-Bylander form. (Our overall Wannier interpolation approach is quite independent of this specific choice and can easily be generalized to other kinds of pseudopotentials or to all-electron methods.) The pseudopotential was constructed using a reference valence configuration of $3d^7 4s^{0.75} 4p^{0.25}$. We treat the overlap of the valence states with the semicore $3p$ states using the nonlinear core correction approach.[28] The pseudopotential core radii for the $3d$, $4s$, and $4p$ states are 1.3, 2.0, and 2.2 Bohr, respectively. We find the small cutoff radius for the $3d$ channel to be necessary in order to reproduce the all-electron band structure accurately.

We obtain the self-consistent ground state using a $16 \times 16 \times 16$ Monkhorst-Pack[29] mesh of k-points and a fictitious Fermi smearing[30] of 0.02 Ry for the Brillouin-zone integration. The magnetization is along the [001] direction, so that the only nonzero component of the integrated Berry curvature, Eq. (7), is the one along z. The spin magnetic moment is found to be 2.22 μ_B, the same as that from an all-electron calculation[17] and close to the experimental value of 2.12 μ_B.

In order to calculate the Wannier functions, we freeze the self-consistent potential and perform a non-self-consistent calculation on a uniform $n\times n\times n$ grid of k-points (the "*ab-initio* mesh"). We tested several grid densities ranging from $n=4$ to $n=10$ and ultimately chose $n=8$ (see end of next section). Since we want to construct 18 WFs (s, p, and d-like for spin up and down), we need to include a sufficient number of extra bands to cover the orbital character of these intended WFs everywhere in the Brillouin zone. With this in mind, we calculate the first 28 bands at each k-point, and then exclude any bands above 58 eV, the "outer window" of Ref. 20. (The choice of outer window is somewhat arbitrary as long as the number of bands it encloses is larger than the number of WFs, and we confirm that the calculated AHC has very little dependence upon this choice. The main effect of choosing a larger outer window is that one obtains slightly more localized WFs in real space, and thus slightly smoother bands in k-space.) The 18 WFs are then disentangled from the remaining bands using the procedure described in the next section.

B. Maximally localized spinor Wannier functions for bcc Fe

The energy bands of interest (extending up to, and just above, the Fermi energy) have mainly mixed s and d character and are entangled with the bands at higher energies. In order to construct maximally localized WFs to describe these bands, we use a modified version of the postprocessing procedure of Ref. 20. We start by reviewing the original two-step procedure from that work, as it applies to iron. In the first ("subspace selection") step, an 18-band subspace (the "projected space") is identified. This is done by minimizing a suitably defined functional, subject to the constraint of including the states within an inner energy window.[20] In the case of iron we choose this window to span an energy range of 30 eV from the bottom of the valence bands (up to E_{win} in Fig. 1). In the second ("gauge selection") step, the gauge freedom within the projected subspace is explored to obtain a set of Bloch-like functions $|u_{n\mathbf{k}}^{(W)}\rangle$ which are optimally smooth as a function of $\mathbf{k}$.[23] They are related to the 18 maximally localized WFs by Eq. (12). Although the method of Refs. 20 and 23 was formulated for the spinless case, it is trivial to adapt it to treat spinor wave functions, in which

case the resulting WFs also have spinor character: each element of the overlap matrix, which is the key input to the WF-generation code, is simply calculated as the sum of two spin components,

$$S_{\mathbf{k},\mathbf{b}}^{nm} = \sum_{\sigma=\uparrow,\downarrow} \langle u_{n\mathbf{k}}^{\sigma} | u_{m,\mathbf{k}+\mathbf{b}}^{\sigma} \rangle. \tag{45}$$

In order to facilitate later analysis (e.g., of the orbital and spin character of various bands), we have used a modified three-step procedure. The initial subspace selection step remains unchanged. The new second step ("subspace division") consists of splitting the 18-dimensional projected space for each **k** on the *ab initio* mesh into two nine-dimensional subspaces, as follows. At each k-point we form the 18×18 matrix representation of the spin operator $\hat{S}_z = (\hbar/2)\hat{\sigma}_z$ in the projected space and diagonalize it. The two nine-dimensional subspaces are then chosen as a mostly spin-up subspace spanned by the eigenstates having S_z eigenvalues close to $+1$, and a mostly spin-down subspace associated with eigenvalues close to -1 (we will use units of $\hbar/2$ whenever we discuss S_z in the remainder of the paper). The third and final step is the gauge-selection step, which is now done separately for each of the two nine-dimensional subspaces. We thus emerge with 18 well-localized WFs divided into two groups: nine that are almost entirely spin-up and nine that are almost entirely spin-down (in practice we find $|\langle \hat{S}_z \rangle| > 0.999$ in all cases). While this procedure results in a total spread that is slightly greater than the original two-step procedure, we find that the difference is very small in practice, and the imposition of these rules makes for a much more transparent analysis of subsequent results. For example, it makes it much easier to track the changes in the WFs before and after the spin-orbit coupling is turned on, or to identify the spin character of various pieces of the Fermi surface.

The subspace-selection step can be initialized[20] by providing 18 trial functions having the form of s, p, and (e_g and t_{2g}) d-like Gaussians of pure spin character (nine up and nine down). In our first attempts at initializing the gauge-selection step, we used these same trial functions. However, we found that the iterative gauge-selection procedure,[23] which projects the nine trial functions of each spin onto the appropriate band subspace and improves upon them, converted the three t_{2g}-like trial functions into t_{2g}-like WFs, while it mixed the e_g, s, and p-like states to form six hybrid WFs of sp^3d^2-type.[31] Having discovered this, we have modified our procedure accordingly: henceforth, we choose three t_{2g}-like trial functions and six sp^3d^2-like ones in each spin channel. With this initialization, we find the convergence to be quite rapid, with only about 100 iterations needed to get a well-converged spread functional.

We have implemented the above procedure in the WANNIER90 code.[32] The resulting WFs are shown in Fig. 2. The up-spin WFs are plotted, but the WFs are very similar for both spins. An example of an sp^3d^2-hybrid WF is shown in Fig. 2(a); this one extends along the $-x$ axis, and the five others are similarly projected along the $+x$, $\pm y$, and $\pm z$ axes. One of the t_{2g}-like WFs is shown in Fig. 2(b); this one has xy
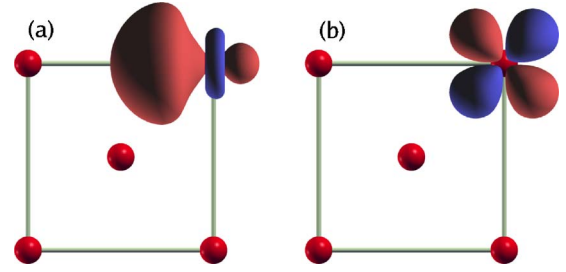

FIG. 2. (Color online) Isosurface contours of maximally localized spin-up WF in bcc Fe (red for positive value and blue for negative value), for the $8 \times 8 \times 8$ k-point sampling. (a) sp^3d^2-like WF centered on a Cartesian axis and (b) d_{xy}-like WF centered on the atom.

symmetry, while the others have xz and yz symmetry. The centers of the sp^3d^2-like WFs are slightly shifted from the atomic center along $\pm x$, $\pm y$, or $\pm z$, while the t_{2g}-like WFs remain centered on the atom.

We studied the convergence of the WFs and interpolated bands as a function of the density $n \times n \times n$ of the Monkhorst-Pack k-mesh used for the initial *ab initio* calculation. We tested $n = 4$, 6, 8, and 10, and found that $n = 8$ provided the best tradeoff between interpolation accuracy and computational cost. This is the mesh that was used in generating the results presented in Sec. V.

C. Variable spin-orbit coupling in the pseudopotential framework

Since the AHE present in ferromagnetic iron is a spin-orbit-induced effect, it is obviously important to understand the role of this coupling as thoroughly as possible. For this purpose, it is very convenient to be able to treat the strength of the coupling as an adjustable parameter. For example, by turning up the spin-orbit coupling continuously from zero and tracking how various contributions to the AHC behave, it is possible to separate out those contributions that are of linear, quadratic, or higher order in the coupling strength. Some results of this kind will be given later in Sec. V.

Because the spin-orbit coupling is a relativistic effect, it is appreciable mainly in the core region of the atom where the electrons have relativistic velocities. In a pseudopotential framework of the kind adopted here, both the scalar relativistic effects and the spin-orbit coupling are included in the pseudopotential construction. For example, in the Bachelet-Hamann semilocal pseudopotential scheme,[33] the construction procedure generates, for each orbital angular momentum l, a scalar-relativistic potential $V_l^{\text{sr}}(r)$ and a spin-orbit difference potential $V_l^{\text{so}}(r)$ which enter the Hamiltonian in the form

$$\hat{V}_{\text{ps}} = \sum_l \hat{P}_l [V_l^{\text{sr}}(r) + \lambda V_l^{\text{so}}(r) \mathbf{L} \cdot \mathbf{S}], \tag{46}$$

where $\hat{P}_l$ is the projector onto states of orbital angular momentum l and λ controls the strength of spin-orbit coupling (with $\lambda = 1$ being the physical value). For the free atom, this correctly leads to eigenstates labeled by total angular momentum $j = l \pm 1/2$.

In our calculations, we employ fully nonlocal pseudopotentials instead of semilocal ones because of their computationally efficient form. In this case, controlling the strength of the spin-orbit coupling requires some algebraic manipulation. We write the norm-conserving nonlocal pseudopotential operator as

$$\hat{V}_{\mathrm{ps}} = |\beta_{lj\mu}\rangle D_{lj} \langle \beta_{lj\mu}|, \tag{47}$$

where there is an implied sum running over the indices (orbital angular momentum l, total angular momentum $j = l \pm 1/2$, and $\mu = -j, \ldots, j$) and species and atomic position indices have been suppressed. The $|\beta_{lj\mu}\rangle$ are radial functions multiplied by appropriate spin-angular functions and the D_{lj} are the channel weights. We introduce the notation $\beta_l^{(+)}(r)$ and $\beta_l^{(-)}(r)$ for the radial parts of $|\beta_{l,l+1/2,\mu}\rangle$ and $\beta_{l,l-1/2,\mu}\rangle$, respectively, and similarly define $D_l^{(\pm)} = D_{l,l\pm1/2}$. Using this notation, we can define the scalar-relativistic (i.e., j-averaged) quantities

$$D_l^{\mathrm{sr}} = \frac{l+1}{2l+1} D_l^{(+)} + \frac{l}{2l+1} D_l^{(-)}, \tag{48}$$

$$\beta_l^{\mathrm{sr}}(r) = \frac{l+1}{2l+1} \sqrt{\frac{D_l^{(+)}}{D_l^{\mathrm{sr}}}} \beta_l^{(+)}(r) + \frac{l}{2l+1} \sqrt{\frac{D_l^{(-)}}{D_l^{\mathrm{sr}}}} \beta_l^{(-)}(r) \tag{49}$$

and the corresponding spin-orbit difference quantities

$$D_{lj}^{\mathrm{so}} = D_{lj} - D_l^{\mathrm{sr}}, \tag{50}$$

$$|\beta_{lj\mu}^{\mathrm{so}}\rangle = |\beta_{lj\mu}\rangle - |\beta_{lj\mu}^{\mathrm{sr}}\rangle, \tag{51}$$

where $|\beta_{lj\mu}^{\mathrm{sr}}\rangle$ is $\beta_l^{\mathrm{sr}}(r)$ multiplied by the spin-angular function with labels $(lj\mu)$. Then the nonlocal pseudopotential can be written as

$$\hat{V}_{\mathrm{ps}} = \hat{V}^{\mathrm{sr}} + \lambda \hat{V}^{\mathrm{so}}, \tag{52}$$

where

$$\hat{V}^{\mathrm{sr}} = |\beta_{lj\mu}^{\mathrm{sr}}\rangle D_l^{\mathrm{sr}} \langle \beta_{lj\mu}^{\mathrm{sr}}| \tag{53}$$

and

$$\begin{aligned} \hat{V}_{\mathrm{so}} = &|\beta_{lj\mu}^{\mathrm{sr}}\rangle D_{lj}^{\mathrm{so}} \langle \beta_{lj\mu}^{\mathrm{sr}}| + |\beta_{lj\mu}^{\mathrm{so}}\rangle (D_l^{\mathrm{sr}} + D_{lj}^{\mathrm{so}}) \langle \beta_{lj\mu}^{\mathrm{sr}}| \\ &+ |\beta_{lj\mu}^{\mathrm{sr}}\rangle (D_l^{\mathrm{sr}} + D_{lj}^{\mathrm{so}}) \langle \beta_{lj\mu}^{\mathrm{so}}| + |\beta_{lj\mu}^{\mathrm{so}}\rangle (D_l^{\mathrm{sr}} + D_{lj}^{\mathrm{so}}) \langle \beta_{lj\mu}^{\mathrm{so}}|. \end{aligned} \tag{54}$$

This clearly reduces to the desired results [Eq. (47)] for $\lambda = 1$ and [Eq. (53)] for $\lambda = 0$.

V. RESULTS

In this section, we present the results of the calculations of the Berry curvature and its integration over the BZ using the formulas presented in Sec. III, for the case of bcc Fe.

A. Berry curvature

We begin by illustrating the very sharp and strong variations that can occur in the total Berry curvature, Eq. (8), near
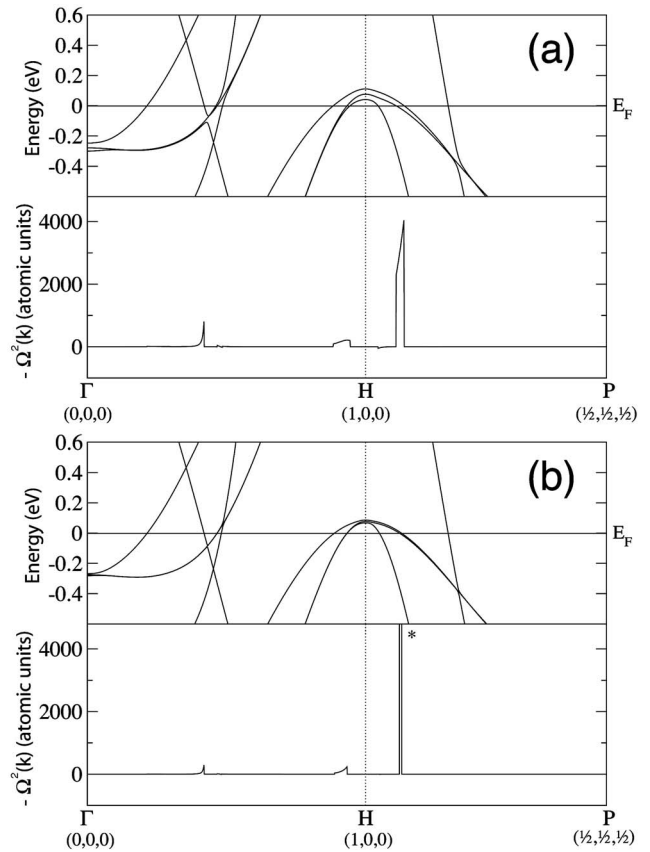

FIG. 3. Band structure and total Berry curvature, as calculated using Wannier interpolation, plotted along the path Γ-H-P in the Brillouin zone. (a) Computed at the full spin-orbit coupling strength $\lambda = 1$. (b) Computed at the reduced strength $\lambda = 0.25$. The peak marked with a star has a height of 5×10^4 a.u.

Fermi-surface features in the band structure.[16] In Fig. 3(a) we plot the energy bands (top subpanel) and the total Berry curvature (bottom subpanel) in the vicinity of the zone-boundary point $H = \frac{2\pi}{a}(1,0,0)$, where three states, split by the spin-orbit interaction, lie just above the Fermi level. The large spike in the Berry curvature between the H and P points arises where two bands, split by the spin orbit interaction, lie on either side of the Fermi level.[17] This gives rise to small energy denominators, and hence large contributions, mainly in Eq. (34). On reducing the strength of the spin-orbit interaction as in Fig. 3(b), the energy separation between these bands is reduced, resulting in a significantly sharper and higher spike in the Berry curvature. A second type of sharp structure is visible in Fig. 4, where one can see two smaller spikes, one at about 40% and another at about 90% of the way from Γ to H, which decrease in magnitude as the spin-orbit coupling strength is reduced. These arise from pairs of bands that straddle the Fermi energy even in the absence of spin-orbit interaction. Thus the small spin-orbit coupling does not shift the energies of these bands significantly, but it does induce an appreciable Berry curvature that is roughly linear in the spin-orbit coupling.

The decomposition of the total Berry curvature into its various contributions in Eq. (32) is illustrated by plotting the first ("$\bar{\Omega}$") term, the second and third ("D-$\bar{A}$") terms, and the
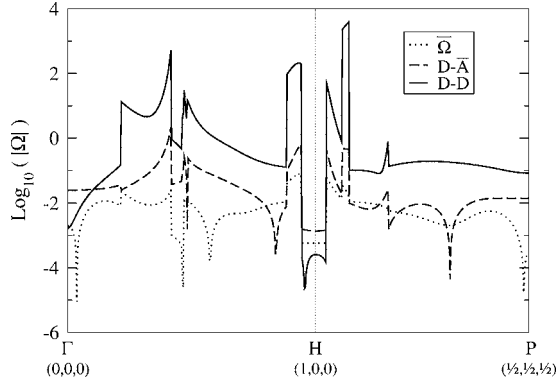
FIG. 4. Decomposition of the total Berry curvature into contributions coming from the three kinds of terms appearing in Eq. (32). The path in k-space is the same as in Fig. 3. The dotted line is the first ($\bar{\Omega}$) term, the dashed line is the sum of second and third (D-$\bar{A}$) terms, and the solid line is the fourth (D-D) term of Eq. (32). Note the log scale on the vertical axis.

fourth ("D-D" or Kubo-like) term of Eq. (32) separately along the line Γ-H-P. Note the logarithmic scale. The results confirm the expectations expressed in Secs. III C and III D, namely that the largest terms would be those reflecting large contributions to D arising from small energy denominators. Thus the $\bar{\Omega}$ term remains small everywhere, the D-$\bar{A}$ terms become one or two orders of magnitude larger at places where small energy denominators occur, and the D-D term, Eq. (34), is another one or two orders larger in those same regions. Scans along other lines in k-space reveal similar behavior. We may therefore expect that the D-D term will make the dominant overall contribution to the AHC. As we shall show in the next section, this is precisely the case.

In order to get a better feel for the connection between Fermi surface features and the Berry curvature, we next inspect these quantities on the $k_y=0$ plane in the Brillouin zone, following Ref. 17. In Fig. 5 we plot the intersection of the Fermi surface with this plane and indicate, using color coding, the S_z component of the spin carried by the corresponding wave functions. The good agreement between the shape of the Fermi surface given here and in Fig. 3 of Ref. 17 is further evidence that the accuracy of our approach matches that of all-electron methods. It is evident that the presence of the spin-orbit interaction, in addition to the exchange splitting, is sufficient to remove all degeneracies on this plane,[34] changing significantly the connectivity of the Fermi surface.

The calculated Berry curvature is shown in Fig. 6. It can be seen that the regions in which the Berry curvature is small (light green regions) fill most of the plane. The largest values occur at the places where two Fermi lines approach one another, consistent with the discussion of Fig. 3. Of special importance are the avoided crossings between two bands having the same sign of spin, or between two bands of opposite spin. Examples of both kinds are visible in the figure, and both tend to give rise to very large contributions in the region of the avoided crossing. Essentially, the spin-orbit interaction causes the character of these bands to change extremely rapidly with **k** near the avoided crossing; this is the
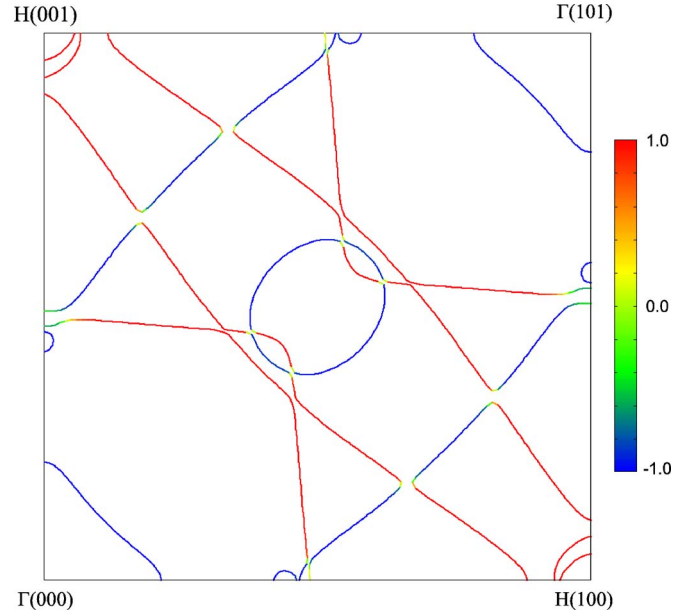

FIG. 5. (Color online) Lines of intersection between the Fermi surface and the plane $k_y=0$. Colors indicate the S_z spin-component of the states on the Fermi surface (in units of $\hbar/2$).

origin of the large Berry curvature. The large contributions near the H points correspond to the peaks that were already mentioned in the discussion of Fig. 3, resulting from mixing of nearly degenerate bands by the spin-orbit interaction.

B. Integrated anomalous Hall conductivity

We now discuss the computation of the AHC as an integral of the Berry curvature over the Brillouin zone, Eq. (7). We first define a nominal $N_0 \times N_0 \times N_0$ mesh that uniformly
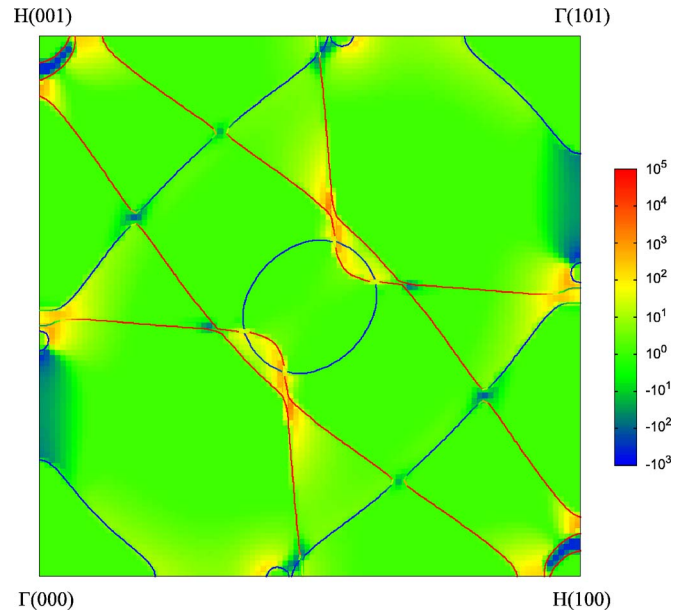

FIG. 6. (Color online) Calculated total Berry curvature $-\Omega_z$ in the plane $k_y=0$ (note log scale). Intersections of the Fermi surface with this plane are again shown.

TABLE I. Convergence of AHC with respect to the density of the nominal k-point mesh (left column) and the adaptive refinement scheme used to subdivide the mesh in regions of large contributions (middle column).

k-point mesh	Adaptive refinement	σ_{xy} $(\Omega \, \text{cm})^{-1}$
$200 \times 200 \times 200$	$3 \times 3 \times 3$	766.94
$250 \times 250 \times 250$	$3 \times 3 \times 3$	767.33
$320 \times 320 \times 320$	$3 \times 3 \times 3$	768.29
$200 \times 200 \times 200$	$5 \times 5 \times 5$	758.35
$250 \times 250 \times 250$	$5 \times 5 \times 5$	758.84
$320 \times 320 \times 320$	$5 \times 5 \times 5$	759.25
$200 \times 200 \times 200$	$7 \times 7 \times 7$	756.25
$250 \times 250 \times 250$	$7 \times 7 \times 7$	757.32
$320 \times 320 \times 320$	$7 \times 7 \times 7$	757.59
$320 \times 320 \times 320$	$9 \times 9 \times 9$	757.08
$320 \times 320 \times 320$	$11 \times 11 \times 11$	756.86
$320 \times 320 \times 320$	$13 \times 13 \times 13$	756.76

fills the Brillouin zone. We next reduce this to a sum over the irreducible wedge that fills $\frac{1}{16}$th of the Brillouin zone, using the tetragonal point-group symmetry (broken from cubic by the onset of ferromagnetism), and calculate Ω_z on each mesh point using Eq. (32). Finally, following Yao *et al.*,[17] we implement an adaptive mesh refinement scheme in which we identify those points of the k-space mesh at which the computed Berry curvature exceeds a threshold value Ω_{cut}, and recompute Ω_z on an $N_a \times N_a \times N_a$ submesh spanning the original cell associated with this mesh point. The AHC is then computed as a sum of Ω_z over this adaptively refined mesh with appropriate weights.

The convergence of the AHC with respect to the choice of mesh is presented in Table I. We have chosen $\Omega_{cut} = 1.0 \times 10^2$ a.u., which causes the adaptive mesh refinement to be triggered at approximately 0.11% of the original mesh points. Based on the results of Table I, we estimate the converged value to be $\sigma_{xy} = 756 (\Omega \, \text{cm})^{-1}$. This agrees to within 1% with the value of 751 $(\Omega \, \text{cm})^{-1}$ reported previously in Ref. 17, where an adaptive mesh refinement was also used. As discussed in Ref. 17, this value is in reasonable agreement with the available measurements,[35,36] which yield a value for σ_{xy} slightly above $1000 (\Omega \, \text{cm})^{-1}$.

It can be seen from Table I that a $200 \times 200 \times 200$ mesh with $3 \times 3 \times 3$ refinement brings us within $\sim 1\%$ of the converged value. It is also evident that the level of refinement is more important than the fineness of the nominal mesh; a $200 \times 200 \times 200$ mesh with $5 \times 5 \times 5$ adaptive refinement yields a result that is within 0.2% of the converged value, better than a $320 \times 320 \times 320$ mesh with a lower level of refinement.

It is interesting to decompose the total AHC into contributions coming from different parts of the Brillouin zone. For example, as we saw in Fig. 6, there is a smooth, low-intensity background that fills most of the volume of the Brillouin zone, and it is hard to know *a priori* whether the total AHC is dominated by these contributions or by the

TABLE II. Contributions to the AHC coming from different regions of the Brillouin zone, as defined in the text.

ΔE (eV)	Like-spin (%)	Opposite-spin (%)	Smooth (%)
0.1	21	26	53
0.2	23	51	26
0.5	30	68	2

much larger ones concentrated in small regions. With this motivation, we have somewhat arbitrarily divided the Brillouin zone into three kinds of regions, which we label as "smooth," "like-spin," and "opposite-spin." To do this, we identify k-points at which there is an occupied band in the interval $[E_f - \Delta E, E_f]$ and an unoccupied band in the interval $[E_f, E_f + \Delta E]$, where ΔE is arbitrarily chosen to be a small energy such as 0.1, 0.2, or 0.5 eV. If so, the k-point is said to belong to the "like-spin" or "opposite-spin" region depending on whether the dominant characters of the two bands below and above the Fermi energy are of the same or of opposite spin. Otherwise, the k-point is assigned to the "smooth" region. As shown in Table II, the results depend strongly on the value of ΔE. Overall, what is clear is that the major contributions arise from the bands within ± 0.5 eV of E_f, and that neither like-spin nor opposite-spin contributions are dominant.

Next, we return to the discussion of the decomposition of the total Berry curvature in Eq. (32) into the $\bar{\Omega}$, D-$\bar{A}$, and D-D terms. We find that these three terms account for -0.39%, 1.36%, and 99.03%, respectively, of the total AHC. [Similarly, for the alternative decomposition of Appendix A, the second term of Eq. (A4) is found to be responsible for more than 99% of the total.] Thus if a 1% accuracy is acceptable, one could actually neglect the $\bar{\Omega}$ and D-$\bar{A}$ terms entirely, and approximate the total AHC by the D-D (Kubo-like) term alone, Eq. (34).

From a computational point of view, the fact that the D-D term is fully specified by the Hamiltonian matrix elements alone means that considerable savings can be obtained by avoiding the evaluation of the Fourier transforms in Eqs. (39) and (40) at every interpolation point (and avoiding the setup of the matrix elements $\langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}m \rangle$, which can be costly in a real-space implementation). More importantly, this observation, if it turns out to hold for other materials as well, could prove to be important for future efforts to derive approximate schemes capable of capturing the most important contributions to the AHC.

Finally, we investigate how the total AHC depends upon the strength of the spin-orbit interaction, following the approach of Sec. IV C to modulate the spin-orbit strength. The result is shown in Fig. 7. We emphasize that our approach is a more specific test of the dependence upon spin-orbit strength than the one carried out in Ref. 17; there, the speed of light c was varied, which entails changing the strength of the various scalar relativistic terms as well. Nevertheless, both studies lead to a similar conclusion: the variation is found to be linear for small values of the spin-orbit coupling
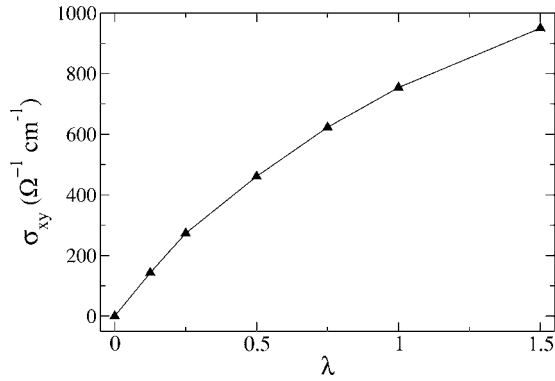
FIG. 7. Anomalous Hall conductivity vs spin-orbit coupling strength.

($\lambda \ll 1$), while quadratic or other higher-order terms also become appreciable when the full interaction is included ($\lambda = 1$).

C. Computational considerations

The computational requirements for this scheme are quite modest. The self-consistent ground state calculation and the construction of the WFs takes 2.5 h on a single 2.2 GHz AMD-Opteron processor. The expense of computing the AHC as a sum over interpolation mesh points depends strongly on the density of the mesh. On the same processor as above, the average CPU time to evaluate Ω_z on each k-point was about 14 ms. We find that the mesh refinement operation does not significantly increase the total number of k-point evaluations until the refinement level N_a exceeds ~10. Allowing for the fact that the calculation only needs to be done in the irreducible $\frac{1}{16}$ of the Brillouin zone, the cost for the AHC evaluation on a $200 \times 200 \times 200$ mesh is about 2 h.

The CPU time per k-point evaluation is dominated (roughly 90%) by the Fourier transform operations needed to construct the objects in Eqs. (37)–(40). The diagonalization of the 18×18 Hamiltonian matrix, and other operations needed to compute Eq. (32), account for only about 10% of the time. The CPU requirement for the Fourier transform step is roughly proportional to the number of **R** vectors kept in Eqs. (37)–(40); it is possible that this number could be reduced by exploring more sophisticated methods for truncating the contributions coming from the more distant **R** vectors.

Of course, the loop over k-points in the AHC calculation is trivial to parallelize, so for dense k-meshes we speed up this stage of the calculation by distributing across multiple processors.

VI. SUMMARY AND DISCUSSION

In summary, we have developed an efficient method for computing the intrinsic contribution to the anomalous Hall conductivity of a metallic ferromagnet as a Brillouin-zone integral of the Berry curvature. Our approach is based on Wannier interpolation, a powerful technique for evaluating properties that require a very dense sampling of the Brillouin zone or Fermi surface. The key idea is to map the low-energy first-principles electronic structure onto an "exact tight-binding model" in the basis of appropriately constructed Wannier functions, which are typically partially occupied. In the Wannier representation the desired quantities can then be evaluated at arbitrary k-points at very low computational cost. All that is needed is to evaluate, once and for all, the Wannier-basis matrix elements of the Hamiltonian and a few other property-specific operators (namely, for the Berry curvature, the three Cartesian position operators).

When evaluating the Berry curvature in this way, the summation over all unoccupied bands and the expensive calculation of the velocity matrix elements needed in the traditional Kubo formula are circumvented. They are replaced by quantities defined strictly within the projected space spanned by the WFs. Our final expression for the total Berry curvature, Eq. (32), consists of three terms, namely, the $\bar{\Omega}$, $D\text{-}\bar{A}$, and $D\text{-}D$ terms.

We have applied this approach to calculate the AHC of bcc Fe. While our Wannier interpolation formalism, with its decomposition (32), is entirely independent of the choice of an all-electron or pseudopotential method, we have chosen here a relativistic pseudopotential approach[24] that includes scalar relativistic effects as well as the spin-orbit interaction. We find that this scheme successfully reproduces the fine details of the electronic structure and of the Berry curvature. The resulting AHC is in excellent agreement with a previous calculation[17] that used an all-electron LAPW method.[37]

Remarkably, we found that more than 99% of the integrated Berry curvature is concentrated in the $D\text{-}D$ term of our formalism. This term, given explicitly in Eq. (34), takes the form of a Kubo-like Berry curvature formula for the "tight-binding states." Unlike the $\bar{\Omega}$ and $D\text{-}\bar{A}$ terms, it depends exclusively on the Hamiltonian matrix elements between the Wannier orbitals, and not on the position matrix elements. Thus we arrive at the very appealing result that a Kubo picture defined within the "tight-binding space" gives an excellent representation of the Berry curvature in the original *ab initio* space. This result merits further investigation.

Several directions for future studies suggest themselves. For example, it would be desirable to obtain a better understanding of how the AHC depends on the weak spin-orbit interaction. As we have seen, this weak interaction causes splittings and avoided crossings that give rise to very large Berry curvatures in very small regions of k-space. There is a kind of paradox here. Our numerical tests, as in Fig. 7, demonstrate that the AHC falls smoothly to zero as the spin-orbit strength λ is turned off, suggesting that a perturbation theory in λ should be applicable. However, in the limit that λ becomes small, the full calculation becomes *more difficult*, not less: the splittings occur in narrower and narrower regions of k-space, energy denominators become smaller, and Berry curvature contributions become larger (see Fig. 3), even if the *integrated* contribution is going to zero. It would be of considerable interest, therefore, to explore ways to reformulate the perturbation theory in λ so that the expansion coefficients can be computed in a robust and efficient fashion.

Because the exchange splitting is much larger than the spin-orbit splitting, it may also be of use to introduce two separate couplings that control the strengths of the spin-flip and spin-conserving parts of the spin-orbit interaction, respectively, and to work out the perturbation theory in these two couplings independently.

Another promising direction is to explore whether the AHC can be computed as a Fermi-surface integral using the formulation of Haldane[12] in which an integration by parts is used to convert the volume integral of the Berry curvature to a Fermi-surface integral involving Berry curvatures or potentials. Such an approach promises to be more efficient than the volume-integration approach, provided that a method can be developed for carrying out an appropriate sampling of the Fermi surface. This is likely to be a delicate problem, however, since the weak spin-orbit splitting causes Fermi sheets to separate and reattach in a complex way at short k-scales, and the dominant contributions to the AHC are likely to come from precisely these portions of the reconstructed Fermi surface that are the most difficult to describe numerically.

Finally, it would be of considerable interest to generalize the Wannier-interpolation techniques developed here for the dc anomalous Hall effect to treat finite-frequency magneto-optical effects.

In any case, even without such further developments, the present approach is a powerful one. It reduces the expense needed to do an extremely fine sampling of Fermi-surface properties to the level where the AHC of a material like bcc Fe can be computed on a workstation in a few hours. This opens the door to realistic calculations of the intrinsic anomalous Hall conductivity of much more complex materials. More generally, the techniques developed here for the AHE are readily applicable to other problems which also require a very dense sampling of the Fermi surface or Brillouin zone. For example, an extension of these ideas to the evaluation of the electron-phonon coupling matrix elements by Wannier interpolation is currently under way.[41]

APPENDIX A: ALTERNATIVE EXPRESSION FOR THE BERRY CURVATURE

In this appendix, we return to Eq. (27) and rewrite it in such a way that all of the large, rapidly varying contributions arising from small energy denominators in the expression for D_α, Eq. (24), are segregated into a single term. We do this by solving Eq. (25) for D_α and substituting into Eq. (27) to obtain

$$\Omega_{\alpha\beta}^{(H)} = \bar{\Omega}_{\alpha\beta}^{(H)} - i[\bar{A}_\alpha^{(H)}, \bar{A}_\beta^{(H)}] + i[A_\alpha^{(H)}, A_\beta^{(H)}]. \quad (A1)$$

Then only the last term will contain the large, rapid variations. This equation could have been anticipated based on the fact that the tensor

$$\tilde{\Omega}_{\alpha\beta} = \Omega_{\alpha\beta} - i[A_\alpha, A_\beta] \quad (A2)$$

is well-known to be a gauge-covariant quantity;[23,38] applying Eq. (21) to $\tilde{\Omega}_{\alpha\beta}$ then leads directly to Eq. (A1).

This formulation provides an alternative route to the calculation of the matrix $\Omega_{\alpha\beta}^{(H)}$: evaluate $\tilde{\Omega}_{\alpha\beta}^{(W)}$ in the Wannier representation using Eqs. (A5) and (A6) below, convert it to $\tilde{\Omega}_{\alpha\beta}^{(H)}$ via Eq. (21), compute $A_\alpha^{(H)}$ using Eq. (25), and assemble

$$\Omega_{\alpha\beta}^{(H)} = \tilde{\Omega}_{\alpha\beta}^{(H)} + i[A_\alpha^{(H)}, A_\beta^{(H)}]. \quad (A3)$$

The large and rapid variations then appear only in the last term involving commutators of the A matrices.

In Sec. III D, we showed how to write the total Berry curvature $\Omega_{\alpha\beta}(\mathbf{k})$ as a sum over bands in such a way that potentially troublesome contributions coming from small energy denominators between pairs of occupied bands are explicitly excluded, leading to Eq. (32). The corresponding expression based on Eq. (A3) is

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_n f_n \tilde{\Omega}_{nn,\alpha\beta}^{(H)} + \sum_{nm} (f_n - f_m) A_{nm,\alpha}^{(H)} A_{mn,\beta}^{(H)}. \quad (A4)$$

Now, in addition to the four quantities given in Eqs. (37)–(40), we need a corresponding equation for $\tilde{\Omega}_{\alpha\beta}$. After some manipulations, we find that

$$\tilde{\Omega}_{nn,\alpha\beta}^{(W)}(\mathbf{k}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} w_{n,\alpha\beta}(\mathbf{R}), \quad (A5)$$

where

$$w_{n,\alpha\beta}(\mathbf{R}) = -i \sum_{\mathbf{R}'m} \langle \mathbf{0}n|\hat{r}_\alpha|\mathbf{R}'m\rangle\langle\mathbf{R}'m|\hat{r}_\beta|\mathbf{R}n\rangle$$
$$+ i \sum_{\mathbf{R}'m} \langle \mathbf{0}n|\hat{r}_\beta|\mathbf{R}'m\rangle\langle\mathbf{R}'m|\hat{r}_\alpha|\mathbf{R}n\rangle. \quad (A6)$$

This formulation again requires the same basic ingredients as before, namely, the Wannier matrix elements of $\hat{H}$ and $\hat{r}_\alpha$. In some respects it is a little more elegant than the formulation of Eq. (32). However, the direct evaluation of $w_{n,\alpha\beta}$ in the Wannier representation, as given in Eq. (A6), is not as convenient because of the extra sum over intermediate WFs appearing there; moreover, $w_{n,\alpha\beta}$ is longer-ranged than the Hamiltonian and coordinate matrix elements. Also, one appealing feature of the formulation of Sec. III, that more than 99% of the effect can be recovered without using the position-operator matrix elements, is lost in this reformulation. We have therefore chosen to base our calculations and analysis on Eq. (32) instead.

It is informative to obtain Eq. (A3) in a different way: define the gauge-invariant band projection operator[23] $\hat{P}_{\mathbf{k}} = \sum_{n=1}^M |u_{n\mathbf{k}}\rangle\langle u_{n\mathbf{k}}|$ and its complement $\hat{Q}_{\mathbf{k}} = \hat{1} - \hat{P}_{\mathbf{k}}$. Inserting $\hat{1} = \hat{Q}_{\mathbf{k}} + \hat{P}_{\mathbf{k}}$ into Eq. (18) in the Hamiltonian gauge then yields directly Eq. (A3) since, as can be easily verified, Eq. (A2) may be written as

$$\tilde{\Omega}_{nm,\alpha\beta} = i\langle \tilde{\partial}_\alpha u_n | \tilde{\partial}_\beta u_m \rangle - i\langle \tilde{\partial}_\beta u_n | \tilde{\partial}_\alpha u_m \rangle, \qquad \text{(A7)}$$

where $\tilde{\partial}_\alpha \equiv \hat{Q}\partial_\alpha$. The gauge-covariance of $\tilde{\Omega}_{\alpha\beta}$ follows directly from the fact that $\tilde{\partial}_\alpha$ is a gauge-covariant derivative, in the sense that $|\tilde{\partial}_\alpha u_n^{(\mathrm{H})}\rangle = \Sigma_{m=1}^M |\tilde{\partial}_\alpha u_m^{(\mathrm{W})}\rangle U_{mn}$ is the same transformation law as Eq. (15) for the Bloch states themselves. It is apparent from this derivation that as the number M of WFs increases and $\hat{P}_{\mathbf{k}}$ approaches $\hat{1}$, the second term on the right-hand side of Eq. (A4) increases at the expense of the first term. Indeed, in the large-M limit the entire Berry curvature is contained in the second term. For the choice Wannier orbitals described in the main text for bcc Fe, that term already accounts for 99.0% of the total AHC.

APPENDIX B: FINITE-DIFFERENCE APPROACH

In this appendix, we outline an alternative scheme for computing the AHC by Wannier interpolation. The essential difference relative to the approaches described in Sec. III and in Appendix A is that the needed k-space derivatives are approximated here by finite differences instead of being expressed analytically in the Wannier representation.

This approach is most naturally applied to the zero-temperature limit where there are exactly $N_{\mathbf{k}}$ occupied states at a given $\mathbf{k}$. Instead of starting from the Berry curvature of each individual band separately, as in Eq. (6), we find it convenient here to work from the outset with the total Berry curvature

$$\Omega_{\alpha\beta}(\mathbf{k}) = \sum_{n=1}^{N_{\mathbf{k}}} \Omega_{nn,\alpha\beta}(\mathbf{k}) \qquad \text{(B1)}$$

of the occupied manifold at $\mathbf{k}$ [the zero-temperature limit of Eq. (19)]. We now introduce a covariant derivative $\tilde{\partial}_\alpha^{(N_{\mathbf{k}})} = \hat{Q}_{\mathbf{k}}^{(N_{\mathbf{k}})}\partial_\alpha$ designed to act on the occupied states only; here $\hat{Q}_{\mathbf{k}}^{(N_{\mathbf{k}})} = \hat{1} - \hat{P}_{\mathbf{k}}^{(N_{\mathbf{k}})}$ and $\hat{P}_{\mathbf{k}}^{(N_{\mathbf{k}})} = \Sigma_{n=1}^{N_{\mathbf{k}}} |u_{n\mathbf{k}}\rangle\langle u_{n\mathbf{k}}|$. The only difference with respect to the definition of $\tilde{\partial}_\alpha$ in Appendix A is that the projection operator here spans the $N_{\mathbf{k}}$ occupied states only, instead of the M states of the full projected space. Accordingly, terms such as "gauge-covariance" and "gauge-invariance" are to be understood here in a restricted sense. For example, the statement that $\tilde{\partial}_\alpha^{(N_{\mathbf{k}})}$ is a gauge-covariant derivative means that under an $N_{\mathbf{k}} \times N_{\mathbf{k}}$ unitary rotation $\mathcal{U}(\mathbf{k})$ between the occupied states at $\mathbf{k}$ it obeys the transformation law

$$|\tilde{\partial}_\alpha^{(N_{\mathbf{k}})} u_{n\mathbf{k}}\rangle \rightarrow \sum_{m=1}^{N_{\mathbf{k}}} |\tilde{\partial}_\alpha^{(N_{\mathbf{k}})} u_{m\mathbf{k}}\rangle \mathcal{U}_{mn}(\mathbf{k}). \qquad \text{(B2)}$$

(We will use calligraphic symbols to distinguish $N_{\mathbf{k}} \times N_{\mathbf{k}}$ matrices such as $\mathcal{U}$ from their $M \times M$ counterparts such as U.) We now define a gauge-covariant curvature $\tilde{\Omega}_{\alpha\beta}^{(N_{\mathbf{k}})}(\mathbf{k})$ by replacing $\tilde{\partial}$ by $\tilde{\partial}^{(N_{\mathbf{k}})}$ in Eq. (A7). Since the trace of a commutator vanishes, it follows from Eq. (A2) that Eq. (B1) can be written as

$$\Omega_{\alpha\beta}(\mathbf{k}) = \mathrm{Tr}^{(N_{\mathbf{k}})}[\tilde{\Omega}_{\alpha\beta}^{(N_{\mathbf{k}})}(\mathbf{k})], \qquad \text{(B3)}$$

where the symbol $\mathrm{Tr}^{(N_{\mathbf{k}})}$ denotes the trace over the occupied states.

The advantage of this expression over Eq. (B1) is that the covariant derivative of a Bloch state can be approximated by a very robust finite-differences formula:[39,40]

$$\tilde{\partial}_{\mathbf{k}}^{(N_{\mathbf{k}})} \rightarrow \sum_{\mathbf{b}} w_b \mathbf{b} \hat{P}_{\mathbf{k},\mathbf{b}}^{(N_{\mathbf{k}})}, \qquad \text{(B4)}$$

where the sum is over shells of neighboring k-points,[23] as in Eq. (44), and we have defined the gauge-invariant operator

$$\hat{P}_{\mathbf{k},\mathbf{b}}^{(N_{\mathbf{k}})} = \sum_{n=1}^{N_{\mathbf{k}}} |\tilde{u}_{n,\mathbf{k}+\mathbf{b}}\rangle\langle u_{n\mathbf{k}}| \qquad \text{(B5)}$$

in terms of the gauge-covariant "dual states"

$$|\tilde{u}_{n,\mathbf{k},\mathbf{b}}\rangle = \sum_{m=1}^{N_{\mathbf{k}}} |u_{m,\mathbf{k}+\mathbf{b}}\rangle(\mathcal{Q}_{\mathbf{k}+\mathbf{b},\mathbf{k}})_{mn}. \qquad \text{(B6)}$$

Here $\mathcal{Q}_{\mathbf{k}+\mathbf{b},\mathbf{k}}$ is the inverse of the $N_{\mathbf{k}} \times N_{\mathbf{k}}$ overlap matrix,

$$\mathcal{Q}_{\mathbf{k}+\mathbf{b},\mathbf{k}} = (\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}})^{-1}, \qquad \text{(B7)}$$

where

$$(\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}})_{nm} = \langle u_{n\mathbf{k}} | u_{m,\mathbf{k}+\mathbf{b}} \rangle. \qquad \text{(B8)}$$

The discretization (B4) is immune to arbitrary gauge phases and unitary rotations among the occupied states; because of that property, the occurrence of band crossings and avoided crossings does not pose any special problems.

Inserting Eqs. (B4)–(B8) into Eq. (B3) and using $\mathcal{Q}_{\mathbf{k},\mathbf{k}+\mathbf{b}} = \mathcal{Q}_{\mathbf{k}+\mathbf{b},\mathbf{k}}^\dagger$, we find that an appropriate finite-difference expression for the total Berry curvature is

$$\Omega_{\alpha\beta}^{(N_{\mathbf{k}})}(\mathbf{k}) = 2 \sum_{\mathbf{b}_1,\mathbf{b}_2} w_{b_1} w_{b_2} b_{1,\alpha} b_{2,\beta} \Lambda_{\mathbf{k},\mathbf{b}_1,\mathbf{b}_2}, \qquad \text{(B9)}$$

where

$$\Lambda_{\mathbf{k},\mathbf{b}_1,\mathbf{b}_2} = -\mathrm{Im}\,\mathrm{Tr}^{(N_{\mathbf{k}})}[\mathcal{Q}_{\mathbf{k},\mathbf{k}+\mathbf{b}_1} \mathcal{S}_{\mathbf{k}+\mathbf{b}_1,\mathbf{k}+\mathbf{b}_2} \mathcal{Q}_{\mathbf{k}+\mathbf{b}_2,\mathbf{k}}]. \qquad \text{(B10)}$$

This expression is manifestly gauge-invariant, since both $\mathcal{S}$ and $\mathcal{Q}$ are gauge-covariant matrices, i.e., $\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}} \rightarrow \mathcal{U}^\dagger(\mathbf{k})\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}\mathcal{U}(\mathbf{k}+\mathbf{b})$, and the same transformation law holds for $\mathcal{Q}_{\mathbf{k},\mathbf{k}+\mathbf{b}}$.

Equations. (B9) and (B10) can be evaluated at an arbitrary point $\mathbf{k}$ once the overlap matrices $\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}$ are known. For that purpose we construct a uniform mesh of spacing Δk in the immediate vicinity of $\mathbf{k}$, set up the needed shells of neighboring k-points $\mathbf{k}+\mathbf{b}$ on that local mesh, and then evaluate $\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}$ by Wannier interpolation. Since the WFs span the entire M-dimensional projected space, at this stage we revert to the full $M \times M$ overlap matrices $S_{\mathbf{k},\mathbf{k}+\mathbf{b}}$. In the Wannier gauge they are given by a Fourier transform of the form

$$(S_{\mathbf{k},\mathbf{k}+\mathbf{b}}^{(\mathrm{W})})_{nm} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle \mathbf{0}n | e^{i\mathbf{b}\cdot(\mathbf{R}-\hat{\mathbf{r}})} | \mathbf{R}m \rangle. \qquad \text{(B11)}$$

For sufficiently small Δk, this can be approximated as

$$(S_{\mathbf{k},\mathbf{k}+\mathbf{b}}^{(\mathrm{W})})_{nm} \simeq \delta_{nm} - i\mathbf{b}\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\langle 0n|\hat{\mathbf{r}}|\mathbf{R}m\rangle. \qquad \text{(B12)}$$

Note that the dependence of the last expression on Δk is trivial, since it only enter as a multiplicative prefactor. In practice one chooses Δk to be quite small, $\sim 10^{-6}$ a.u.$^{-1}$, so as to reduce the error of the finite-differences expression.

In the Wannier gauge the occupied and empty states are mixed with one another because the WFs are partially occupied. In order to decouple the two subspaces we perform the unitary transformation

$$S_{\mathbf{k},\mathbf{k}+\mathbf{b}}^{(\mathrm{H})} = U^{\dagger}(\mathbf{k})S_{\mathbf{k},\mathbf{k}+\mathbf{b}}^{(\mathrm{W})}U(\mathbf{k}+\mathbf{b}). \qquad \text{(B13)}$$

This produces the full $M \times M$ overlap matrix in the Hamiltonian gauge. The $N_{\mathbf{k}} \times N_{\mathbf{k}}$ submatrix in the upper left corner is precisely the matrix $\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}}^{(\mathrm{H})}$ needed in Eq. (B10).

Like the approach described in the main text, this approach still only requires the WF matrix elements of the four operators $\hat{H}$ and $\hat{r}_{\alpha}$ ($\alpha = x$, y, and z). We have implemented it, and have checked that the results agree closely with those obtained using using the method of the main text. Although not as elegant, this approach has the interesting feature of

circumventing the evaluation of the matrix $D_{\alpha}^{(\mathrm{H})}$, Eq. (24). This may be advantageous in certain special situations. For example, if a parameter such as pressure is tuned in such a way that a k-space Dirac monopole[16] drifts to the Fermi surface, the vanishing of the energy denominator in Eq. (24) may result in a numerical instability when trying to find the monopole contribution to the AHC.

We conclude by noting that Eq. (B10) is but one of many possible finite-differences expressions,[42] and may not even be the most convenient one to use in practice. By recalling that the Berry curvature is the Berry phase per unit area, one realizes that in the small-Δk limit of interest, the quantity $\Lambda_{\mathbf{k},\mathbf{b}_1,\mathbf{b}_2}$ in Eq. (B9) can be viewed as the discrete Berry phase ϕ accumulated along the small loop $\mathbf{k} \to \mathbf{k}+\mathbf{b}_1 \to \mathbf{k} + \mathbf{b}_2 \to \mathbf{k}$. As is well-known, the Berry phase around a discrete loop is defined as[14]

$$\phi = -\operatorname{Im}\ln\det[\mathcal{S}_{\mathbf{k},\mathbf{k}+\mathbf{b}_1}\mathcal{S}_{\mathbf{k}+\mathbf{b}_1,\mathbf{k}+\mathbf{b}_2}\mathcal{S}_{\mathbf{k}+\mathbf{b}_2,\mathbf{k}}]. \qquad \text{(B14)}$$

It can be shown that $\phi = \Lambda_{\mathbf{k},\mathbf{b}_1,\mathbf{b}_2} + O(\Delta k^2)$, so that for small loops the two formulas agree. Equation (B14) has the practical advantage over Eq. (B10) that it does not require inverting the overlap matrix.

---

[1] C. M. Hurd, *The Hall Effect in Metals and Alloys* (Plenum, New York, 1972).

[2] A. Gerber *et al.*, J. Magn. Magn. Mater. **242**, 90 (2002).

[3] R. Karplus and J. M. Luttinger, Phys. Rev. **95**, 1154 (1954).

[4] J. Smit, Physica (Amsterdam) **24**, 39 (1958).

[5] L. Berger, Phys. Rev. B **2**, 4559 (1970).

[6] J. Sinova, T. Jungwirth, and J. Černe, Int. J. Mod. Phys. B **18**, 1083 (2004).

[7] H. R. Leribaux, Phys. Rev. **150**, 384 (1966).

[8] M.-C. Chang and Q. Niu, Phys. Rev. B **53**, 7010 (1996).

[9] G. Sundaram and Q. Niu, Phys. Rev. B **59**, 14915 (1999).

[10] M. Onoda and N. Nagaosa, J. Phys. Soc. Jpn. **71**, 19 (2002).

[11] T. Jungwirth, Q. Niu, and A. H. MacDonald, Phys. Rev. Lett. **88**, 207208 (2002).

[12] F. D. M. Haldane, Phys. Rev. Lett. **93**, 206602 (2004).

[13] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).

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

[15] E. N. Adams and E. I. Blount, J. Phys. Chem. Solids **10**, 286 (1959).

[16] Z. Fang *et al.*, Science **302**, 92 (2003).

[17] 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).

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

[19] E. I. Blount, Solid State Phys. **13**, 305 (1962).

[20] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001).

[21] M. Graf and P. Vogl, Phys. Rev. B **51**, 4940 (1995).

[22] T. G. Pedersen, K. Pedersen, and T. B. Kriestensen, Phys. Rev. B **63**, 201101(R) (2001).

[23] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997).

[24] S. Baroni, A. Dal Corso, S. de Gironcoli, P. Giannozzi, C. Cavazzoni, G. Ballabio, S. Scandolo, G. Chiarotti, P. Focher, A. Pasquarello, K. Laasonen, A. Trave, R. Car, N. Marzari, and A. Kokalj, http://www.pwscf.org/.

[25] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[26] G. Theurich and N. A. Hill, Phys. Rev. B **64**, 073106 (2001).

[27] A. Dal Corso and A. M. Conte, Phys. Rev. B **71**, 115106 (2005).

[28] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[29] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[30] N. Marzari, D. Vanderbilt, A. De Vita, and M. C. Payne, Phys. Rev. Lett. **82**, 3296 (1999).

[31] L. Pauling, J. Am. Chem. Soc. **53**, 1367 (1931).

[32] A. A. Mostofi, J. R. Yates, N. Marzari, I. Souza, and D. Vanderbilt, http://www.wannier.org/.

[33] G. B. Bachelet and M. Schlüter, Phys. Rev. B **25**, 2103 (1982).

[34] M. Singh, C. S. Wang, and J. Callaway, Phys. Rev. B **11**, 287 (1975).

[35] P. N. Dheer, Phys. Rev. **156**, 637 (1967).

[36] R. W Klaffky and R. V. Coleman, Phys. Rev. B **10**, 2915 (1974).

[37] D. J. Singh, *Planewaves, Pseudopotentials and the LAPW Method* (Kluwer Academic, Boston, 1994).

[38] C. A. Mead, Rev. Mod. Phys. **64**, 51 (1992).

[39] N. Sai, K. M. Rabe, and D. Vanderbilt, Phys. Rev. B **66**, 104108 (2002).

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

[41] F. Giustino, J. R. Yates, I. Souza, M. L. Cohen, and S. G. Louie (unpublished).

[42] T. Fukui, Y. Hatsugai, and H. Suzuki, J. Phys. Soc. Jpn. **74**, 1674 (2005).