

***Ab initio* calculation of the shift photocurrent by Wannier interpolation**Julen Ibañez-Azpiroz,¹ Stepan S. Tsirkin,¹ and Ivo Souza^{1,2}¹*Centro de Física de Materiales, Universidad del País Vasco (UPV/EHU), 20018 San Sebastián, Spain*²*Ikerbasque Foundation, 48013 Bilbao, Spain*

(Received 11 April 2018; published 26 June 2018)

We describe and implement a first-principles algorithm based on maximally localized Wannier functions for calculating the shift-current response of piezoelectric crystals in the independent-particle approximation. The proposed algorithm presents several advantages over existing ones, including full gauge invariance, low computational cost, and a correct treatment of the optical matrix elements with nonlocal pseudopotentials. Band-truncation errors are avoided by a careful formulation of $k \cdot p$ perturbation theory within the subspace of wannierized bands. The needed ingredients are the matrix elements of the Hamiltonian and of the position operator in the Wannier basis, which are readily available at the end of the wannierization step. If the off-diagonal matrix elements of the position operator are discarded, our expressions reduce to the ones that have been used in recent tight-binding calculations of the shift current. We find that this “diagonal” approximation can introduce sizable errors, highlighting the importance of carefully embedding the tight-binding model in real space for an accurate description of the charge transfer that gives rise to the shift current.

DOI: [10.1103/PhysRevB.97.245143](https://doi.org/10.1103/PhysRevB.97.245143)**I. INTRODUCTION**

Under homogeneous illumination, noncentrosymmetric crystals exhibit the bulk photovoltaic effect (BPVE), a nonlinear optical response that consists of the generation of a photovoltage (open circuit) or photocurrent (closed circuit) when light is absorbed via intrinsic or extrinsic processes [1–3]. Contrary to the conventional photovoltaic effect in p - n junctions, the BPVE occurs in homogeneous systems, and the attained photovoltage is not limited by the band gap of the material. The BPVE comprises a “circular” part that changes sign with the helicity of light, and a “linear” part that also occurs with linearly polarized or unpolarized light. The former is symmetry-allowed in the gyrotropic crystal classes, and the latter in the piezoelectric ones [1–3].

The present work deals with the intrinsic contribution to the linear BPVE due to interband absorption, known as “shift current.” This phenomenon was intensively studied in the 1960s and 1970s, particularly in ferroelectric oxides such as BaTiO₃ [4]. In recent years it has attracted renewed interest in view of potential applications in novel solar-cell designs [5–7], and in connection with topological insulators [8–10] and Weyl semimetals [11–13].

In a simplified picture, the shift current arises from a coordinate shift accompanying the photoexcitation of electrons from one band to another. Like the intrinsic anomalous Hall effect [14], the shift current originates from interband velocity matrix elements, depending not only on their magnitudes but also on their phases [15–18].

Over the years, the understanding of the shift current has greatly benefited from model calculations [6,7,19,20]. Tight-binding models have been used to analyze various aspects of the problem, including the possible correlation with electric polarization, the role of virtual transitions, and the sensitivity to the wave functions. Recently, density-functional theory methods started being employed to calculate the shift-current

responsivity in specific materials [6,21–23]. The results are generally in good agreement with experimental measurements, proving the predictive power of the *ab initio* approach.

The first-principles evaluation of the shift current (and of other nonlinear optical responses) is technically challenging, due to the intricate form of the matrix elements involved [15–18]. Two basic approaches have been devised. One is to express those matrix elements as an infinite sum over intermediate virtual states [15,16,18]. In practice this requires calculating a large number of unoccupied bands, to minimize truncation errors [21,23]. Alternatively, the matrix elements can be recast in terms of derivatives with respect to the crystal momentum \mathbf{k} of the initial and final band states [15–18]. This strategy circumvents the summation over intermediate states, but its practical implementation requires a careful treatment of the derivatives on a finite k -point grid in order to retain gauge invariance and handle degeneracies [22]. Finally, it has been found that the shift current tends to converge slowly with respect to the number of k points used for the Brillouin zone (BZ) integration [23]. All these factors render the shift current more challenging and expensive to calculate than the ordinary linear optical conductivity.

In this work, we develop an accurate and efficient *ab initio* scheme for calculating the shift current and related nonlinear optical responses in the independent-particle approximation. The proposed methodology, based on localized Wannier functions [24], is closely related to the Wannier interpolation method of calculating the Berry curvature and the intrinsic anomalous Hall conductivity [25]. In essence, it consists of evaluating the matrix elements by $k \cdot p$ perturbation theory within the subspace of wannierized bands. This strategy inherits the practical advantages of the sum-over-states approach in the complete space of Bloch eigenstates, but without introducing truncations errors. In addition, it has a very low computational cost thanks to the compact basis set. We will

comment on the relation between our methodology and a recent proposal with similar characteristics [26].

Our Wannier-interpolation scheme distinguishes itself in two aspects. First, it provides a physically transparent connection to tight-binding approaches [7]. This is achieved by adopting a phase convention for the Bloch sums that includes the Wannier centers in the phase factors, such that the resulting expressions cleanly separate into two parts: an “internal” part that only depends on the Hamiltonian matrix elements and Wannier centers (the only ingredients in a typical tight-binding calculation), and an “external” part containing the off-diagonal position matrix elements. We find that the latter can give a sizable contribution to the shift current; moreover, its inclusion removes an artificial symmetry of the shift-current matrix elements in two-band tight-binding models [7]. These findings highlight the importance of carefully embedding the tight-binding model in real space—via the position matrix elements—when calculating the shift current. The other salient feature of our formulation is that it is fully gauge invariant. This is in contrast to previous Wannier-based schemes, where a parallel-transport gauge was assumed when calculating the interband matrix elements [25,26].

The paper is organized as follows. In Sec. II we provide some background on the microscopic theory of the shift current. In Sec. III we first review the Wannier-interpolation scheme for calculating the energy bands and the interband dipole matrix elements; the same interpolation approach is then applied to the generalized derivative of the interband dipole matrix, completing the list of ingredients needed for evaluating the shift current. The technical details of our electronic-structure and Wannier-function calculations are described in Sec. IV, and the resulting shift-current spectra of GaAs and monolayer GeS are presented and discussed in Sec. V. We provide some concluding remarks in Sec. VI, and leave additional technical discussions to the appendices.

II. PRELIMINARIES

A. Definitions and background

Our starting point is the formalism of Sipe and Shkrebtiĭ for calculating second-order interband optical responses of bulk crystals within the independent-particle approximation [18]. The basic ingredients are the interband dipole matrix, and its “generalized derivative” with respect to the crystal momentum \mathbf{k} . They are given by

$$r_{knm}^a = (1 - \delta_{nm})A_{knm}^a \quad (1)$$

and

$$r_{knm}^{a;b} = \partial_b r_{knm}^a - i(A_{knn}^b - A_{kmm}^b)r_{knm}^a \quad (2)$$

respectively, where

$$A_{knm}^a = i \langle u_{kn} | \partial_a u_{km} \rangle \quad (3)$$

is the Berry connection matrix, where $|u_{km}\rangle$ denotes the cell-periodic part of a Bloch eigenstate and ∂_a stands for $\partial/\partial k_a$.

The three equations above define Hermitian matrices in the band indices n and m . Importantly, the first two transform

covariantly under band-diagonal gauge transformations,

$$|u_n\rangle \rightarrow e^{i\beta_n} |u_n\rangle \Rightarrow \begin{cases} r_{nm}^a \rightarrow e^{i(\beta_m - \beta_n)} r_{nm}^a, \\ r_{nm}^{a;b} \rightarrow e^{i(\beta_m - \beta_n)} r_{nm}^{a;b}, \end{cases} \quad (4)$$

where the subscript \mathbf{k} has been dropped for brevity. As a result, the combination

$$I_{mn}^{abc} = r_{mn}^b r_{nm}^{c;a} \quad (5)$$

appearing in Eq. (8) below is gauge invariant.

Consider a monochromatic electric field of the form

$$\mathcal{E}(t) = \mathcal{E}(\omega)e^{-i\omega t} + \mathcal{E}(-\omega)e^{i\omega t}, \quad (6)$$

with $\mathcal{E}(-\omega) = \mathcal{E}^*(\omega)$. Phenomenologically, the dc photocurrent density from the linear BPVE reads [1–3]

$$j^a = 2\sigma^{abc}(0; \omega, -\omega) \text{Re}[\mathcal{E}_b(\omega)\mathcal{E}_c(-\omega)]. \quad (7)$$

The third-rank response tensor is symmetric under $b \leftrightarrow c$, and transforms like the piezoelectric tensor. According to Eqs. (38) and (41) in Ref. [18], the interband (shift-current) part of the response is given by

$$\begin{aligned} \sigma^{abc}(0; \omega, -\omega) = & -\frac{i\pi e^3}{4\hbar^2} \int [d\mathbf{k}] \sum_{n,m} f_{nm} (I_{mn}^{abc} + I_{mn}^{acb}) \\ & \times [\delta(\omega_{mn} - \omega) + \delta(\omega_{nm} - \omega)]. \end{aligned} \quad (8)$$

Here $f_{nm} = f_n - f_m$ and $\hbar\omega_{nm} = E_m - E_n$ are differences between occupation factors and band energies, respectively, and the integral is over the first BZ, with $[d\mathbf{k}] = d^d k / (2\pi)^d$ in d dimensions. Because I_{mn}^{abc} is Hermitian, the right-hand side of Eq. (8) is real. Its transformation properties under inversion and time-reversal symmetry are summarized in Appendix A.

For comparison, we also calculate the joint density of states (JDOS) per crystal cell,

$$D_{\text{joint}}(\omega) = \frac{v_c}{\hbar} \int [d\mathbf{k}] \sum_{n,m} f_{nm} \delta(\omega_{mn} - \omega) \quad (9)$$

(v_c is the cell volume), and the interband contribution to the absorptive (abs) part of the dielectric function [18],

$$\epsilon_{\text{abs}}^{ab}(\omega) = \frac{i\pi e^2}{\hbar} \int [d\mathbf{k}] \sum_{n,m} f_{nm} r_{nm}^a r_{mn}^b \delta(\omega_{mn} - \omega). \quad (10)$$

In nonmagnetic crystals $\epsilon_{\text{abs}}^{ab}$ is purely imaginary and symmetric, and we report values for $\text{Im} \epsilon_r^{ab} = \text{Im} \epsilon_{\text{abs}}^{ab} / \epsilon_0$, the imaginary part of the relative permittivity.

B. Sum rule for the generalized derivative

The matrix elements r_{nm}^a and $r_{nm}^{a;b}$ appearing in Eq. (8) satisfy the identities

$$r_{nm}^a = \frac{v_{nm}^a}{i\omega_{nm}} \quad (m \neq n) \quad (11)$$

and

$$\begin{aligned} r_{nm}^{a;b} = & \frac{i}{\omega_{nm}} \left[\frac{v_{nm}^a \Delta_{nm}^b + v_{nm}^b \Delta_{nm}^a}{\omega_{nm}} - w_{nm}^{ab} \right. \\ & \left. + \sum_{p \neq n,m} \left(\frac{v_{np}^a v_{pm}^b}{\omega_{pm}} - \frac{v_{np}^b v_{pm}^a}{\omega_{np}} \right) \right] \quad (m \neq n), \end{aligned} \quad (12)$$

where

$$v_{nm}^a = \frac{1}{\hbar} \langle u_n | \partial_a \hat{H} | u_m \rangle, \quad (13a)$$

$$\Delta_{nm}^a = \partial_a \omega_{nm} = v_{nn}^a - v_{mm}^a, \quad (13b)$$

$$w_{nm}^{ab} = \frac{1}{\hbar} \langle u_n | \partial_{ab}^2 \hat{H} | u_m \rangle. \quad (13c)$$

Equation (11) can be obtained by differentiating the identity $\langle u_n | \hat{H} | u_m \rangle = E_n \delta_{nm}$ with respect to k_a for $m \neq n$. Differentiating once more with respect to k_b and inserting a complete set of states yields the sum rule in Eq. (12) [7,18]. For Hamiltonians of the form $\hat{H}_k = (\hat{\mathbf{p}} + \hbar \mathbf{k})^2 / 2m_e + V(\mathbf{r})$, the term w_{nm}^{ab} therein has no off-diagonal components and does not contribute to the sum rule. That term should however be included in tight-binding calculations [7], and in first-principles calculations with nonlocal pseudopotentials [26].

Equation (12) has been used in *ab initio* calculations of the shift current [21,23], with a truncated summation over intermediate states $p \neq n, m$. An exact (truncation-free) expression for $r_{nm}^{a;b}$ that only requires summing over a *finite* number of wannierized bands, Eq. (36) below, constitutes a central result of the present work.

III. WANNIER INTERPOLATION SCHEME

The needed quantities for calculating the shift-current response from Eq. (8) are the energy eigenvalues, and the matrix elements r_{nm}^a and $r_{nm}^{a;b}$ defined by Eqs. (1) and (2). In this section we describe how to evaluate each of them in a Wannier-function basis.

Consider a set of M well-localized Wannier functions per cell $w_j(\mathbf{r} - \mathbf{R}) = \langle \mathbf{r} | \mathbf{R} j \rangle$ spanning the initial and final states involved in interband absorption processes up to some desired frequency ω . (In practice we shall construct them by post-processing a first-principles calculation, using the method of maximally localized Wannier functions [27,28].) Starting from these orbitals, we define a set of Blochlike basis states as

$$|u_{kj}^{(W)}\rangle = \sum_{\mathbf{R}} e^{-i\mathbf{k} \cdot (\hat{\mathbf{r}} - \mathbf{R} - \boldsymbol{\tau}_j)} | \mathbf{R} j \rangle, \quad (14)$$

where the superscript (W) stands for ‘‘Wannier gauge’’ [25]. Note that at variance with Ref. [25], we have chosen to include the Wannier center

$$\boldsymbol{\tau}_j = \langle \mathbf{0} j | \hat{\mathbf{r}} | \mathbf{0} j \rangle \quad (15)$$

in the phase factor of Eq. (14). This phase convention, often used in tight-binding calculations, is the most natural one for expressing the Berry connection and related geometric quantities in reciprocal space [29].

A. Energy eigenvalues

The matrix elements of the first-principles Hamiltonian $\hat{H}_k = e^{-i\mathbf{k} \cdot \hat{\mathbf{r}}} \hat{H} e^{i\mathbf{k} \cdot \hat{\mathbf{r}}}$ between the Blochlike states (14) read

$$\begin{aligned} H_{kij}^{(W)} &= \langle u_{ki}^{(W)} | \hat{H}_k | u_{kj}^{(W)} \rangle \\ &= \sum_{\mathbf{R}} e^{i\mathbf{k} \cdot (\mathbf{R} + \boldsymbol{\tau}_j - \boldsymbol{\tau}_i)} \langle \mathbf{0} i | \hat{H} | \mathbf{R} j \rangle. \end{aligned} \quad (16)$$

Diagonalization of this $M \times M$ matrix yields the Wannier-interpolated energy eigenvalues,

$$(U_k^\dagger H_k^{(W)} U_k)_{nm} = E_{kn} \delta_{nm}, \quad (17)$$

where U_k is the unitary matrix taking from the Wannier gauge to the Hamiltonian gauge. This Slater-Koster type of interpolation, with the Wannier functions acting as an orthogonal tight-binding basis, has been shown in practice to provide a smooth k -space interpolation of the *ab initio* eigenvalues. (With disentangled Wannier functions, the interpolation is faithful only within the so-called ‘‘inner’’ or ‘‘frozen’’ energy window [28].)

B. Berry connection and interband dipole

The same interpolation strategy can be applied to other k -dependent quantities. In particular, the Hamiltonian-gauge Bloch states

$$|u_{kn}\rangle = \sum_{j=1}^M |u_{kj}^{(W)}\rangle U_{kjn} \quad (18)$$

interpolate the *ab initio* Bloch eigenstates, allowing us to treat wave-function-derived quantities.

As a first example, consider the Berry connection matrix defined by Eq. (3). Inserting the above expression for $|u_{kn}\rangle$ in that equation yields [25]

$$A_{nm}^a = \mathbb{A}_{nm}^a + \mathfrak{a}_{nm}^a, \quad (19a)$$

$$\mathbb{A}_{nm}^a = i(U^\dagger \partial_a U)_{nm}, \quad (19b)$$

$$\mathfrak{a}_{nm}^a = (U^\dagger A_a^{(W)} U)_{nm}, \quad (19c)$$

where $A_a^{(W)}$ in Eq. (19c) denotes a Cartesian component of the Berry connection matrix in the Wannier gauge,

$$\begin{aligned} A_{kij}^{(W)} &= i \langle u_{ki}^{(W)} | \partial_k u_{kj}^{(W)} \rangle \\ &= \sum_{\mathbf{R}} e^{i\mathbf{k} \cdot (\mathbf{R} + \boldsymbol{\tau}_j - \boldsymbol{\tau}_i)} \langle \mathbf{0} i | \hat{\mathbf{r}} - \boldsymbol{\tau}_j | \mathbf{R} j \rangle. \end{aligned} \quad (20)$$

The term \mathbb{A}_{nm}^a in Eq. (19) carries the interpretation of a Berry connection for the eigenvectors of $H^{(W)}$ (the column vectors of U). Introducing the notation $\|u_n\rangle$ for those vectors,¹ Eq. (19b) becomes $\mathbb{A}_{nm}^a = i \langle u_n | \partial_a u_m \rangle$. This is the ‘‘internal’’ Berry connection for the tight-binding model defined by Eq. (16) in terms of the Hamiltonian matrix elements and Wannier centers.

The extra term \mathfrak{a}_{nm}^a in Eq. (19) arises from off-diagonal matrix elements of the position operator in the Wannier basis, as can be seen by inspecting the matrix element in Eq. (20) together with Eq. (15). In tight-binding formulations, it is customary to postulate a diagonal representation for $\hat{\mathbf{r}}$ [29–33],

$$\langle \mathbf{0} i | \hat{\mathbf{r}} | \mathbf{R} j \rangle \doteq \boldsymbol{\tau}_i \delta_{\mathbf{R}, \mathbf{0}} \delta_{ji}, \quad (21)$$

¹When the Wannier centers are included in the phase factors of the Bloch sums as in Eq. (14), the eigenvectors of $H^{(W)}$ can be thought of as tight-binding analogs of the *cell-periodic* Bloch states, hence the notation $\|u_n\rangle$. The fact that Berry-phase-type quantities are defined in terms of the cell-periodic Bloch states is the reason why that phase convention is the most natural one for dealing with such quantities in tight binding [29].

where we have introduced the symbol “ \doteq ” to denote equalities that only hold only within this “diagonal tight-binding approximation” (diagonal TBA). Thus, \mathfrak{a}_{nm}^a is the part of the Berry connection matrix A_{nm}^a that is discarded when making the diagonal TBA, and we will refer to it as the “external” part.

For the interband dipole matrix of Eq. (1) we get

$$r_{nm}^a = \begin{cases} r_{nm}^a + \mathfrak{a}_{nm}^a, & \text{if } m \neq n, \\ 0, & \text{if } m = n, \end{cases} \quad (22)$$

where

$$r_{nm}^a = (1 - \delta_{nm})\mathbb{A}_{nm}^a = \begin{cases} \frac{v_{nm}^a}{i\omega_{nm}}, & \text{if } m \neq n, \\ 0, & \text{if } m = n, \end{cases} \quad (23a)$$

$$v_{nm}^a = \frac{1}{\hbar}[U^\dagger(\partial_a H^{(W)})U]_{nm}, \quad (23b)$$

with $\partial_a H^{(W)}$ obtained by differentiating the right-hand side of Eq. (16). Equation (23a) is the “internal” counterpart of Eq. (11) for r_{nm}^a . It can be derived in a similar manner, by differentiating Eq. (17) with $m \neq n$.

C. Generalized derivative of the interband dipole

The energy eigenvalues and interband dipole matrix elements r_{nm}^a are the only ingredients entering Eq. (10) for the dielectric function, which has been previously evaluated by Wannier interpolation [34]. Equation (8) for the shift current contains in addition the generalized derivative $r_{nm}^{a;b}$, and in the following we describe how to evaluate it within the same framework.

1. Useful definitions and identities

Our strategy will be to evaluate Eq. (2) for $r_{nm}^{a;b}$ starting from Eqs. (19) and (22) for A_{nm}^a and r_{nm}^a , respectively. Inspection of those equations reveals that we need to differentiate with respect to k_b the matrices v_{nm}^a and \mathfrak{a}_{nm}^a . Noting that both of them are of the form

$$\bar{\mathcal{O}} = U^\dagger \mathcal{O}^{(W)} U \quad (24)$$

and using the identity

$$\partial_b U = -iU\mathbb{A}^b, \quad (25)$$

we find

$$\partial_b \bar{\mathcal{O}} = U^\dagger(\partial_b \mathcal{O}^{(W)})U + i[\mathbb{A}^b, \bar{\mathcal{O}}]. \quad (26)$$

Writing \mathbb{A}_{nm}^b in the commutator as $\delta_{nm}\mathbb{A}_{nn}^b + r_{nm}^b$ and then expanding $[\mathbb{A}^b, \bar{\mathcal{O}}]$ as a sum over states yields

$$\begin{aligned} \partial_b(\bar{\mathcal{O}})_{nm} &= [U^\dagger(\partial_b \mathcal{O}^{(W)})U]_{nm} - i(\bar{\mathcal{O}}_{nn} - \bar{\mathcal{O}}_{mm})r_{nm}^b \\ &+ i \sum_{p \neq n, m}^M (r_{np}^b \bar{\mathcal{O}}_{pm} - \bar{\mathcal{O}}_{np} r_{pm}^b) \\ &+ i(\mathbb{A}_{nn}^b - \mathbb{A}_{mm}^b)\bar{\mathcal{O}}_{nm}, \end{aligned} \quad (27)$$

where the contribution from intermediate states $p \neq n, m$ has been separated out.

We find it convenient to define an “internal generalized derivative” of the matrix $\bar{\mathcal{O}}$ in analogy with Eq. (2),

$$(\bar{\mathcal{O}})_{nm}^{a;b} = \partial_b(\bar{\mathcal{O}})_{nm} - i(\mathbb{A}_{nn}^b - \mathbb{A}_{mm}^b)\bar{\mathcal{O}}_{nm}. \quad (28)$$

Note that this is equal to the sum of the first three terms in Eq. (27). Before proceeding, let us also define the following internal quantities in analogy with Eq. (13),

$$\mathbb{b}_{nm}^{ab} = [U^\dagger(\partial_b A_a^{(W)})U]_{nm}, \quad (29a)$$

$$\Delta_{nm}^a = v_{nn}^a - v_{mm}^a, \quad (29b)$$

$$\mathbb{w}_{nm}^{ab} = \frac{1}{\hbar}[U^\dagger(\partial_{ab}^2 H^{(W)})U]_{nm}. \quad (29c)$$

2. Derivation

We begin by differentiating the term r_{nm}^a in Eq. (22) for r_{nm}^a . From Eq. (23a) we get

$$\partial_b r_{nm}^a = \frac{i}{\omega_{nm}^2} v_{nm}^a \Delta_{nm}^b - \frac{i}{\omega_{nm}} \partial_b v_{nm}^a \quad (m \neq n). \quad (30)$$

Evaluating $\partial_b v_{nm}^a$ with the help of Eq. (27) and expressing the result in the form of Eq. (28),

$$\partial_b r_{nm}^a = r_{nm}^{a;b} + i(\mathbb{A}_{nn}^b - \mathbb{A}_{mm}^b)r_{nm}^a \quad (m \neq n), \quad (31)$$

we find

$$\begin{aligned} r_{nm}^{a;b} &= \frac{i}{\omega_{nm}} \left[\frac{v_{nm}^a \Delta_{nm}^b + v_{nm}^b \Delta_{nm}^a}{\omega_{nm}} - \mathbb{w}_{nm}^{ab} \right. \\ &\left. + \sum_{p \neq n, m}^M \left(\frac{v_{np}^a v_{pm}^b}{\omega_{pm}} - \frac{v_{np}^b v_{pm}^a}{\omega_{np}} \right) \right] \quad (m \neq n). \end{aligned} \quad (32)$$

This is the internal counterpart of the sum rule (12), written in terms of the tight-binding eigenvectors, eigenvalues, and Hamiltonian, instead of the *ab initio* ones.

The same procedure can be used to differentiate the term \mathfrak{a}_{nm}^a in Eq. (22), given by Eq. (19c). The result is

$$\partial_b \mathfrak{a}_{nm}^a = \mathfrak{a}_{nm}^{a;b} + i(\mathbb{A}_{nn}^b - \mathbb{A}_{mm}^b)\mathfrak{a}_{nm}^a, \quad (33)$$

where

$$\begin{aligned} \mathfrak{a}_{nm}^{a;b} &= \mathbb{b}_{nm}^{ab} - (\mathfrak{a}_{nn}^a - \mathfrak{a}_{mm}^a) \frac{v_{nm}^b}{\omega_{nm}} \\ &+ \sum_{p \neq n, m}^M \left(\frac{v_{np}^b \mathfrak{a}_{pm}^a}{\omega_{np}} - \frac{\mathfrak{a}_{np}^a v_{pm}^b}{\omega_{pm}} \right) \quad (m \neq n). \end{aligned} \quad (34)$$

Adding $\partial_b r_{nm}^a$ and $\partial_b \mathfrak{a}_{nm}^a$ from Eqs. (31) and (33) to form $\partial_b r_{nm}^{a;b}$, and then subtracting the amount $i(\mathbb{A}_{nn}^b - \mathbb{A}_{mm}^b)r_{nm}^a$ in the form

$$i(\mathbb{A}_{nn}^b + \mathfrak{a}_{nn}^b - \mathbb{A}_{mm}^b - \mathfrak{a}_{mm}^b)(r_{nm}^a + \mathfrak{a}_{nm}^a) \quad (35)$$

to obtain $r_{nm}^{a;b}$ as per Eq. (2), we arrive at

$$\begin{aligned} r_{nm}^{a;b} &= r_{nm}^{a;b} + \mathfrak{a}_{nm}^{a;b} - (\mathfrak{a}_{nn}^b - \mathfrak{a}_{mm}^b) \frac{v_{nm}^a}{\omega_{nm}} \\ &- i(\mathfrak{a}_{nn}^b - \mathfrak{a}_{mm}^b)\mathfrak{a}_{nm}^a \quad (m \neq n). \end{aligned} \quad (36)$$

This expression for the generalized derivative in the Wannier representation is a central result of the present work. An alternative expression that is equally valid was obtained in

Ref. [26], and the precise relation between the two formulations is established in Appendix B.

D. Discussion

1. Summary of the interpolation algorithm

To summarize, the response tensor $\sigma^{abc}(0; \omega, -\omega)$ is given by Eq. (8) in terms of the energy eigenvalues and of the matrix elements I_{mn}^{abc} defined by Eq. (5). At each \mathbf{k} , the former are interpolated using Eq. (17), and the latter using Eqs. (22) and (36) for r_{nm}^a and $r_{nm}^{a;b}$, respectively. These equations depend on a small number of ingredients: the matrices $H^{(W)}$ [Eq. (16)] and $A^{(W)}$ [Eq. (20)], their first and second mixed derivatives with respect to k_a and k_b , and the unitary matrix U that diagonalizes $H^{(W)}$. The needed real-space matrix elements, $\langle 0n | \hat{H} | Rm \rangle$ and $\langle 0n | \hat{r} | Rm \rangle$, can be evaluated as described in Ref. [25].

2. Independence of the Berry connection matrix from the choice of phase convention for the Bloch sums

It is well known that the tight-binding expression for an operator depends on the phase convention used for the Bloch sums [35]. Let us discuss how this plays out for the Berry connection matrix (similar remarks apply to the interband dipole matrix and its generalized derivative).

The phase convention we have adopted in this work is that of Eq. (14). The other commonly used convention is to drop τ_j from that equation [29,35], in which case the Berry connection matrix is still given by Eq. (19) but τ_i and τ_j should be removed from Eqs. (16) and (20). As a result, the term \mathbb{A}_{nm}^a in Eq. (19) becomes a function of the Hamiltonian matrix elements only and not of the Wannier centers, whose contributions to the Berry connection are absorbed by \mathbb{a}_{nm}^a . The total Berry connection A_{nm}^a remains the same as before, but the term \mathbb{a}_{nm}^a is now nonzero under the diagonal TBA of Eq. (21).

3. Gauge covariance of the generalized derivative

Although Eq. (2) for $r_{nm}^{a;b}$ is gauge covariant in the sense of Eq. (4), its individual terms are not, leading to numerical difficulties. Instead, the individual terms in the Wannier-based expression (36) for $r_{nm}^{a;b}$ transform covariantly under band-diagonal gauge transformations. As a result, its numerical implementation is very robust.

Contrary to Ref. [26], we did not impose the parallel-transport condition $\mathbb{A}_{nn}^b = 0$ in our derivation of a Wannier-based expression for $r_{nm}^{a;b}$. The gauge-dependent quantities \mathbb{A}_{nn}^b appear in intermediate steps of our derivation, only to drop out in the final step leading to Eq. (36). (A parallel-transport gauge was also assumed in Ref. [25] when deriving a Wannier-based expression for the Berry curvature, and in Appendix C we indicate how to remove that unnecessary assumption.)

4. Generalized derivative versus the effective-mass sum rule: The role of position matrix elements

As remarked in Sec. IIB, Eq. (12) for $r_{nm}^{a;b}$ follows from differentiating the identity $\langle u_n | \hat{H} | u_m \rangle = E_n \delta_{nm}$ once with respect to k_a and once with respect to k_b , for $m \neq n$. Doing so for $m = n$ yields the effective-mass sum rule.

For tight-binding models with a finite number of bands, the effective-mass sum rule can be formulated exactly. The modified sum-rule expression, which only depends on the Hamiltonian matrix elements, includes an intraband term \mathbb{w}_{nm}^{ab} given by Eq. (29c) [30,34,36].

The effect of the basis truncation on the calculation of nonlinear optical responses has been the subject of several recent investigations [7,37,38]. In particular, it was suggested in Ref. [7] that Eq. (32) for $\mathbb{r}_{nm}^{a;b}$, which includes an interband term \mathbb{w}_{nm}^{ab} , is the correct expression for $r_{nm}^{a;b}$ in tight-binding models. In fact, that expression only accounts for part of the wave-function dependence of $r_{nm}^{a;b}$, via the diagonal position matrix elements. The full expression, Eq. (36), has additional terms that depend on the off-diagonal position matrix elements. Those should be included in order to completely describe the wave-function dependence, and to render the result independent of the choice of Wannier basis orbitals [32].

In the diagonal TBA of Eq. (21), Eqs. (22) and (36) for r_{nm}^a and $r_{nm}^{a;b}$ reduce to their internal terms, $r_{nm}^a \doteq \mathbb{r}_{nm}^a$ and $r_{nm}^{a;b} \doteq \mathbb{r}_{nm}^{a;b}$. In this approximation the shift current only depends on the Hamiltonian matrix elements and on the Wannier centers, and a strong dependence on the latter was found in Ref. [7]. As we will see in Sec. V (and also noted in Ref. [26]), the additional contributions from off-diagonal position matrix elements can modify appreciably the calculated shift-current spectrum.

5. The two-band limit

The shift-current response of two-band tight-binding models has been considered in Refs. [7,19]. In that limit the three-band terms in Eq. (36) (those containing intermediate states) vanish identically, and $r_{nm}^{a;b}$ is completely specified by the two-band terms, which pick up the missing contributions (the importance of the \mathbb{w}_{nm}^{ab} term in this regard was emphasized in Ref. [7]). It appears to have gone unnoticed that the diagonal TBA introduces a qualitative error for two-band models, as we now discuss.

In the diagonal TBA, Eq. (36) for a two-band model reduces to the first two terms in Eq. (32),

$$r_{nm}^{a;b} \doteq \frac{i}{\omega_{nm}} \left[\frac{\mathbb{v}_{nm}^a \Delta_{nm}^b + \mathbb{v}_{nm}^b \Delta_{nm}^a}{\omega_{nm}} - \mathbb{w}_{nm}^{ab} \right]. \quad (37)$$

This expression is symmetric under $a \leftrightarrow b$, and when used in Eq. (5) for I_{mn}^{abc} it renders Eq. (8) for $\sigma^{abc}(0; \omega, -\omega)$ totally symmetric, irrespective of crystal symmetry. This unphysical behavior is not an artifact of two-band models, but of the diagonal TBA applied to such models. The shift current arising from the photoexcitation of carriers between the two bands can be calculated exactly, without adding more bands to the model, by including the additional two-band terms in Eq. (36) associated with off-diagonal position matrix elements. These considerations appear relevant to the ongoing discussion on the shift-current response of Weyl semimetals [11–13].

IV. COMPUTATIONAL DETAILS

In this section we describe the various steps of the calculations that we have carried out for two test systems, bulk GaAs and single-layer GeS. In a first step, we performed density-functional theory calculations using the QUANTUM

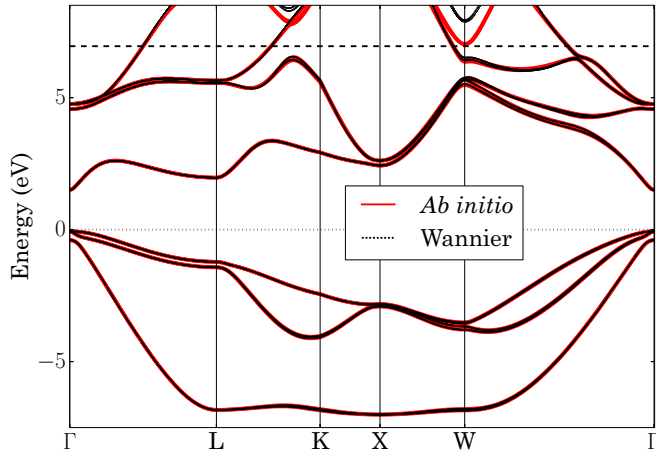


FIG. 1. The *ab initio* and Wannier-interpolated energy bands of GaAs, including a scissors correction of 1.15 eV (energies are measured from the valence-band maximum). The horizontal dashed line at 6.9 eV denotes the upper limit of the inner energy window used in the disentangling step of the Wannier construction procedure.

ESPRESSO code package [39]. The core-valence interaction was treated by means of fully relativistic projector-augmented-wave pseudopotentials (taken from the QUANTUM ESPRESSO website) that had been generated with the Perdew-Burke-Ernzerhof exchange-correlation functional [40], and the energy cutoff for the plane-wave basis expansion was set at 60 Ry. Maximally localized Wannier functions were then constructed in a post-processing step, using the WANNIER90 code package [41]. Finally, the shift-current spectrum [Eq. (8)], the JDOS [Eq. (9)], and the dielectric function [Eq. (10)] were calculated in the Wannier basis as described in Sec. III.

In the case of zinc-blende GaAs, the self-consistent calculation was carried out on a $10 \times 10 \times 10$ k -point mesh, using the experimental lattice constant of $a = 10.68a_0$. Starting from the converged self-consistent Kohn-Sham potential, the 24 lowest bands and Bloch wave functions were then calculated on the same mesh. Finally, a set of 16 disentangled Wannier functions spanning the eight valence bands and the eight low-lying conduction bands were constructed using s and p atom-centered orbitals as trial orbitals. The Wannier-interpolated energy bands are shown in Fig. 1 together with the *ab initio* bands (including in both cases a “scissors correction”). The agreement between the two is excellent inside the inner energy window [28], which spans the energy range from the bottom of the figure up to the dashed horizontal line.

The calculations for monolayer GeS were done in a slab geometry, with a supercell of length 15 \AA along the non-periodic direction and a $1 \times 12 \times 12$ k -point mesh for both the self-consistent and for the band structure calculation. The parameters for the structure with an in-plane polar distortion were taken from Table II in the Supplemental Material of Ref. [23]. Starting from a manifold of 46 bands, we constructed 32 disentangled Wannier functions spanning the 20 highest valence bands and the 12 lowest conduction bands. For the initial projections, we again chose s and p trial orbitals centered on each atom. The *ab initio* and Wannier-interpolated energy bands are shown in Fig. 2.

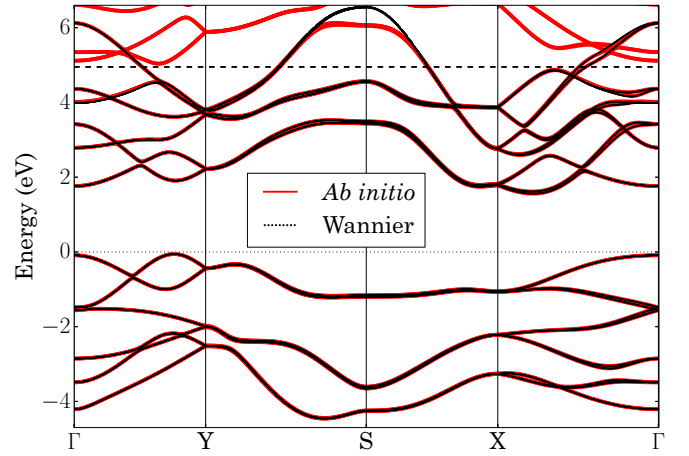


FIG. 2. The *ab initio* and Wannier-interpolated energy bands of monolayer GeS. The horizontal dashed line denotes the upper limit of the inner energy window.

To obtain well-converged shift-current spectra, we used dense k -point interpolation meshes of $100 \times 100 \times 100$ for GaAs and $1 \times 1000 \times 1000$ for GeS. In the case of GaAs, we employed an adaptive scheme [34] for choosing the width of the broadened delta functions in Eq. (8). For GeS we used a fixed width of 0.02 eV, as it was found to handle better the strong Van Hove singularities characteristic of two-dimensional (2D) systems.

In the sum-over-states expression for $\sigma^{abc}(0; \omega, -\omega)$, the energy denominators involving intermediate states should be interpreted as principal values [15]. In our formalism such denominators appear in Eqs. (32) and (34), and in practice we make the replacement

$$\frac{1}{\omega_{np}} \rightarrow \frac{\omega_{np}}{\omega_{np}^2 + (\eta/\hbar)^2}, \quad (38)$$

and similarly for $1/\omega_{pm}$. Such a regularization procedure is needed to avoid numerical problems caused by near degeneracies. Following Ref. [21], we choose η in a range where the calculated spectrum remains stable. In the calculations reported below, we have used $\eta = 0.04$ eV for both GaAs and GeS.

As mentioned earlier, a scissors correction was applied to the calculated band structure of GaAs in Fig. 1, in order to cure the underestimation of the gap. The conduction bands were rigidly shifted by 1.15 eV and the spectral quantities plotted in Fig. 3 were modified accordingly as described below, facilitating comparison with Ref. [21] where a scissors correction was also applied.

It is clear from Eq. (9) that the scissors correction leads to a rigid shift of the JDOS. Although less obvious, the shift-current spectrum [Eq. (8)] and the dielectric function [Eq. (10)] also undergo rigid shifts. The reason is that Eqs. (8) and (10) do not contain any frequency prefactors, and the matrix elements therein are intrinsic properties of the Bloch eigenstates [see Eqs. (1) and (2)], which are unaffected by the scissors correction (only the eigenvalues change). The eigenvalues do appear in Eqs. (11) and (12) that are used in practice to evaluate the optical matrix elements, but a careful

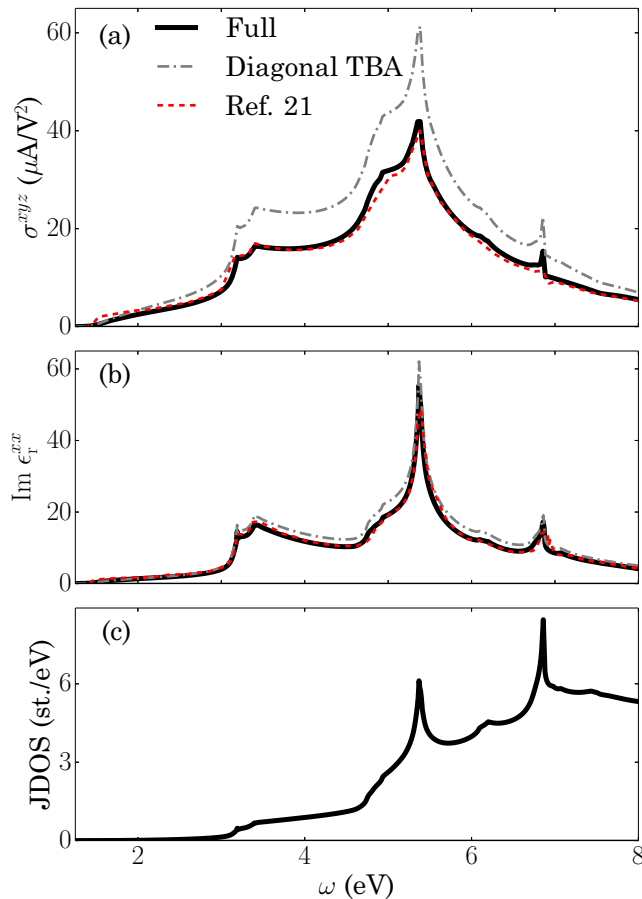


FIG. 3. (a) Shift-current spectrum, (b) imaginary part of the dielectric function, and (c) joint density of states of GaAs, calculated by Wannier interpolation including a scissors correction. “Diagonal TBA” denotes spectra calculated making the diagonal tight-binding approximation of Eq. (21) for the optical matrix elements. Data adapted from Ref. [21] are also shown.

analysis reveals that those equations remain invariant under a scissors correction [42].

V. RESULTS

A. Bulk GaAs

The zinc-blende semiconductor GaAs was the first piezoelectric crystal whose shift-current spectrum was evaluated using modern band structure methods. The original calculation [18] suffered from a computational error, and a corrected spectrum was reported later [21]. Given the existence of this benchmark calculation, we have chosen GaAs as the first test case for our implementation.

Figure 3(a) shows the calculated $\sigma^{xyz}(0; \omega, -\omega)$, which is equal to $\sigma^{abc}(0; \omega, -\omega)$ for any permutation abc of xyz , and all other components vanish by symmetry [18]. The imaginary part of the dielectric function is shown in panel (b) of the same figure, and the JDOS in panel (c). For comparison, we have included in panels (a) and (b) the spectra calculated in Ref. [21].

The dielectric function and the shift-current spectrum share similar peak structures, inherited from the JDOS. The level of agreement with Ref. [21] is excellent for $\text{Im } \epsilon_r^{xx}(\omega)$ and also

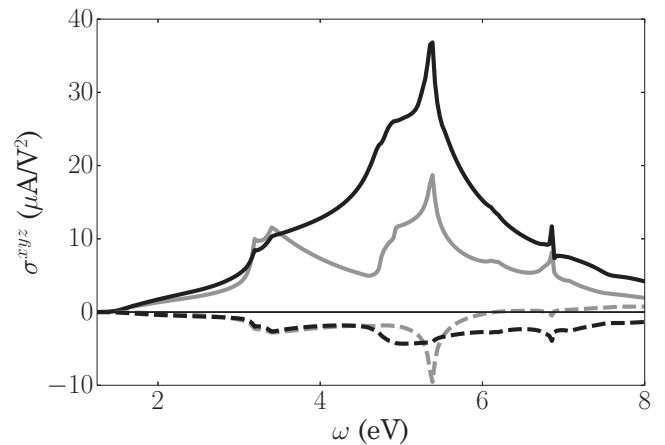


FIG. 4. Decomposition of the shift-current spectrum of GaAs shown in Fig. 3(a) into “internal” (solid lines) and “external” (dashed lines) terms on one hand, and into “three-band” (black lines) and “two-band” (gray lines) terms on the other.

very good for $\sigma^{xyz}(0; \omega, -\omega)$, with only minor deviations. The presence of small discrepancies is not surprising, given that the shift current is rather sensitive to the wave functions [15–17] and that the two calculations differ on several technical aspects. For example, we use pseudopotentials instead of an all-electron method, and a generalized gradient approximation for the exchange-correlation potential instead of the local-density approximation. The BZ integration methods are also different, and the spin-orbit contribution to the velocity matrix elements was not included in Ref. [21].

The dash-dotted gray lines in panels (a) and (b) of Fig. 3 show the spectra calculated in the diagonal TBA of Eq. (21). While in the case of $\text{Im } \epsilon_r^{xx}(\omega)$ the changes are quite small, they are more significant for $\sigma^{abc}(0; \omega, -\omega)$. This reflects the strong wave-function dependence of the shift current, encoded not only in the Wannier centers [7] but also in the off-diagonal position matrix elements $\langle \mathbf{0}n | \hat{r} | \mathbf{R}m \rangle$. Those matrix elements are usually discarded in tight-binding calculations, but they should be included to fully embed the tight-binding model in real space. The sensitivity of the shift current to those matrix elements can be understood from the charge-transfer nature of the photoexcitation process in piezoelectric crystals [15,16,21].

It is instructive to break down the shift-current spectrum calculated by Wannier interpolation into different types of contribution. Inserting Eqs. (22) and (36) for r_{nm}^a and $r_{nm}^{a;b}$ into Eq. (5) for I_{nm}^{abc} generates a number of terms. Each can be classified as “external” or “internal” depending on whether or not it contains off-diagonal position matrix elements: the term $r_{mn}^b r_{nm}^{c;a}$ is internal, and all others are external. In addition, we classify each term as “two-band” or “three-band” depending on whether it only involves states n and m , or intermediate states p as well. This gives a total of four types of terms, whose contributions to the shift current are shown in Fig. 4.

The dominant contribution comes from internal three-band terms, which by themselves provide a reasonable approximation to the full spectrum shown in Fig. 3(a). They are followed by the internal two-band terms, while the two external terms are somewhat smaller. Over most of the spectral range, the external

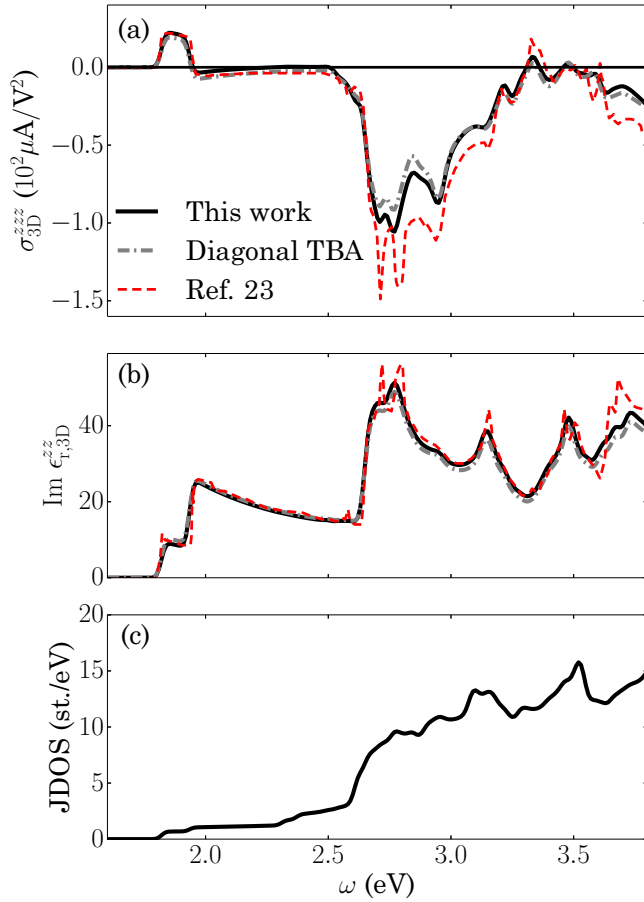


FIG. 5. (a), (b), and (c) show the calculated zzz component of the shift-current spectrum, the zz component of the dielectric function, and the joint density of states of monolayer GeS, respectively. The first two have been rescaled according to Eq. (39) to become 3D-like quantities.

terms have the opposite sign compared to the internal ones. Since the diagonal TBA amounts to discarding the external terms, that explains why the dash-dotted gray line in Fig. 3(a) overestimates the magnitude of the full spectrum given by the solid black line. We emphasize that the decomposition of the shift-current spectrum in Fig. 4 depends on the choice of Wannier functions.

B. Monolayer GeS

GeS is a member of the group-IV monochalcogenides, which in bulk form are centrosymmetric, but become polar—and hence piezoelectric—when synthesized as a single layer. The point group of monolayer GeS is $mm2$, which allows for seven tensorial components of $\sigma^{abc}(0; \omega, -\omega)$ to be nonzero [23]. With the same choice of coordinate axis as in Fig. 1 of Ref. [23] (the in-plane directions are \hat{y} and \hat{z} , with the spontaneous polarization along \hat{z}), the nonzero components are zxx , zyy , zzz , $yyz = yzy$, and $xxz = xzx$.

The zzz component of the shift-current spectrum is displayed in Fig. 5(a). Following Ref. [23], we report a 3D-like response obtained assuming an active single-layer thickness of 2.56 Å. This is achieved by rescaling the calculated response

of the slab of thickness 15 Å as follows,

$$\sigma_{3D}^{zzz} = \frac{15}{2.56} \sigma_{\text{slab}}^{zzz}. \quad (39)$$

In Figs. 5(b) and 5(c) we plot the dielectric function [also rescaled according to Eq. (39)] and the JDOS. As in the case of GaAs, the main peak structures of the optical spectra in panels (a) and (b) are inherited from the JDOS. The diagonal TBA (dash-dotted gray lines) changes the calculated spectra only slightly, consistent with what is found in Ref. [26] for monolayer WS_2 .

Our calculated spectra in Fig. 5 are in reasonable agreement with those reported in Ref. [23] (dashed red lines), including on the positions of the main peaks and on the sign change of the shift current taking place at around 2 eV. However, the agreement is not as good as that seen in Fig. 3 for GaAs. This may be due in part to some differences in computational details between the two calculations, namely the use of different k -point meshes and BZ integration methods: we have sampled the BZ on a uniform mesh of 10^6 k points, while in Ref. [23] a more sophisticated tetrahedron method was used for the integration, but with far fewer k points (4900). There is however another source of disagreement, which was not present in Fig. 3: the approximate treatment in Ref. [23] of the optical matrix elements within the nonlocal pseudopotential framework. This source of error is discussed further in Appendix D.

C. Analysis of computational time

Here we compare the computational requirements of our numerical scheme with a direct calculation of the shift-current spectrum without Wannier interpolation (e.g., using the method outlined in Appendix D). The spectrum is evaluated by discretizing the BZ integral in Eq. (8) over a mesh containing N k points, and we wish to see how the computational times of the two approaches scale with N .

For that purpose, let us define the following time scales per k point: t_w and t_d are the times to evaluate the integrand in Eq. (8) by Wannier interpolation and using the direct method, respectively, and t_{nscf} is the time to carry out a non-self-consistent calculation to obtain the *ab initio* Bloch eigenfunctions and energy eigenvalues. Further, we define T_{scf} as the total time needed to carry out the self-consistent ground-state calculation, and T_{wf} as the total time needed to construct the Wannier functions on a grid of M k points. The total time of a Wannier-based calculation of the shift current is then

$$T_{\text{scf}} + M t_{\text{nscf}} + T_{\text{wf}} + N t_w, \quad (40)$$

while the total time of a direct calculation is

$$T_{\text{scf}} + N(t_{\text{nscf}} + t_d) \approx T_{\text{scf}} + N t_{\text{nscf}}, \quad (41)$$

where we used $t_d \ll t_{\text{nscf}}$.

Let us take as a concrete example a calculation for monolayer GeS done on a single Intel Xeon E5-2680 processor with 24 cores running at 2.5 GHz. For the choice of parameters indicated in Sec. IV we find $t_w \simeq 21$ ms, $t_{\text{nscf}} \simeq 46$ s, $T_{\text{scf}} \simeq 0.5$ hours, and $T_{\text{wf}} \simeq 1$ hour. In Fig. 6 we plot as a function of N the total times obtained from Eqs. (40) and (41), for $M = 12^2$. The use of Wannier interpolation is already quite advantageous

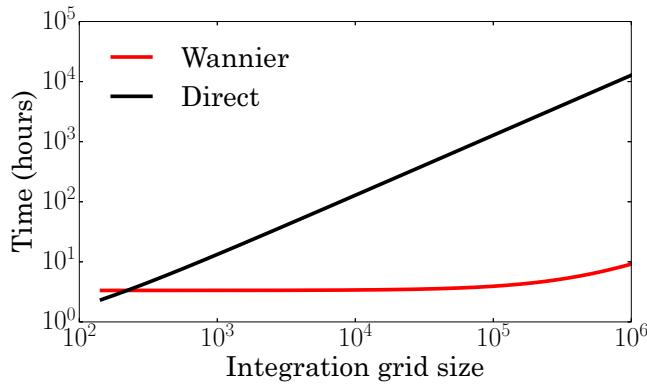


FIG. 6. Time estimates for calculating the shift-current spectrum of monolayer GeS on a single processor with and without Wannier interpolation [Eqs. (40) and (41), respectively], as a function of the size N of the BZ integration grid.

for $N \sim 500$, and the speedup increases very rapidly with N . If a dense k -point sampling with $N \sim 10^6$ is required, the speedup reaches three orders of magnitude. (The absolute times reported in Fig. 6 can be reduced by parallelizing the loop over the $N k$ points, which is trivial to do both with and without Wannier interpolation.)

VI. SUMMARY

In summary, we have described and validated a Wannier-interpolation scheme for calculating the shift-current spectrum of piezoelectric crystals, starting from the output of a conventional electronic-structure calculation. The method is both accurate and efficient; this is achieved by using a truncated Wannier-function basis, but without incurring in truncation errors when evaluating the optical matrix elements. The same approach can be applied to other nonlinear optical responses, such as second-harmonic generation, that involve the same matrix elements [18,26].

Our work was motivated in part by the growing interest in the calculation of nonlinear optical properties of novel materials such as Weyl semimetals and 2D materials. We hope that the proposed methodology, and its implementation in the WANNIER90 code package, will help turn such calculations into a fairly routine task.

When describing the formalism, we tried to emphasize the notion that Wannier functions provide an essentially exact (in some chosen energy range) tight-binding parametrization of the *ab initio* electronic structure. Thus, we chose our notation and conventions so as to facilitate comparison with the expressions for nonlinear optical responses found in the tight-binding literature. Our numerical results suggest that it should be possible to systematically improve the tight-binding description of such responses by including off-diagonal position matrix elements as additional model parameters. In Ref. [31], an attempt was made along those lines to improve the tight-binding parametrization of semiconductors for the calculation of Born effective charges, but with limited success. Clearly more work is needed in this direction, and the shift current, with its strong sensitivity to the wave functions, is particularly well-suited for such investigations.

ACKNOWLEDGMENTS

The authors gratefully acknowledge stimulating discussions with Fernando de Juan, Jianpeng Liu, Cheol-Hwan Park, and David Vanderbilt. They also thank Cheol-Hwan Park for bringing Ref. [26] to their attention, and Chong Wang for elucidating the relation between the formalism of Ref. [26] and the one presented here. The work was supported by Grant No. FIS2016-77188-P from the Spanish Ministerio de Economía y Competitividad, and by Elkartek Grant No. KK-2016/00025. Computing time was granted by the JARA-HPC Vergabegremium and provided on the JARA-HPC Partition part of the supercomputer JURECA at Forschungszentrum Jülich.

APPENDIX A: SYMMETRY CONSIDERATIONS

As mentioned in the Introduction, the shift current vanishes in centrosymmetric crystals. To verify that Eq. (8) behaves correctly in that limit, note that the presence of inversion symmetry implies the relations

$$E_n(-\mathbf{k}) = E_n(\mathbf{k}), \quad (\text{A1a})$$

$$I_{mn}^{abc}(-\mathbf{k}) = -I_{mn}^{abc}(\mathbf{k}). \quad (\text{A1b})$$

Hence \mathbf{k} and $-\mathbf{k}$ give equal and opposite contributions to the BZ integral in Eq. (8), leading to $\sigma^{abc}(0; \omega, -\omega) = 0$.

The shift current has been mostly studied in acentric crystals without magnetic order. The presence of time-reversal symmetry in such systems implies

$$E_n(-\mathbf{k}) = E_n(\mathbf{k}), \quad (\text{A2a})$$

$$I_{mn}^{abc}(-\mathbf{k}) = -[I_{mn}^{abc}(\mathbf{k})]^*. \quad (\text{A2b})$$

The points \mathbf{k} and $-\mathbf{k}$ now give equal contributions to the BZ integral, and Eq. (8) reduces to

$$\begin{aligned} \sigma^{abc}(0; \omega, -\omega) = & -\frac{i\pi e^3}{2\hbar^2} \int [d\mathbf{k}] \sum_{n,m} f_{nm} (I_{mn}^{abc} + I_{mn}^{acb}) \\ & \times \delta(\omega_{mn} - \omega), \end{aligned} \quad (\text{A3})$$

which is Eq. (57) in Ref. [18]. For $b = c$, this form remains equivalent to Eq. (8) even without time-reversal symmetry.

APPENDIX B: COMPARISON WITH REFERENCE [26]

In Ref. [26], a similar Wannier-interpolation scheme for calculating the shift current was proposed independently. The expression given in that work for the generalized derivative in the Wannier basis is however different from Eq. (36). In this Appendix, we show that the two formulations are in fact consistent with one another. Below their Eq. (7), the authors of Ref. [26] write

$$\begin{aligned} \partial_b A^a = & (\partial_b U^\dagger) A_a^{(W)} U + U^\dagger (\partial_b A_a^{(W)}) U \\ & + U^\dagger A_a^{(W)} \partial_b U + i(\partial_b U^\dagger) \partial_a U + iU^\dagger \partial_{ab}^2 U, \end{aligned} \quad (\text{B1})$$

which follows from differentiating Eq. (19). The last term can be expressed in terms of $D^a = U^\dagger \partial_a U = -i\mathbb{A}^a$ as

$$iU^\dagger \partial_{ab}^2 U = i\partial_b D^a + iD^b D^a. \quad (\text{B2})$$

The non-Hermitian term $iD^b D^a$ cancels the fourth term in Eq. (B1), leaving an expression for $\partial_b A^a$ that is correctly Hermitian, term by term. Let us now evaluate the term $\partial_b D^a$ assuming $D_{nm}^a = 0$ (parallel-transport) [26]. The off-diagonal

matrix elements of the matrix D^a read

$$D_{nm}^a = -\frac{\mathbb{V}_{nm}^a}{\omega_{nm}} \quad (m \neq n), \quad (\text{B3})$$

where \mathbb{V}_{nm}^a was defined in Eq. (23b). Invoking Eq. (26) we find

$$\partial_b D_{nm}^a = -\frac{1}{\omega_{nm}} \left(\mathbb{W}_{nm}^{ab} - \sum_{l \neq m} \frac{\mathbb{V}_{nl}^a \mathbb{V}_{lm}^b}{\omega_{lm}} - \sum_{l \neq n} \frac{\mathbb{V}_{nl}^b \mathbb{V}_{lm}^a}{\omega_{ln}} + \frac{\mathbb{V}_{mm}^b \mathbb{V}_{nm}^a}{\omega_{nm}} - \frac{\mathbb{V}_{nn}^b \mathbb{V}_{nm}^a}{\omega_{nm}} \right) \quad (m \neq n), \quad (\text{B4})$$

with \mathbb{W}_{nm}^{ab} given by Eq. (29c). Substituting the term $\partial_b r_{nm}^a$ in Eq. (2) by Eq. (B1) combined with Eqs. (B2) and (B4), Eq. (36) for $r_{nm}^{a:b}$ is eventually recovered (after using $\mathbb{A}^b = \mathbb{r}^b$, which holds in a parallel-transport gauge). We can now proceed to compare with Ref. [26]. Combining Eqs. (B2)–(B4) we obtain

$$(U^\dagger \partial_{ab}^2 U)_{nm} = -\frac{1}{\omega_{nm}} \left(\mathbb{W}_{nm}^{ab} - \sum_{l \neq m} \frac{\mathbb{V}_{nl}^a \mathbb{V}_{lm}^b}{\omega_{lm}} - \sum_{l \neq n} \frac{\mathbb{V}_{nl}^b \mathbb{V}_{lm}^a}{\omega_{ln}} + \frac{\mathbb{V}_{mm}^b \mathbb{V}_{nm}^a}{\omega_{nm}} - \frac{\mathbb{V}_{nn}^b \mathbb{V}_{nm}^a}{\omega_{nm}} \right) + \sum_{l \neq n,m} \frac{\mathbb{V}_{nl}^b \mathbb{V}_{lm}^a}{\omega_{nl} \omega_{lm}}. \quad (\text{B5})$$

The first two terms in this equation agree with those in Eq. (8) of Ref. [26], and in the following we show that the remaining terms in both equations can also be brought into agreement. Dropping the first two terms of Eq. (B5) and using $\omega_{nm}/(\omega_{nl}\omega_{lm}) = 1/\omega_{nl} - 1/\omega_{lm}$ in the last term, we find²

$$\begin{aligned} & -\frac{1}{\omega_{nm}} \left(-\sum_{l \neq n} \frac{\mathbb{V}_{nl}^b \mathbb{V}_{lm}^a}{\omega_{ln}} + \frac{\mathbb{V}_{mm}^b \mathbb{V}_{nm}^a}{\omega_{nm}} - \frac{\mathbb{V}_{nn}^b \mathbb{V}_{nm}^a}{\omega_{nm}} - \sum_{l \neq n,m} \frac{\mathbb{V}_{nl}^b \mathbb{V}_{lm}^a}{\omega_{nl}} - \sum_{l \neq n,m} \frac{\mathbb{V}_{nl}^b \mathbb{V}_{lm}^a}{\omega_{lm}} \right) \\ & = -\frac{1}{\omega_{nm}} \left(-\sum_{l \neq n,m} \frac{\mathbb{V}_{nl}^b \mathbb{V}_{lm}^a}{\omega_{lm}} + \frac{\mathbb{V}_{nm}^b \mathbb{V}_{nm}^a}{\omega_{nm}} + \frac{\mathbb{V}_{mm}^b \mathbb{V}_{nm}^a}{\omega_{nm}} - \frac{\mathbb{V}_{nn}^b \mathbb{V}_{nm}^a}{\omega_{nm}} \right) \\ & = -\frac{1}{\omega_{nm}} \left(-\sum_{l \neq m} \frac{\mathbb{V}_{nl}^b \mathbb{V}_{lm}^a}{\omega_{lm}} + \frac{\mathbb{V}_{nm}^b \mathbb{V}_{nm}^a}{\omega_{nm}} + \frac{\mathbb{V}_{mm}^b \mathbb{V}_{nm}^a}{\omega_{nm}} \right), \end{aligned} \quad (\text{B6})$$

which is indeed identical to the last three terms in Eq. (8) of Ref. [26]. It is worth mentioning that in this formulation the Hermiticity of $r_{nm}^{a:b}$ is only satisfied globally, not term by term as in the case of Eq. (36).

APPENDIX C: BERRY CURVATURE IN THE WANNIER BASIS: REMOVAL OF THE PARALLEL-TRANSPORT ASSUMPTION

In Ref. [25], around Eqs. (23) and (24), a parallel-transport gauge was imposed on the U matrices while evaluating the Berry curvature in a Wannier basis. Should one then enforce the parallel-transport condition when choosing those matrices at neighboring k points? This is in fact not necessary, as we now show.

The Berry curvature of band n is given by the $m = n$ element of the matrix

$$\Omega_{knn}^{ab} = i \langle \partial_a u_{kn} | \partial_b u_{kn} \rangle - i \langle \partial_b u_{kn} | \partial_a u_{kn} \rangle. \quad (\text{C1})$$

Using

$$|\partial_a u_n\rangle = \sum_j |\partial_a u_j^{(w)}\rangle U_{jn} - i \sum_m |u_m\rangle \mathbb{A}_{mn}^a, \quad (\text{C2})$$

which follows from Eqs. (18) and (25), we find

$$\Omega^{ab} = \bar{\Omega}_{ab} + i[\mathbb{A}^a, \bar{\mathbb{A}}_b] - i[\mathbb{A}^b, \bar{\mathbb{A}}_a] + i[\mathbb{A}^a, \mathbb{A}^b]. \quad (\text{C3})$$

This is Eq. (27) of Ref. [25], in a slightly different notation. Recall from Eq. (19b) that \mathbb{A}^a is the Berry connection for the U matrices; instead of imposing the parallel-transport condition $\mathbb{A}_{nn}^a = 0$ as done in Ref. [25], we let \mathbb{A}_{nn}^a be nonzero and write $\mathbb{A}_{nm}^a = \delta_{nm} \mathbb{A}_{nn}^a + \mathbb{r}_{nm}^a$, in accordance with Eq. (23a). The first commutator in Eq. (C3), for example, becomes

$$i[\mathbb{r}^a, \bar{\mathbb{A}}_b]_{nm} - i\bar{\mathbb{A}}_{b,nn}(\mathbb{A}_{mm}^a - \mathbb{A}_{nn}^a). \quad (\text{C4})$$

Since the second term vanishes for $m = n$, we conclude that the Berry curvature, given by the band-diagonal entries in Eq. (C3), is insensitive to the value of the gauge-dependent quantity \mathbb{A}_{nn}^a . This is consistent with the fact that the Berry curvature is gauge invariant.

²Equation (B6) was obtained by Chong Wang, commenting on an earlier version of the paper (private communication).

APPENDIX D: APPROXIMATE TREATMENT OF THE OPTICAL MATRIX ELEMENTS WITH NONLOCAL PSEUDOPOTENTIALS

In some previous *ab initio* calculations of the shift current [21,23], the velocity operator was approximated as

$$\hat{v} = \frac{\hat{p}}{m_e} = -\frac{i\hbar}{m_e} \nabla_r. \tag{D1}$$

The interband velocity matrix elements v_{nm} in the Bloch basis were then inserted into Eqs. (11) and (12) (dropping the term w_{nm}^{ab} in the latter) to obtain the interband dipole matrix r_{nm}^a and its generalized derivative $r_{nm}^{a,b}$.

When using either an all-electron method (as in the GaAs calculation of Ref. [21]) or local pseudopotentials, the above procedure is exact, at least when spin-orbit coupling is neglected.³ However, modern pseudopotential calculations employ nonlocal pseudopotentials, for which that procedure introduces some errors: the velocity operator is not simply given by Eq. (D1) [44,45], and as a result the term w_{nm}^{ab} in Eq. (12) for $r_{nm}^{a,b}$ becomes nonzero (see Appendix B in Ref. [26]).

In this Appendix we perform additional calculations for single-layer GeS employing the same computational setup as used in Ref. [23] (ABINIT code [46] with Hartwigsen-Goedecker-Hutter pseudopotentials [47]), in order to estimate the errors arising from the use of the approximate procedure outlined above.

As a first step, we switched off by hand the nonlocal terms in the pseudopotentials. For a given k -point sampling and delta-function smearing, the resulting spectra $\text{Im } \epsilon_r^{zz}(\omega)$ and $\sigma^{zz}(0; \omega, -\omega)$ (not shown) were found to be in perfect agreement with those calculated by Wannier interpolation using the same local pseudopotentials. This provided a strong numerical check of our Wannier interpolation scheme, which does not depend on whether an all-electron or a pseudopotential method has been used, or on whether the pseudopotentials are local or nonlocal.

We then redid both calculations using the full nonlocal pseudopotentials. The results obtained by sampling the 2D BZ on a relatively coarse 70×70 grid with a fairly large delta-

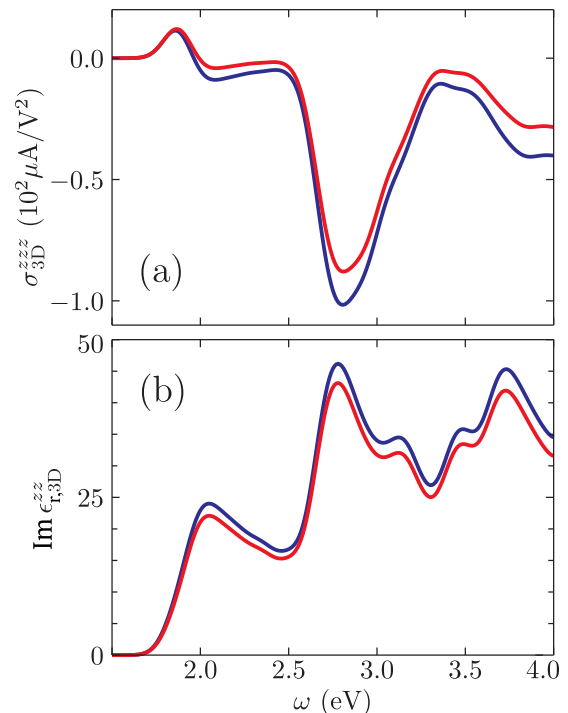


FIG. 7. (a) Shift-current spectrum, and (b) dielectric function of single-layer GeS calculated using an exact (red) and an approximate (blue) treatment of the optical matrix elements within the nonlocal-pseudopotential approach. The red curve was obtained with Wannier interpolation, while for the blue curve the optical matrix elements were calculated directly in the plane-wave basis using Eq. (D1).

function broadening of 0.1 eV are shown in Fig. 7 (as a result of the coarse k -point sampling and of the large broadening, the spectral features are broadened compared to Fig. 5). There are clear differences between the spectra calculated in the manner of Ref. [23], and those obtained using the Wannier interpolation scheme: the positions of the peaks are the same, but their heights are somewhat different, as expected from a small change in the matrix elements. Given the perfect agreement that had been found with local pseudopotentials, these differences must arise exclusively from the approximate treatment of the optical matrix elements in the approach of Ref. [23] combined with nonlocal pseudopotentials. Since the level of disagreement seen in Fig. 7 is comparable to that seen in Figs. 5(a) and 5(b), it seems plausible that there the discrepancies may also arise in part from these small errors in the matrix elements.

³The spin-orbit interaction gives an additional contribution to the velocity operator [43]. That contribution is typically small and can be safely neglected, as done in Ref. [21]. In our formulation, that contribution is automatically included.

[1] V. M. Fridkin, Bulk photovoltaic effect in noncentrosymmetric crystals, *Crystallogr. Rep.* **46**, 654 (2001).
 [2] B. I. Sturman and V. M. Fridkin, *The Photovoltaic and Photo-refractive Effects in Noncentrosymmetric Materials* (Gordon and Breach, Philadelphia, 1992).
 [3] E. L. Ivchenko and G. E. Pikus, *Superlattices and Other Heterostructures* (Springer, Berlin, 1997), Chap. 10.5.

[4] W. T. H. Koch, R. Munser, W. Ruppel, and P. Würfel, Anomalous photovoltage in BaTiO₃, *Ferroelectrics* **13**, 305 (1976).
 [5] K. T. Butler, J. M. Frost, and A. Walsh, Ferroelectric materials for solar energy conversion: Photoferroics revisited, *Energy Environ. Sci.* **8**, 838 (2015).
 [6] L. Z. Tan, F. Zheng, S. M. Young, F. Wang, S. Liu, and A. M. Rappe, Shift current bulk photovoltaic effect in polar materials—

- hybrid and oxide perovskites and beyond, *npj Comput. Mater.* **2**, 16026 (2016).
- [7] A. M. Cook, B. M. Fregoso, F. de Juan, S. Coh, and J. E. Moore, Design principles for shift current photovoltaics, *Nat. Commun.* **8**, 14176 (2017).
- [8] L. Z. Tan and A. M. Rappe, Enhancement of the Bulk Photovoltaic Effect in Topological Insulators, *Phys. Rev. Lett.* **116**, 237402 (2016).
- [9] L. Braun, G. Mussler, A. Hruban, M. Konczykowski, T. Schumann, M. Wolf, Ma. Müntzenberg, L. Perfetti, and T. Kampfrath, Ultrafast photocurrents at the surface of the three-dimensional topological insulator Bi_2Se_3 , *Nat. Commun.* **7**, 13259 (2016).
- [10] D. A. Bas, R. A. Muniz, S. Babakiray, D. Lederman, J. E. Sipe, and A. D. Bristow, Identification of photocurrents in topological insulators, *Opt. Express* **24**, 23583 (2016).
- [11] G. B. Osterhoudt, L. K. Diebel, X. Yang, J. Stanco, X. Huang, B. Shen, N. Ni, P. Moll, Y. Ran, and K. S. Burch, Colossal photovoltaic effect driven by the singular Berry curvature in a Weyl semimetal, [arXiv:1712.04951](https://arxiv.org/abs/1712.04951).
- [12] X. Yang, K. Burch, and Y. Ran, Divergent bulk photovoltaic effect in Weyl semimetals, [arXiv:1712.09363](https://arxiv.org/abs/1712.09363).
- [13] Y. Zhang, H. Ishizuka, J. van den Brink, C. Felser, B. Yan, and N. Nagaosa, Photogalvanic effect in Weyl semimetals from first principles, [arXiv:1803.00562](https://arxiv.org/abs/1803.00562).
- [14] N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, and N. P. Ong, Anomalous Hall effect, *Rev. Mod. Phys.* **82**, 1539 (2010).
- [15] R. von Baltz and W. Kraut, Theory of the bulk photovoltaic effect in pure crystals, *Phys. Rev. B* **23**, 5590 (1981).
- [16] V. I. Belinicher, E. L. Ivchenko, and B. I. Sturman, Kinetic theory of the displacement photovoltaic effect in piezoelectrics, *Zh. Eksp. Teor. Fiz.* **83**, 649 (1982) [*Sov. Phys. JETP* **56**, 359 (1982)].
- [17] N. Kristoffel, R. von Baltz, and D. Hornung, On the intrinsic bulk photovoltaic effect: Performing the sum over intermediate states, *Z. Phys. B* **47**, 293 (1982).
- [18] J. E. Sipe and A. I. Shkrebtii, Second-order optical response in semiconductors, *Phys. Rev. B* **61**, 5337 (2000).
- [19] H. Presting and R. Von Baltz, Bulk photovoltaic effect in a ferroelectric crystal: A model calculation, *Phys. Status Solidi (b)* **112**, 559 (1982).
- [20] B. M. Fregoso, T. Morimoto, and J. E. Moore, Quantitative relationship between polarization differences and the zone-averaged shift photocurrent, *Phys. Rev. B* **96**, 075421 (2017).
- [21] F. Nastos and J. E. Sipe, Optical rectification and shift currents in GaAs and GaP response: Below and above the band gap, *Phys. Rev. B* **74**, 035201 (2006).
- [22] S. M. Young and A. M. Rappe, First Principles Calculation of the Shift Current Photovoltaic Effect in Ferroelectrics, *Phys. Rev. Lett.* **109**, 116601 (2012).
- [23] T. Rangel, B. M. Fregoso, B. S. Mendoza, T. Morimoto, J. E. Moore, and J. B. Neaton, Large Bulk Photovoltaic Effect and Spontaneous Polarization of Single-Layer Monochalcogenides, *Phys. Rev. Lett.* **119**, 067402 (2017).
- [24] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Maximally localized Wannier functions: Theory and applications, *Rev. Mod. Phys.* **84**, 1419 (2012).
- [25] X. Wang, J. R. Yates, I. Souza, and D. Vanderbilt, *Ab initio* calculation of the anomalous Hall conductivity by Wannier interpolation, *Phys. Rev. B* **74**, 195118 (2006).
- [26] C. Wang, X. Liu, L. Kang, B.-L. Gu, Y. Xu, and W. Duan, First-principles calculation of nonlinear optical responses by Wannier interpolation, *Phys. Rev. B* **96**, 115147 (2017).
- [27] N. Marzari and D. Vanderbilt, Maximally localized generalized Wannier functions for composite energy bands, *Phys. Rev. B* **56**, 12847 (1997).
- [28] I. Souza, N. Marzari, and D. Vanderbilt, Maximally localized Wannier functions for entangled energy bands, *Phys. Rev. B* **65**, 035109 (2001).
- [29] T. Yusufaly, D. Vanderbilt, and S. Coh, Tight-Binding Formalism in the Context of the PythTB Package, <http://physics.rutgers.edu/pythtb/formalism.html>.
- [30] M. Graf and P. Vogl, Electromagnetic fields and dielectric response in empirical tight-binding theory, *Phys. Rev. B* **51**, 4940 (1995).
- [31] J. Benetto and D. Vanderbilt, Semiconductor effective charges from tight-binding theory, *Phys. Rev. B* **53**, 15417 (1996).
- [32] I. Paul and G. Kotliar, Thermal transport for many-body tight-binding models, *Phys. Rev. B* **67**, 115131 (2003).
- [33] T. B. Boykin, M. Luisier, and G. Klimeck, Current density and continuity in discretized models, *Eur. J. Phys.* **31**, 1077 (2010).
- [34] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, Spectral and Fermi surface properties from Wannier interpolation, *Phys. Rev. B* **75**, 195121 (2007).
- [35] C. Bena and G. Montambaux, Remarks on the tight-binding model of graphene, *New J. Phys.* **11**, 095003 (2009).
- [36] T. B. Boykin, Incorporation of incompleteness in the $k \cdot p$ perturbation theory, *Phys. Rev. B* **52**, 16317 (1995).
- [37] G. B. Ventura, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, Gauge covariances and nonlinear optical responses, *Phys. Rev. B* **96**, 035431 (2017).
- [38] A. Taghizadeh, F. Hipolito, and T. G. Pedersen, Linear and nonlinear optical response of crystals using length and velocity gauges: Effect of basis truncation, *Phys. Rev. B* **96**, 195413 (2017).
- [39] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [40] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [41] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, wannier90: A tool for obtaining maximally-localised Wannier functions, *Comput. Phys. Commun.* **178**, 685 (2008).
- [42] F. Nastos, B. Olejnik, K. Schwarz, and J. E. Sipe, Scissors implementation within length-gauge formulations of the frequency-dependent nonlinear optical response of semiconductors, *Phys. Rev. B* **72**, 045223 (2005).
- [43] E. I. Blount, Formalisms of band theory, *Solid State Phys.* **13**, 305 (1962).
- [44] S. Baroni and R. Resta, *Ab initio* calculation of the macroscopic dielectric constant in silicon, *Phys. Rev. B* **33**, 7017 (1986).
- [45] M. S. Hybertsen and S. G. Louie, *Ab initio* static dielectric matrices from the density-functional approach. I. Formulation

- and application to semiconductors and insulators, [Phys. Rev. B](#) **35**, 5585 (1987).
- [46] X. Gonze, B. Amadon, P.-M. Anglade, J.-M. Beuken, F. Bottin, P. Boulanger, F. Bruneval, D. Caliste, R. Caracas, M. Côté, T. Deutsch, L. Genovese, Ph. Ghosez, M. Giantomassi, S. Goedecker, D. R. Hamann, P. Hermet, F. Jollet, G. Jomard, S. Leroux, M. Mancini, S. Mazevet, M. J. T. Oliveira, G. Onida, Y. Pouillon, T. Rangel, G.-M. Rignanese, D. Sangalli, R. Shaltaf, M. Torrent, M. J. Verstraete, G. Zerah, and J. W. Zwanziger, ABINIT: First-principles approach to material and nanosystem properties, [Comput. Phys. Commun.](#) **180**, 2582 (2009).
- [47] C. Hartwigsen, S. Goedecker, and J. Hutter, Relativistic separable dual-space Gaussian pseudopotentials from H to Rn, [Phys. Rev. B](#) **58**, 3641 (1998).