





**Basic formulation and first-principles implementation of nonlinear magneto-optical effects**Haowei Chen <sup>1</sup>, Meng Ye <sup>1,\*</sup>, Nianlong Zou <sup>1</sup>, Bing-Lin Gu <sup>3</sup>, Yong Xu <sup>1,2,4,5,6,†</sup> and Wenhui Duan <sup>1,2,3,4,5</sup><sup>1</sup>State Key Laboratory of Low Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing 100084, China<sup>2</sup>Tencent Quantum Laboratory, Tencent, Shenzhen, Guangdong 518057, China<sup>3</sup>Institute for Advanced Study, Tsinghua University, Beijing 100084, China<sup>4</sup>Frontier Science Center for Quantum Information, Beijing 100084, China<sup>5</sup>Beijing Academy of Quantum Information Sciences, Beijing 100193, China<sup>6</sup>RIKEN Center for Emergent Matter Science (CEMS), Wako, Saitama 351-0198, Japan

(Received 25 October 2021; revised 27 January 2022; accepted 28 January 2022; published 10 February 2022)

First-principles calculation of nonlinear magneto-optical effects has become an indispensable tool to reveal the geometric and topological nature of electronic states and to understand light-matter interactions. While intriguingly rich physics could emerge in magnetic materials, further methodological developments are required to deal with time-reversal symmetry breaking, due to the degeneracy and gauge problems caused by symmetry and the low-frequency divergence problem in the existing calculation formalism. Here we present a gauge-covariant and low-frequency convergent formalism for the first-principles computation. Remarkably, this formalism generally works for both nonmagnetic and magnetic materials with or without band degeneracy. Reliability and capability of our method are demonstrated by studying example materials (i.e., bilayers of  $\text{MnBi}_2\text{Te}_4$  and  $\text{CrI}_3$ ) and comparing with published results. Moreover, an importance correction term that ensures gauge covariance of degenerate states is derived, whose influence on physical responses is systematically checked. Our method enables computation of nonlinear magneto-optical effects in magnetic materials and paves the way for exploring rich physics created by the interplay of light and magnetism.

DOI: [10.1103/PhysRevB.105.075123](https://doi.org/10.1103/PhysRevB.105.075123)**I. INTRODUCTION**

Optical materials and phenomena play indispensable roles in modern science and technology. Along with the development of laser technology, nonlinear optical materials in which the polarization responds nonlinearly to the electric field of light are discovered and nonlinear optical phenomena such as the bulk photovoltaic effect (BPVE) [1–3] and the second-harmonic generation (SHG) [4] have been widely investigated. The BPVE is the generation of a rectification current driven by light in a single-phase material lacking inversion symmetry and it originates from the shift current mechanism [5]. SHG, which is the generation of photons with twice the frequency of incident photons in noncentrosymmetric materials, is commonly used as a tool for symmetry characterization and laser frequency conversion. Remarkably, the intrinsic connections between optical responses and the quantum geometry in momentum space have been revealed recently, giving rise to anomalous optical responses in topological materials [6–11].

Light-matter interaction in magnetic materials brings up profound physics due to the interplay between magnetism and the electromagnetic wave. In the linear regime, the plane of polarization can be rotated when light is reflected by or transmitted through magnetic materials, which is called

the linear magneto-optical Kerr effect or Faraday effect. In the nonlinear regime, the magnetization-sensitive rectification current has recently been proposed in antiferromagnetic (AFM) bilayers of  $\text{CrI}_3$  and  $\text{MnBi}_2\text{Te}_4$  with parity-time ( $\mathcal{PT}$ ) symmetry [12–14], whose nonmagnetic counterparts are absent. Magnetization-sensitive SHG has long been a tool to probe surface and interface magnetization in metallic materials since 1990s [15–18] and later used to investigate complex magnetic orders in bulk magnetic oxides, such as in  $\text{Cr}_2\text{O}_3$  [19]. Due to the spectral and spatial resolution, SHG is an indispensable tool to detect magnetic symmetries, orders, domains and phase transitions, especially in low dimensions [20]. A recent SHG measurement discovered a gigantic signal from AFM bilayer  $\text{CrI}_3$  where the microscopic origin of the large response remains unclear [21].

Theoretical calculations based on first-principles methods are powerful tools to reveal microscopic correlations between the energy spectrum, the momentum space geometry and optical effects. Moreover, nonlinear magneto-optical effects are especially powerful in studying AFM insulators with  $\mathcal{PT}$  symmetry, where linear magneto-optical effects, such as magneto-optic Kerr and Faraday effects, are absent. The nonlinear magneto-optical effects, e.g., magnetization-sensitive SHG, have been theoretically investigated since 1980s [22–27]. But as the position operator is ill-defined in periodic systems (e.g., bulk crystals), it is approximated based on cluster models which is applicable for finite systems (e.g., molecules) but could be problematic in periodic systems. Although the problem of position operator in periodic

\*mengye@mail.tsinghua.edu.cn

†yongxu@mail.tsinghua.edu.cn

systems has been solved along with modern investigations of the polarization of solids [28,29], several problems still hinder the theoretical investigation of nonlinear magneto-optical effects. First, the conventional formalism of nonlinear magneto-optical effects is ill-defined in the presence of  $\mathcal{PT}$  symmetry due to the Kramers-like degeneracy at each  $k$  point in the band structure. The conventional expressions of nonlinear conductivity and electric susceptibility [12,13,30] only consider nondegenerate bands with an arbitrary phase freedom for eigenstates at each  $k$  point, which is also called the U(1)-gauge freedom. However, due to the Kramers-like degeneracy, formulations under  $\mathcal{PT}$  symmetry should be invariant under an arbitrary  $2 \times 2$  unitary transformation in the subspace spanned by the doubly degenerate eigenstates, corresponding to a U(2)-gauge freedom. Although Watanabe and Yanase [31] have proposed to directly replace the U(1)-covariant derivative by the U(2)-covariant derivative in the presence of  $\mathcal{PT}$  symmetry for magnetization-sensitive shift current conductivity, detailed derivations are missing.

Second, the conventional expression of second-order electric susceptibility including SHG has a low-frequency divergent problem by including leading term factors of  $\omega^{-1}$  and  $\omega^{-2}$  for gapped insulators with the incident light frequency  $\omega$  [30,32,33]. While the low-frequency divergence is frequently observed in physical properties of metallic materials due to the response from the Fermi surface [8], the low-frequency divergence is unphysical in gapped insulators. Even if the expression has a 0/0 form and is not divergent at  $\omega \rightarrow 0$ , the 0/0 form expression can cause sizable numerical instability in computation [30,32,33]. Fortunately, for SHG susceptibility, it has been shown that the divergence can be eliminated in the presence of time-reversal ( $\mathcal{T}$ ) symmetry [30,32,34–36]. However, the divergence could be problematic for magnetic materials. The  $\mathcal{T}$ -symmetry simplified ( $\mathcal{T}$ -smp) expression in which  $\mathcal{T}$  symmetry is only applied to the divergent terms have been implemented in many computational codes [37–43] and reasonable agreement between theoretical predictions and experiments has been reached for nonmagnetic materials [44]. While  $\mathcal{T}$ -smp expressions can still be used to estimate the SHG susceptibility in magnetic materials [45], the divergent terms that are unique to magnetic materials are missing. More importantly, the same divergence appears in the general second-order susceptibility (including sum-frequency generation and difference-frequency generation), which cannot be removed even under  $\mathcal{T}$  symmetry. Therefore, it is imperative to solve the above mentioned problems to facilitate and advance the first-principles studies of nonlinear magneto-optical effects.

In this work, we present general solutions to complete the computational methods for second-order magneto-optical effects in which the physical consequence of  $\mathcal{PT}$  symmetry is derived and investigated, and a full-frequency convergent formalism of second-order susceptibility is proposed. The remainder of this article is organized as follows. In Sec. II, we derive the expressions for second-order magneto-optical responses based on density matrix perturbation method for degenerate bands and a full-frequency convergent formalism for second-order susceptibility including SHG. Section III discusses the symmetry properties of second-order magneto-optical responses under  $\mathcal{T}$  and  $\mathcal{PT}$  operations, including

charge/spin current and susceptibility. Section IV introduces the implementation of our methods in first-principles calculations and computational details. In Sec. V, we investigate magnetization-sensitive conductivity and susceptibility in two prototypical two-dimensional AFM materials with  $\mathcal{PT}$  symmetry, which are bilayer CrI<sub>3</sub> and MnBi<sub>2</sub>Te<sub>4</sub>.

## II. BASIC FORMULATION

### A. General formulas with Kramers-like degeneracy

The conventional formulas for second-order photocurrent and electric polarization only consider nondegenerate bands, and therefore are only U(1)-gauge invariant [13,33]. However, in magnetic materials with  $\mathcal{PT}$  symmetry, Kramers-like degeneracy appears at every  $k$  point and the choice of eigenstates in the doubly degenerate subspace has a gauge freedom determined by a  $2 \times 2$  unitary matrix. As physical results are invariant in spite of an arbitrary unitary rotation in the degenerate subspace, the formulas for physical observables under  $\mathcal{PT}$  symmetry should be U(2)-gauge invariant. Clearly, certain terms are missing in the conventional formula in the presence of  $\mathcal{PT}$  symmetry and in what follows, we will demonstrate that the U(2)-invariant formulas can be retrieved with special treatment at degeneracy [33].

In the length gauge and under long-wavelength approximation, the electromagnetic wave is introduced through an  $\mathbf{E}(t) \cdot \hat{\mathbf{x}}$  term to the unperturbed Hamiltonian  $H_0$  where a monochromatic light at frequency  $\omega$  is written as  $\mathbf{E}(t) = \mathbf{E}(\omega)e^{-i\omega t} + \text{c.c.}$  The Hamiltonian in second quantization form can be found in Appendix A 1. The evaluation of the position operator  $\hat{\mathbf{x}}$  is the key to the calculation of polarization and current responses, and the matrix elements of position operator can be written as [33]

$$\begin{aligned} \langle n_\mu \mathbf{k} | \hat{\mathbf{x}} | m_\nu \mathbf{k}' \rangle &= \delta_{nm} \delta_{\mu\nu} \left[ \delta(\mathbf{k} - \mathbf{k}') \xi_{n_\mu m_\nu}(\mathbf{k}) + i \frac{\partial}{\partial \mathbf{k}} \delta(\mathbf{k} - \mathbf{k}') \right] \\ &\quad + (1 - \delta_{nm} \delta_{\mu\nu}) \xi_{n_\mu m_\nu}(\mathbf{k}) \delta(\mathbf{k} - \mathbf{k}') \\ &= (\hat{\mathbf{x}}_{\text{intra}})_{n_\mu \mathbf{k}, m_\nu \mathbf{k}'} + (\hat{\mathbf{x}}_{\text{inter}})_{n_\mu \mathbf{k}, m_\nu \mathbf{k}'}. \end{aligned} \quad (1)$$

$|n_\mu \mathbf{k}\rangle = e^{i\mathbf{k} \cdot \hat{\mathbf{x}}} |u_{n_\mu}(\mathbf{k})\rangle$  is the eigenstate of  $H_0$ , and  $|u_{n_\mu}(\mathbf{k})\rangle$  is the periodic part of the Bloch function. We introduced two subscripts  $n$  and  $\mu$  for bands, where  $n$  denotes bands with different energy, and  $\mu$  is used in the subspace of degenerate bands.  $\mu$  runs from 1 to 2 under  $\mathcal{PT}$  symmetry, while the following derivations hold for an arbitrary  $\mu$  value. The diagonal and off-diagonal parts of  $\xi_{n_\mu m_\nu}(\mathbf{k}) = i \langle u_{n_\mu}(\mathbf{k}) | \nabla_{\mathbf{k}} | u_{m_\nu}(\mathbf{k}) \rangle$  are the single band and two-band Berry connections  $\mathbf{A}_{n_\mu}(\mathbf{k})$  and  $\mathbf{r}_{n_\mu m_\nu}(\mathbf{k})$  defined as

$$\xi_{n_\mu m_\nu}(\mathbf{k}) \equiv \begin{cases} \mathbf{A}_{n_\mu}(\mathbf{k}) & \text{if } n = m \text{ and } \mu = \nu, \\ \mathbf{r}_{n_\mu m_\nu}(\mathbf{k}) & \text{otherwise.} \end{cases} \quad (2)$$

In the following formulas, we omit the  $\mathbf{k}$  dependence for simplicity. The first (second) term of Eq. (1) is the intra-band (interband) position matrix element in an infinite crystal and the position operator matrix element between degenerate bands ( $n = m$ ,  $\mu \neq \nu$ ) are considered in the interband term. As a result, the polarization can be divided into  $\mathbf{P} = e\mathbf{x} = \mathbf{P}_{\text{inter}} + \mathbf{P}_{\text{intra}}$  in which  $e = -|e|$  is the charge of an electron. Similarly, the current density along  $a$  direction which is

defined as the time derivative of electric polarization is divided into two parts as

$$\begin{aligned} \langle \hat{j}^a(t) \rangle &= \frac{1}{i\hbar} \langle [\hat{P}_{\text{intra}}^a(t), \hat{H}(t)] \rangle + \frac{d}{dt} \langle \hat{P}_{\text{inter}}^a(t) \rangle \\ &= \langle \hat{j}_{\text{intra}}^a(t) \rangle + \frac{d}{dt} \langle \hat{P}_{\text{inter}}^a(t) \rangle, \end{aligned} \quad (3)$$

where  $\hat{H}(t)$  is the total Hamiltonian including the oscillating electric field.

With the density matrix derived with perturbation expansion in Appendices A 1 and A 2, the zeroth- and first-order responses including the finite frequency anomalous conductivity are derived in Appendix A 3. The second-order electric current and polarization with incoming electric field of frequency  $\omega_\beta$  and  $\omega_\gamma$  and the response frequency  $\omega_\Sigma = \omega_\beta + \omega_\gamma$  are

$$\begin{aligned} \langle \hat{j}_{\text{intra}}^a(t) \rangle^{(2)} &= \frac{e^3}{2\hbar^2} \int [d\mathbf{k}] \left\{ E_\beta^b E_\gamma^c e^{-i\omega_\Sigma t} \left[ \frac{f_{nm} \Delta_{mn}^a r_{n_\mu m_\nu}^c r_{m_\nu n_\mu}^b}{\omega_\Sigma (\omega_{mn} - \omega_\beta - i/\tau)} \right. \right. \\ &\quad \left. \left. - \frac{f_{nm} r_{m_\nu n_\mu}^b r_{n_\mu m_\nu}^c}{\omega_{mn} - \omega_\beta - i/\tau} \right] + (b\beta \leftrightarrow c\gamma) \right\} \end{aligned} \quad (4)$$

and

$$\begin{aligned} \langle \hat{P}_{\text{inter}}^a(t) \rangle^{(2)} &= \frac{e^3}{2\hbar^2} \int [d\mathbf{k}] \sum_{n,l,m} \left[ \frac{[\dots] E_\beta^b E_\gamma^c e^{-i\omega_\Sigma t}}{\omega_{ln} - \omega_\Sigma - i/\tau} + (b\beta \leftrightarrow c\gamma) \right], \end{aligned} \quad (5)$$

where the full expression of  $\langle \hat{P}_{\text{inter}}^a(t) \rangle^{(2)}$  can be found in Appendix A 3 and the Einstein summation convention is adopted in all formulas for repeated indices.  $\hbar$  is the reduced Planck's constant and  $[d\mathbf{k}] = d\mathbf{k}/(2\pi)^d$  is the unit volume in reciprocal space of  $d$  dimension. An infinitesimal  $1/\tau$  is introduced to avoid divergence at resonance, which can also be regarded as a relaxation rate.  $a, b, c$  represent Cartesian directions, and  $E_\beta$  is the shorthand notation for  $E(\omega_\beta)$  which represents the amplitude of an electric field with frequency  $\omega_\beta$ .  $(b\beta \leftrightarrow c\gamma)$  is the interchange of  $b\beta$  and  $c\gamma$  indices which reflects the fact that physical responses are irrelevant to the sequence of electric fields  $E_\beta^b$  and  $E_\gamma^c$  in the expression.  $f_{nm}$ ,  $\hbar\omega_{nm} = (E_n - E_m)$  and  $\Delta_{mn} = v_{mm} - v_{nn}$  are the occupation, energy and group velocity differences between the  $n$ th and the  $m$ th bands. In gapped semiconductor at zero temperature,  $f = 1$  for valence band and  $f = 0$  for conduction band.  $r_{n_\mu m_\nu}^b = \partial_{k_a} r_{n_\mu m_\nu}^b - i(A_{n_\mu}^a r_{n_\mu m_\nu}^b - r_{n_\mu m_\nu}^b A_{m_\nu}^a)$  is the U(1)-covariant derivative of  $r_{n_\mu m_\nu}^b$ , where the U(1)-covariant derivative of a nondegenerate eigenstate  $|u_n(\mathbf{k})\rangle$  is  $|u_n(\mathbf{k})\rangle_{;a} = \partial_{k_a} |u_n(\mathbf{k})\rangle + iA_n^a |u_n(\mathbf{k})\rangle = (1 - |u_n(\mathbf{k})\rangle \langle u_n(\mathbf{k})|) \partial_{k_a} |u_n(\mathbf{k})\rangle$ , which represents the projection of  $\partial_{k_a} |u_n(\mathbf{k})\rangle$  onto all states except itself.

It is convenient to use Fourier transformation to convert results into the frequency domain and define the conductivity and susceptibility in the frequency domain as

$$\begin{aligned} j^a(\omega_\Sigma) &= \sigma^{abc}(-\omega_\Sigma; \omega_\beta, \omega_\gamma) E^b(\omega_\beta) E^c(\omega_\gamma), \\ P^a(\omega_\Sigma) &= \chi^{abc}(-\omega_\Sigma; \omega_\beta, \omega_\gamma) E^b(\omega_\beta) E^c(\omega_\gamma). \end{aligned} \quad (6)$$

In the following, we will write  $j_{\text{intra}}^a(\omega)^{(2)}$  instead of  $\langle \hat{j}_{\text{intra}}^a(\omega) \rangle^{(2)}$  as abbreviation to represent the expectation value.

## B. Rectification current from degeneracy

A changing electric polarization gives rise to a current, and the rectification current ( $\omega_\beta = -\omega_\gamma = \pm\omega$ ,  $\omega_\Sigma = 0$ ) is of particular interest. If bands are nondegenerate, which is equivalent to dropping the  $\mu, \nu$  indices, and  $\omega_{ln} \neq 0$  in Eq. (5), then  $\mathbf{P}_{\text{inter}}(\omega_\Sigma = 0)^{(2)}$  is finite, with  $\omega_\Sigma \mathbf{P}_{\text{inter}}(\omega_\Sigma = 0)^{(2)} = 0$ , and  $\mathbf{j}(\omega_\Sigma = 0)^{(2)} = \mathbf{j}_{\text{intra}}(\omega_\Sigma = 0)^{(2)}$  according to Eq. (3), which can reduce to the well-known U(1)-invariant formulas of rectification current densities [13,33]. However, if band degeneracy occurs, e.g.,  $\omega_{ln} = 0$ , then degenerate bands contribute a polarization that is diverging as  $\frac{1}{\omega_\Sigma}$  when  $\omega_\Sigma$  approaches the static limit according to Eq. (5).

The divergent polarization at  $\omega_\Sigma \rightarrow 0$  corresponds to a DC current as

$$\begin{aligned} \lim_{\omega_\Sigma \rightarrow 0} [-i\omega_\Sigma P_{\text{inter}}^a(\omega_\Sigma)^{(2)}] &= \frac{ie^3}{2\hbar^2} \int [d\mathbf{k}] \left[ E_\beta^b E_\gamma^c \lim_{\omega_\Sigma \rightarrow 0} \left( \frac{-\omega_\Sigma}{-\omega_\Sigma - i/\tau} \right) f_{nm} r_{m_\nu n_\mu}^b \right. \\ &\quad \left. \times \frac{r_{n_\mu n_\lambda}^a r_{n_\lambda m_\nu}^c - r_{n_\mu m_\lambda}^c r_{m_\lambda m_\nu}^a}{\omega_{mn} - \omega_\beta - i/\tau} + (b\beta \leftrightarrow c\gamma) \right], \end{aligned} \quad (7)$$

which contains the Berry connection between degenerate bands, e.g.,  $r_{n_\mu n_\lambda}$ . The total rectification current is

$$\begin{aligned} j^a(\omega_\Sigma = 0)^{(2)} &= j_{\text{intra}}^a(\omega_\Sigma = 0)^{(2)} + \lim_{\omega_\Sigma \rightarrow 0} [-i\omega_\Sigma P_{\text{inter}}^a(\omega_\Sigma)^{(2)}] \\ &= \frac{e^3}{2\hbar^2} \int [d\mathbf{k}] \left\{ E_\beta^b E_\gamma^c \left[ \frac{f_{nm} \Delta_{mn}^a r_{n_\mu m_\nu}^c r_{m_\nu n_\mu}^b}{\omega_\Sigma (\omega_{mn} - \omega_\beta - i/\tau)} \right. \right. \\ &\quad \left. \left. - \frac{f_{nm} r_{m_\nu n_\mu}^b D^a[r_{n_\mu m_\nu}^c]}{\omega_{mn} - \omega_\beta - i/\tau} \right] + (b\beta \leftrightarrow c\gamma) \right\}, \end{aligned} \quad (8)$$

where  $D^a[r_{n_\mu m_\nu}^b] = \partial_{k_a} r_{n_\mu m_\nu}^b - i \sum_\lambda (r_{n_\mu n_\lambda}^a r_{n_\lambda m_\nu}^b - r_{n_\mu m_\lambda}^b r_{m_\lambda n_\mu}^a)$  is the U( $N$ )-covariant derivative of  $r_{n_\mu m_\nu}^b$  and  $N$  is the number of degeneracy determined by the index  $\mu$ . Similarly, the U( $N$ )-covariant derivative of degenerate eigenstates just projects  $\partial_{k_a} |u_{n_\mu}(\mathbf{k})\rangle$  out of the  $N$ -dimensional degenerate subspace such that  $D^a |u_{n_\mu}(\mathbf{k})\rangle = \partial_{k_a} |u_{n_\mu}(\mathbf{k})\rangle + iA_{n_\mu}^a |u_{n_\mu}(\mathbf{k})\rangle + i \sum_{\nu \neq \mu} r_{n_\nu n_\mu}^a |u_{n_\nu}(\mathbf{k})\rangle = (1 - \sum_\nu |u_{n_\nu}(\mathbf{k})\rangle \langle u_{n_\nu}(\mathbf{k})|) \partial_{k_a} |u_{n_\mu}(\mathbf{k})\rangle$ . Up to now, we have presented a thorough derivation of the second-order electric current and have revealed the microscopic origin of the U( $N$ )-covariant derivative. In Appendix B, we demonstrated an alternative derivation of the above formulas by including the position operator matrix between degenerate bands into the intraband term in Eq. (1).

In the intermediate region where the double degeneracy protected by  $\mathcal{PT}$  symmetry is weakly broken with the energy splitting  $E_{\text{deg}}$ , the fundamental formalisms in Eqs. (3)–(5) should be used to calculate the rectification current. As  $E_{\text{deg}}$  increases from 0 to finite, the rectification current turn to zero

TABLE I. Transformation rules of basic gauge-invariant quantities under  $\mathcal{P}$ ,  $\mathcal{T}$ , and  $\mathcal{PT}$ , respectively.

	$I_{mn}^{abc}(\mathbf{k})$	$\Omega_{mn}^{ab}(\mathbf{k})$	$g_{mn}^{ab}(\mathbf{k})$	$\Delta_{mn}^a(\mathbf{k})$
$\mathcal{P}$	$-I_{mn}^{abc}(-\mathbf{k})$	$\Omega_{mn}^{ab}(-\mathbf{k})$	$g_{mn}^{ab}(-\mathbf{k})$	$-\Delta_{mn}^a(-\mathbf{k})$
$\mathcal{T}$	$-I_{mn}^{abc}(-\mathbf{k})$	$-\Omega_{mn}^{ab}(-\mathbf{k})$	$g_{mn}^{ab}(-\mathbf{k})$	$-\Delta_{mn}^a(-\mathbf{k})$
$\mathcal{PT}$	$I_{mn}^{abc}(\mathbf{k})$	$-\Omega_{mn}^{ab}(\mathbf{k})$	$g_{mn}^{ab}(\mathbf{k})$	$\Delta_{mn}^a(\mathbf{k})$

which is consistent with the U(1)-invariant result and an AC current peaked at  $E_{\text{deg}}$  appears.

Using the Sokhotski–Plemelj theorem, the denominator can be simplified as  $(\omega_{mn} - \omega - i/\tau)^{-1} = \text{P}(\omega_{mn} - \omega)^{-1} + i\pi\delta(\omega_{mn} - \omega)$ , where P indicates principal part. Therefore, the total rectification current can be divided into four parts as normal injection current (NIC), magnetic injection current (MIC), normal shift current (NSC), and magnetic shift current (MSC):

$$\begin{aligned}
j^a(\omega_\Sigma = 0)^{(2)} &= [\sigma^{abc}(0; \omega, -\omega) + \sigma^{abc}(0; -\omega, \omega)] \text{Re}\{E_\omega^b E_{-\omega}^c\} \\
&\quad + i[\sigma^{abc}(0; \omega, -\omega) - \sigma^{abc}(0; -\omega, \omega)] \text{Im}\{E_\omega^b E_{-\omega}^c\} \\
&= 2[\sigma_{\text{NSC}}^{abc}(\omega) + \tau\eta_{\text{MIC}}^{abc}(\omega)] \text{Re}\{E_\omega^b E_{-\omega}^c\} \\
&\quad + 2i[\sigma_{\text{MSC}}^{abc}(\omega) + \tau\eta_{\text{NIC}}^{abc}(\omega)] \text{Im}\{E_\omega^b E_{-\omega}^c\}, \quad (9)
\end{aligned}$$

with expressions

$$\begin{aligned}
\sigma_{\text{NSC}}^{abc}(\omega) &= \frac{-i\pi e^3}{4\hbar^2} \int [d\mathbf{k}] f_{nm} (I_{mn}^{abc} + I_{mn}^{acb}) \\
&\quad \times [\delta(\omega_{nm} - \omega) + \delta(\omega_{mn} - \omega)], \quad (10)
\end{aligned}$$

$$\begin{aligned}
\sigma_{\text{MSC}}^{abc}(\omega) &= \frac{-i\pi e^3}{4\hbar^2} \int [d\mathbf{k}] f_{nm} (I_{mn}^{abc} - I_{mn}^{acb}) \\
&\quad \times [\delta(\omega_{mn} - \omega) - \delta(\omega_{nm} - \omega)], \quad (11)
\end{aligned}$$

$$\eta_{\text{NIC}}^{abc}(\omega) = -\frac{\pi e^3}{2\hbar^2} \int [d\mathbf{k}] f_{mn} \Delta_{mn}^a (-i\Omega_{mn}^{bc}) \delta(\omega_{nm} - \omega), \quad (12)$$

$$\eta_{\text{MIC}}^{abc}(\omega) = -\frac{\pi e^3}{2\hbar^2} \int [d\mathbf{k}] f_{mn} \Delta_{mn}^a (2g_{mn}^{bc}) \delta(\omega_{nm} - \omega). \quad (13)$$

The integrand  $I_{mn}^{abc} = \sum_{\mu\nu} r_{m_\nu n_\mu}^b D^a[r_{n_\mu m_\nu}^c]$ , the two-band Berry curvature  $\Omega_{mn}^{bc} = -2 \sum_{\mu\nu} \text{Im}\{r_{m_\nu n_\mu}^b r_{n_\mu m_\nu}^c\}$ , the two-band quantum metric  $g_{mn}^{bc} = \sum_{\mu\nu} \text{Re}\{r_{m_\nu n_\mu}^b r_{n_\mu m_\nu}^c\}$  [7,8,46], and the group velocity difference  $\Delta_{mn}^a$  are all gauge invariant and their symmetry properties are listed in Table I. The above results are consistent with literature [13,33], and additionally we generalized the formulas of NSC and MSC to a U(N)-invariant form in the presence of  $N$ -fold degeneracy.

### C. Low-frequency divergent problem of susceptibility

The general formula for second-order polarization can be derived through  $\mathbf{P}(\omega)^{(2)} = \mathbf{j}_{\text{intra}}(\omega)^{(2)}/(-i\omega) + \mathbf{P}_{\text{inter}}(\omega)^{(2)}$  in which Berry connection terms containing degenerate bands are already included in  $\mathbf{P}_{\text{inter}}(\omega)^{(2)}$  in the conventional expression. Therefore, the expressions of polarization and

susceptibility are compatible with degenerate bands and no additional modifications are required to fulfill the gauge invariant requirement.

However, two terms in the susceptibility with leading factors  $\omega_\Sigma^{-1}$  and  $\omega_\Sigma^{-2}$  originated from  $\mathbf{j}_{\text{intra}}^{(2)}$  give rise to a divergent problem in the  $\omega_\Sigma \rightarrow 0$  limit with the divergent terms

$$\begin{aligned}
\chi_{\text{divg}}^{abc}(-\omega_\Sigma; \omega_\beta, \omega_\gamma) &= \frac{-ie^3}{2\hbar^2 \omega_\Sigma} \int [dk] f_{nm} \left[ \left( \frac{r_{mn}^b r_{nm;a}^c}{\omega_{mn} - \omega_\beta} + \frac{r_{mn}^c r_{nm;a}^b}{\omega_{mn} - \omega_\gamma} \right) \right. \\
&\quad \left. - \frac{\Delta_{mn}^a}{\omega_\Sigma} \left( \frac{r_{nm}^c r_{mn}^b}{\omega_{mn} - \omega_\beta} + \frac{r_{nm}^b r_{mn}^c}{\omega_{mn} - \omega_\gamma} \right) \right], \quad (14)
\end{aligned}$$

where the imaginary part  $i/\tau$  in the denominator is not written out explicitly. In the case  $\omega_\beta = -\omega_\gamma$ , the two terms are related to the shift current and injection current discussed in Eqs. (10)–(13). In general, the divergence is unphysical in an insulator with a finite energy gap. Even if  $\chi_{\text{divg}}$  takes a 0/0 form and the divergence can be removed eventually, the expression in Eq. (14) is unfriendly to numerical calculations at low frequency. This problem has been noticed in SHG susceptibility ( $\omega_\beta = \omega_\gamma = \omega$ ,  $\omega_\Sigma = 2\omega$ ) by Sipe and Ghahramani [30] and they discovered that the divergence in SHG can be removed by applying  $\mathcal{T}$  symmetry. However, the simplification does not work in magnetic materials and as a result, the current SHG expressions might be problematic in studying magnetic materials. Furthermore, applying  $\mathcal{T}$  symmetry fails to remove the divergence in the general susceptibility in Eq. (14). Therefore, the second-order susceptibility for arbitrary frequencies still suffers from an unphysical divergent problem even for nonmagnetic materials.

In what follows, we present an alternate expression for the divergent term which is given by

$$\begin{aligned}
\chi_{\text{divg}}^{abc}(-\omega_\Sigma; \omega_\beta, \omega_\gamma) &= \frac{-ie^3}{2\hbar^2} \int [dk] \frac{f_{nm}}{\omega_{mn}} \left[ \left( \frac{\rho_\beta r_{nm}^b r_{nm;a}^c}{\omega_{mn} - \omega_\beta} + \frac{\rho_\gamma r_{nm}^c r_{nm;a}^b}{\omega_{mn} - \omega_\gamma} \right) \right. \\
&\quad \left. - \frac{\Delta_{mn}^a}{\omega_{mn}} \left( \frac{\rho_\beta^2 r_{nm}^b r_{nm}^c}{\omega_{mn} - \omega_\beta} + \frac{\rho_\gamma^2 r_{nm}^c r_{nm}^b}{\omega_{mn} - \omega_\gamma} \right) \right], \quad (15)
\end{aligned}$$

where  $\rho_\beta = \omega_\beta/\omega_\Sigma$  and  $\rho_\gamma = \omega_\gamma/\omega_\Sigma$ . As  $\omega_\Sigma$  is completely removed from the denominator, the above expression does not have divergence and numerical instability in the zero-frequency limit and can be applied to both nonmagnetic and magnetic materials. Equation (15) is derived from Eq. (14) by the following techniques without assuming  $\mathcal{T}$  symmetry: exchange of dummy band indices  $n$  and  $m$ , integration by parts, and the equality  $-\omega^{-1}[(\omega_{mn} - \omega)^{-1} + (\omega_{nm} - \omega)^{-1}] = \omega_{nm}^{-1}[(\omega_{mn} - \omega)^{-1} - (\omega_{nm} - \omega)^{-1}]$ . Although the U(1)-derivative appears in the first term of Eq. (15), the full susceptibility can be rearranged in a way that terms with Berry connection between degenerate bands are grouped with terms with the U(1)-covariant derivative to form the U(N)-covariant derivative.

As SHG is one of the most commonly used tool for symmetry detection and frequency conversion, we take SHG as an example to discuss properties of the ‘‘divergent’’ terms. The full expression of SHG is given in the Appendix C and the



“divergent” terms are

$$\begin{aligned} \chi_{\text{divg}}^{abc}(-2\omega; \omega, \omega) &= -\frac{ie^3}{4\hbar^2} \int [d\mathbf{k}] f_{nm} \left[ \frac{r_{mn}^c r_{nm;a}^b + r_{mn}^b r_{nm;a}^c}{\omega_{mn}(\omega_{mn} - \omega)} \right. \\ &\quad \left. - \frac{\Delta_{mn}^a (r_{nm}^b r_{mn}^c + r_{nm}^c r_{mn}^b)}{2\omega_{mn}^2 (\omega_{mn} - \omega)} \right]. \end{aligned} \quad (16)$$

The first term in Eq. (16) survives under  $\mathcal{T}$  symmetry and is included in the conventional SHG expression [30,32,35–43]. In addition, the first term is purely real in the static limit. However, the second term in Eq. (16) is present only when  $\mathcal{T}$  symmetry is broken and therefore it is a unique term for magnetic materials. We denote the second term by  $\chi_{\text{new}}(-2\omega; \omega, \omega)$  as it has not been calculated previously. Unexpectedly, we discovered that although  $\chi_{\text{new}}$  has a low-frequency ‘divergent’ form in Eq. (14),  $\chi_{\text{new}}(2\omega = 0; \omega = 0, \omega = 0)$  does not contribute to static SHG  $\chi(0; 0, 0)$  at all, which can be concluded by permutation of the dummy band indices in the summation in Eq. (16). Therefore, the low-frequency ‘divergent’ term  $\chi_{\text{new}}$  can be observed away from low-frequency regime but negligible in low-frequency regime.

### III. SYMMETRY ANALYSIS

#### A. Charge and spin rectification currents

The coefficients of second-order electro-optical responses are third-rank tensors, and their nonzero and independent elements are determined by their magnetic point group [47]. As the electric field, polarization and current are all odd under  $\mathcal{P}$ , the second-order electro-optical responses are only present in  $\mathcal{P}$ -breaking materials, according to Eq. (6). Additionally, special attentions are paid to  $\mathcal{T}$  and  $\mathcal{PT}$  symmetry and the symmetry properties of quantum metric, two-band Berry curvature, and *etc.* are summarized in Table I. As a result, both NSC and NIC are even under  $\mathcal{T}$  while MSC and MIC are even under  $\mathcal{PT}$  symmetry according to Eqs. (10)–(13). Furthermore, the current coupled to  $\text{Re}\{E^b(\omega)E^c(-\omega)\}$  can be induced both by a linearly polarized light (LPL) and a circularly polarized light (CPL), while the current coupled to  $\text{Im}\{E^b(\omega)E^c(-\omega)\}$  can only be induced by CPL with the direction of current determined by the helicity of light [13,31].

Up to now, we have focused on the charge photocurrent which is the collective motion of both spin-up and spin-down electrons. As electrons also carry the spin degree of freedom, the spin current can also exist. Although it is still challenging to properly define the spin current in materials with spin-orbit coupling (SOC) as spin is not a good quantum number in this circumstance [48], the conventional definition  $j^{ab} = 1/2(v^a s^b + s^b v^a)$  with  $v$  and  $s$  the velocity and spin operators respectively is still appropriate without SOC [49,50].

We adopted the conventional definition to analyze the symmetry of spin current. As spin is  $\mathcal{T}$  odd and  $\mathcal{P}$  even, the symmetry of spin current and charge current are opposite under  $\mathcal{T}$  and  $\mathcal{PT}$  symmetry. For instance, while the charge current from the NSC and NIC mechanisms are  $\mathcal{T}$  even, the spin current from those mechanisms are  $\mathcal{T}$  odd. As a result, spin current from NSC and NIC mechanisms only exists in magnetic materials. The symmetry analysis of both charge

TABLE II. Existence of charge and spin current under  $\mathcal{T}$  and  $\mathcal{PT}$  symmetry.  $\checkmark$  means present while  $\times$  means absent.

	$\sigma_{\text{NSC}}$ LPL	$\sigma_{\text{MSC}}$ CPL	$\eta_{\text{NIC}}$ CPL	$\eta_{\text{MIC}}$ LPL
Light				
Current	charge/spin	charge/spin	charge/spin	charge/spin
$\mathcal{T}$	$\checkmark/\times$	$\times/\checkmark$	$\checkmark/\times$	$\times/\checkmark$
$\mathcal{PT}$	$\times/\checkmark$	$\checkmark/\times$	$\times/\checkmark$	$\checkmark/\times$

and spin currents are summarized in Table II. Additionally, in whatever situation, the spin current and charge current are both present. For instance, the charge NSC and spin MIC are present simultaneously in  $\mathcal{T}$ -symmetric materials under LPL.

Although the four mechanisms listed in Table II can generate both charge and spin currents, the microscopic origins are different.  $\mathcal{P}$  breaking is the precondition of second-order charge and spin currents and it usually happens in the crystallographic structure. However, if it is the spin order instead of the geometric structure that breaks  $\mathcal{P}$ , the presence of SOC is imperative for a nonzero charge current but not necessary for generating spin current.

#### B. Second-order susceptibility

Despite the intricate expression of second-order susceptibility, certain symmetry conditions are fulfilled. First, second-order susceptibility has the intrinsic permutation symmetry which states that  $\chi_2^{abc}(-\omega_\Sigma; \omega_\beta, \omega_\gamma)$  is unchanged by the simultaneous interchange of its last two frequency arguments and its last two Cartesian as  $\chi_2^{abc}(-\omega_\Sigma; \omega_\beta, \omega_\gamma) = \chi_2^{acb}(-\omega_\Sigma; \omega_\gamma, \omega_\beta)$ . The intrinsic permutation symmetry is introduced in Eqs. (4) and (5) as a convention. Second, when all frequencies are detuned from resonance and the small imaginary part in the denominator can be ignored, the nonlinear susceptibility possesses full permutation symmetry which states that all of the frequency arguments of the nonlinear susceptibility can be freely interchanged, as long as the corresponding Cartesian indices are interchanged simultaneously:  $\chi_2^{abc}(-\omega_\Sigma; \omega_\beta, \omega_\gamma) = \chi_2^{bca}(\omega_\beta; \omega_\gamma, -\omega_\Sigma)$ . The full permutation symmetry can be deduced from a consideration of the form of the electromagnetic field energy within a lossless medium [51]. Third, in the static limit, the nonzero components of the susceptibility tensor are all equal  $\chi_2^{abc}(0; 0, 0) = \chi_2^{bca}(0; 0, 0) = \chi_2^{cab}(0; 0, 0)$ , which is an extension of the full permutation symmetry in the static limit, also called the Kleinman’s symmetry. Moreover, the static susceptibility is a purely real quantity with the expression given in Appendix C.

Furthermore, the second-order susceptibility which is odd in  $\mathcal{P}$  can be decoupled into two contributions, a nonmagnetic response  $\chi_N$  which is even in  $\mathcal{T}$  and a magnetic response  $\chi_M$  which is odd in  $\mathcal{T}$  with their expressions given in Appendix C.  $\chi_N$  reflects the inversion symmetry breaking from the lattice and charge density while  $\chi_M$  is sensitive to the inversion symmetry breaking by the magnetic ordering. The principal part and  $\delta$ -function part of the nonmagnetic  $\chi_N$  are purely real and imaginary respectively, while the real and imaginary parts are opposite in the magnetic  $\chi_M$  [24,52]. In other words,  $\chi_N$  is purely real and  $\chi_M$  is purely imaginary far away from resonance. As the susceptibility in the static limit is purely

real,  $\chi_M$  does not contribute to the static SHG ( $\chi(2\omega = 0; \omega = 0, \omega = 0)$ ). Away from the static limit,  $\chi_M$  increases linearly with  $\omega$  until a resonance.

Moreover, there is an intuitive picture explaining the presence of the imaginary  $\chi_M$  away from resonance as a consequence of the motion of the optically driven electrons under the Lorentz force exerted by the magnetic field  $\mathbf{B}$  of the ordered state [15]. As the Lorentz force is proportional to velocity which is the time derivative of the electron displacement, it introduces an imaginary term in the equation of motion of optically driven electrons. As a result, the oscillation amplitude is purely real along the direction of the driven electric field independent of  $\mathbf{B}$ , and is purely imaginary with amplitude proportional to  $\omega|\mathbf{B}|$  (to the lowest order) in the direction perpendicular to both the driven electric field and the magnetic field  $\mathbf{B}$ . Eventually, the imaginary amplitude leads to the imaginary off-diagonal linear susceptibility and the imaginary SHG that are proportional to  $\omega|\mathbf{B}|$  away from resonance. Therefore, the symmetry argument based on our formula is consistent with the intuitive picture.

Generally, the imaginary part of the susceptibility is responsible for energy dissipation of the electromagnetic field. However, in magnetic materials,  $\chi_M$  is imaginary even away from resonance, therefore, the meaning of dissipation needs to be scrutinized. The derivation of energy dissipation from the first-order process is in Appendix D. The dissipation rate of the electromagnetic energy due to the second-order polarization  $\mathbf{P}^{(2)}$  is given by

$$\begin{aligned} \mathbf{E} \cdot \frac{d}{dt} \mathbf{P}^{(2)} &= -2i[\omega_\Sigma \chi_2^{abc}(-\omega_\Sigma; \omega_\beta, \omega_\gamma) \\ &\quad - \omega_\beta \chi_2^{bac}(\omega_\beta; -\omega_\Sigma, \omega_\gamma) \\ &\quad - \omega_\gamma \chi_2^{cab}(\omega_\gamma; -\omega_\Sigma, \omega_\beta)] E_{(-\Sigma)}^a E_\beta^b E_\gamma^c. \end{aligned} \quad (17)$$

For frequencies detuned from resonance, the full-permutation symmetry of  $\chi_2$  guarantees that the term in the bracket is zero. Therefore, only the  $\delta$ -function part of susceptibility participates in the dissipation no matter whether it is real or imaginary.

Moreover,  $\chi_N$  and  $\chi_M$  also corresponds to different measurable quantities. The sum-frequency generation (SFG) intensity ( $I$ ) measured in experiments can be decoupled into a nonmagnetic term, a magnetic term and interference terms as

$$\begin{aligned} I &= |\chi_N^{abc} E^b E^c + \chi_M^{abc} E^b E^c|^2 \\ &= |\chi_N^{abc} E^b E^c|^2 + |\chi_M^{abc} E^b E^c|^2 \\ &\quad + (\chi_N^{abc} E^b E^c)(\chi_M^{aij} E^i E^j)^* \\ &\quad + (\chi_N^{abc} E^b E^c)^*(\chi_M^{aij} E^i E^j). \end{aligned} \quad (18)$$

For materials with  $\mathcal{T}$  or  $\mathcal{PT}$  symmetry, the SFG intensity is proportional to either the  $\chi_N$  term or the  $\chi_M$  term. When both  $\mathcal{T}$  and  $\mathcal{PT}$  symmetries are broken, e.g., in a polar ferromagnetic, the appearance of the third and the fourth terms reflects the interference between the magnetic and nonmagnetic signals and the sign of the interference is determined by the direction of the magnetic and the polar orders. As a result, domains with opposite magnetic/polar orders could show different SFG intensities which makes SFG an important tool for multiferroics detection.

## IV. COMPUTATIONAL METHODS

### A. Implementation of U(2)-covariant derivative

In the following, we present the implementation of the U(2)-invariant formulas using both orthogonal and nonorthogonal basis. The general derivative is usually evaluated through a sum-over state method [33,53] or the Wannier interpolation method [43,54]. We adopted the Wannier interpolation method proposed in Ref. [43] for orthogonal basis and the scheme proposed in Ref. [55] for nonorthogonal basis, while the Berry connection in both methods is evaluated based on nondegenerate perturbation theory. Although methods with orthogonal and nonorthogonal basis sets have different gauge choices, as the optical responses are gauge invariant, the two methods are equivalent in principle. While the Wannier interpolation method is more computationally efficient, the method with nonorthogonal basis is more suitable for high-throughput calculations [55]. For SHG calculations, the three-band summation in Eq. (C1) requires a large number of bands to achieve convergence, therefore, the method with nonorthogonal basis is a more suitable choice to evaluate SHG responses.

We generalized the above mentioned methods by including degenerate perturbation to evaluate the Berry connection and the general derivative, which are unavoidable in the presence of  $\mathcal{PT}$  symmetry. For the Wannier interpolation method, we chose a specific gauge which satisfies that  $[U^\dagger(\mathbf{k}_0)U(\mathbf{k}_0 + \delta\mathbf{k})]_{n_\mu n_\nu}$  is purely real for degenerate bands  $n_\mu$  and  $n_\nu$ . Therefore, the non-Abelian connection  $U^\dagger \partial_{k_a} U$  between two degenerate states vanishes, while the non-Abelian connections between nondegenerate states can be calculated as usual [43],

$$(U^\dagger \partial_{k_a} U)_{n_\mu m_\nu} = \begin{cases} -\frac{(U^\dagger \partial_{k_a} (H^{(W)} U))_{n_\mu m_\nu}}{E_n - E_m} & E_n \neq E_m, \\ 0 & E_n = E_m, \end{cases}$$

where  $H^{(W)}$  is the Hamiltonian matrix in the representation of local Wannier functions and  $U$  is a matrix of which each column is an eigenvector of  $H^{(W)}$ . For the method of nonorthogonal basis, we again chose a specific gauge so that the matrix elements between degenerate bands in Ref. [55] are modified as

$$v_{n_\mu}^\dagger S \partial_{k_a} v_{n_\nu} = -\frac{1}{2} v_{n_\mu}^\dagger (\partial_{k_a} S) v_{n_\nu}.$$

Here  $S$  is the overlap matrix of the nonorthogonal basis,  $v$  is a matrix with each column a generalized eigenvector of  $H^{(W)}$  which is in the representation of a set of complete but not orthogonal basis. More details can be found in Ref. [55].

### B. First-principles methods

First-principles calculations were performed both by Vienna *Ab initio* Simulation Package (VASP) [56] which is a plane-wave basis package and by OpenMX which is a pseudoatomic basis package [57,58]. The exchange-correlation functional was parameterized in the PBE form [59]. PAW pseudopotentials [60] and norm-conserving pseudopotentials [61] were used in VASP and OpenMX, respectively. We have included the effect of SOC. For the 3d orbitals in magnetic ions Mn and Cr, the Hubbard  $U$  of 4 eV and 3 eV were used, respectively [62]. For layered materials, we used DFT-D3 form van der Waals correction without damping [63].

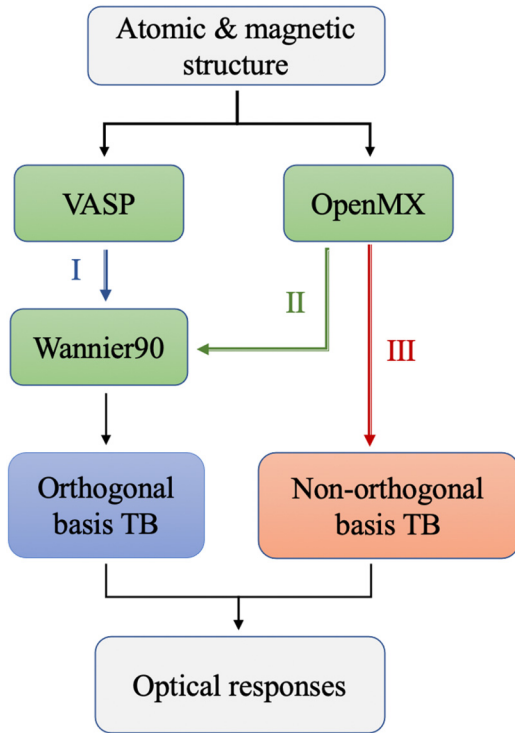


FIG. 1. Calculation workflow of optical responses with three distinct routes. In route I, after first-principles calculations by VASP, a TB Hamiltonian in orthogonal basis is constructed using Wannier90. For route II, the difference from route I is that DFT calculations are done by OpenMX. In route III, a TB Hamiltonian in nonorthogonal basis is constructed directly from OpenMX.

In VASP, the cut-off energy of plane waves was set to 350 eV and 450 eV for  $\text{MnBi}_2\text{Te}_4$  and  $\text{CrI}_3$ , respectively. In OpenMX, Cr6.0-s3p3d2 and I9.0-s3p3d2f1 were chosen as our basis for Cr and I. The convergence criterion for force and electronic calculations were  $10 \text{ meV}/\text{\AA}$  and  $10^{-6} \text{ eV}$ .  $k$ -point samplings of  $15 \times 15 \times 1$  and  $13 \times 13 \times 1$  were used for  $\text{MnBi}_2\text{Te}_4$  and  $\text{CrI}_3$  in VASP, and a  $7 \times 7 \times 1$   $k$ -mesh was adopted for  $\text{CrI}_3$  in OpenMX. The bilayer structures of  $\text{CrI}_3$  and  $\text{MnBi}_2\text{Te}_4$  are simulated in slab supercells with additional vacuum thickness of  $15.8 \text{ \AA}$  and  $20.9 \text{ \AA}$ , respectively.

After getting the converged electronic structures, we either generated the maximally localized Wannier functions using Wannier90 [64] or used the nonorthogonal atomic basis from OpenMX to build the tight-binding (TB) Hamiltonian and calculate the optical responses [43,55]. We obtained 100 maximally localized orbitals for  $\text{MnBi}_2\text{Te}_4$  and 112 maximally localized orbitals for  $\text{CrI}_3$ . In the calculation of MIC and MSC, we adopted a  $400 \times 400 \times 1$  kmesh. For SHG susceptibility, the relaxation rate  $\hbar/\tau$  was set to be 0.05 eV, and the  $k$ -mesh was set to be  $50 \times 50 \times 1$  which leads to identical results as  $100 \times 100 \times 1$   $k$ -point sampling. The degenerate perturbation was applied when the energy difference between two bands are smaller than  $E_{\text{deg}}$  which was set to be 0.5 meV in our calculations.

The flowchart of our computations including the electronic structure and optical response calculations is illustrated in Fig. 1. Three different routes were adopted to calculate optical

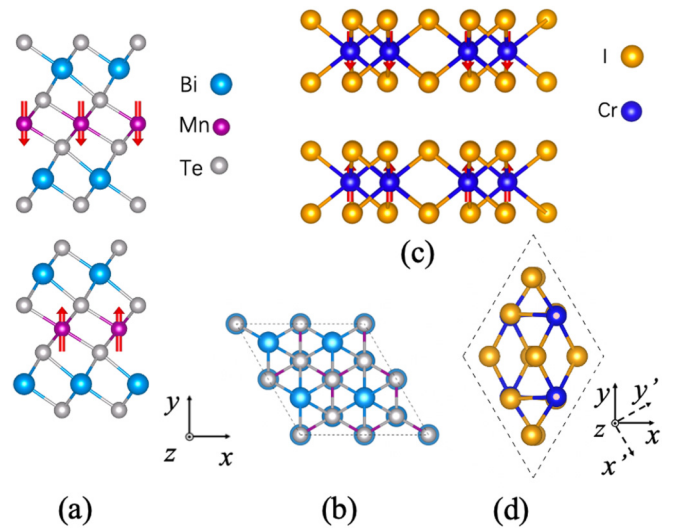


FIG. 2. Atomic and magnetic structures of bilayer  $\text{MnBi}_2\text{Te}_4$  and  $\text{CrI}_3$ . (a, b) The side and top view of bilayer AFM- $z$   $\text{MnBi}_2\text{Te}_4$ . (c, d) The side and top view of AFM- $z$  bilayer  $\text{CrI}_3$ . The Cartesian coordinates adopted in our calculations are marked by solid lines and the one adopted in Ref. [67] are in the dashed line.

responses. In route I, the electronic structure is calculated by VASP and then maximally localized Wannier functions are constructed to obtain the TB Hamiltonian in an orthogonal basis set. The difference between route II and I is that the electronic structure is calculated by OpenMX rather than VASP. In route III, after the self-consistent calculation by OpenMX, we directly get the TB Hamiltonian in nonorthogonal basis. The purpose of using multiple packages (VASP and OpenMX) and different basis sets (orthogonal and nonorthogonal) for TB Hamiltonian is to examine the influence of different electronic structures and gauge choices for optical responses separately.

## V. EXAMPLE STUDIES AND RESULTS

### A. Candidate materials

We considered two prototypical materials: bilayer  $\text{MnBi}_2\text{Te}_4$  [65] and bilayer  $\text{CrI}_3$  [66], both of which have interlayer A-type AFM magnetic orders. Both materials exhibit  $\mathcal{PT}$  symmetry, which implies that the spatial inversion symmetry is broken only through the AFM spin order between two layers while the geometric structure is centrosymmetric as illustrated in Fig. 2. In this case, only magnetic responses including MIC, MSC and magnetization-sensitive SHG are present while normal responses NIC, NSC and nonmagnetic SHG are forbidden, which makes them a simple platform to investigate the nonlinear magneto-optical responses. Additionally, both materials are feasible in experiments [21,66] and their MIC has been studied theoretically in literature [12,13,67], which provides valuable results for comparison. For bilayer  $\text{MnBi}_2\text{Te}_4$ , we considered AB-type stacking (same stacking pattern as in bulk) in which each layer laterally shifted by  $[1/3, 1/3, 0]$ . For bilayer  $\text{CrI}_3$ , multiple stacking orders have been investigated previously [66], and we focused on the one with AB' stacking in which each layer laterally shifted by  $[1/3, 0, 0]$  as both MIC



TABLE III. The stacking order, magnetic point group, magnetic order, and independent tensor components of MIC, MSC, and MSH in bilayer  $\text{MnBi}_2\text{Te}_4$  and bilayer  $\text{CrI}_3$ .

Material	Stacking	Magnetic point group	Magnetic order	MIC	MSC	MSHG
Bilayer $\text{MnBi}_2\text{Te}_4$	AB	$\bar{3}'m'$	AFM- $z$	$xxx$	none	$xxx$
Bilayer $\text{CrI}_3$	AB'	$2/m'$	AFM- $z$	$xyx, yxx, yyy$	$xyy$	$xyx, yxx, yyy$

and MSC are allowed. The stacking order, magnetic order, magnetic symmetry and independent tensor components are summarized in Table III.

### B. Magnetic injection current

The MIC of bilayer  $\text{MnBi}_2\text{Te}_4$  and bilayer  $\text{CrI}_3$  have been calculated recently [13,67]. Although previous works have not considered the  $U(2)$ -gauge problem under  $\mathcal{PT}$  symmetry, fortunately, MIC is not subjected to modifications in  $\mathcal{PT}$  symmetric case as demonstrated in Eq. (13). In the following, we performed calculations and compared with previous works to validate our method and results.

In bilayer  $\text{MnBi}_2\text{Te}_4$ , its interlayer coupling is AFM and the intralayer coupling is FM. The energy band is highly dispersive near  $\Gamma$  point with a small gap of 0.076 eV at  $\Gamma$  point as shown in Fig. 3(a). In addition, the comparison between our band structure (red solid line) and the one from Ref. [13] (blue dotted line) along high-symmetry line showed excellent agreement near band edge. Fig. 3(b) shows the MIC conductivity of bilayer  $\text{MnBi}_2\text{Te}_4$  along  $xxx$  direction. Our result again shows an excellent agreement with the Ref. [13].

For bilayer  $\text{CrI}_3$ , to compare the band structure and MIC, we adopted the same Hubbard  $U = 1$  eV and the same Cartesian coordinates shown by dashed line in Fig. 2(d) as Ref. [67]. Figures 3(c) and 3(d) shows the calculated band

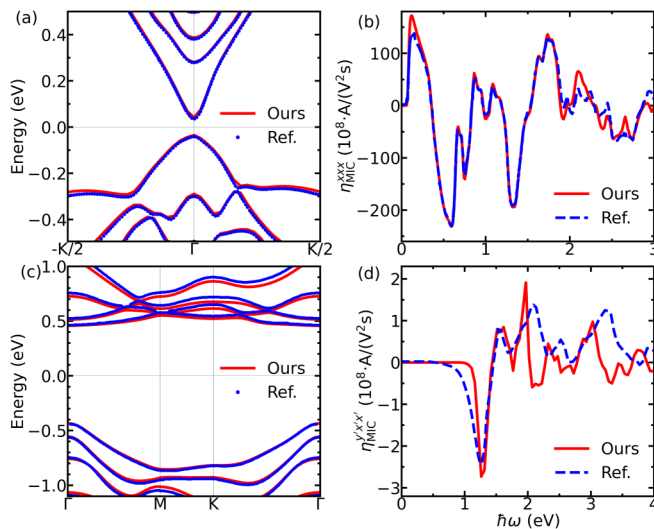


FIG. 3. Band structure and MIC conductivity of bilayer  $\text{MnBi}_2\text{Te}_4$  and  $\text{CrI}_3$ . (a, b) Band structure and MIC conductivity of  $\text{MnBi}_2\text{Te}_4$ . Our results are in red solid line and the results in Ref. [13] are in blue dotted line. (c, d) Band structure and MIC conductivity of  $\text{CrI}_3$ . Our results are in red solid line, while the results in Ref. [67] are in blue dashed line. The coordinate was set to be the same as Ref. [67], as illustrated in Fig. 2.

structure and MIC conductivity. The band gap of bilayer  $\text{CrI}_3$  is 0.89 eV at  $\Gamma$  point in our calculation and 0.78 eV in the reference, therefore we applied a 0.11 eV scissor operation to results in the reference to align the band gap in Fig. 3(c) to 0.89 eV. While the shape of bands is similar near band edge, noticeable differences exist in many bands. In addition, the magnetic moment of Cr atom along  $z$  axis is  $3.12 \mu_B$  in our calculation, while  $3.21 \mu_B$  in the reference. As we adopted the same parameters (Hubbard  $U$ , exchange-correlation functional, pseudopotentials, Van der Waals correction) in the electronic structure calculation, the above discrepancies might result from small differences in atomic structures. For the MIC shown in Fig. 3(d), the main characters of our results (red solid line) are similar to the reference (blue dashed line) including the location and height of the first peak. However, due to discrepancies in band structures away from the band edge, the location and height of other peaks are shifted and modified. Therefore, combining the results of bilayer  $\text{MnBi}_2\text{Te}_4$  and  $\text{CrI}_3$ , we can conclude that we are able to reproduce MIC results in literature while the detailed features of MIC are sensitive to geometric and electronic structures.

### C. Magnetic shift current

The correction to ensure  $U(2)$ -gauge invariant is essential in the presence of  $\mathcal{PT}$  symmetry which hasn't been considered in the previous computational works. Here we calculated the MSC conductivity of AB' stacking bilayer  $\text{CrI}_3$  through three different routes shown in Fig. 1 and alternating between  $U(1)$ -covariant derivative and  $U(2)$ -covariant derivative to demonstrate the influence of the gauge choice. As the results are very sensitive to details in electronic structure (see Fig. 6 in the Appendix), we stay with the electronic structure obtained from OpenMX for comparison.

Figure 4(a) shows the results from orthogonal basis method (blue dashed line) versus nonorthogonal basis method (red solid line) (route II versus route III in Fig. 1) with the same  $U(2)$ -gauge-invariant formula. Since the two methods have different gauge choices, the excellently agreed results in Fig. 4(a) confirm that our formula of MSC is truly gauge independent.

In contrast, Fig. 4(b) shows the results of orthogonal basis method (blue dashed line) versus nonorthogonal basis method (red solid line) (route II versus route III) both using  $U(1)$ -form formula adopted in previous works. Although the same electronic structure guarantees that the peak positions resulted from the delta function in Eq. (11) are exactly the same, differences in peak height are prominent, especially at low-frequency region. For instance, they differ by more than three times at 1.2 eV and there is a shoulder in the red solid line at 0.8 eV while absent in the blue dashed line. More importantly, the differences demonstrate that the previous



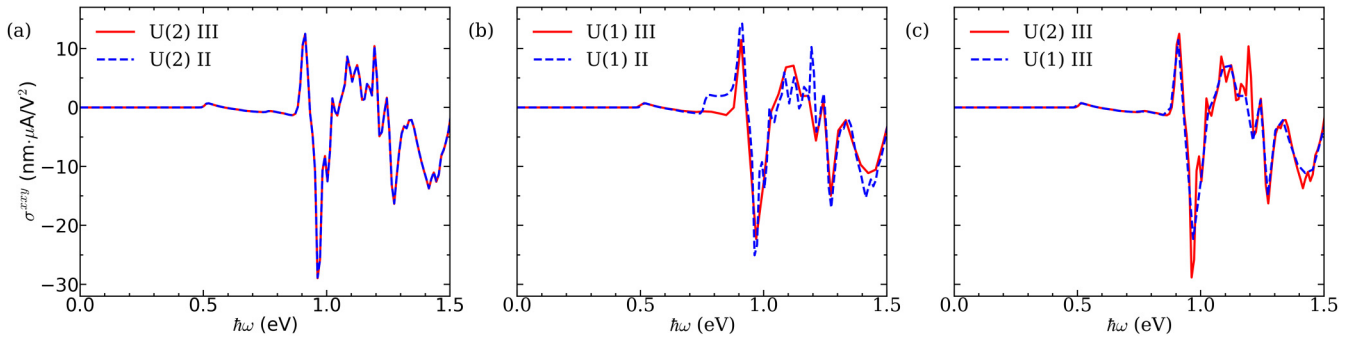


FIG. 4. Magnetic shift current of bilayer  $\text{CrI}_3$  calculated by (a) route II (blue dashed line) compared with route III (red solid line), both using the U(2)-invariant formula. The gap has been scissored to the same value. (b) The result of nonorthogonal basis (red solid line, route III) compared with that of orthogonal basis (blue dashed line, route II), both using U(1)-invariant formula. (c) Results of U(2)-invariant formula (red solid line) versus U(1)-invariant formula (blue dashed line), both using nonorthogonal basis (route III).

formula with U(1)-covariant derivative is gauge dependent and the results are not reproducible due to different gauge choices. Therefore, the previous MSC formula with U(1)-covariant derivative cannot be used to describe physical observables in  $\mathcal{PT}$ -symmetric materials.

Furthermore, Fig. 4(c) shows results from U(1)-invariant formula (blue dashed line) versus U(2)-invariant formula (red solid line) using nonorthogonal basis (both are in route III). Noticeable differences are observed at peak values due to the improper implementation of the covariant derivative in  $\mathcal{PT}$ -symmetric materials, implying the necessity of using U(2)-invariant formula.

#### D. Magnetic second-harmonic generation

In terms of the SHG response in magnetic materials, there are three sets of formulas that are available to use: the “DVG” (short for “divergent”) expression which is the original expression in Ref. [30] with low-frequency divergent problem shown in Eq. (14), the “ $\mathcal{T}$ -SMP” expression in which the low-frequency divergent terms are simplified by  $\mathcal{T}$  symmetry and have been implemented in many calculation packages [37–43], and the full-frequency convergent expression “CVG” (short for “convergent”) proposed in this paper as demonstrated in Eq. (15). The complete formulas of these three sets of formulas can be found in Appendix C. As the gigantic SHG susceptibility in  $\text{AB}'\text{CrI}_3$  has been measured [21] and calculated using the  $\mathcal{T}$ -SMP expression recently [45], we compare the SHG response of bilayer  $\text{CrI}_3$  along  $xy$  direction calculated using the DVG,  $\mathcal{T}$ -SMP and CVG expressions.

Figure 5(a) shows the real and imaginary parts of SHG calculated using the DVG and the CVG expressions. As the CVG expression is derived from the DVG expression, results from both expressions are exactly the same away from the low-frequency region ( $>0.1$  eV). However, in the low-frequency region (0–0.1 eV), both the real and imaginary parts calculated from the DVG expression exhibit divergent problem which is absent in the CVG expression. Additionally, in the static limit, the CVG expression converges to exactly zero which is consistent with our symmetry analysis in Sec. III.

Figure 5(b) shows the real and imaginary parts of SHG calculated using the  $\mathcal{T}$ -SMP and the CVG expressions. As the SHG response in bilayer  $\text{CrI}_3$  is gigantic, the difference

between the two expressions is hardly observed in Fig. 5(b). Detailed analyses in Fig. 5(c) demonstrates that the differences between the two expressions, which is  $\chi_{\text{new}}$ , are roughly  $10^4$   $\text{pm}^2/\text{V}$ . The materials and frequency range in which  $\chi_{\text{new}}$  is considerable is subject to further investigation with more magnetic materials.

## VI. CONCLUSION

In summary, we have given a thorough derivation of second-order magneto-optical effects, including rectification current and second-order susceptibility using density matrix perturbation method. Especially, the formalism of rectification current conductivity is compatible with both nondegenerate and degenerate bands and is applicable to both nonmagnetic and magnetic materials. Additionally, the second-order susceptibility with  $\omega^{-1}$  and  $\omega^{-2}$  terms, which suffers from low-frequency divergent problem, numerical instability and is dropped conventionally, has been reformulated into a full-frequency convergent form without any symmetry assumptions. Furthermore, we have implemented the above formulas to first-principles calculations, which has not been done for real materials as far as we know, using both orthogonal and nonorthogonal basis methods. Two  $\mathcal{PT}$ -symmetric magnetic materials, bilayer AFM  $\text{MnBi}_2\text{Te}_4$  and  $\text{CrI}_3$  are used as examples. We confirmed that our formulas for rectification current including the MIC and MSC are truly invariant under U(2)-gauge freedom. Meanwhile, for SHG in magnetic materials, we found noticeable modifications of SHG responses from the divergent term. With the vast variety of magnetic materials, it is expected that our reformulated formalism can greatly facilitate and advance the theoretical understandings on nonlinear magneto-optical effects.

## ACKNOWLEDGMENTS

We thank Chong Wang and Shiqiao Du for helpful discussion. This work was supported by the Basic Science Center Project of NSFC (Grant No. 51788104), the Ministry of Science and Technology of China (Grants No. 2018YFA0307100 and No. 2018YFA0305603), the National Science Fund for Distinguished Young Scholars (Grant No. 12025405), the National Natural Science Foundation of China (Grant No.

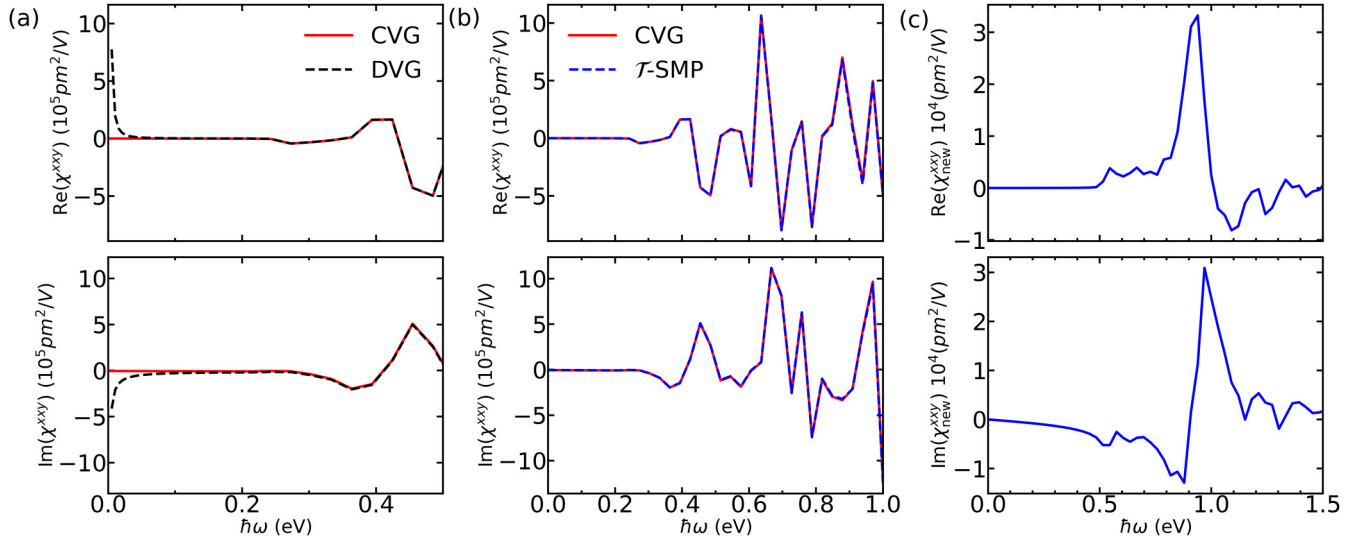


FIG. 5. SHG of bilayer AB' stacking CrI<sub>3</sub> along *xxy* direction. Upper and lower panels (a) are the real and imaginary part of SHG, respectively. The black dashed line shows the results of the previous divergent formulas marked as “DVG,” the red solid line shows the results of complete and modified formulas marked as “CVG.” Panel (b) shows the comparison between the complete and modified formulas (“CVG,” red solid line) and the conventional used but not complete formulas (*T*-SMP, blue dashed line). Panel (c) shows the difference between them, where the blue line is the value of the new term.

11874035), and the Beijing Advanced Innovation Center for Future Chip (ICFC). M.Y. is supported by Shuimu Tsinghua Scholar Program and Postdoctoral International Exchange Program.

## APPENDIX A: BASIC HAMILTONIAN, CURRENT OPERATOR, AND POLARIZATION OPERATOR

### 1. Basic Hamiltonian

The Hamiltonian of a semiconductor in an electromagnetic field, whose magnetic field is negligible and the electric field is treated in the long-wavelength limit, can be written as [33]

$$\hat{H}(t) = \int \tilde{\psi}^\dagger(\mathbf{x}) [H_0 - e\mathbf{x} \cdot \mathbf{E}(t)] \tilde{\psi}(\mathbf{x}) d\mathbf{x}, \quad (\text{A1})$$

where  $H_0$  is the unperturbed single particle Hamiltonian in the coordinate representation with Bloch eigenfunctions  $\psi_{n\mu}(\mathbf{k}; \mathbf{x})$  and eigenvalues  $\hbar\omega_n(\mathbf{k})$ , and  $\tilde{\psi}(\mathbf{x}) = \sum_n \sum_\mu \int d\mathbf{k} \psi_{n\mu}(\mathbf{k}; \mathbf{x}) \hat{a}_{n\mu}(\mathbf{k})$  is the field operator with the annihilation operator for Bloch state as  $\hat{a}_{n\mu}(\mathbf{k})$ .  $n$  and  $\mu$  are the band indices and  $e = -|e|$ . The interaction of electric field and electrons is represented using the length gauge by the  $-e\mathbf{x} \cdot \mathbf{E}(t)$  term while the results are equivalent to the velocity gauge formula [32,68]. The long-wavelength limit is valid as long as the wavelength of light is much larger than the length scale of interests.

Combining Eqs. (1) and (A1), we get

$$\begin{aligned} \hat{H}(t) = & \int d\mathbf{k} \hbar\omega_{n\mu}(\mathbf{k}) \hat{a}_{n\mu}^\dagger(\mathbf{k}) \hat{a}_{n\mu}(\mathbf{k}) \\ & - e \int d\mathbf{k} \mathbf{E}(t) \cdot \mathbf{r}_{n\mu m_\nu}(\mathbf{k}) \hat{a}_{n\mu}^\dagger(\mathbf{k}) \hat{a}_{m_\nu}(\mathbf{k}) \end{aligned}$$

$$\begin{aligned} & - e \int d\mathbf{k} \mathbf{E}(t) \cdot \hat{a}_{n\mu}^\dagger(\mathbf{k}) i \partial_{\mathbf{k}} \hat{a}_{n\mu}(\mathbf{k}) \\ & - e \int d\mathbf{k} \mathbf{E}(t) \cdot \mathbf{A}_{n\mu}(\mathbf{k}) \hat{a}_{n\mu}^\dagger(\mathbf{k}) \hat{a}_{n\mu}(\mathbf{k}). \end{aligned} \quad (\text{A2})$$

The intraband electric current according to Eq. (3) is

$$\begin{aligned} \hat{J}_{\text{intra}}^a = & e \int d\mathbf{k} v_{n\mu n\mu}^a(\mathbf{k}) \hat{a}_{n\mu}^\dagger(\mathbf{k}) \hat{a}_{n\mu}(\mathbf{k}) \\ & - \frac{e^2}{\hbar} \int d\mathbf{k} E^b r_{n\mu m_\nu; a}^b \hat{a}_{n\mu}^\dagger(\mathbf{k}) \hat{a}_{m_\nu}(\mathbf{k}) \\ & - \frac{e^2}{\hbar} \int d\mathbf{k} E^b \Omega_{n\mu}^{ab} \hat{a}_{n\mu}^\dagger(\mathbf{k}) \hat{a}_{n\mu}(\mathbf{k}), \end{aligned} \quad (\text{A3})$$

where  $v_{n\mu n\mu}^a = (1/\hbar) \partial_{k_a} E_n(\mathbf{k})$  is the group velocity and  $\Omega_{n\mu}^{ab} = \partial_{k_a} A_{n\mu}^b(\mathbf{k}) - \partial_{k_b} A_{n\mu}^a(\mathbf{k})$  is Berry curvature. The interband polarization of Eq. (3) is

$$\hat{P}_{\text{inter}}^a = e \int d\mathbf{k} r_{n\mu m_\nu}^a(\mathbf{k}) \hat{a}_{n\mu}^\dagger(\mathbf{k}) \hat{a}_{m_\nu}(\mathbf{k}), \quad (\text{A4})$$

where  $r_{n\mu m_\nu; a}^b = \partial_a r_{n\mu m_\nu}^b - i r_{n\mu m_\nu}^b (A_{n\mu}^a - A_{m_\nu}^a)$  is the U(1)-covariant derivative [33].

### 2. Dynamics of density operators

The matrix element of density operator is

$$\langle n_\lambda \mathbf{k} | \hat{\rho} | m_\mu \mathbf{k}' \rangle = \langle \hat{a}_{m_\mu}^\dagger(\mathbf{k}') \hat{a}_{n_\lambda}(\mathbf{k}) \rangle, \quad (\text{A5})$$

and the initial density operator without perturbation is

$$\hat{\rho}_0 = \frac{1}{Z} e^{-\beta H_0}, \quad (\text{A6})$$

with the matrix element  $\langle n_\lambda \mathbf{k} | \hat{\rho}_0 | m_\mu \mathbf{k}' \rangle = f_{n_\lambda} \delta_{nm} \delta_{\lambda\mu} \delta(\mathbf{k} - \mathbf{k}')$ , where the prefactor  $f_{n_\lambda}$  is the band occupation number.

Switching on illumination at  $t = -\infty$ , the Heisenberg equation of motion of the density operator matrix element reads [51,69,70]

$$\frac{d}{dt}\hat{\rho}_{n_\lambda m_\mu}(t) = \frac{1}{i\hbar}[H(t), \hat{\rho}(t)]_{n_\lambda m_\mu} - \frac{\hat{\rho}_{n_\lambda m_\mu}(t) - \hat{\rho}_{n_\lambda m_\mu}(-\infty)}{\tau_{n_\lambda m_\mu}}, \quad (\text{A7})$$

where the last term is a phenomenological term describing scattering processes of electrons with the relaxation time  $\tau_{n_\lambda m_\mu}$ . Most frequently, the relaxation time is assumed to be independent of the initial and final states, and therefore the subscript of  $\tau$  is dropped.

Expanding the density operator in power of the electric field as  $\hat{\rho}(t) = \hat{\rho}_0 + \hat{\rho}_1(t) + \hat{\rho}_2(t) + \dots$  and inserting it into Eq. (A7), the equation of motion of density operator can be rewritten as

$$i\hbar \frac{d\hat{\rho}_{i+1}(t)}{dt} = [H_0, \hat{\rho}_{i+1}(t)] + [H', \hat{\rho}_i(t)] - i\hbar \frac{\hat{\rho}_{i+1}(t)}{\tau}, \quad (\text{A8})$$

which can be solved using the iteration method. The first-order term of the density operator matrix element is

$$(\hat{\rho}_1)_{n_\lambda \mathbf{k}, m_\mu \mathbf{k}'} = \frac{eE_\beta^b e^{-i\omega_\beta t}}{\hbar} \frac{f_{mn} r_{n_\lambda m_\mu}^b(\mathbf{k})}{\omega_{nm} - \omega_\beta - i/\tau} \delta(\mathbf{k} - \mathbf{k}'), \quad (\text{A9})$$

and the second-order term is

$$(\hat{\rho}_2)_{n_\lambda \mathbf{k}, m_\mu \mathbf{k}'} = \frac{-e^2}{\hbar^2} \frac{(\Gamma_2)_{n_\lambda \mathbf{k}, m_\mu \mathbf{k}'} \delta(\mathbf{k} - \mathbf{k}')}{\omega_{nm}(\mathbf{k}) - \omega_\Sigma - i/\tau} E_\beta^b E_\gamma^c e^{-i\omega_\Sigma t}, \quad (\text{A10})$$

with

$$\begin{aligned} (\Gamma_2^L)_{n_\lambda \mathbf{k}, m_\mu \mathbf{k}'} &= \frac{f_{ln}(\mathbf{k}) r_{n_\lambda l_\nu}^b(\mathbf{k}) r_{l_\nu m_\mu}^c(\mathbf{k})}{\omega_{nl}(\mathbf{k}) - \omega_\beta - i/\tau} - \frac{f_{ml}(\mathbf{k}) r_{l_\nu m_\mu}^b(\mathbf{k}) r_{n_\lambda l_\nu}^c(\mathbf{k})}{\omega_{lm}(\mathbf{k}) - \omega_\beta - i/\tau} \\ &\quad - i \left[ \frac{f_{mn}(\mathbf{k}) r_{n_\lambda m_\mu}^b(\mathbf{k})}{\omega_{nm}(\mathbf{k}) - \omega_\beta - i/\tau} \right]_{:c}. \end{aligned} \quad (\text{A11})$$

In derivation of Eq. (A10), we have assumed that the system is a gapped system, in other words,  $\partial_{\mathbf{k}} f_n(\mathbf{k}) = 0$ .

### 3. First- and second-order response functions

The zeroth-order polarization response can be obtained through Eqs. (1) and (A6) as

$$P_0^a = -e \int [d\mathbf{k}] f_n A_{n_\mu}^a(\mathbf{k}), \quad (\text{A12})$$

which reproduces the result of modern theory of polarization.

The first-order electric susceptibility is acquired by combining Eqs. (1) and (A9) as

$$\begin{aligned} \chi_1^{ab}(-\omega; \omega) &= \frac{e^2}{\hbar} \int [d\mathbf{k}] \frac{f_{nm} r_{n_\mu m_\nu}^a r_{m_\nu n_\mu}^b}{\omega_{mn} - \omega - i/\tau} \\ &\quad - \frac{e^2}{i\hbar\omega} \int [d\mathbf{k}] f_n \left[ \frac{\partial}{\partial k_a} A_{n_\mu}^b - \frac{\partial}{\partial k_b} A_{n_\mu}^a \right]. \end{aligned} \quad (\text{A13})$$

While the first term provides a finite susceptibility, the second term shows a  $\omega^{-1}$  divergence in the static limit ( $\omega \rightarrow 0$ ) which indicates the presence of a DC current in the static limit. Using the continuity relationship, the first-order electric current

conductivity is

$$\begin{aligned} \sigma_1^{ab}(-\omega; \omega) &= \frac{-ie^2\omega}{\hbar} \int [d\mathbf{k}] \frac{f_{nm} r_{n_\mu m_\nu}^a r_{m_\nu n_\mu}^b}{\omega_{mn} - \omega - i/\tau} \\ &\quad + \frac{e^2}{\hbar} \int [d\mathbf{k}] f_n \Omega_{n_\mu}^{ab}. \end{aligned} \quad (\text{A14})$$

In the static limit, the first term vanishes and the second term is just the anomalous Hall current [71]. For two-dimensional insulators with nonzero Chern number, this term contributes a quantized anomalous Hall current.

For second-order responses,  $\langle j_{\text{intra}}^a(t) \rangle^{(2)}$  has been given in the main text and the full expression of  $\langle P_{\text{inter}}^a(t) \rangle^{(2)}$  is given by

$$\begin{aligned} \langle \mathbf{P}_{\text{inter}}(t) \rangle^{(2)} &= \frac{e^3}{2\hbar^2} \int [d\mathbf{k}] \frac{[\dots] E_\beta^b E_\gamma^c e^{-i\omega_\Sigma t}}{\omega_{ln} - \omega_\Sigma - i/\tau} \\ &= \frac{e^3}{2\hbar^2} \int [d\mathbf{k}] \frac{r_{n_\mu l_\nu}^a}{\omega_{ln} - \omega_\Sigma - i/\tau} \left[ \frac{f_{nm} r_{l_\nu m_\lambda}^c r_{m_\lambda n_\mu}^b}{\omega_{mn} - \omega_\beta - i/\tau} \right. \\ &\quad - \frac{f_{ml} r_{l_\nu m_\lambda}^b r_{m_\lambda n_\mu}^c}{\omega_{lm} - \omega_\beta - i/\tau} + \frac{if_{nl} r_{l_\nu n_\mu}^b}{\omega_{ln} - \omega_\beta - i/\tau} \\ &\quad \left. - \frac{if_{nl} r_{l_\nu n_\mu}^b \Delta_{ln}^c}{(\omega_{ln} - \omega_\beta - i/\tau)^2} + (b\beta \leftrightarrow c\gamma) \right] E_\beta^b E_\gamma^c e^{-i\omega_\Sigma t}. \end{aligned} \quad (\text{A15})$$

## APPENDIX B: ALTERNATIVE DERIVATION OF THE U(2)-INVARIANT FORMULATION

Starting from Eq. (1) and considering the position operator matrix between degenerate bands in the intraband term, an alternate division of the interband and intraband position operator matrix elements gives

$$\begin{aligned} \langle n_\mu \mathbf{k} | \hat{\mathbf{x}}_{\text{intra}} | m_\nu \mathbf{k}' \rangle &= \delta_{nm} \delta_{\mu\nu} \left[ \mathbf{A}_{n_\mu}(\mathbf{k}) + i \frac{\partial}{\partial \mathbf{k}} \delta(\mathbf{k} - \mathbf{k}') \right] \\ &\quad + \delta_{nm} (1 - \delta_{\mu\nu}) \mathbf{r}_{n_\mu m_\nu}(\mathbf{k}), \end{aligned} \quad (\text{B1})$$

and

$$\langle n_\mu \mathbf{k} | \hat{\mathbf{x}}_{\text{inter}} | m_\nu \mathbf{k}' \rangle = (1 - \delta_{nm}) \mathbf{r}_{n_\mu m_\nu}(\mathbf{k}). \quad (\text{B2})$$

The newly added second term in Eq. (B1) is the origin of the diverging  $P_{\text{inter}}(\omega_\Sigma)$  in the  $\omega_\Sigma \rightarrow 0$  limit. Following the standard treatment of the main text, we retrieved the same results regardless of how the position operator matrix elements are divided. In the new division, DC current is only from  $J_{\text{intra}}$  term regardless of the presence of the degeneracy condition.

## APPENDIX C: FULL EXPRESSION OF SHG AND THE STATIC LIMIT

In the main text, we focused on the divergent terms in the second-order susceptibility, while the full expressions consist of both the nondivergent term originated from Eq. (A15) and the ‘‘apparent’’ diverging terms in Eq. (15). In addition, as SHG is frequently measured and computed, we provided the

full expression of ‘‘CVG’’ SHG as

$$\chi_2^{abc}(-2\omega; \omega, \omega) = \chi_e^{abc}(-2\omega; \omega, \omega) + \chi_i^{abc}(-2\omega; \omega, \omega), \quad (\text{C1})$$

with

$$\begin{aligned} & \chi_e^{abc}(-2\omega; \omega, \omega) \\ &= \frac{e^3}{2\hbar^2} \int [d\mathbf{k}] \sum' \frac{r_{n_\mu m_\nu}^a (r_{m_\nu l_\lambda}^b r_{l_\lambda n_\mu}^c + r_{m_\nu l_\lambda}^c r_{l_\lambda n_\mu}^b)}{\omega_{ln} - \omega_{ml}} \\ & \quad \times \left( \frac{2f_{nm}}{\omega_{mn} - 2\omega} + \frac{f_{ln}}{\omega_{ln} - \omega} + \frac{f_{ml}}{\omega_{ml} - \omega} \right) \end{aligned} \quad (\text{C2})$$

is purely the contribution from interband process, and

$$\begin{aligned} & \chi_i^{abc}(-2\omega; \omega, \omega) \\ &= \frac{ie^3}{2\hbar^2} \int [d\mathbf{k}] f_{nm} \left[ \frac{2(I_{nm}^{cb} + I_{nm}^{bc})}{\omega_{mn}(\omega_{mn} - 2\omega)} + \frac{I_{mn}^{cb} + I_{mn}^{bc}}{\omega_{mn}(\omega_{mn} - \omega)} \right. \\ & \quad + \frac{(g_{nm}^{ab} - \frac{i}{2}\Omega_{nm}^{ab})\Delta_{mn}^c + (g_{nm}^{ac} - \frac{i}{2}\Omega_{nm}^{ac})\Delta_{mn}^b}{\omega_{mn}^2} \\ & \quad \times \left( \frac{1}{\omega_{mn} - \omega} - \frac{4}{\omega_{mn} - 2\omega} \right) \\ & \quad \left. - \frac{I_{mn}^{cb} + I_{mn}^{bc}}{2\omega_{mn}(\omega_{mn} - \omega)} - \frac{\Delta_{mn}^a g_{nm}^{bc}}{2\omega_{mn}^2(\omega_{mn} - \omega)} \right] \end{aligned} \quad (\text{C3})$$

is the contribution of the mixed interband and intraband processes where the last two terms are shown in Eq. (16). If we remove the last term, then we get the ‘‘ $\mathcal{T}$ -smp’’ formula, which is derived assuming time-reversal symmetry. If we replace the last two terms by Eq. (14), then we obtain the ‘‘DVG’’ formula which suffered from the low frequency divergent problem. Again, the small imaginary part  $i/\tau$  in the denominator is not written out explicitly. The  $\sum'$  in Eq. (C2) means that the in the summation,  $n, m, l$  are not equal to each other. In addition to the new term derived in Eq. (16), our full expression also satisfies that each term is gauge-covariant regardless of the degeneracy condition. This is achieved by restricting  $n, m, l$  not equal to each other in the three-band summation term and rearranging the rest terms not obeying this restriction (terms including connections in degenerate sub space) with the U(1)-covariant derivative to construct the U(2)-covariant derivative.

The SHG susceptibility described in Eqs. (C2) and (C3) is made up of terms of the form  $\int \frac{A}{\omega_{mn} - \omega - i/\tau}$ . The  $\mathcal{T}$ -even part of the susceptibility  $\chi_N$  takes the form of  $\int \frac{\text{Re}A}{\omega_{mn} - \omega - i/\tau}$  and the  $\mathcal{T}$ -odd part  $\chi_M$  takes the form of  $\int \frac{\text{Im}A}{\omega_{mn} - \omega - i/\tau}$ . Therefore, in the limit  $\tau \rightarrow \infty$ , the principal part and  $\delta$ -function part of the nonmagnetic  $\chi_N$  are purely real and imaginary respectively, while the real and imaginary parts are opposite in the magnetic  $\chi_M$  [24,52].

The SHG susceptibility at the static limit is an important criterion for the application of nonlinear crystals and we rearranged the static susceptibility in a form that directly represent

the Kleinman’s symmetry as

$$\begin{aligned} \chi_e^{abc}(0; 0, 0) &= \frac{e^3}{6\hbar^2} \int [dk] \sum' P(abc) \text{Re}\{r_{n_\mu m_\nu}^a r_{m_\nu l_\lambda}^b r_{l_\lambda n_\mu}^c\} \\ & \quad \times \frac{\omega_m f_{nl} + \omega_p f_{mn} + \omega_n f_{lm}}{\omega_{mn}\omega_{ln}\omega_{ml}} \end{aligned} \quad (\text{C4})$$

and

$$\chi_i^{abc}(0; 0, 0) = -\frac{e^3}{4\hbar^2} \int [dk] \frac{f_{nm}}{\omega_{mn}^2} P(abc) \text{Im}\{I_{nm}^{abc}\}, \quad (\text{C5})$$

where  $P(abc)$  denotes full permutation of indices  $a, b, c$ .

#### APPENDIX D: THE RELATIONSHIP BETWEEN ENERGY DISSIPATION AND SUSCEPTIBILITY

In  $\mathcal{T}$ -symmetric systems, the imaginary part of susceptibility which only contains the  $\delta$ -function part characterizes the dissipation of the electromagnetic field energy, while the real part does not contain  $\delta$ -function terms. In  $\mathcal{T}$ -breaking systems, the susceptibility is complex in general, and the imaginary part of the susceptibility is not related to dissipation. For example, the power of dissipation in the first order

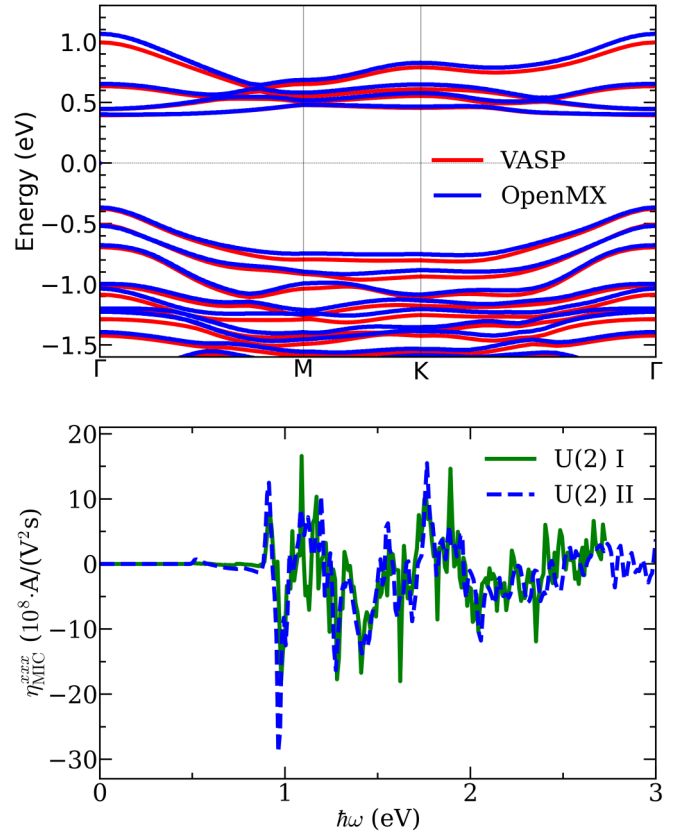


FIG. 6. Band structure and MSC conductivity of bilayer  $\text{CrI}_3$  calculated from that from route II compared with route I, both using U(2) invariant formulas with first-principles packages VASP and OpenMX. We have applied a scissor operation to the results of OpenMX to increase the gap to the same value as that of VASP.



gives

$$\begin{aligned} \mathbf{E} \cdot \frac{d\mathbf{P}_1}{dt} &= \sum_{\omega} i\omega [-\chi_1^{ab}(-\omega; \omega) + \chi_1^{ba}(\omega; -\omega)] E_{-\omega}^a E_{\omega}^b + \text{c.c.} \\ &= \sum_{\omega} \frac{\pi e^2 \omega}{\hbar} \int [d\mathbf{k}] f_{nm} [r_{nm}^a r_{mn}^b E_{-\omega}^a E_{\omega}^b + \text{c.c.}] \\ &\quad \times \delta(\omega_{mn} - \omega). \end{aligned} \quad (\text{D1})$$

Therefore, in  $\mathcal{T}$ -breaking systems, the dissipation is not related to the imaginary part of  $\chi_1$ , rather, it depends only on the  $\delta$ -function part which is not purely imaginary.

## APPENDIX E: BAND STRUCTURE OF BILAYER CrI<sub>3</sub>

Figure 6 shows the band structure and MSC conductivity calculated from routes I and II using the U(2)-invariant formula, where the electronic structures are obtained from different first-principles packages while other parameters are essentially the same. The electronic structures calculated from the two packages are slightly different in many aspects including bandgap values and band dispersion as shown in Fig. 6(a). For instance, the band gap is 0.89 eV from VASP and 0.51 eV from OpenMX. Even though the scissor operator has been applied to align the bandgap in Fig. 4(b), different features in MSC are still observed re-emphasizing that the optical current is sensitive to the electronic structure and it is difficult to compare results from different packages.

- 
- [1] V. M. Fridkin, *Crystallogr. Rep. (Transl. Kristallografiya)* **46**, 654 (2001).
- [2] A. M. Glass, D. von der Linde, and T. J. Negran, *Appl. Phys. Lett.* **25**, 233 (1974).
- [3] W. Koch, R. Munser, W. Ruppel, and P. Würfel, *Solid State Commun.* **17**, 847 (1975).
- [4] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich, *Phys. Rev. Lett.* **7**, 118 (1961).
- [5] S. M. Young and A. M. Rappe, *Phys. Rev. Lett.* **109**, 116601 (2012).
- [6] T. Morimoto and N. Nagaosa, *Sci. Adv.* **2**, e1501524 (2016).
- [7] N. Nagaosa and T. Morimoto, *Adv. Mater.* **29**, 1603345 (2017).
- [8] J. Ahn, G.-Y. Guo, and N. Nagaosa, *Phys. Rev. X* **10**, 041041 (2020).
- [9] G. B. Osterhoudt, L. K. Diebel, M. J. Gray, X. Yang, J. Stanco, X. Huang, B. Shen, N. Ni, P. J. W. Moll, Y. Ran, and K. S. Burch, *Nat. Mater.* **18**, 471 (2019).
- [10] F. de Juan, A. G. Grushin, T. Morimoto, and J. E. Moore, *Nat. Commun.* **8**, 15995 (2017).
- [11] D. Rees, K. Manna, B. Lu, T. Morimoto, H. Borrmann, C. Felser, J. E. Moore, D. H. Torchinsky, and J. Orenstein, *Sci. Adv.* **6**, eaba0509 (2020).
- [12] Y. Zhang, T. Holder, H. Ishizuka, F. de Juan, N. Nagaosa, C. Felser, and B. Yan, *Nat. Commun.* **10**, 3783 (2019).
- [13] H. Wang and X. Qian, *npj Comput. Mater.* **6**, 199 (2020).
- [14] R. Fei, W. Song, and L. Yang, *Phys. Rev. B* **102**, 035440 (2020).
- [15] J. Reif, J. C. Zink, C.-M. Schneider, and J. Kirschner, *Phys. Rev. Lett.* **67**, 2878 (1991).
- [16] J. Reif, C. Rau, and E. Matthias, *Phys. Rev. Lett.* **71**, 1931 (1993).
- [17] A. Kirilyuk, *J. Phys. D* **35**, R189 (2002).
- [18] A. Kirilyuk and T. Rasing, *J. Opt. Soc. Am. B* **22**, 148 (2005).
- [19] M. Fiebig, D. Fröhlich, B. B. Krichevstov, and R. V. Pisarev, *Phys. Rev. Lett.* **73**, 2127 (1994).
- [20] M. Fiebig, V. V. Pavlov, and R. V. Pisarev, *J. Opt. Soc. Am. B* **22**, 96 (2005).
- [21] Z. Sun, Y. Yi, T. Song, G. Clark, B. Huang, Y. Shan, S. Wu, D. Huang, C. Gao, Z. Chen, M. McGuire, T. Cao, D. Xiao, W.-T. Liu, W. Yao, X. Xu, and S. Wu, *Nature (London)* **572**, 497 (2019).
- [22] W. Hübner and K.-H. Bennemann, *Phys. Rev. B* **40**, 5973 (1989).
- [23] W. Hübner, *Phys. Rev. B* **42**, 11553 (1990).
- [24] R.-P. Pan, H. D. Wei, and Y. R. Shen, *Phys. Rev. B* **39**, 1229 (1989).
- [25] V. N. Muthukumar, R. Valentí, and C. Gros, *Phys. Rev. B* **54**, 433 (1996).
- [26] Y. Tanabe, M. Muto, and E. Hanamura, *Solid State Commun.* **102**, 643 (1997).
- [27] M. Muto, Y. Tanabe, T. Iizuka-Sakano, and E. Hanamura, *Phys. Rev. B* **57**, 9586 (1998).
- [28] R. D. King-Smith and D. Vanderbilt, *Phys. Rev. B* **47**, 1651 (1993).
- [29] R. Resta, *Rev. Mod. Phys.* **66**, 899 (1994).
- [30] J. E. Sipe and E. Ghahramani, *Phys. Rev. B* **48**, 11705 (1993).
- [31] H. Watanabe and Y. Yanase, *Phys. Rev. X* **11**, 011001 (2021).
- [32] C. Aversa and J. E. Sipe, *Phys. Rev. B* **52**, 14636 (1995).
- [33] J. E. Sipe and A. I. Shkrebtii, *Phys. Rev. B* **61**, 5337 (2000).
- [34] E. Ghahramani, D. J. Moss, and J. E. Sipe, *Phys. Rev. B* **43**, 8990 (1991).
- [35] S. Sharma, J. K. Dewhurst, and C. Ambrosch-Draxl, *Phys. Rev. B* **67**, 165332 (2003).
- [36] S. N. Rashkeev, W. R. L. Lambrecht, and B. Segall, *Phys. Rev. B* **57**, 3905 (1998).
- [37] X. Gonze, F. Jollet *et al.*, *Comput. Phys. Commun.* **205**, 106 (2016).
- [38] J. Lin, M.-H. Lee, Z.-P. Liu, C. Chen, and C. J. Pickard, *Phys. Rev. B* **60**, 13380 (1999).
- [39] Z. Fang, J. Lin, R. Liu, P. Liu, Y. Li, X. Huang, K. Ding, L. Ning, and Y. Zhang, *CrystEngComm* **16**, 10569 (2014).
- [40] H. Wang and X. Qian, *Nano Lett.* **17**, 5027 (2017).
- [41] W. Song, G.-Y. Guo, S. Huang, L. Yang, and L. Yang, *Phys. Rev. Appl.* **13**, 014052 (2020).
- [42] G. Y. Guo and J. C. Lin, *Phys. Rev. B* **72**, 075416 (2005).
- [43] C. Wang, X. Liu, L. Kang, B.-L. Gu, Y. Xu, and W. Duan, *Phys. Rev. B* **96**, 115147 (2017).
- [44] B. Zhang, X. Zhang, J. Yu, Y. Wang, K. Wu, and M.-H. Lee, *Chem. Mater.* **32**, 6772 (2020).
- [45] W. Song, R. Fei, L. Zhu, and L. Yang, *Phys. Rev. B* **102**, 045411 (2020).
- [46] J. P. Provost and G. Vallee, *Commun. Math. Phys.* **76**, 289 (1980).
- [47] S. V. Gallego, J. Etxebarria, L. Elcoro, E. S. Tasci, and J. M. Perez-Mato, *Acta Crystallogr. Sect. A* **75**, 438 (2019).

- [48] J. Shi, P. Zhang, D. Xiao, and Q. Niu, *Phys. Rev. Lett.* **96**, 076604 (2006).
- [49] S. M. Young, F. Zheng, and A. M. Rappe, *Phys. Rev. Lett.* **110**, 057201 (2013).
- [50] R. Fei, X. Lu, and L. Yang, [arXiv:2006.10690](https://arxiv.org/abs/2006.10690).
- [51] R. W. Boyd, *Nonlinear Optics* (Academic Press, San Diego, CA, 2020).
- [52] P. S. Pershan, *Phys. Rev.* **130**, 919 (1963).
- [53] A. M. Cook, B. M. Fregoso, F. de Juan, S. Coh, and J. E. Moore, *Nat. Commun.* **8**, 14176 (2017).
- [54] J. Ibañez-Azpiroz, S. S. Tsirkin, and I. Souza, *Phys. Rev. B* **97**, 245143 (2018).
- [55] C. Wang, S. Zhao, X. Guo, X. Ren, B.-L. Gu, Y. Xu, and W. Duan, *New J. Phys.* **21**, 093001 (2019).
- [56] G. Kresse and J. Furthmüller, *Phys. Rev. B* **54**, 11169 (1996).
- [57] T. Ozaki, *Phys. Rev. B* **67**, 155108 (2003).
- [58] T. Ozaki and H. Kino, *Phys. Rev. B* **69**, 195113 (2004).
- [59] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [60] G. Kresse and D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).
- [61] I. Morrison, D. M. Bylander, and L. Kleinman, *Phys. Rev. B* **47**, 6728 (1993).
- [62] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, *Phys. Rev. B* **57**, 1505 (1998).
- [63] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010).
- [64] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, *Comput. Phys. Commun.* **185**, 2309 (2014).
- [65] J. Li, Y. Li, S. Du, Z. Wang, B.-L. Gu, S.-C. Zhang, K. He, W. Duan, and Y. Xu, *Sci. Adv.* **5**, eaaw5685 (2019).
- [66] N. Sivadas, S. Okamoto, X. Xu, C. J. Fennie, and D. Xiao, *Nano Lett.* **18**, 7658 (2018).
- [67] V. K. Gudelli and G.-Y. Guo, *Chin. J. Phys.* **68**, 896 (2020).
- [68] G. B. Ventura, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, *Phys. Rev. B* **96**, 035431 (2017).
- [69] R. von Baltz and W. Kraut, *Phys. Rev. B* **23**, 5590 (1981).
- [70] W. Kraut and R. von Baltz, *Phys. Rev. B* **19**, 1548 (1979).
- [71] N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, and N. P. Ong, *Rev. Mod. Phys.* **82**, 1539 (2010).