Editors' Suggestion

Anomalous shift and optical vorticity in the steady photovoltaic current

Penghao Zhu [1,2] and A. Alexandradinata [3]

[1]*Department of Physics and Institute for Condensed Matter Theory, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

[2]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*

[3]*Department of Physics and Santa Cruz Materials Center, University of California Santa Cruz, Santa Cruz, California 95064, USA*

(Received 14 October 2023; accepted 20 August 2024; published 5 September 2024)

Steady illumination of a noncentrosymmetric semiconductor results in a bulk photovoltaic current, which is contributed by real-space displacements ("shifts") of charged quasiparticles as they transit between Bloch states. The shift induced by interband excitation via absorption of photons has received the prevailing attention. However, this excitation-induced shift can be far outweighed ($\ll$) by the shift induced by intraband relaxation, or by the shift induced by radiative recombination of electron-hole pairs. This outweighing ($\ll$) is attributed to (i) time-reversal-symmetric, intraband Berry curvature, which results in an anomalous shift of quasiparticles as they scatter with phonons, as well as to (ii) topological singularities in the interband Berry phase ("optical vortices"), which makes the photovoltaic current extraordinarily sensitive to the linear polarization vector of the light source. Both (i) and (ii) potentially lead to nonlinear conductivities of order mA V^{-2}, without fine-tuning of the incident radiation frequency, band gap, or joint density of states. A case study of BiTeI showcases the anomalous shift and optical vorticity in a realistic material.

DOI: 10.1103/PhysRevB.110.115108

I. INTRODUCTION

Light that is harvested for large-scale power transmission needs to be rectified, i.e., converted from electromagnetic waves at solar frequencies to a direct or low-frequency current. Rectification in a noncentrosymmetric, nonmagnetic semiconductor results in a bulk direct current that is proportional to the radiation intensity, in the lowest-order response. This *bulk photovoltaic current* has a contribution attributed to an asymmetry in the fermionic quasiparticle distribution [1–3] and a second contribution attributed with the real-space displacements (or "shifts") of quasiparticles as they transit between Bloch states; cf. Fig. 1 [4,5].

The shift induced by interband *exc*itation via absorption of photons—in short, $\text{shift}_{\text{exc}}$—has received the prevailing attention. However, the steady photovoltaic current is also contributed by a shift induced by *rec*ombination of electron-hole pairs, as well as a shift induced by *intra*band relaxation via scattering with phonons or impurities. Both $\text{shift}_{\text{rec}}$ and $\text{shift}_{\text{intra}}$ have been emphasized by Belinicher-Ivchenko-Sturman (BIS) in their kinetic theory of the shift current, which accounts for the steady, nonequilibrium quasiparticle distribution [5]. In contrast, $\text{shift}_{\text{rec}}$ and $\text{shift}_{\text{intra}}$ have been ignored in all recent literature, which either (a) disregarded relaxation completely [6–12], (b) were agnostic about the nature of the relaxation mechanism, e.g., by naive relaxation-time approximations [13–16], or (c) adopted relaxation mechanisms that are unrealizable in experiments, e.g., by scattering with a "fermionic bath" [17–19].[1] The one-sided interpretation of shift currents as a dissipationless "hot carrier effect" [9] cannot explain the vanishing photocurrent in the low-temperature polar phase of organic charge-transfer complexes [21]. To recapitulate, excitation, recombination, and relaxation induce shifts that may counteract or synergize, and a complete model of the kinetic processes is required to quantitatively predict the steady photovoltaic current [22].

While the $\text{shift}_{\text{intra}}/\text{shift}_{\text{rec}}$ currents have been explored for simplified models of piezoelectrics and pyroelectrics [5]. there has not been an attempt to relate the $\text{shift}_{\text{intra}}/\text{shift}_{\text{rec}}$ currents to notions of quantum wave-function geometry that have revitalized the condensed-matter field. Here, we identify scenarios (unimagined by BIS) in which the $\text{shift}_{\text{intra}}$ or $\text{shift}_{\text{rec}}$ current dominates over the $\text{shift}_{\text{exc}}$ current by an order of magnitude. Such dominance is attributed to two quantum geometric properties of the Bloch wave function, namely (i) time-reversal-symmetric, *intra*band Berry curvature [cf. Fig. 2(a)], which results in an anomalous shift of quasiparticles as they are scattered with phonons, and (ii) topological singularities in the *inter*band Berry phase known as "optical vortices" [cf. Fig. 2(c)], which make the photovoltaic current extraordinarily sensitive to the linear polarization vector of the light source.

Both effects (i) and (ii) will be demonstrated in model Hamiltonians with generic values of the joint density of states and without assuming a semimetallic band gap. The nonlinear conductivities in our models are of order mA V^{-2} without fine-tuning of the incident radiation frequency,[2] as illustrated

[1](d) Barik and Sau have considered electron-phonon scattering as a relaxation mechanism for the shift current [20]. However, they assumed without justification that phonon-mediated scattering does not result in a shift. A detailed criticism of the recent literature is provided in Appendix C 3.

[2]In contrast, a number of proposals for large shift currents (at low frequencies) have relied on small [23] or vanishing [24–26] band gaps, which makes for a questionable application to solar cells.

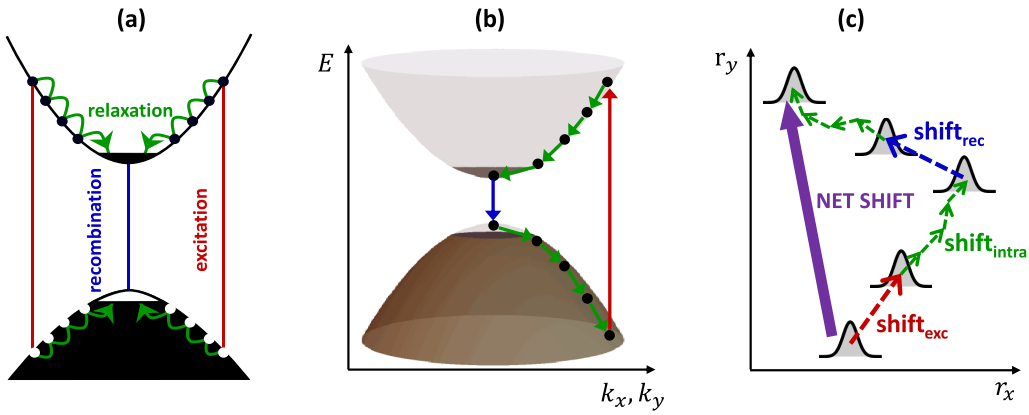

FIG. 1. (a) The kinetic processes of excitation, relaxation, and recombination in a steady state of a homogeneously illuminated semiconductor. (b) These kinetic processes can be recast as loops in energy-momentum space if one views a hole as an electron going backward in time. To each transition between Bloch states we associate a shift of a wave packet in real space, as illustrated representatively by the dashed arrows in panel (c); the net shift over all possible loops results in the steady shift current.

in Figs. 2(b) and 2(d). For comparison with a prototypical ferroelectric, the nonlinear conductivity of $PbTiO_3$ has a maximum (over frequency) of 0.05 mA V^{-2} when only $\text{shift}_{\text{exc}}$ is accounted for [8].

The paper is outlined as follows: As a preliminary step to substantiating these results, Sec. II first reviews BIS's kinetic theory for the shift current [5] and several salient properties of the nonequilibrium distribution of photoexcited carriers [27]. In addition, we will formalize an underappreciated distinction between the transient and steady shift currents; in particular, the transient shift current in intrinsic semiconductors will be shown to be identical to the current calculated by Kubo-type perturbation theories (e.g., by Kraut and Baltz [4,13] and by Sipe and Shkrebtii [6]), which assume a weak perturbation from thermal equilibrium. The difference between the steady and transient shift currents will turn out to be the sum of the $\text{shift}_{\text{intra}}$ and $\text{shift}_{\text{rec}}$ currents. Section III demonstrates the relevance of $\text{shift}_{\text{intra}}$ in the presence of time-reversal-symmetric intraband Berry curvature. Section IV demonstrates the relevance of $\text{shift}_{\text{rec}}$ in the presence of optical vortices. Because Sec. III also introduces our method of calculating the shift current via loops, we recommend that Sec. III be read before Sec. IV. Section V showcases the importance of both $\text{shift}_{\text{intra}}$ and $\text{shift}_{\text{rec}}$ in the 3D polar semiconductor BiTeI, which has an appreciable Berry curvature as well as optical vorticity. Finally, Sec. VI summarizes our results, gives directions to finding photovoltaic material with the desired wave-function geometry, comments on experimental discrepancies between the transient and steady photovoltaic current, elaborates on the notion of loop currents, and discusses the potential of shift-current materials for solar-cell applications.

II. KINETIC THEORY OF THE SHIFT CURRENT

The BIS kinetic theory presupposes that carrier-optical-phonon scattering (rather than carrier-carrier scattering) is the dominant mechanism of energy relaxation for photoexcited carriers in the "active region." A carrier is said to be in the active region if its energy (defined with respect to the conduction-band/valence-band extremum for an electron/hole) exceeds the optical phonon threshold: $E > \hbar\Omega_o$, as illustrated by the yellow energy intervals in Fig. 3(b). The dominance of carrier-optical-phonon scattering over carrier-carrier scattering occurs for not-too-high carrier densities, which is typical of most continuous-wave laser experiments [27,28].

The BIS formula for the shift current can be compactly expressed as

$$\boldsymbol{j} = -\frac{|e|}{\mathcal{V}} \sum_{B,B',m} \boldsymbol{S}^m_{B'\leftarrow B}\left(\mathcal{A}^m_{B'\leftarrow B} - \mathcal{E}^m_{B\leftarrow B'}\right);$$
$$B = (b\boldsymbol{k}), \quad B' = (b'\boldsymbol{k}') \quad m = (\boldsymbol{q}p), \tag{1}$$

with $|e|$ the absolute value of the electron charge, $\mathcal{V}$ the volume of the medium, and $B = (b\boldsymbol{k})$ a collective label for a Bloch state in band b with wave vector $\boldsymbol{k}$. $\sum_{B,B',m}$ sums over all possible quasiparticle transitions $(B' \leftarrow B)$ mediated by a boson of mode m; $m = (\boldsymbol{q}p)$ is specified by a bosonic wave vector $\boldsymbol{q}$ and a bosonic branch/band p. $\boldsymbol{S}$ and $\mathcal{A} - \mathcal{E}$, which appear in the summand of $\sum_{B,B',m}$, will be explained in turn.

The shift vector $\boldsymbol{S}^m_{B'\leftarrow B}$ is the real-space displacement of a Bloch quasiparticle as it transits from B to B', by way of absorbing/emitting a boson of mode m. For phonons,

$$\text{Phonon}: \quad \boldsymbol{S}^m_{B'\leftarrow B} = -(\nabla_{\boldsymbol{k}} + \nabla_{\boldsymbol{k}'}) \arg V^m_{B',B} + \boldsymbol{A}_{b'b'\boldsymbol{k}'} - \boldsymbol{A}_{bb\boldsymbol{k}}$$
$$= -\boldsymbol{S}^{-m}_{B\leftarrow B'}, \tag{2}$$

with $V^m_{B'B}$ being the electron-phonon matrix element [cf. Eq. (C17)], $\boldsymbol{A}_{bb'\boldsymbol{k}} = \langle u_{b\boldsymbol{k}} | i\nabla_{\boldsymbol{k}} u_{b'\boldsymbol{k}}\rangle_{\text{cell}}$ the Berry connection,[3] and $-m = (-\boldsymbol{q}p)$ being the momentum-reversed partner of $m = (\boldsymbol{q}p)$.[4] For a photonic mode m with linear polarization

[3] $u_{b\boldsymbol{k}}$ denotes the cell-periodic component of the Bloch function; the inner product $\langle x|y\rangle_{\text{cell}}$ involves an integral over the intracell coordinate; cf. Eq. (C3).

[4] In the BIS paper, all mode indices were omitted from their phononic shift, and their electron-phonon-matrix element was never explicitly defined, but one may guess that $\boldsymbol{S}^{\boldsymbol{k}'-\boldsymbol{k},p}_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}}$ in our notation corresponds to $\boldsymbol{R}_b(\boldsymbol{k}', \boldsymbol{k})$ in their notation.

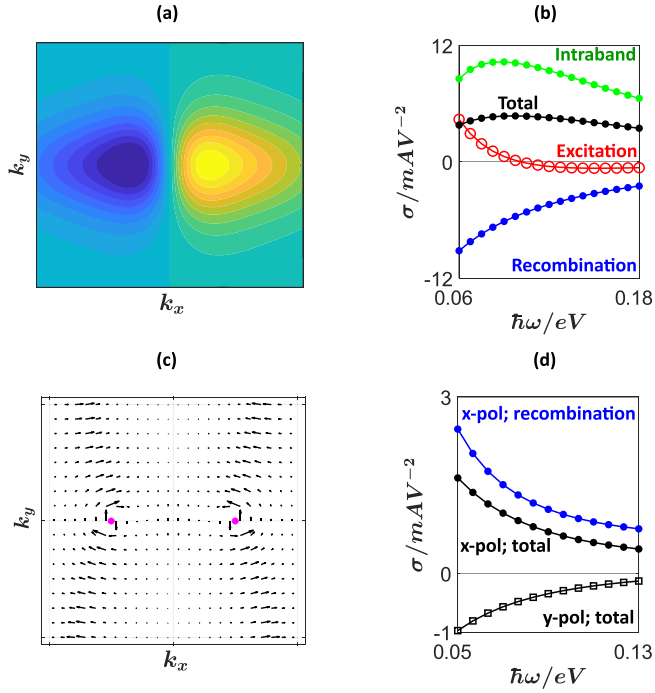

FIG. 2. Panels (a) and (b) derive from a model Hamiltonian with intraband Berry curvature, and (c) and (d) derive from a different model Hamiltonian with optical vortices. In panel (a), a time-reversal-symmetric distribution of intraband Berry curvature is plotted as a color field, with yellow (blue) representing positive (negative) curvature. Panel (b) plots the nonlinear conductivity, which is defined though $j_a = \boldsymbol{\sigma}^a_\omega |\mathcal{E}_\omega|^2$, with $\mathcal{E}_\omega$ the amplitude of an incident electric wave of frequency ω, assuming that the light source is unpolarized. The black curve represents the net conductivity $\boldsymbol{\sigma}^y_\omega$, while the red, green, and blue curves represent the components of $\boldsymbol{\sigma}^y_\omega$ contributed by interband excitation, intraband relaxation, and interband recombination, respectively. The intraband component is manifestly dominant. (c) When a Bloch quasiparticle (with wave vector $\boldsymbol{k}$) is optically excited from the valence to the conduction band, the quasiparticle is displaced by a $\boldsymbol{k}$-dependent $\text{shift}_{\text{exc}}$ vector [Eq. (3)]; the circulation of this vector field has a quantized contribution attributed to optical vortices. (d) The black curves represent $\boldsymbol{\sigma}^y_{\vec{x},\omega}$ for an $\vec{x}$-polarized light source vs $\boldsymbol{\sigma}^y_{\vec{y},\omega}$ for a $\vec{y}$-polarized light source; the blue curve represents the recombination component of $\boldsymbol{\sigma}^y_{\vec{x},\omega}$, which dominates the total current.

vector $\boldsymbol{\epsilon}_m$,

$$\text{Photon}: \quad \boldsymbol{S}^m_{b'\boldsymbol{k}\leftarrow b\boldsymbol{k}} = -\nabla_{\boldsymbol{k}} \arg \boldsymbol{\epsilon}_m \cdot \boldsymbol{A}_{b'b\boldsymbol{k}} + \boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}} = -\boldsymbol{S}^{-m}_{b\boldsymbol{k}\leftarrow b'\boldsymbol{k}}. \tag{3}$$

We have assumed that the photon wavelength greatly exceeds the lattice period; within the dipole approximation, photon-mediated transitions are *vertical* $\equiv \boldsymbol{k}$-preserving [cf. the red and blue lines in Fig. 3(b)], and the shift vector depends on $m = (\boldsymbol{q}p)$ only through $\boldsymbol{\epsilon}_m$. For this reason, we often use $\boldsymbol{S}^m_{b'\boldsymbol{k}\leftarrow b\boldsymbol{k}} \equiv \boldsymbol{S}^{\boldsymbol{\epsilon}_m}_{b'\boldsymbol{k}\leftarrow b\boldsymbol{k}}$ synonymously, when m is photonic. Equation (3) is henceforth referred to as the *photonic shift vector*, and Eq. (2) as the *phononic shift vector*. In either case, the sign of the shift vector is inverted if the Bloch labels are interchanged and the bosonic wave vector simultaneously inverted: $\boldsymbol{q} \to -\boldsymbol{q}$, reflecting that forward and backward transitions (between the same pair of Bloch states) result in opposite shifts.

For either type of boson, $\mathcal{A}^m_{B'\leftarrow B}$ ($\mathcal{E}^m_{B\leftarrow B'}$) is the transition probability rate for absorbing (emitting) a boson of mode m. As explicitly written in Eqs. (A3) and (A5), both $\mathcal{A}$ and $\mathcal{E}$ have the Golden Rule forms that are familiar from Dirac's perturbation theory [29]. In particular, $\mathcal{A}^m \propto N_m$ and $\mathcal{E}^m \propto (N_m + 1)$, with N_m the average occupancy of the boson m. For phonons, N_m is assumed to follow the Planck distribution with lattice temperature T_l; for photons, N_m is a sum of thermal and nonthermal contributions, with the latter being generated by the light source. Additionally, both $\mathcal{A}$ and $\mathcal{E}$ depend on the quasiparticle distribution functions f_B in a manner consistent with Pauli's exclusion principle: $\mathcal{A}^m_{B'\leftarrow B} \propto (1 - f_{B'})f_B$ and $\mathcal{E}^m_{B\leftarrow B'} \propto (1 - f_B)f_{B'}$. Consequently, $\boldsymbol{j} = \boldsymbol{j}[f_B, N_m]$ depends on f_B and N_m through $\mathcal{A}$ and $\mathcal{E}$; however, the dependence on N_m will subsequently be made implicit: $\boldsymbol{j}[f_B, \cancel{N_m}]$, to simplify notation.

Our expression for the shift current is derived in Appendix C 1 and is slightly more general than the expression presented in the BIS paper [5], in that ours allows for interband, phonon-mediated transitions while theirs does not.

Let us consider three scenarios for the quasiparticle and bosonic distributions:

(I) *No source.* Without photoexcitation by a source, quasiparticles, phonons, and photons are all thermalized with an equilibrium temperature T_0, and Eq. (1) manifestly vanishes due to detailed balance: $\mathcal{A}^m_{B'\leftarrow B} = \mathcal{E}^m_{B\leftarrow B'}$; cf. Eq. (A15).

(II) *Transient state.* This balancing is disrupted when the light source is switched on. At the onset of radiation, the quasiparticle distribution retains its equilibrium value (the Fermi-Dirac function $f^{T_0}_B$), but the nonthermal photons drive a transient current $\boldsymbol{j}_{\text{tran}} = \boldsymbol{j}[f^{T_0}_B]$, which is purely attributed to vertical, interband transitions throughout the *excitation surface*, as illustrated in Fig. 3(a). The excitation surface (ES) is defined as the surface in the Brillouin zone where the difference in conduction- and valence-band energies equals the source photon energy: $E_{c\boldsymbol{k}} - E_{v\boldsymbol{k}} = \hbar\omega$. The meaning of being "purely attributed" is that the sum over all quasiparticle transitions [in Eq. (1)] is contributed nontrivially only by vertical transitions throughout the excitation surface, i.e., the value of $\boldsymbol{j}[f^{T_0}_B]$ does not change if the summation $\sum_{\boldsymbol{k},\boldsymbol{k}'}$ is restricted such that $\boldsymbol{k} = \boldsymbol{k}'$ lies on the excitation surface:[5]

$$\boldsymbol{j}_{\text{tran}} = \boldsymbol{j}\big[f^{T_0}_B\big] = \boldsymbol{j}\big[f^{T_0}_B\big]_{\boldsymbol{k}=\boldsymbol{k}'\in\text{ES}}. \tag{4}$$

If the nonthermal photons are well approximated by a classical electromagnetic wave, then $\boldsymbol{j}[f^{T_0}_B]$ reduces to the Kraut-Baltz-Sipe-Shkrebtii formula [4,6,13] calculated by Kubo-type perturbation theory, as demonstrated in Appendix C 3.

(III) *Steady state.* For $t \gg \tau_{\text{rec}} \sim 1$ ns (a typical timescale for radiative interband recombination [27,30]), the transient current evolves to a steady current: $\boldsymbol{j}[f_B]$, with the nonequilibrium distribution f_B being the steady solution to a kinetic equation encoding all the processes in Fig. 1(a); cf. Appendix B 2. The difference between the equilibrium $f^{T_0}_B$ and the nonequilibrium f_B is caricatured in Fig. 3(a) versus 3(b).

[5]This is proven more elaborately in Appendix E.

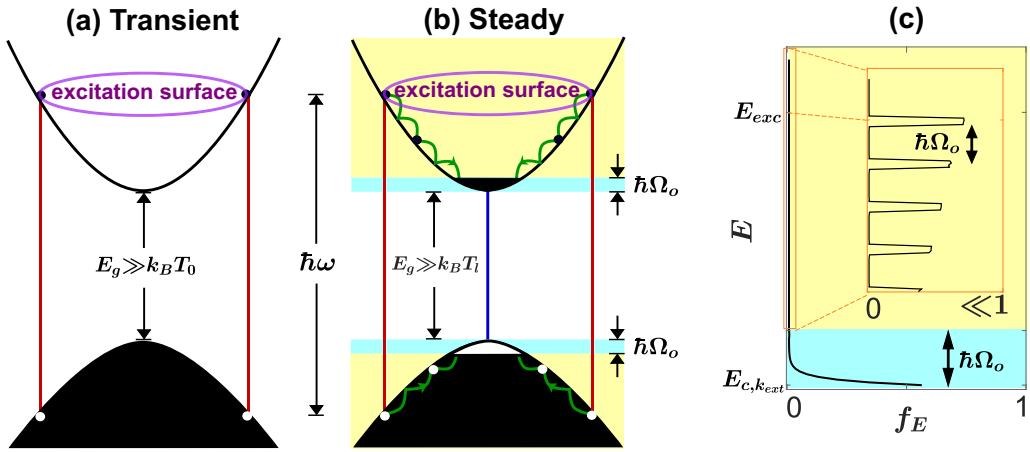

FIG. 3. (a) The excitation process in a transient state of a homogeneously illuminated semiconductor. White dots against a black background represent hole carriers. The purple ellipse should be understood as a cross-section of an ellipsoidal excitation surface. Panel (b) represents the steady state; yellow and blue energy intervals indicate the active and passive regions respectively. (c) A representative illustration of the quasiparticle distribution $f_{c\boldsymbol{k}}$ in the conduction band, with f_E being the average of $f_{c\boldsymbol{k}}$ for all $\boldsymbol{k}$ satisfying $E_{c\boldsymbol{k}} = E$. The inset displays the same function f_E in the active region, but with a much finer scale for the horizontal axis. The shape of this distribution is supported by theoretical models (cf. Appendix B 2) and hot-carrier photoluminescence spectroscopy (e.g., Fig. 25 in Ref. [28]).

Henceforth, $\boldsymbol{j}[f]$ and $\boldsymbol{j}[f^{T_0}]$ will be our shorthand for the steady and transient currents, respectively.

While in principle the BIS formula [Eq. (1)] for $\boldsymbol{j}[f]$ sums over all possible quasiparticle transitions, it is worthwhile in practice to identify the *predominant* transitions that make an outsized contribution to the summation; throughout this work, our use of "predominant" should be understood as significantly contributing to the *steady*, nonequilibrium shift current.

It is simplest to consider the predominant transitions in an intrinsic, direct-gap semiconductor, with $E_{c\boldsymbol{k}} - E_{v\boldsymbol{k}}$ minimized at a single wave vector: $E_{c\boldsymbol{k}_{\rm ext}} - E_{v\boldsymbol{k}_{\rm ext}} = E_g$. The band gap is assumed to exceed the optical phonon energy ($E_g \gg \hbar\Omega_o$), such that phonon-mediated transitions are intraband; cf. the green curves in Fig. 3(b). A typical electron–optical-phonon scattering time is $\tau^o \sim 100$ fs [31,32]. The lattice temperature is assumed to be small ($k_B T_l \ll E_g, \hbar\Omega_o$), so that the emission of optical phonons outweighs the absorption. Supposing that carriers are optically excited into the active region with energy $E_{\rm exc}$, then the transitions illustrated in Fig. 3(b) predominate. Indeed, the vast difference in relaxation timescales—$\tau^o \lll \tau_{\rm rec}$ [27,28]—favors fast, intraband transitions by emission of optical phonons with energy $\geqslant \hbar\Omega_o$; carriers quickly relax into a *passive region*, defined as the energy interval near a band extremum where the carrier energy $E < \hbar\Omega_o$ [cf. the blue interval in Fig. 3(a)]; carriers in the passive region can no longer relax via optical phonons, and they remain in the passive region until they are annihilated in the slower process of radiative recombination.[6] The majority of photoexcited carriers are thus contained within the passive region, with a steady distribution that depends on whether electron-electron scattering or electron–acoustic-phonon scattering is the dominant mechanism for energy relaxation in the passive region [Fig. 3(c)]. However, fine-grained details about the carrier distribution within the passive region do not matter when estimating the shift current, because the optical phonon threshold $\hbar\Omega_o$ is typically a small fraction of the band gap E_g, and E_g is the energy scale for significant variations of the energy-dependent shift vectors.[7] For additional details on the predominant relaxation mechanisms in a direct-gap semiconductor, we refer the reader to Appendixes A and B.

By decomposing the BIS formula [Eq. (1)] according to the three classes of transitions sketched in Fig. 1(a), one obtains $\boldsymbol{j}[f] = \boldsymbol{j}_{\rm exc} + \boldsymbol{j}_{\rm intra} + \boldsymbol{j}_{\rm rec}$, which is the precise meaning of the shift$_{\rm exc}$, shift$_{\rm intra}$, and shift$_{\rm rec}$ currents mentioned colloquially in the Introduction. To clarify, the *intraband current* is extracted from Eq. (1) by restricting the band summations $\sum_{b,b'}$ by the condition $b = b'$: $\boldsymbol{j}_{\rm intra} = \boldsymbol{j}[f]_{b=b'}$; the *excitation-induced current* is extracted by restricting the wave-vector summations $\sum_{\boldsymbol{k},\boldsymbol{k}'}$ with the condition that $\boldsymbol{k} = \boldsymbol{k}'$ lies on the excitation surface: $\boldsymbol{j}_{\rm exc} = \boldsymbol{j}[f]_{\boldsymbol{k}=\boldsymbol{k}'\in{\rm ES}}$; and the *recombination-induced current* $\boldsymbol{j}_{\rm rec}$ restricts $\boldsymbol{k} = \boldsymbol{k}'$ to lie outside the excitation surface. Explicit expressions for the threefold decomposition of $\boldsymbol{j}$ are given in Appendix A 4.

[6] Auger recombination empirically occurs at much higher photoexcited carrier densities than the present consideration [27].

[7] The majority of recombination transitions occur at $\boldsymbol{k}$ points close to $\boldsymbol{k}_{\rm ext}$ and contained within the passive region. Each recombination transition is associated with a photonic shift $\boldsymbol{S}^{\epsilon}_{v\boldsymbol{k}\leftarrow c\boldsymbol{k}}$ [Eq. (3)], which may as well be approximated as $\boldsymbol{S}^{\epsilon}_{v\boldsymbol{k}_{\rm ext}\leftarrow c\boldsymbol{k}_{\rm ext}}$, because the variation of the photonic shift vector within the passive region is small. Likewise, the current induced by phonon-mediated transitions within the passive region is outweighed by the current induced by phonon-mediated transitions outside the passive region, assuming that the active region is much bigger than the passive region. This assumption holds for most radiation frequencies, because $\hbar\Omega_o$ is a tiny fraction of the bandwidth. To formalize this discussion, one may split the line integral in Eq. (9) to a short-line integral within the passive region and a long-line integral within the active region; the long-line integral dominates, because the Berry curvature typically varies on the scale of $E_g \gg \hbar\Omega_o$.

Let us argue that $\boldsymbol{j}_{\rm exc}$ in a steady state is well approximated by $\boldsymbol{j}_{\rm tran}$ in a transient state. Our argument relies on the following property of the nonequilibrium quasiparticle distributions: for all $\boldsymbol{k} \in$ ES, $f_{c\boldsymbol{k}} \ll 1$ and $(1 - f_{v\boldsymbol{k}}) \ll 1$, as illustrated by the inset in Fig. 3(c). The smallness of $f_{c\boldsymbol{k}}$ and $(1 - f_{v\boldsymbol{k}})$ originates from the slowness in optical excitations compared to the fastness of inelastic collisions by carrier-carrier and carrier-phonon scatterings. In other words, despite the continuous generation of electron-hole pairs by photon absorption, inelastic scattering processes are so efficient that the nonequilibrium carrier distribution (over the excitation surface) never builds up to significance; this statement is derived rigorously in Appendix E. Thus for the purpose of computing the excitation-induced current ($\boldsymbol{j}_{\rm exc} = \boldsymbol{j}[f]_{\boldsymbol{k}=\boldsymbol{k}'\in \rm ES}$) in a nonequilibrium state, one may as well input the equilibrium distribution: $\boldsymbol{j}_{\rm exc} \approx \boldsymbol{j}[f^{T_0}]_{\boldsymbol{k}=\boldsymbol{k}'\in \rm ES}$, since it also holds that $f^{T_0}_{c\boldsymbol{k}} \ll 1$ and $(1 - f^{T_0}_{v\boldsymbol{k}}) \ll 1$, assuming $k_B T_0 \ll E_g$. Recalling a similar expression for the transient current in Eq. (4), we deduce that $\boldsymbol{j}_{\rm tran} \approx \boldsymbol{j}_{\rm exc}$, implying that $\boldsymbol{j}_{\rm intra} + \boldsymbol{j}_{\rm rec}$ is precisely what is missed from previous Kubo-type theories [4,6,13] that purport to calculate a steady shift current.

III. ANOMALOUS SHIFT

In connection to $\boldsymbol{j}_{\rm intra}$, our first main result is that the phononic shift induced by small-angle, intraband scattering is expressible in terms of the intraband Berry curvature:

$$\text{Phonon}: \quad \boldsymbol{S}^{m}_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}} = \boldsymbol{S}^{\rm ano}_{b;\boldsymbol{k}'\leftarrow\boldsymbol{k}} + O(\delta k^3),$$
$$\boldsymbol{S}^{\rm ano}_{b;\boldsymbol{k}'\leftarrow\boldsymbol{k}} = \boldsymbol{\Omega}_{b\boldsymbol{k}_{\rm ave}} \times \delta\boldsymbol{k}, \tag{5}$$

with the curvature defined as $\boldsymbol{\Omega}_{b\boldsymbol{k}} = \nabla \times \boldsymbol{A}_{bb\boldsymbol{k}}$; $\boldsymbol{k}_{\rm ave} = (\boldsymbol{k} + \boldsymbol{k}')/2$ and $\delta\boldsymbol{k} = \boldsymbol{k}' - \boldsymbol{k}$ are the average and difference in quasiparticle wave vectors, and $\delta k = ||\delta\boldsymbol{k}||$. The *anomalous shift* ($\boldsymbol{S}^{\rm ano}$) is purely a geometric property of the quasiparticle wave function, and is insensitive to the nature of the electron-phonon coupling; such coupling affects the shift current only through the transition rate; cf. Eq. (A3).

Our use of "anomalous" evokes a comparison with the anomalous velocity correction in the semiclassical equation of motion [33,34], which gives an anomalous displacement: $\delta\boldsymbol{r}_{\rm ano} = \Omega_{b\boldsymbol{k}} \times \delta\boldsymbol{k}$ for a wave packet of Bloch states in band b. In the photovoltaic context, $\delta\boldsymbol{k}$ is driven by a phonon-induced electric field rather than an externally applied field. Indeed, phonons in noncentrosymmetric semiconductors induce macroscopic electric fields, which cause the electron-phonon matrix element $V^{m}_{b\boldsymbol{k}',b\boldsymbol{k}}$ to diverge as $\delta\boldsymbol{k} \to \boldsymbol{0}$ [35]. In the self-consistent-field approximation [36], $V^{m}_{b\boldsymbol{k}',b\boldsymbol{k}} = f^{m}_{\delta\boldsymbol{k}} \langle u_{b\boldsymbol{k}'}|u_{b\boldsymbol{k}}\rangle_{\rm cell}$ plus asymptotically irrelevant terms; $f^{m}_{\delta\boldsymbol{k}}$ diverges as $1/\delta k$ for "polarization scattering" [37,38] with optical phonons, and as $1/\delta k^{1/2}$ for "piezoacoustic scattering" [39] with acoustic phonons.[8] (With the possible exception of small-gap semiconductors, the polarization and piezoacoustic scatterings typically dominate [35] over the deformation scattering [40].) Let us substitute $V^{m}_{b\boldsymbol{k}',b\boldsymbol{k}}$ in Eq. (2) with its asymptotically dominant contribution. Since the symmetrized derivative of any function of $\delta\boldsymbol{k}$ vanishes, we are led to evaluate $(\nabla_{\boldsymbol{k}} + \nabla_{\boldsymbol{k}'}) \arg \langle u_{b\boldsymbol{k}'}|u_{b\boldsymbol{k}}\rangle_{\rm cell} = \nabla_{\boldsymbol{k}_{\rm ave}}(\boldsymbol{A}_{bb\boldsymbol{k}_{\rm ave}} \cdot \delta\boldsymbol{k}) + O(\delta k^3)$. Subsequently applying the identities $(\nabla_{\boldsymbol{k}_{\rm ave}} \times \boldsymbol{A}) \times \delta\boldsymbol{k} = (\delta\boldsymbol{k} \cdot \nabla_{\boldsymbol{k}_{\rm ave}})\boldsymbol{A} - \nabla_{\boldsymbol{k}_{\rm ave}}(\boldsymbol{A} \cdot \delta\boldsymbol{k})$ and $\boldsymbol{A}_{\boldsymbol{k}'} - \boldsymbol{A}_{\boldsymbol{k}} = (\delta\boldsymbol{k} \cdot \nabla_{\boldsymbol{k}_{\rm ave}})\boldsymbol{A} + O(\delta k^3)$, one obtains the anomalous shift in Eq. (5).[9]

The anomalous shift induces a large $\boldsymbol{j}_{\rm intra}$ if the excitation surface encloses a time-reversal-symmetric distribution of Berry curvature. The minimal model to demonstrate this effect is quasi-two-dimensional, meaning that the band energies and cell-periodic wave functions $|u_B\rangle_{\rm cell}$ are approximately independent of one component of $\boldsymbol{k}$, say, k_z. Let us consider a quasi-2D excitation surface that encircles a 2π-quantum of Berry flux ($2\pi = \iint \Omega^{z}_{c\boldsymbol{k}} dk_x dk_y$) in the positive-$k_x$ half-plane [yellow region in Fig. 4(d)]; by time-reversal symmetry, the same excitation surface must encircle a (-2π)-quantum of Berry flux in the negative-k_x half-plane [cyan region in Fig. 4(d)]. In short, we simply say that the excitation surface encloses a *time-reversal-symmetric Berry flux* of 2π.[10]

How do $\boldsymbol{j}_{\rm intra}, \boldsymbol{j}_{\rm exc}$, and $\boldsymbol{j}_{\rm rec}$ compare in this minimal model? Before getting too quantitative, one may gain some qualitative insight from comparatively evaluating shift$_{\rm intra}$, shift$_{\rm exc}$, and shift$_{\rm rec}$ for the representative electron-hole trajectory in Fig. 4(a), which describes the photoexcitation of an electron-hole pair at $\boldsymbol{k}_{\rm exc}$ on the excitation surface, the relaxation of the excited electron (hole) in the conduction (valence) band, and recombination at $\boldsymbol{k}_{\rm ext}$. Viewing a forward-moving hole as a backward-moving electron, this electron-hole trajectory becomes an oriented electron loop, which we denote as loop[$\boldsymbol{k}_{\rm exc}$]. This loop concatenates two interband links with two intraband pathways, $p_c(\boldsymbol{k}_{\rm exc})$ and $p_v(\boldsymbol{k}_{\rm exc})$, which correspond, respectively, to the conduction and valence band; cf. Fig. 4(b). The net shift associated with this loop, which we call the *shift loop* $\equiv \boldsymbol{S}_{\rm loop[\boldsymbol{k}_{\rm exc}]}$, is the summation of shift

[8]The explicit expressions for f^m can be found in equation (3.12) of Ref. [36] in the case of "polarization scattering," and in the sum of (3.15) and (3.16) for the case of "piezoacoustic scattering." In the general case, f^m may have an anisotropic dependence on $\delta\boldsymbol{k}$, but this does not affect the power exponent of the divergence. In the case of optical phonons, the divergence is cut off by a minimal $\delta k_{\rm cut}$ which is determined by the minimal optical phonon energy $\hbar\Omega_o$; because $\hbar\Omega_o$ is much smaller than typical bandwidths, $\delta k_{\rm cut}$ is much smaller than the Brillouin-zone period. The author of Ref. [36], Vogl, dropped the factor $\langle u_{b\boldsymbol{k}'}|u_{b\boldsymbol{k}}\rangle_{\rm cell}$ from all their long-wavelength expressions for the electron-phonon matrix element [including (3.12), (3.15), and (3.16)], based on the fallacious belief that $\langle u_{b\boldsymbol{k}'}|u_{b\boldsymbol{k}}\rangle_{\rm cell} = 1 + O(\delta k^2)$ can be chosen as a gauge choice for the wave function. The error in this belief is explained in Appendix J.

[9]Appendix I describes an alternative derivation of the anomalous shift by identifying $-\nabla_{\boldsymbol{k}_{\rm ave}}(\boldsymbol{A}_{\boldsymbol{k}_{\rm ave}} \cdot \delta\boldsymbol{k}) + \boldsymbol{A}_{\boldsymbol{k}'} - \boldsymbol{A}_{\boldsymbol{k}}$ as a line integral of the Berry connection over an infinitesimally thin parallelogram centered at $\boldsymbol{k}_{\rm ave}$. This somewhat demystifies the appearance of the Berry curvature. Just as the intraband anomalous shift vector is expressible in terms of geometric quantities over an intraband loop in momentum space, so is the excitation shift vector (generalized to nonvertical transitions) expressible in terms of geometric quantities over an interband loop [41,42].

[10]In the quasi-2D context, the time-reversal-symmetric Berry flux is defined as $\iint \Omega^{z}_{c}\Theta(k_x > 0)dk_x dk_y$, with $\Theta(k_x > 0)$ a projector to positive values of k_x.

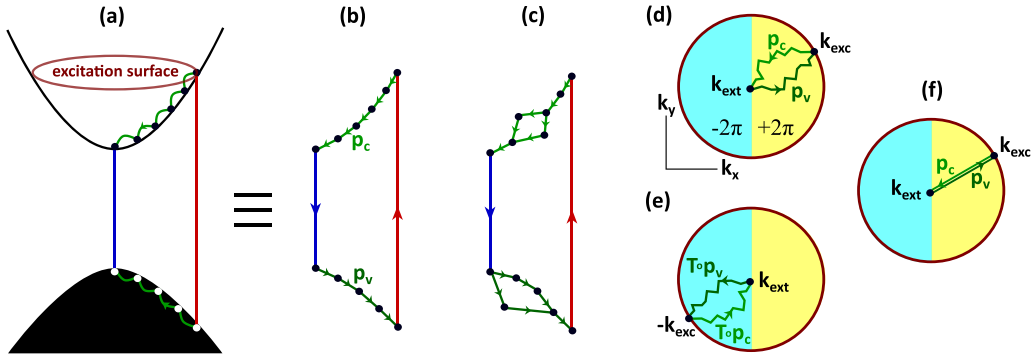

FIG. 4. Viewing a closed electron-hole trajectory [panel (a)] as an electron loop [panel (b)]. (c) An electron loop with a nonunique relaxation pathway. (d) Top-down view of the same electron loop in panel (b). The excitation surface encloses a 2π flux of Berry curvature ($\Omega^z_{c\boldsymbol{k}}$) in the yellow region, and a negative 2π flux in the cyan region. (e) The time-reversed counterpart to the loop in panel (d). (f) Approximating the relaxation pathways as geodesic paths.

vectors over all one-electron transitions that make up the loop:[11]

$$\boldsymbol{S}^{\boldsymbol{\epsilon}}_{\text{loop}[\boldsymbol{k}]} = \boldsymbol{S}^{\boldsymbol{\epsilon}}_{\text{exc},\boldsymbol{k}} + \boldsymbol{S}_{\text{rec}} + \boldsymbol{S}_{\text{intra},\boldsymbol{k}}, \tag{6}$$

$$\boldsymbol{S}^{\boldsymbol{\epsilon}}_{\text{exc},\boldsymbol{k}} = \boldsymbol{S}^{\boldsymbol{\epsilon}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}, \tag{7}$$

$$\boldsymbol{S}_{\text{rec}} = \left.\frac{\int d\lambda_{\hat{\boldsymbol{q}}} \sum_{p=1}^{2} |\boldsymbol{\epsilon}_{\boldsymbol{q}p}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2\, \boldsymbol{S}^{\boldsymbol{\epsilon}_{\boldsymbol{q}p}}_{v\boldsymbol{k}\leftarrow c\boldsymbol{k}}}{\int d\lambda_{\hat{\boldsymbol{q}}} \sum_{p=1}^{2} |\boldsymbol{\epsilon}_{\boldsymbol{q}p}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2}\right|_{\boldsymbol{k}=\boldsymbol{k}_{\text{ext}}}, \tag{8}$$

$$\boldsymbol{S}_{\text{intra},\boldsymbol{k}} = \int_{p_c(\boldsymbol{k})} \boldsymbol{\Omega}_c \times d\boldsymbol{k} + \int_{p_v(\boldsymbol{k})} \boldsymbol{\Omega}_v \times d\boldsymbol{k}. \tag{9}$$

The first line [Eq. (6)] represents the threefold decomposition of the shift loop into its excitation, recombination, and intraband components. Assuming the light source is linearly polarized with polarization vector $\boldsymbol{\epsilon}_s$, the shift loop depends on $\boldsymbol{\epsilon}_s$ through the excitation-induced $\boldsymbol{S}_{\text{exc}}$; cf. Eq. (7) with Eq. (3). The recombination shift $\boldsymbol{S}_{\text{rec}}$ [Eq. (8)] is an average of the photonic shift vector over all possible modes of the spontaneously emitted photon: that is to say, fixing the photon energy by $\hbar c|\boldsymbol{q}| = E_{c\boldsymbol{k}_{\text{ext}}} - E_{v\boldsymbol{k}_{\text{ext}}}$, one averages over all directions for $\boldsymbol{q}$ (parametrized by the solid angle $\lambda_{\hat{\boldsymbol{q}}}$) and over all transverse polarizations $\boldsymbol{\epsilon}_{\boldsymbol{q}p}$; this average is weighted by the transition rate, which is proportional to the square of the interband Berry connection by the Golden Rule; cf. Eq. (A5). Finally, we have taken the liberty of approximating the summation (over small momentum jumps) as line integrals over p_c and p_v.

Let us argue for our minimal model that no symmetry enforces $\boldsymbol{S}_{\text{intra}} = 0$. Being quasi-two-dimensional implies that only the z component of $\boldsymbol{\Omega}_{c/v,\boldsymbol{k}}$ is nonzero. As illustrated in Fig. 4(d), both relaxation pathways (p_c and p_v) lie in the $\boldsymbol{k}$-region with positive Berry curvature (Ω^z_c) for the conduction-band states. In the two-band approximation, the Berry curvatures of conduction- and valence-band states sum to zero: $\Omega^z_{c\boldsymbol{k}} = -\Omega^z_{v\boldsymbol{k}}$ [43], but this does not imply a cancellation in Eq. (9), because p_c and p_v are oppositely oriented. There is also no cancellation with the time-reversed loop, which is indicated by $T\circ p_c$ and $T\circ p_v$ in Fig. 4(e): under time reversal, $\Omega^z_c \to -\Omega^z_c$, but the orientation of $T\circ p_c$ is opposite to that of p_c. In the absence of symmetry-enforced cancellations, one expects that an anomalous shift current is a generic consequence of enclosed, time-reversal-symmetric Berry flux. This does not violate any symmetry principle, because optical excitation creates a nonequilibrium state with an arrow of time; this arrow manifests in the orientation of our loops. Our argument for the anomalous shift is widely generalizable: one may imagine a greater variety of enclosed time-reversal-symmetric Berry flux for which the line integrals in Eq. (9) are nonvanishing, and such imagination need not be restricted to semiconductors.

Moving beyond qualitative arguments, we would like to quantify the current for our quasi-2D semiconducting model. However, a brief diversion is required to explain the calculational method we invented. Our method introduces the *affinity shift loop* as a figure of merit for the shift current:

$$\text{ASL}_{\boldsymbol{\epsilon}_s,\omega} = \left\langle f_{vc\boldsymbol{k}} |\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2 \boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{\text{loop}[\boldsymbol{k}]}\right\rangle_\omega; \quad f_{vc\boldsymbol{k}} = f_{v\boldsymbol{k}} - f_{c\boldsymbol{k}}, \tag{10}$$

with $\langle\cdots\rangle_\omega$ meaning to average over all $\boldsymbol{k}$ on the excitation surface:

$$\langle \Xi(\boldsymbol{k})\rangle_\omega = \int \frac{d^3k}{(2\pi)^3} \frac{\delta(E_{cv\boldsymbol{k}} - \hbar\omega)}{\text{JDOS}_\uparrow}\, \Xi(\boldsymbol{k});$$
$$E_{cv\boldsymbol{k}} = E_{c\boldsymbol{k}} - E_{v\boldsymbol{k}}. \tag{11}$$

$\text{JDOS}_\uparrow$ stands for the joint density of states for quasiparticles of one spin orientation:

$$\text{JDOS}_\uparrow = \int \frac{d^3k}{(2\pi)^3}\, \delta(E_{cv\boldsymbol{k}} - \hbar\omega). \tag{12}$$

What is being averaged in Eq. (10) is the shift loop weighted by the rate of optical excitations, with $\boldsymbol{\epsilon}_s$ the polarization vector of the light source; by Fermi's Golden Rule, this rate is proportional to $|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2$, which we will refer to as the *optical affinity*. The shift loop is defined in Eqs. (6) and (9), with $p_c(\boldsymbol{k}_{\text{exc}})$ [$p_v(\boldsymbol{k}_{\text{exc}})$] chosen to be the unique oriented path that (i) connects $\boldsymbol{k}_{\text{exc}} \to \boldsymbol{k}_{\text{ext}}$ ($\boldsymbol{k}_{\text{ext}} \to \boldsymbol{k}_{\text{exc}}$), and (ii) is tangential to $\boldsymbol{v}_{c\boldsymbol{k}} = \nabla_{\boldsymbol{k}} E_{c\boldsymbol{k}}$ ($\boldsymbol{v}_{v\boldsymbol{k}} = \nabla_{\boldsymbol{k}} E_{v\boldsymbol{k}}$) at all points along

[11] A general definition of the shift loop is given in Eq. (D14), which applies beyond direct-gap semiconductors.

the path. We refer to p_b (with $b = c$ or v) as a *geodesic path*.[12] If E_B is an isotropic function of $\boldsymbol{k}$, then the geodesic path is simply the Euclidean-straight path connecting $\boldsymbol{k}_{\rm exc}$ to $\boldsymbol{k}_{\rm ext}$; cf. Fig. 4(f). The motivation for geodesic paths is that the predominant relaxation pathways [Fig. 4(d)] do not deviate far from being geodesic [Fig. 4(f)]: each time a quasiparticle in the conduction band emits an optical phonon, the likeliest transition involves the smallest wave-number change δk [Fig. 4(a)], since the electron-phonon matrix element diverges as $1/\delta k^2$ [36]; minimizing $\delta k = ||\boldsymbol{k}' - \boldsymbol{k}||$ with the constraint $E_{c\boldsymbol{k}} - E_{c\boldsymbol{k}'} = \hbar\Omega_o$ is approximately equivalent to $\boldsymbol{k} - \boldsymbol{k}'$ being parallel to $\boldsymbol{v}_{c\boldsymbol{k}}$, given that the optical phonon energy $\hbar\Omega_o$ is small compared to typical bandwidths.

Defining the *shift conductivity* through $\boldsymbol{j}[f] = \boldsymbol{\sigma}_{\epsilon_s,\omega}|\mathcal{E}_\omega|^2$, with $\boldsymbol{j}[f]$ the steady shift current given by the BIS formula [Eq. (1)], and $[\boldsymbol{\mathcal{E}}(\boldsymbol{r},t) = \boldsymbol{\epsilon}_s\mathcal{E}_\omega e^{i(\boldsymbol{q}\cdot\boldsymbol{r}-i\omega t)} +$ complex conjugate] being the incident electric wave, the shift conductivity relates to our figure of merit through

$$\boldsymbol{\sigma}_{\boldsymbol{\epsilon}_s,\omega} \approx -1.53\ \text{mA V}^{-2}\ \frac{\text{ASL}_{\boldsymbol{\epsilon}_s,\omega}}{\mathcal{V}_{\rm cell}}\ \frac{2\text{JDOS}_\uparrow}{(\mathcal{V}_{\rm cell}eV)^{-1}}, \tag{13}$$

with $1.53\ \text{mA V}^{-2} = 2\pi|e|^2/\hbar V$ in SI units, and $\mathcal{V}_{\rm cell}$ being the volume of the primitive unit cell.[13] Because the shift loop is threefold decomposable according to Eq. (6), one may likewise decompose

$$\boldsymbol{\sigma}_{\boldsymbol{\epsilon},\omega} = \boldsymbol{\sigma}^{\rm exc}_{\boldsymbol{\epsilon},\omega} + \boldsymbol{\sigma}^{\rm rec}_{\boldsymbol{\epsilon},\omega} + \boldsymbol{\sigma}^{\rm intra}_{\boldsymbol{\epsilon},\omega}, \tag{14}$$

with $\boldsymbol{\sigma}^{\rm exc}_{\boldsymbol{\epsilon},\omega}$ matching the Kraut-Baltz-Sipe-Shkrebtii formula [4,6,13] from Kubo-type perturbation theory; cf. Appendix C 3.

Our relation between the shift conductivity and the affinity shift loop [Eq. (13)] holds at low temperature ($k_BT_l \ll \hbar\Omega_o, E_g$) and for small optical phonon energy (compared to the band gap and the largest energy of photoexcited carriers).[14] The right-hand side of Eq. (13) should be understood as an approximation to the BIS formula [Eq. (1)]; the major error in this approximation originates from fixing p_b to be a geodesic path, hence we refer to Eq. (13) as the *geodesic approximation* to the shift conductivity. In reality, an electron excited at $\boldsymbol{k}_{\rm exc}$ follows multiple relaxation pathways [as caricatured in Fig. 4(c)] which deviate from being geodesic and narrow. The geodesic approximation is therefore justified to the extent that small-angle scattering dominates over large-angle scattering.[15] A benchmarking of the approximation will be presented.

While other groups have attempted to optimize the shift conductivity by maximizing the JDOS [44], we adopt a wave-function-centric approach in maximizing the affinity shift loop. Assuming a generic value for $\text{JDOS}_\uparrow \approx (\mathcal{V}_{\rm cell}eV)^{-1}$, $||\text{ASL}_{\boldsymbol{\epsilon}_s,\omega}|| \sim \mathcal{V}_{\rm cell}$ implies a conductivity of order mA V^{-2}.

Returning to our quasi-2D model, we now demonstrate that the intraband component [cf. Eq. (9)] of the affinity shift loop is indeed comparable in magnitude to $\mathcal{V}_{\rm cell}$:

$$||\langle(f_{v\boldsymbol{k}} - f_{c\boldsymbol{k}})|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2\boldsymbol{S}_{\text{intra},\boldsymbol{k}}\rangle_\omega|| \sim \mathcal{V}_{\rm cell}. \tag{15}$$

We adopt two heuristic approximations for a back-of-the-envelope calculation, namely that (i) the excitation surface is circular with radius $k_r = \pi/2a$ (assuming $\mathcal{V}_{\rm cell} = a^3$), and (ii) in the absence of unusually small band gaps,[16] the 2π Berry flux is roughly homogeneous over the yellow semicircle enclosed by the excitation surface; cf. Fig. 4(d). Because $\Omega^z_{c\boldsymbol{k}} = -\Omega^z_{v\boldsymbol{k}}$ (in the two-band approximation) and $p_c = -p_v$ (presuming an electron-hole symmetry $E_{c\boldsymbol{k}} = -E_{v\boldsymbol{k}}$), the anomalous contribution to the shift loop simplifies to $\boldsymbol{S}_{\text{intra},\boldsymbol{k}_{\rm exc}} = 2\int_{p_c(\boldsymbol{k}_{\rm exc})}\Omega^z_{c\boldsymbol{k}}\vec{z}\times d\boldsymbol{k}$, with $p_c(\boldsymbol{k}_{\rm exc})$ a straight path connecting $\boldsymbol{k}_{\rm exc}$ to $\boldsymbol{k}_{\rm ext} = (0,0,0)$. $\boldsymbol{S}_{\rm intra}(\boldsymbol{k}_{\rm exc}) = -16a\vec{y}/\pi$ if $\boldsymbol{k}_{\rm exc} = (k_r,0,0)$ and vanishes if $\boldsymbol{k}_{\rm exc} = (0,k_r,0)$; from this, one deduces that the average of $\boldsymbol{S}_{\text{intra},\boldsymbol{k}_{\rm exc}}$ over all $\boldsymbol{k}_{\rm exc}$ on the excitation surface is comparable to $-a\vec{y}$. Assuming that $f_{v\boldsymbol{k}_{\rm exc}} - f_{c\boldsymbol{k}_{\rm exc}} \approx 1$ (cf. Sec. II), and that the interband connection is generic-valued, $|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}_{\rm exc}}|^2 \sim a^2$, one finds $\langle|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv}|^2\boldsymbol{S}_{\rm intra}\rangle_\omega \sim -\mathcal{V}_{\rm cell}\vec{y}$, leading to Eq. (15).

It is of interest to demonstrate that our crudely derived conclusion holds true in a precise calculation for a model Hamiltonian:[17]

$$H(\boldsymbol{k}) = -E_o(z^\dagger\boldsymbol{\sigma}z)\cdot\boldsymbol{\sigma},\quad z = \begin{pmatrix}\tilde{k}_x - \tilde{k}_x^3/6\\ \tilde{k}_y + i\big(\tilde{Q} - \tilde{k}_x^2 - \tilde{k}_y^2\big)/2\end{pmatrix},$$
$$E_o = \frac{1}{\tilde{P}}\frac{\hbar^2}{m_f a^2}. \tag{16}$$

$\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2,\boldsymbol{\sigma}_3)$ is the vector of Pauli matrices; $\tilde{k}_j = k_j a$ (for $j = x,y,z$) is a dimensionless wave number, with $a = 5$ Å a generic value for the lattice period; m_f is the free-electron mass; $\tilde{P}$ and $\tilde{Q}$ are dimensionless Hamiltonian parameters. Assuming that $\tilde{Q} < 2$, Eq. (16) is the Hamiltonian of a direct-gap, quasi-2D semiconductor with band gap $E_g = E_o\tilde{Q}^2/2$, effective masses $m_x = m_y = \tilde{P}m_f/(2-\tilde{Q})$, $m_z = \infty$ (for both electrons and holes), and $\text{JDOS}_\uparrow \approx \tilde{P}/(8-4\tilde{Q})$ in units of $(\mathcal{V}_{\rm cell}eV)^{-1}$. We choose $\tilde{Q} = 1$ and $\tilde{P} = 4$, such that $\text{JDOS}_\uparrow \approx 1$ is generic-valued.

[12] Construct the four-momentum $(P^0,P^1,P^2,P^3) = (E_{c\boldsymbol{k}}/||\boldsymbol{v}_{c\boldsymbol{k}}||, \hbar k_x, \hbar k_y, \hbar k_z)$ and introduce the Lorentzian metric $g_{\mu\nu}$, with $g_{11} = g_{22} = g_{33} = -g_{00} = -1$. Then $p_c(\boldsymbol{k}_{\rm exc})$ can be viewed as a path $P^\mu(\lambda)$ that minimizes the action $\int_0^1\sqrt{[-g_{\mu\nu}\frac{dP^\mu}{d\lambda}\frac{dP^\nu}{d\lambda}]}d\lambda$, given that the end points of the path are fixed to $\boldsymbol{k}_{\rm exc}$ and $\boldsymbol{k}_{\rm ext}$.

[13] Equation (13) applies to direct-gap semiconductors in which a single conduction and a single valence band are optically excited in the vicinity of a single extremal wave vector $\boldsymbol{k}_{\rm ext}$. In direct-gap semiconductors with multiple valleys/pockets (indexed by v), such as transition-metal dichalcogenides, the total shift conductivity is proportional to $\sum_v \text{ASL}_v\text{JDOS}_v$.

[14] The largest energy of photoexcited electrons is represented as $E_{\rm exc} - E_{c,\boldsymbol{k}_{\rm ext}}$ in Fig. 3(c).

[15] The BIS formula [Eq. (1)] reduces asymptotically to the geodesic approximation [Eq. (13)] in the limit of vanishing scattering angle, as proven in Appendix D 5.

[16] The Berry curvature only exhibits significant variations over a lengthscale that is comparable to the inverse of the band gap.

[17] This $\boldsymbol{k}\cdot\boldsymbol{p}$ model can be extended to a tight-binding model by recognizing certain terms as Taylor-series coefficients for trigonometric functions.

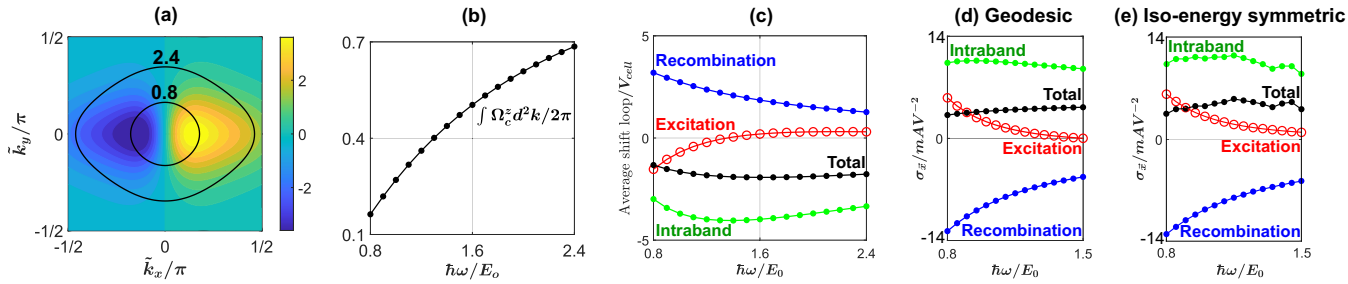

FIG. 5. Characterization of a quasi-2D model [Eq. (16) with $\tilde{Q} = 1$] with a time-reversal-symmetric Berry flux. Panel (a) plots the Berry curvature as a colored background, with a color legend (on the right) specifying the value of $\Omega^z_{c\boldsymbol{k}}/(\mathcal{V}_{\rm cell})^{2/3}$; the black ellipses represent excitation surfaces for two source photon energies: $\hbar\omega = 2.4E_0$ and $0.8E_0$, with $E_o \approx 76$ meV. Panel (b) shows the ω-dependence of the time-reversal-symmetric Berry flux enclosed by the excitation surface. The black curve in panel (c) represents the polarization-averaged affinity shift loop $[(1/2)\sum_{\boldsymbol{\epsilon}_s\in\{\vec{x},\vec{y}\}} \mathrm{ASL}^y_{\boldsymbol{\epsilon}_s,\omega}/\mathcal{V}_{\rm cell}]$ vs ω; the colored curves represent the threefold decomposition of the polarization-averaged affinity shift loop: excitation (red), intraband relaxation (green), and recombination (blue). The shift conductivity $\boldsymbol{\sigma}^y_{\vec{x},\omega}$ (and its threefold decomposition) is calculated in the geodesic approximation in panel (d), and in the isoenergy symmetric approximation in panel (e).

The model Hamiltonian has been chosen because it realizes a time-reversal-symmetric Berry flux of $2\pi = \int_{-\infty}^{\infty}\int_{0}^{\infty} \Omega^z_c dk_x dk_y$ [45–47], as illustrated in Fig. 5(a). The time-reversal-symmetric Berry flux enclosed by the excitation surface varies from roughly $0.7(2\pi)$ (for $\hbar\omega = 2.4E_0$) to $0.2(2\pi)$ (for $\hbar\omega = 0.8E_0$); cf. Fig. 5(b). The case of $\hbar\omega = 2.4E_0$ is not unlike the caricature we drew in Fig. 4(d).

The quasi-two-dimensionality and reflection symmetry ($x \to -x$) of our model imply that only the y-component of the shift current can be nontrivial. We have numerically computed the y-component of the affinity shift loop via Eqs. (6) and (10) for the chosen model parameters.[18] Figure 5(c) plots the polarization-averaged affinity shift loop $[(1/2)\sum_{\boldsymbol{\epsilon}_s\in\{\vec{x},\vec{y}\}} \mathrm{ASL}^y_{\boldsymbol{\epsilon}_s,\omega}]$ with respect to $\hbar\omega$ in the interval $[0.8E_0, 2.4E_0]$; the threefold decomposition of $(1/2)\sum_{\boldsymbol{\epsilon}_s} \mathrm{ASL}^y$ is also illustrated. For a broad range of frequencies ($\hbar\omega > 1.3E_0$) where the encircled time-reversal-symmetric Berry flux exceeds $0.4(2\pi)$, the intraband component not only exceeds the excitation component by an order of magnitude, but it also carries an opposite sign. These values for the affinity shift loop translate [via Eq. (13)] to a shift conductivity ≈ 4 mA V^{-2} for an unpolarized light source, as illustrated in Fig. 2(b) for the same frequency range.

To benchmark the geodesic approximation that has been used in all conductivity calculations thus far, we also computed $\boldsymbol{\sigma}^y_{\vec{x},\omega}$ via the more traditional method of numerically simulating a steady quasiparticle distribution f_B that sets the collisional integral to zero, and then inputting f_B into Eq. (1); the detailed procedure is described in Appendix C 2. Because this procedure is numerically intensive, we resorted to approximating f_B as an *iso-energy-symmetric* function of $\boldsymbol{k}$, meaning $f_{b\boldsymbol{k}}$ is constant over isoenergy $\boldsymbol{k}$-surfaces of $E_{b\boldsymbol{k}}$. Figures 5(d) and 5(e) show the same quantity $\boldsymbol{\sigma}^y_{\vec{x},\omega}$ calculated in the geodesic and iso-energy-symmetric approximations, respectively. It is reassuring to see semiquantitative consistency in the values of the shift conductivity and all its components, especially at lower photon frequencies where the isoenergy symmetric approximation is better justified.[19]

IV. OPTICAL VORTICITY

Having demonstrated the dominance of the intraband current $\boldsymbol{j}_{\rm intra}$ in the presence of time-reversal-symmetric *intra*band Berry curvature, this section will demonstrate the dominance of the recombination-induced current $\boldsymbol{j}_{\rm rec}$ in the presence of optical vortices—topological singularities in the *inter*band Berry phase.

Before discussing vortices properly, we first consider a vortexless scenario in which the photon-mediated current components cancel out: $\boldsymbol{j}_{\rm exc} + \boldsymbol{j}_{\rm rec} \approx 0$. In understanding how this cancellation happens, it will become apparent that vorticity is one route to prevent such a cancellation. Let us then hypothesize a scenario in which the photonic shift vector is roughly independent of the light polarization, and roughly homogeneous in the $\boldsymbol{k}$-region enclosed by the excitation surface. This would imply that the shift vector at excitation [Eq. (7)] is opposite to the shift vector at recombination [Eq. (8)], leading to a cancellation of the excitation- and recombination-induced currents; cf. Eqs. (10) and (14).

The contrapositive implication of this thought experiment is that for $\boldsymbol{j}_{\rm exc} + \boldsymbol{j}_{\rm rec}$ to be significant, the photonic shift at excitation must differ from the photonic shift at recombination. With great circulation comes great differences. The circulation of the photonic shift vector is defined by integrating Eq. (3) over a loop $\partial\boldsymbol{\Sigma}$ in $\boldsymbol{k}$-space:

$$\oint_{\partial\boldsymbol{\Sigma}} \boldsymbol{S}^{\boldsymbol{\epsilon}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}} \cdot \frac{d\boldsymbol{k}}{2\pi} = -\oint_{\partial\boldsymbol{\Sigma}} \nabla_{\boldsymbol{k}} \arg[\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}] \cdot \frac{d\boldsymbol{k}}{2\pi} + \int_{\boldsymbol{\Sigma}} (\boldsymbol{\Omega}_{c\boldsymbol{k}} - \boldsymbol{\Omega}_{v\boldsymbol{k}}) \cdot \frac{d^2\boldsymbol{\Sigma}}{2\pi}. \quad (17)$$

[18]The computation was simplified by setting $f_{v\boldsymbol{k}} - f_{c\boldsymbol{k}} = 1$ in Eq. (10) for reasons explained at the end of Sec. II. We have also approximated the geodesic paths as straight [cf. Fig. 4(f)], since the energy-momentum dispersion is roughly isotropic.

[19]The $\boldsymbol{k}$-dependent transition rate for optical excitation [cf. Eq. (A5)] becomes increasingly isoenergy-asymmetric at higher frequencies: the standard deviation of $|A^x_{cv}|^2$ (over the excitation surface at frequency ω) increases from 8.6% (of $\langle|A^x_{cv}|^2\rangle_\omega$) at $\hbar\omega = 0.8E_0$ to 23.8% at $\hbar\omega = 1.5E_0$.

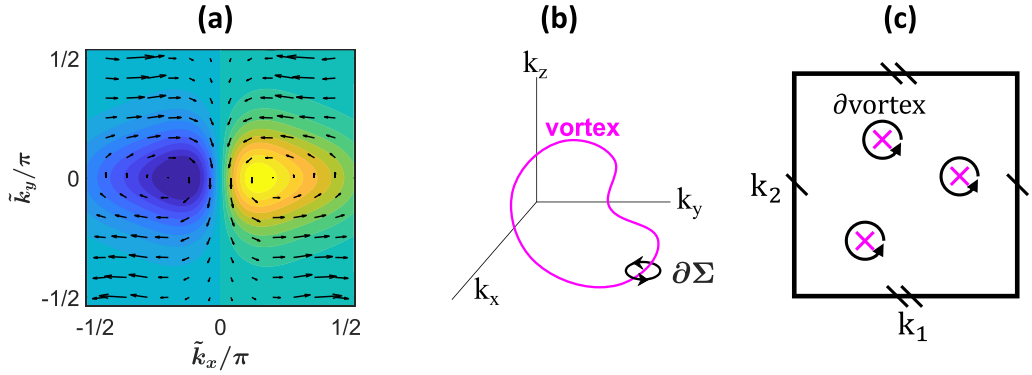

FIG. 6. (a) Circulation of the photonic shift vector $\boldsymbol{S}^{\vec{x}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$ associated with the intraband Berry curvature for our model Hamiltonian in Eq. (16), with $\tilde{Q} = 1$. The curvature is represented with the color legend in Fig. 5(a). (b) The magenta line represents a singularity in the interband Berry phase, i.e., an optical vortex. (c) Optical vortices appear as points (represented by magenta crosses) in a 2D closed $\boldsymbol{k}$-manifold parametrized by k_1 and k_2.

The last term is derived by Stokes' theorem, and it corresponds to a generically nonquantized[20] circulation associated with the intraband Berry phase, as illustrated in Fig. 6(a) for our model Hamiltonian in Eq. (16), with $\tilde{Q} = 1$. Beyond our model Hamiltonian, the possibility exists for an integer-quantized circulation stemming from the preceding term, which involves the interband Berry connection [48,49]. One is led to consider an optical vortex—a line in three-dimensional $\boldsymbol{k}$-space where (i) $\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}} = 0$, and (ii) the argument/phase of $\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}$ winds nontrivially as $\boldsymbol{k}$ is varied along any infinitesimal loop linked to the vortex line, as representatively illustrated by the black curve in Fig. 6(b). It is worth distinguishing between $\vec{x}$-vortices (where $A^x_{cv\boldsymbol{k}} = 0$) and $\vec{y}$-vortices (where $A^y_{cv\boldsymbol{k}} = 0$). Because $A^x_{cv\boldsymbol{k}} = 0$ does not generally imply $A^y_{cv\boldsymbol{k}} = 0$, one type of vortex may occur independently of the other.

Consider the shift current in a quasi-2D model where $\vec{y}$-vortices are absent [Fig. 7(b)], but the excitation surface encircles a pair of $\vec{x}$-vortices that are mutually related by time reversal and mirror reflection [Fig. 7(a)].[21] Figure 7 is derived from the model Hamiltonian in Eq. (16) with a different set of parameters: $\tilde{Q} = -1$ and $\tilde{P} = 12$, but fixed $\mathrm{JDOS}_{\uparrow} \approx (eV\mathcal{V}_{\mathrm{cell}})^{-1}$.

Just as for the model studied in Sec. III, only the y-component of the shift conductivity ($\boldsymbol{\sigma}^y_{\vec{x},\omega}$) is symmetry-allowed to be nonzero. $\boldsymbol{\sigma}^y_{\vec{x},\omega}$ for an $\vec{x}$-polarized light source is calculated via Eqs. (6)–(13), and is shown in Fig. 2(d) to be dominated by its recombination component. The same figure shows the difference between $\boldsymbol{\sigma}^y_{\vec{x},\omega}$ and $\boldsymbol{\sigma}^y_{\vec{y},\omega}$ to be of order $\mathrm{mA\,V^{-2}}$. This represents an extraordinary sensitivity of the shift current to the source polarization, and motivates $(\boldsymbol{\sigma}_{\vec{x},\omega} - \boldsymbol{\sigma}_{\vec{y},\omega})$ as an experimental indicator of $\vec{x}$-vorticity. More generally, for a quantity $B(\boldsymbol{\epsilon}_s)$ that depends on the source

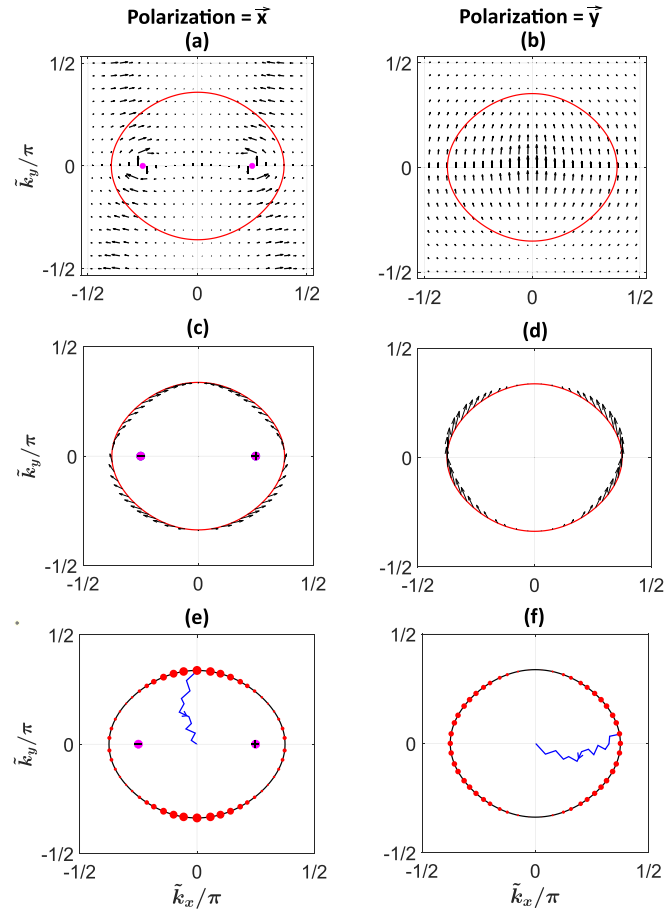

FIG. 7. Characterization of a quasi-2D model with a pair of $\vec{x}$-vortices, whose $\boldsymbol{k}$-locations are indicated by pink dots in panels (a), (c), and (e). Panels (a) and (b) depict the photonic shift vector field $\boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$, with $\boldsymbol{\epsilon}_s = \vec{x}$ and $\vec{y}$, respectively. In panels (c) and (d), the red ellipse represents the same excitation surface; the arrows on the ellipse represent the vectors $|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2\boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$ for $\boldsymbol{k}$ on the excitation surface. In panels (e) and (f), the size of the red dots indicates the magnitude of $|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2$ for $\boldsymbol{k}$ on the excitation surface; the jagged blue line represents a predominant relaxation pathway from the excitation surface to the extremal wave vector: $\boldsymbol{k}_{\mathrm{ext}} = \boldsymbol{0}$.

[20]With the inclusion of crystallographic symmetry and for a symmetric choice of $\partial\boldsymbol{\Sigma}$, it is possible for the Berry-flux term to be integer-quantized, which makes the circulation of the shift vector a topological invariant [48].

[21]That optical vortices come in time-reversed pairs was proven in Ref. [48].

polarization, we refer to $B(\vec{x}) \neq B(\vec{y})$ as a *linear disparity* in B.[22] The remainder of this section aims to demystify the large linear disparity in $\boldsymbol{\sigma}^{y}_{\boldsymbol{\epsilon}_s,\omega}$, and the dominance of the recombination-induced current when $\boldsymbol{\epsilon}_s = \vec{x}$.

Let us first understand the linear disparity in the excitation component of the conductivity; cf. Eq. (14). Recall that $\boldsymbol{\sigma}^{\text{exc}}_{\boldsymbol{\epsilon}_s,\omega}$ is proportional to a weighted average of the photonic shift vector over the excitation surface, which we have denoted as $\langle|\boldsymbol{\epsilon}_s \cdot \boldsymbol{A}_{cv\boldsymbol{k}}|^2 \boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}\rangle_\omega$; cf. Eqs. (6)–(14). A major effect of $\vec{x}$-vorticity without $\vec{y}$-vorticity is that the orientation of $|\boldsymbol{A}^{x}_{cv}|^2 \boldsymbol{S}^{\vec{x}}_{c\leftarrow v}$, viewed as a vector field over $\boldsymbol{k}$-space, tends to be more disordered than the vector field $|\boldsymbol{A}^{y}_{cv}|^2 \boldsymbol{S}^{\vec{y}}_{c\leftarrow v}$.[23] In particular, along the excitation surface, the orientation of $|\boldsymbol{A}^{x}_{cv}|^2 \boldsymbol{S}^{\vec{x}}_{c\leftarrow v}$ exhibits rotations, which are more pronounced than that of $|\boldsymbol{A}^{y}_{cv}|^2 \boldsymbol{S}^{\vec{y}}_{c\leftarrow v}$, as comparatively illustrated in Figs. 7(c) and 7(d). *Ceteris paribus*, the average of a rotational vector field is smaller than the average of an irrotational vector field, hence $||\langle|\boldsymbol{A}^{x}_{cv}|^2 \boldsymbol{S}^{\vec{x}}_{c\leftarrow v}\rangle_\omega|| < ||\langle|\boldsymbol{A}^{y}_{cv}|^2 \boldsymbol{S}^{\vec{y}}_{c\leftarrow v}\rangle_\omega||$ and $|\vec{y} \cdot \boldsymbol{\sigma}^{\text{exc}}_{\vec{x},\omega}| < |\vec{y} \cdot \boldsymbol{\sigma}^{\text{exc}}_{\vec{y},\omega}|$.

A different argument is needed to understand the linear disparity of the intraband conductivity: $\boldsymbol{\sigma}^{\text{intra},y}_{\boldsymbol{\epsilon}_s,\omega}$ in Eq. (14). For an $\vec{x}$-polarized source, the photon-mediated transition rate depends anisotropically on the orientation of $\boldsymbol{k}_{\text{exc}}$; in particular, $|A^{x}_{cv}|^2$ is suppressed on segments of the excitation surface that are closer to the vortex, where $A^{x}_{cv} = 0$; cf. Fig. 7(e). For a $\vec{y}$-polarized source, the photon-mediated transition rate is also anisotropic but in the opposite sense: $|A^{y}_{cv}|^2$ is suppressed near the mirror-invariant line ($k_x = 0$), where $A^{y}_{cv} = 0$ by a dipole selection rule; cf. Fig. 7(f).[24] The opposite senses of anisotropy imply that the predominant intraband relaxation pathways are roughly parallel to the k_y axis for an $\vec{x}$-polarized source [Fig. 7(e)], and parallel to the k_x axis for a $\vec{y}$-polarized source [Fig. 7(f)]. *Ceteris paribus*, this implies a larger $\vec{y} \cdot \boldsymbol{\sigma}^{\text{intra}}_{\vec{y},\omega}$ for a $\vec{y}$-polarized source, because the y-component of the anomalous shift [Eq. (5)] is proportional to the x-component of the momentum transfer: $\vec{y} \cdot \boldsymbol{S}^{\text{ano}}_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}} = \Omega^{z}_{b\boldsymbol{k}_{\text{ave}}} \delta k^x$.

A final argument explains the dominance of the recombination-induced current $\boldsymbol{j}_{\text{rec}}$ over the excitation-induced current $\boldsymbol{j}_{\text{exc}}$ for an $\vec{x}$-polarized source. Unlike $\boldsymbol{j}_{\text{exc}}$, $\boldsymbol{j}_{\text{rec}}$ is insensitive to the vortex-induced disordering in the orientation of $\boldsymbol{S}^{\vec{x}}_{c\leftarrow v}$. To understand why, recall that the majority of recombination transitions occur at $\boldsymbol{k}$ points close to the extremal wave vector $\boldsymbol{k}_{\text{ext}}$ and contained within the passive region; cf. Sec. II and Eq. (A32). Each recombination transition is associated with a photonic shift $\boldsymbol{S}^{\boldsymbol{\epsilon}_m}_{v\boldsymbol{k}\leftarrow c\boldsymbol{k}}$, which may as well be approximated as $\boldsymbol{S}^{\boldsymbol{\epsilon}_m}_{v\boldsymbol{k}_{\text{ext}}\leftarrow c\boldsymbol{k}_{\text{ext}}}$, because the passive region typically occupies a tiny fraction of the Brillouin-zone volume. For the same reason, it is presumed that the optical vortex does not intersect the passive region. It follows that $\boldsymbol{j}_{\text{rec}}$ depends on $\boldsymbol{S}^{\boldsymbol{\epsilon}_m}_{v\boldsymbol{k}_{\text{ext}}\leftarrow c\boldsymbol{k}_{\text{ext}}}$ but not on the vortex-induced disorder in $\boldsymbol{S}^{\vec{x}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$. To wrap up the argument, the vortex-induced orientational disorder in $\boldsymbol{S}^{\vec{x}}_{c\leftarrow v}$ diminishes $\boldsymbol{\sigma}^{\text{exc}}_{\vec{x},\omega}$ but not $\boldsymbol{\sigma}^{\text{rec}}_{\vec{x},\omega}$; *ceteris paribus*, $||\boldsymbol{\sigma}^{\text{rec}}_{\vec{x},\omega}|| > ||\boldsymbol{\sigma}^{\text{exc}}_{\vec{x},\omega}||$. This explains how "great differences" (between the excitation and recombination shifts) result from a "great circulation" (induced by a vortex).

To recapitulate, we have qualitatively argued that $\vec{x}$-vorticity leads to $\boldsymbol{\sigma}^{\text{rec}}_{\vec{x},\omega}$ dominating over $\boldsymbol{\sigma}^{\text{exc}}_{\vec{x},\omega}$, as well as bringing about a linear disparity of both $\boldsymbol{j}_{\text{exc}}$ and $\boldsymbol{j}_{\text{intra}}$.[25] These arguments are quantitatively supported by model calculations detailed in Appendix F; here, we will just summarize the salient conclusions: $\boldsymbol{\sigma}_{\vec{x},\omega}$ is dominated by the recombination-induced current; $\boldsymbol{\sigma}_{\vec{y},\omega}$ is dominated by the excitation-induced and intraband currents; the signs of $\boldsymbol{\sigma}^{y}_{\vec{x},\omega}$ and $\boldsymbol{\sigma}^{y}_{\vec{y},\omega}$ differ over a broad range of frequencies; the linear disparity in the conductivity is large: $|\boldsymbol{\sigma}^{y}_{\vec{x},\omega} - \boldsymbol{\sigma}^{y}_{\vec{y},\omega}| \sim \text{mA V}^{-2}$ [cf. Fig. 2(d)], assuming a generic value for the JDOS; the current response to unpolarized light is slightly smaller: $|\boldsymbol{\sigma}^{y}_{\vec{x},\omega} + \boldsymbol{\sigma}^{y}_{\vec{y},\omega}|/2 \sim 0.1\ \text{mA V}^{-2}$; all these results hold without fine-tuning of the incident radiation frequency.

To find optical vortices in model Hamiltonians and realistic materials, let us develop the close relationship between vorticity and Berry curvature that has been suggested by Eq. (17): for any closed 2D $\boldsymbol{k}$-manifold $\boldsymbol{\Sigma}$ (which can be a two-toroidal or two-spherical cut of the 3D Brillouin zone), we establish a general theorem relating the Chern numbers (C_v, C_c) of the valence and conduction bands to the net optical vorticity $(\text{Vort}_{\boldsymbol{\epsilon}})$:

$$\text{Chern-vorticity theorem:}\quad C_c - C_v = \text{Vort}_{\boldsymbol{\epsilon}} = \sum_{\text{vortex}} \oint_{\partial\ \text{vortex}} \nabla_{\boldsymbol{k}} \arg[\boldsymbol{\epsilon} \cdot \boldsymbol{A}_{cv\boldsymbol{k}}] \cdot \frac{d\boldsymbol{k}}{2\pi}. \tag{18}$$

$\text{Vort}_{\boldsymbol{\epsilon}}$ is the net circulation of the interband Berry phase over all $\boldsymbol{\epsilon}$-vortex points in $\boldsymbol{\Sigma}$, and ∂ vortex is an infinitesimal loop surrounding each $\boldsymbol{\epsilon}$-vortex point as illustrated in Fig. 6(c).[26] This theorem is derived by setting $\boldsymbol{\Sigma}$ to be a closed manifold in Eq. (17), such that the area integral of $\boldsymbol{\Omega}_c$ simplifies to the Chern number C_c of the conduction band, and

[22] In contrast, linear birefringence results from a linear disparity in the first-order-in-$\mathcal{E}_\omega$ conductivity, which is associated with an alternating current.

[23] The vector norm of $|\boldsymbol{A}^{x}_{cv}|^2 \boldsymbol{S}^{\vec{x}}_{c\leftarrow v}$ is not singular: as $\boldsymbol{k}$ approaches the vortex center, the quantized circulation implies $\boldsymbol{S}^{\vec{x}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}} \to \infty$, but this divergence is compensated by $|\boldsymbol{A}^{x}_{cv\boldsymbol{k}}|^2 \to 0$ [48].

[24] The conduction-band states transform in a different representation of mirror than the valence-band states, as detailed in Appendix F.

[25] Vorticity also results in a linear disparity of $\boldsymbol{j}_{\text{rec}}$ due to the absorption coefficient being proportional to $\langle|\boldsymbol{\epsilon}_s \cdot \boldsymbol{A}_{cv}|^2\rangle_\omega$, but this is not a large effect in our model.

[26] An equivalent and manifestly gauge-invariant expression is $\text{Vort}_{\boldsymbol{\epsilon}} = \sum_{\text{vortex}} \oint_{\partial\ \text{vortex}} \boldsymbol{S}^{\boldsymbol{\epsilon}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}} \cdot d\boldsymbol{k}/2\pi$, with the photonic shift vector defined in Eq. (3). This expression differs from Eq. (18) only in the line integral of $\boldsymbol{A}_{cc} - \boldsymbol{A}_{vv}$ over ∂ vortex; this integral vanishes because ∂ vortex is an infinitesimal loop and $\boldsymbol{A}_{cc} - \boldsymbol{A}_{vv}$ is smoothly defined at the vortex point; cf. Fig. 17 in Appendix G.

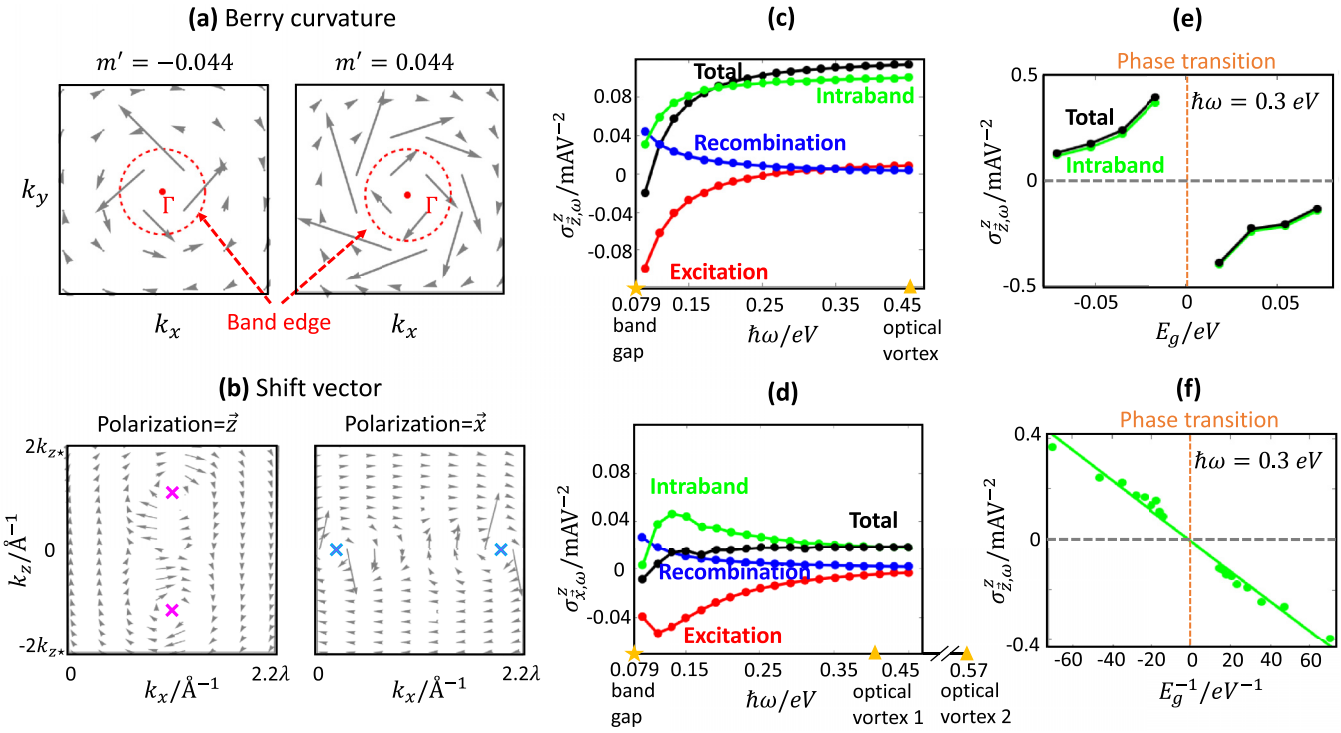

FIG. 8. Panel (a) shows the Berry curvature vector field in the $k_z = 0$ plane for the model Hamiltonian in Eq. (19) with Hamiltonian parameter $m' = m - A\lambda^2 = \pm 0.044$ eV. Panel (b) plots the photonic shift vector fields ($\boldsymbol{S}^{\vec{z}}_{c\leftarrow v}$ and $\boldsymbol{S}^{\vec{x}}_{c\leftarrow v}$) in the $k_{x+} - k_z$ half-plane for $m' = -0.044$. The shift vectors circulate around optical $\vec{z}$-vortices ($\vec{x}$-vortices), which are represented by magenta (navy) crosses. (c),(d) For $m' = -0.044$ eV, we plot the shift conductivities ($\vec{z}\cdot\boldsymbol{\sigma}_{\vec{z}}$ and $\vec{z}\cdot\boldsymbol{\sigma}_{\vec{x}}$) and their threefold decomposition. Panel (e) plots the $\vec{z}\cdot\boldsymbol{\sigma}_{\vec{z}}$ and its intraband component versus E_g. Panel (f) plots $\vec{z}\cdot\boldsymbol{\sigma}^{\text{intra}}_{\vec{z}}$ against $1/E_g$, with negative E_g corresponding to the trivial side of the topological phase transition.

that of $\boldsymbol{\Omega}_v$ to C_v; the line integral of the shift vector over $\partial\boldsymbol{\Sigma}$ vanishes, but the line integral of the interband Berry phase is contributed by the circulation around each vortex point, as elaborated in Appendix G. The next section employs the Chern-vorticity theorem to identify vortices in BiTeI.

V. CASE STUDY OF BiTeI

To demonstrate the effects of the anomalous shift and optical vorticity in a realistic material, we present a case study of BiTeI, a 3D polar, layered semiconductor with $P3m1$ space-group symmetry [50,51]. The large atomic number of Bi correlates with a large Rashba-type spin-orbit coupling [51] and the proximity of BiTeI to a $\mathbb{Z}_2$ topological insulator [52]. A previous study [23] of BiTeI by Tan and Rappe exhibited the enhancement of $\vec{z}\cdot\boldsymbol{\sigma}^{\text{exc}}_{\vec{z}}$ for $\vec{z}$ parallel to the polar axis, assuming the band gap $|E_g|$ was made small (by hydrostatic pressure), and further assuming the photon frequency was fine-tuned to be comparably small: $\omega \approx |E_g|/\hbar$. This enhancement of $\vec{z}\cdot\boldsymbol{\sigma}^{\text{exc}}_{\vec{z}}$ originates from the divergence of the band-edge optical affinity at the phase transition ($E_g = 0$) between the trivial and topological insulator; across this transition, $\vec{z}\cdot\boldsymbol{\sigma}^{\text{exc}}_{\vec{z}}$ changes sign. Our case study demonstrates the following:

(i) The just-mentioned topological phase transition guarantees the existence of large Berry curvature [cf. Fig. 8(a)] and optical vortices [Fig. 8(b)], as per the Chern-vorticity theorem in Eq. (18).

(ii) For photon frequencies such that the excitation surface is close to the optical vortex, the phonon-mediated $\vec{z}\cdot\boldsymbol{\sigma}^{\text{intra}}_{\boldsymbol{\epsilon}}$ dominates over the photon-mediated $\vec{z}\cdot(\boldsymbol{\sigma}^{\text{rec}}_{\boldsymbol{\epsilon}} + \boldsymbol{\sigma}^{\text{exc}}_{\boldsymbol{\epsilon}})$ [Fig. 8(c)], due to the excitation surface enclosing a larger volume of one-quasiparticle Bloch states with nontrivial Berry curvature.

(iii) Conversely for smaller frequencies ($\omega \approx |E_g|/\hbar$) such that the excitation surface encloses a negligible amount of Berry curvature, it is $\vec{z}\cdot(\boldsymbol{\sigma}^{\text{rec}}_{\boldsymbol{\epsilon}} + \boldsymbol{\sigma}^{\text{exc}}_{\boldsymbol{\epsilon}})$, which dominates over $\vec{z}\cdot\boldsymbol{\sigma}^{\text{intra}}_{\boldsymbol{\epsilon}}$. The net current is nonvanishing despite $\vec{z}\cdot\boldsymbol{\sigma}^{\text{rec}}_{\boldsymbol{\epsilon}}$ and $\vec{z}\cdot\boldsymbol{\sigma}^{\text{intra}}_{\boldsymbol{\epsilon}}$ opposing each other [Fig. 8(d)], due to an asymmetry of the photon polarizations in the excitation and recombination processes.

(iv) The recombination shift depends strongly on the symmetry of the Hamiltonian at the wave vectors of recombination. For BiTeI, a chiral symmetry reduces $|\vec{z}\cdot\boldsymbol{S}_{\text{rec}}|$ to about a third of the lattice period, which makes the recombination shift current smaller than the other two components.

(v) Because the $\boldsymbol{k}$-locations of optical vortices depend on the light polarization [Fig. 8(b)], we find that $\vec{x}$-vortices suppress the anomalous shift more effectively than would $\vec{z}$-vortices, resulting in a linear disparity of the shift conductivity ($|\vec{z}\cdot\boldsymbol{\sigma}_{\vec{z}} - \vec{z}\cdot\boldsymbol{\sigma}_{\vec{x}}| \sim 0.1\ \text{mA/V}^2$) at higher frequencies.

(vi) If the topological phase transition is induced by tuning the band gap $|E_g|$ to zero at a fixed photon frequency, the discrete change in wave-function topology manifests as a sign change of the steady shift current $\boldsymbol{j}[f]$. The reason for this sign change is that $\boldsymbol{j}[f]$ is dominated by the phonon-mediated $\boldsymbol{j}_{\text{intra}}$, which changes sign [Fig. 8(e)]; the previously calculated

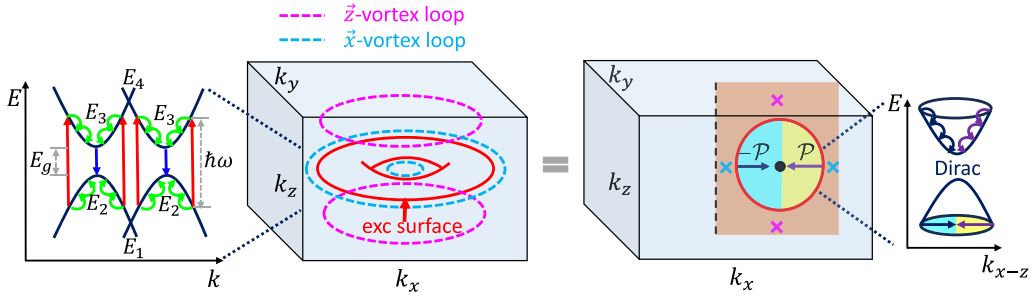

FIG. 9. On the left: the box is a 3D Brillouin zone containing a toroidal excitation surface (outlined in red), a pair of $\vec{z}$-optical vortex loops (magenta dotted lines), and a pair of $\vec{x}$-optical vortex loops (navy dotted lines); the leftmost E-vs-k plot shows the band structure of BiTeI with the kinetic processes of excitation (red arrows), intraband relaxation (green), and recombination (blue). On the right: an O(2) rotation symmetry allows us to focus on the $k_{x^+} - k_z$ half-plane, which contains a massive Dirac fermion in two momentum dimensions; the cross-section of the toroidal excitation surface is a circle (colored red) in the half-plane; the cross-section of vortex loops is vortex points (indicated by magenta and navy crosses). The dark blue and purple arrows represent diametrically opposed geodesic paths for intraband relaxation.

sign change of $\boldsymbol{j}_{\rm exc}$ (by Tan and Rappe [23]) is irrelevant. The sign change of $\boldsymbol{j}_{\rm intra}$ is concomitant with a divergence of the band-edge intraband Berry curvature at the phase transition, which results in an approximate $(1/E_g)$ divergence of the low-temperature $\vec{z}\cdot\boldsymbol{\sigma_\epsilon}$ [Fig. 8(f)], with negative (positive) E_g referring to a semiconductor (with band gap $|E_g|$) on the topologically trivial (nontrivial) side of the phase transition. The $(1/E_g)$ divergence of $\vec{z}\cdot\boldsymbol{\sigma_\epsilon}$ is cut off when E_g becomes comparable to the thermal energy (k_BT_l) or to an energy scale representing trigonal warping; the latter scale is estimated [23] to be about 10 meV.

It may be seen from (ii) and (iii) that the winner in the competition (between $\text{shift}_{\rm intra}$, $\text{shift}_{\rm rec}$, and $\text{shift}_{\rm exc}$) depends sensitively on the photon frequency, to the extent that the net shift conductivity changes sign in the transition from a photon-dominated shift current (at low frequency) to a phonon-dominated shift current (at high frequency); cf. the black curve in Figs. 8(c) and 8(d). This exemplifies a general principle: because band wave functions can depend strongly on energy, the shift current can depend sensitively on the photon frequency. Point (v) exemplifies a general principle that optical vorticity makes the shift current sensitive to changes in light polarization. Point (vi) suggests the bulk photovoltaic effect can provide smoking-gun evidence of the topological phase transition in BiTeI. Such evidence is presently lacking: though it has been alleged that BiTeI is pressure-tunable to a phase transition, the experimental corroborations of this allegation (namely, a minimum of the resistivity [53] or variations of the quantum oscillation frequency [54]) cannot be directly interpreted as a change in wave-function topology.

To substantiate our results, we employ an effective Hamiltonian for the four low-energy, spin-split bands of BiTeI:

$$H_{\rm BiTeI} = \hbar v(\lambda\tau_1\sigma_3 - \tau_2\boldsymbol{k}\cdot\boldsymbol{\sigma}) + M\tau_3\sigma_0;$$
$$M = (m - Ak^2); \quad k^2 = k_x^2 + k_y^2 + k_z^2, \tag{19}$$

with $\tau_{1,2,3}$ and $\sigma_{1,2,3}$ being Pauli matrices for the orbital and spin degrees of freedom, respectively; τ_0 and σ_0 are identity matrices, and $A = 0.5$ eVÅ^2, $\hbar v = 0.7$ eVÅ, $\lambda = 0.25/(\hbar v) = 0.357$ Å^{-1} are *ab initio*–derived [23] parameters.

The spectrum of this model is given by

$$E = \pm\sqrt{M^2 + (\hbar v)^2\left[k_z^2 + (k_\perp \pm \lambda)^2\right]}, \quad k_\perp \equiv \sqrt{k_x^2 + k_y^2}, \tag{20}$$

with each choice of either $\pm$ determining four energy levels $E_1 \leqslant E_2 < E_3 \leqslant E_4$ as illustrated in Fig. 9, with corresponding eigenstates $|u_{1\mathbf{k}}\rangle$, $|u_{2\mathbf{k}}\rangle$, $|u_{3\mathbf{k}}\rangle$, and $|u_{4\mathbf{k}}\rangle$. When m is tuned to the critical value $m_c = A\lambda^2$ (possibly by hydrostatic pressure [23]), the band gap vanishes with a concomitant energy degeneracy $(E_2 = E_3)$ along a loop defined by $k_\perp = \lambda$ and $k_z = 0$. The circular shape of this loop reflects the O(2) rotational symmetry of the Hamiltonian.[27]

Away from the topological phase transition $(m \neq m_c)$, the Berry curvature of $|u_{2\mathbf{k}}\rangle$ is a circulating vector field illustrated in Fig. 8(a), with the circulation flipping in orientation when $m' = m - m_c$ changes sign. Because this circulation is O(2)-rotation symmetric, it may be understood by focusing on a single 2D slice of the Brillouin zone, say, the slice with $k_y = 0$ and $k_x > 0$, which we henceforth call the $k_{x^+} - k_z$ half-plane (see Fig. 9). This half-plane is intersected by the energy-degenerate loop at $\mathbf{k}_0 = (\lambda, 0, 0)$; by Taylor-expanding $H_{\rm BiTeI}$ around $\boldsymbol{k}_0$ and projecting onto bands 2 and 3, we obtain an effective Hamiltonian for a massive Dirac fermion in two momentum dimensions:

$$H' = (m' - \hbar v' q_x)\gamma_3 + \hbar v(q_x\gamma_1 + q_z\gamma_2);$$
$$\hbar v' = 2A\lambda; \quad q_x = k_x - \lambda, \quad q_z = k_z, \tag{21}$$

with $\gamma_{1,2,3}$ being Pauli matrices of the reduced Hilbert space. Detailed derivations of H' can be found in Appendix H. It is

[27] The SO(2) rotational symmetry of the Hamiltonian manifests as $\hat{R}_\theta H_{\rm BiTeI}(\boldsymbol{k})\hat{R}_\theta^{-1} = H_{\rm BiTeI}(g_\theta\boldsymbol{k})$, with $\hat{R}_\theta = e^{i\theta\tau_0\sigma_3}$, and with $g_\theta\boldsymbol{k}$ obtained from $\boldsymbol{k}$ by a rotation of angle θ around the z axis. This SO(2) rotational symmetry is approximate; we neglect a trigonal warping whose energy scale is estimated [23] to be about 10 meV. Time-reversal symmetry is represented by $\hat{T} = i\tau_3\sigma_2K$, with K implementing complex conjugation. The Hamiltonian term proportional to $\tau_1\sigma_3$ breaks both mirror $[z \to -z; \hat{M}_z = i\tau_3\sigma_3]$ and $[(x, y, z) \to (-x, -y, -z); \hat{P} = \tau_3]$ parity symmetries.

known that the massive Dirac fermion is characterized by a large Berry curvature $\Omega^y_{c\boldsymbol{k}} = -\Omega^y_{v\boldsymbol{k}}$ for $\boldsymbol{k}$ in a "hot spot" of width (band gap)$/\hbar v$; assuming this width is small compared to the linear dimension of the Brillouin zone, the Berry flux through the half-plane changes by 2π when m' changes sign: $\int_{\text{half-plane}} \Omega^y_{c\boldsymbol{k}} dk_x dk_z = \pi\, \text{sgn} m'$ [55].

This jump of the Berry flux indicates the presence of optical vorticity on at least one side of the phase transition, meaning for either $m' > 0$ or $m' < 0$. This is because as m' is tuned from 0^- (negative infinitesimal) to 0^+ (positive infinitesimal), the net vorticity $\text{Vort}_{\boldsymbol{\epsilon}}$ of the half-plane must change discontinuously by $+2$ for any $\boldsymbol{\epsilon}$, according to the Chern-vorticity theorem in Eq. (18).[28] This topological argument is verified by our numerical calculations: As shown in Fig. 8(b), for $m' < 0$, two $\vec{z}$-vortices ($\vec{x}$-vortices) are observed on the $k_{x^+} - k_z$ plane, as highlighted by magenta (navy) crosses. This indicates the presence of a pair of $\vec{z}$-vortex ($\vec{x}$-vortex) loops as illustrated in Fig. 9 when $m' < 0$.[29] While the locations of $\vec{x}$-vortex loops can only be determined numerically, direct calculations reveal that the $\vec{z}$-vortex loops reside at

$$\lambda = k_\perp, \quad k_z = \pm k_{z\star} \quad \text{with } k_{z\star} = \sqrt{\lambda^2 - m/A}. \tag{22}$$

In contrast, there is no vorticity for $m' > 0$. The circulation of the photonic shift vectors in Fig. 8(b) implies that both $\vec{z}$-vortex ($\vec{x}$-vortex) loops have vorticity -1, which is consistent with the theorem's prediction that $\text{Vort}_{\boldsymbol{\epsilon}}$ changes by $+2$.

The combination of Berry curvature and optical vorticity results in the threefold decomposition of the shift current being highly sensitive to the photon frequency and polarization. Focusing first on z-polarized light, the only nonzero component of the shift current [allowed by O(2) symmetry] is the z component. Figure 8(e) illustrates the threefold decomposition of the shift conductivity $\vec{z} \cdot \boldsymbol{\sigma}_{\vec{z}}$ for a range of frequencies, including low frequencies that are comparable to $|E_g|/\hbar$, as well as higher frequencies where the excitation surface approaches the pair of $\vec{z}$-vortex loops, as illustrated in Fig. 9.

We focus first on the high-frequency regime, where the optical affinity $|\vec{z} \cdot \mathbf{A}_{cv}|^2$ is reduced in the vicinity of the zeros of $\vec{z} \cdot \mathbf{A}_{cv}$. *Ceteris paribus*, a reduction of the affinity would depress each of $\vec{z} \cdot \boldsymbol{\sigma}^{\text{exc}}_{\vec{z}}$, $\vec{z} \cdot \boldsymbol{\sigma}^{\text{rec}}_{\vec{z}}$, and $\vec{z} \cdot \boldsymbol{\sigma}^{\text{intra}}_{\vec{z}}$, according to Eq. (10). This depression is observed in Fig. 8(e) for both $\vec{z} \cdot \boldsymbol{\sigma}^{\text{exc}}_{\vec{z}}$ and $\vec{z} \cdot \boldsymbol{\sigma}^{\text{rec}}_{\vec{z}}$; in contrast, $\vec{z} \cdot \boldsymbol{\sigma}^{\text{intra}}_{\vec{z}}$ is enhanced rather than depressed, for two reasons:

(a) The nonuniformity of the optical affinity (over the excitation surface) favors $\vec{z} \cdot \boldsymbol{\sigma}^{\text{intra}}_{\vec{z}}$. On the one hand, Fig. 9 shows that the optical affinity is more greatly reduced for large $|k_z|$ (closer to the vortex) than it is for small $|k_z|$ (further from the vortex), implying that the predominant relaxation paths are proximate to the k_x axis. On the other hand, only those relaxation paths (for which $d\boldsymbol{k}$ is perpendicular to $\vec{z}$) result in a large anomalous shift: $\vec{z} \cdot \boldsymbol{\Omega} \times d\boldsymbol{k}$. Bringing both hands together, $\vec{z}$-vorticity preserves the horizontal relaxation paths, which have a large anomalous shift, and deactivates the vertical relaxation paths which nevertheless have a negligible anomalous shift.

(b) A higher photon frequency implies that the excitation surface encloses a larger volume of quasiparticle Bloch states with nontrivial Berry curvature, and this results in a larger anomalous shift. For an intuitive understanding, consider reformulating the intra-conduction-band[30] contribution to the affinity shift loop [Eq. (10)]

$$\sum_{\mathcal{P}} \left(|\vec{z} \cdot \boldsymbol{A}_{cv}|^2_{-\mathcal{P}} \int_{-\mathcal{P}} + |\vec{z} \cdot \boldsymbol{A}_{cv}|^2_{\mathcal{P}} \int_{\mathcal{P}} \right) \vec{z} \cdot \boldsymbol{\Omega}_{c\boldsymbol{k}} \times d\boldsymbol{k}, \tag{23}$$

with $\mathcal{P}$ and $-\mathcal{P}$ representing diametrically opposed geodesic paths in a cross section of the torus enclosed by the excitation surface (as representatively illustrated by arrows in the blue and yellow semicircles of Fig. 9); $|\vec{z} \cdot \boldsymbol{A}_{cv}|^2_{\pm\mathcal{P}}$ is the optical affinity evaluated at the intersection of $\pm\mathcal{P}$ with the excitation surface, and $\sum_{\mathcal{P}}$ integrates over all pairs of $\pm\mathcal{P}$ such as to entirely fill the torus. From Eq. (23), one deduces that the anomalous interband contribution (to the affinity shift loop) increases with increasing photon frequency, because one integrates the Berry curvature over increasingly wider paths.[31]

Conversely, for smaller frequencies ($\omega \approx |E_g|/\hbar$), the excitation surface lies closer to the extremal wave vectors $\boldsymbol{k}_{\text{ext}}$ but further away from the vortex loops; then it is the photon-mediated $\vec{z} \cdot (\boldsymbol{\sigma}^{\text{rec}}_{\vec{z}} + \boldsymbol{\sigma}^{\text{exc}}_{\vec{z}})$ that dominates over the phonon-mediated $\vec{z} \cdot \boldsymbol{\sigma}^{\text{intra}}_{\vec{z}}$ [cf. the trends in Fig. 8(e)], due to the Berry dipole moment vanishing as the volume (enclosed by the excitation surface) shrinks [Fig. 8(g)]. $\vec{z} \cdot \boldsymbol{\sigma}^{\text{exc}}_{\vec{z}}$ and $\vec{z} \cdot \boldsymbol{\sigma}^{\text{rec}}_{\vec{z}}$ oppose each other but do not cancel out, because the excitation shift $\vec{z} \cdot \boldsymbol{S}^{\vec{z}}_{c \leftarrow v}$ is larger in magnitude than the recombination shift $\vec{z} \cdot \boldsymbol{S}_{\text{rec}}$.[32]

In comparing the shift conductivities for $\vec{z}$- versus $\vec{x}$-polarized light [Fig. 8(e) versus Fig. 8(f)], the starkest

[28] Strictly speaking, the half-plane is not a closed 2D $\boldsymbol{k}$-manifold, which precludes a direct application of the Chern-vorticity theorem. However, nearly the same logical considerations apply: the eigenstates $|u_{b\mathbf{k}}\rangle$ continuously depend on m' except for $\boldsymbol{k} = \boldsymbol{k}_0$ (the band-touching point), thus when m' is tuned from 0^- to 0^+, the photonic shift vector $\boldsymbol{S}^{\boldsymbol{\epsilon}}_{c\boldsymbol{k} \leftarrow v\boldsymbol{k}}$ is invariant for $\boldsymbol{k}$ on the boundary of the half-plane. This implies that the change in $2\int_{\text{half-plane}} \Omega^y_{c\boldsymbol{k}} dk_x dk_z/(2\pi)$ equals the change in $\text{Vort}_{\boldsymbol{\epsilon}}$, across the topological phase transition.

[29] Due to the SO(2) symmetry, $\vec{y}$-vortex loops are related to $\vec{x}$-vortex loops because $|\vec{y} \cdot \boldsymbol{A}_{cv}(R_4 \boldsymbol{k})| = |\vec{x} \cdot \boldsymbol{A}_{cv}(\boldsymbol{k})|$, where R_4 is the fourfold rotation in momentum space.

[30] The total interband contribution to the affinity shift loop is twice of Eq. (23), due to $\boldsymbol{\Omega}_{c\boldsymbol{k}} \approx -\boldsymbol{\Omega}_{v\boldsymbol{k}}$.

[31] The increase of $\vec{z} \cdot \boldsymbol{\sigma}^{\text{intra}}$ (with respect to frequency) saturates when the excitation energy (measured from the conduction-band minimum) becomes comparable to the band gap: $E_{\text{exc}} - E_{c,\boldsymbol{k}_{\text{ext}}} \sim |E_g|$, as illustrated in Figs. 8(c) and 8(d). After all, this energy interval contains the Berry curvature "hot spot."

[32] This is explained by $\boldsymbol{S}_{\text{rec}}$ being a weighted average of $\boldsymbol{S}^{\boldsymbol{\epsilon}}_{v\boldsymbol{k}_{\text{ext}} \leftarrow c\boldsymbol{k}_{\text{ext}}}$ over all possible polarization vectors $\boldsymbol{\epsilon}$ of the spontaneously emitted photons [cf. Eq. (8)]. Moreover, $\vec{z} \cdot \boldsymbol{S}^{\vec{x}}_{v\boldsymbol{k}_{\text{ext}} \leftarrow c\boldsymbol{k}_{\text{ext}}} = \vec{z} \cdot \boldsymbol{S}^{\vec{y}}_{v\boldsymbol{k}_{\text{ext}} \leftarrow c\boldsymbol{k}_{\text{ext}}} = 0$ due to a chiral symmetry $[\tau_2 \sigma_3 H_{\text{BiTeI}}(k_x, k_y, 0) \tau_2 \sigma_3 = -H_{\text{BiTeI}}(k_x, k_y, 0)]$, which is elaborated in Appendix H. This exemplifies a general principle that the recombination shift strongly depends on the symmetry of the Hamiltonian at the wave vectors of recombination. For BiTeI, chiral symmetry reduces $|\vec{z} \cdot \boldsymbol{S}_{\text{rec}}|$ to about a third of the lattice period, whereas for the two-band model in Eq. (16), the symmetries of reflection and quasi-two-dimensionality

difference is that $\vec{z}\cdot\boldsymbol{\sigma}^{\rm intra}_{\vec{x}}$ is a nonmonotonic function that is suppressed at higher photon frequency.[33] This is because, in contrast to $\vec{z}$-vortices and the above-mentioned point (a), $\vec{x}$-vortices lie on the k_x axis (cf. Fig. 9) and reduce the optical affinity $|\vec{x}\cdot\boldsymbol{A}_{cv}|^2$ for small $|k_z|$. Thus for high frequencies, $\vec{x}$-vortices deactivate the horizontal relaxation paths that have the largest anomalous shift ($\vec{z}\cdot\boldsymbol{\Omega}\times d\boldsymbol{k}$); this effectively suppresses the intraband shift current and leads to a linear disparity ($|\vec{z}\cdot\boldsymbol{\sigma}_{\vec{x}}-\vec{z}\cdot\boldsymbol{\sigma}_{\vec{z}}|$) of order 0.1 mA V^{-2}.

Let us close this section by explaining the $(1/E_g)$-divergence (and concomitant sign change) of $\vec{z}\cdot\boldsymbol{\sigma}^{\rm intra}_{\boldsymbol{\epsilon}}$ across the topological phase transition, as illustrated in Figs. 8(e) and 8(f). It suffices to show that the affinity shift loop also has a $(1/E_g)$-divergence, according to the proportionality relation in Eq. (13). The contribution to the affinity shift loop [cf. Eq. (23)] by a pair of diametrically opposed geodesic paths $(\pm\mathcal{P})$ can be further decomposed as

$$\begin{aligned}&{\rm Aff}_{\rm ave}\left(\int_{-\mathcal{P}}+\int_{\mathcal{P}}\right)\vec{z}\cdot\boldsymbol{\Omega}_{c\boldsymbol{k}}\times d\boldsymbol{k}\\&+\delta{\rm Aff}\left(\int_{-\mathcal{P}}-\int_{\mathcal{P}}\right)\vec{z}\cdot\boldsymbol{\Omega}_{c\boldsymbol{k}}\times d\boldsymbol{k},\end{aligned}\tag{24}$$

with ${\rm Aff}_{\rm ave}\equiv(|\vec{z}\cdot\boldsymbol{A}_{cv}|^2_{-\mathcal{P}}+|\vec{z}\cdot\boldsymbol{A}_{cv}|^2_{\mathcal{P}})/2$ and $\delta{\rm Aff}\equiv(|\vec{z}\cdot\boldsymbol{A}_{cv}|^2_{-\mathcal{P}}-|\vec{z}\cdot\boldsymbol{A}_{cv}|^2_{\mathcal{P}})/2$. The term proportional to ${\rm Aff}_{\rm ave}$ is asymptotically irrelevant as $|E_g|$ approaches zero, due to an emergent left-right symmetry of the massive Dirac fermion [Eq. (21)] about the extremal wave vector.[34] What remains is to evaluate the asymptotic behavior of the term proportional to δAff: the integral $(\int_{-\mathcal{P}}-\int_{\mathcal{P}})$ of the anomalous shift vector diverges as $1/E_g$, because (i) $\vec{z}\cdot\boldsymbol{\Omega}_{c\boldsymbol{k}}\times d\boldsymbol{k}/|d\boldsymbol{k}|$ diverges as $E_g/|E_g|^3$ at the band extremum, which is a well-known type of divergence for massive Dirac fermions [cf. Eq. (H14)], and (ii) the width of the Berry curvature hot spot is of order $|m'|/\hbar v\propto|E_g|$. Combining both (i) and (ii), the second integral in Eq. (24) is estimated as (extremal value of curvature) $\times$ (hot-spot width), which is proportional to $E_g/|E_g|^3\times|E_g|=1/E_g$. Because this divergence applies to any pair of diametrically opposed geodesic paths, the affinity shift loop must likewise diverge as $1/E_g$, and thus also $\vec{z}\cdot\boldsymbol{\sigma}^{\rm intra}_{\boldsymbol{\epsilon}}\approx\vec{z}\cdot\boldsymbol{\sigma}_{\boldsymbol{\epsilon}}$.

There are two reasons why this divergence will be cut off in a more realistic model of BiTeI, meaning that the $1/E_g$ behavior breaks down in a narrow energy interval: $|E_g|<E_g^{\rm cut}$:

(i) The first reason is that not all photoexcited quasiparticles will relax all the way down to the conduction-band bottom (where the Berry curvature diverges), but instead they will relax to a Maxwell-Boltzmann distribution with a characteristic thermal energy k_BT_l.[35]

(ii) The second reason is that the O(2) symmetry of our effective model of BiTeI is only approximate; in real BiTeI, the topological phase transition (between two topologically distinct semiconductors) is not intermediated by an O(2)-symmetric nodal-loop band touching, but by a C_{3v}-symmetric Weyl-semimetallic phase; the energy scale of the C_{3v}-symmetric trigonal warping is estimated to be about 10 meV [23].

Both reasons suggest the $1/E_g$ behavior of $\vec{z}\cdot\boldsymbol{\sigma}_{\boldsymbol{\epsilon}}$ to be precluded with a cutoff $E_g^{\rm cut}$ that is comparable to either k_BT_l or 10 meV, whichever is larger.

VI. DISCUSSION AND OUTLOOK

A. The threefold way

The steady shift current density in a direct-gap semiconductor has a threefold decomposition: $\boldsymbol{j}=\boldsymbol{j}_{\rm exc}+\boldsymbol{j}_{\rm rec}+\boldsymbol{j}_{\rm intra}$, corresponding, respectively, to current contributions by interband *exc*itation [cf. Eq. (7)], interband *rec*ombination [Eq. (8)], and *intra*band relaxation [Eq. (9)]. While this threefold decomposition has been studied for simplified models of pyroelectrics and piezoelectrics [5], it is here that $\boldsymbol{j}_{\rm rec}+\boldsymbol{j}_{\rm intra}$ acquires a new dimension of understanding through the lens of wave-function geometry. Geometrical notions (such as the Berry phase) transcend the traditional classification of piezoelectrics versus pyroelectrics, and they provide overarching principles to guide our interpretation of the out-of-equilibrium, many-body dynamics of photoexcited matter.

One of our main results is that the excitation-induced current density $\boldsymbol{j}_{\rm exc}$ can be outweighed by either $\boldsymbol{j}_{\rm intra}$ or $\boldsymbol{j}_{\rm rec}$, especially in semiconductors characterized by large intraband Berry curvature or optical vortices (topological singularities in the interband Berry phase). Model semiconductors with large Berry curvature exhibit a shift-current conductivity that is of order mA V^{-2} without fine-tuning of the incident radiation frequency; in the presence of optical vortices, the conductivity can change by $\sim$mA V^{-2} if the linear polarization vector flips by 90°. These estimates of the conductivity assumed a generic value of the joint density of states, but in principle the joint density of states can be further optimized [44] for a synergistic enhancement. To our knowledge, no measurement of the short-circuit conductivity in shift-current materials has reached the mA V^{-2} range.

B. Wave-function approach to photovoltaic materials

Establishing the steady shift current in the broader framework of wave-function geometry confers an advantage: we acquire a Rosetta stone to translate our vast body of knowledge (on topological materials) to concrete predictions of photovoltaic materials. Here are two cases in point:

(i) Intraband relaxation due to electron-phonon scattering results in an anomalous shift that is proportional to the intraband Berry curvature; cf. Eq. (5).[36] Let us juxtapose this

enhance $|\vec{y}\cdot\boldsymbol{S}_{\rm rec}|$ by precluding an orientational-disordered average, as explained in Appendix F.

[33] A minor difference between Figs. 8(c) and 8(d) is that $\vec{z}\cdot\boldsymbol{\sigma}^{\rm exc}_{\vec{x}}$ is also nonmonotonic and remains small at low frequencies. This occurs because $\vec{z}\cdot\boldsymbol{S}^{\vec{x}}_{v\boldsymbol{k}_{\rm ext}\leftarrow c\boldsymbol{k}_{\rm ext}}=0$, due to an emergent chiral symmetry at the band edge, as elaborated in Appendix H.

[34] This left-right symmetry is explained in App. H 3.

[35] The preceding calculation of the $1/E_g$ divergence assumed that as $|E_g|\to0$, k_BT_l must likewise $\to0$; indeed, the geodesic approximation relied on $k_BT_l\ll|E_g|$, as was explained in Sec. II.

[36] Electron-phonon scattering is not the only mechanism for intraband relaxation in a direct-gap semiconductor; electron-impurity scattering also results in a shift [5], which may substantially contribute to $\boldsymbol{j}_{\rm intra}$ for dirtier samples. The impurity-mediated shift is

anomalous shift with the nonlinear Hall effect predicted by Sodemann and Fu [57]. What matters to the anomalous shift current is the Berry curvature of all Bloch states enclosed by the excitation surface; this contrasts with the nonlinear Hall effect, which depends (at low temperature) on the Berry curvature of the Fermi surface. However, the two effects are not completely divorced: a semiconductor with a large anomalous shift current is continuously tunable (e.g., by doping) to a metal with a large nonlinear Hall effect. This is evident from Fig. 5(a) if one imagines the excitation surface to be a Fermi surface.

(ii) While the topological-matter community is well-versed in finding materials with large intraband Berry curvature, it is presently unclear which materials have optical vorticity. On the one hand, a high-throughput *ab initio* algorithm has been proposed in Ref. [48] to search for materials with optical vorticity. On the other hand, it would also be advantageous to identify general, topological principles that guarantee the existence of optical vorticity in certain classes of materials. One such principle is the Chern-vorticity theorem in Eq. (18), which relates the Chern number (of a 2D cross-section of a 3D Brillouin zone, or of a 2D Brillouin zone) to the net vorticity (of the same 2D cross-section or 2D Brillouin zone). This theorem has broad implications for the vorticity in topological semimetals and topological insulators, one of which is the necessary existence of optical vorticity in BiTeI.

Our case study of the linear photogalvanic effect (LPGE) in semiconducting BiTeI (Sec. V) illustrates four principles:

(a) Because the $\boldsymbol{k}$-locations of optical vortices depend on the light polarization, the steady shift current depends sensitively on the light polarization. If the excitation surface is proximate to an optical vortex, the photonic shift vector is orientationally disordered over the excitation surface, which tends to reduce the excitation shift current.

(b) The recombination shift strongly depends on the symmetry of the Hamiltonian at the wave vectors of recombination. For BiTeI, chiral symmetry reduces the recombination shift vector to about a third of the lattice period.

(c) Because band wave functions can depend strongly on energy, so can the steady shift current depend sensitively on the photon frequency. A rule of thumb is that the net shift conductivity changes sign in the transition from a photon-mediated shift current (at low frequencies corresponding to band-edge excitation) to a phonon-mediated anomalous shift current (at higher frequencies).

(d) The steady shift current is sensitive to discrete changes of the wave-function topology. In particular, the steady shift current changes sign across the $\mathbb{Z}_2$ topological phase transition, and the magnitude of said current is extraordinarily large in the vicinity of the transition.

closely analogous to the "side jump" in the anomalous Hall effect of magnetic metals [56]. In their study of the "side jump," Sinitsyn *et al.* have argued that the impurity-mediated shift reduces to the anomalous shift [Eq. (5)] under two assumptions: (i) the small-angle scattering is dominant, and (ii) the cell-periodic component of the Bloch function is spatially homogeneous. The improbability of either assumption makes for a tenuous relation between the impurity-mediated shift and the anomalous shift.

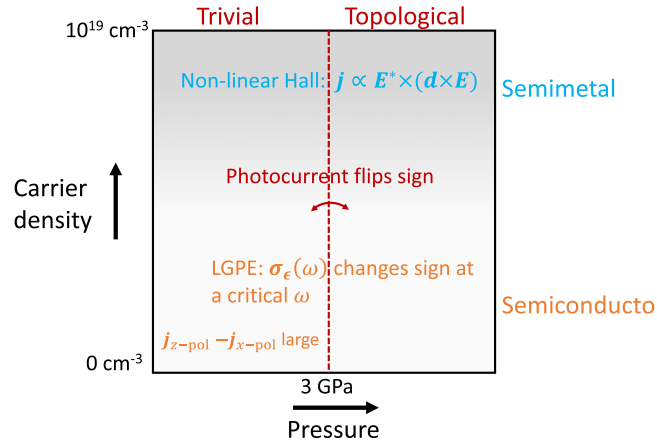

FIG. 10. A photovoltaic phase diagram of BiTeI. The carrier density can be tuned by varying the growth method [59,60] or by doping with Cu [61]. Pressure can be applied within a diamond anvil cell [53] or by chemical substitution.

The experimental implications of (a)–(d) are summarized in Fig. 10. The figure also illustrates the bulk photovoltaic current in semimetallic BiTeI: a previous theory [58] has predicted a nonlinear Hall current that depends on the Berry curvature dipole $\boldsymbol{d}$ of the Fermi surface and also changes sign across the topological phase transition. For light sources in the 100 THz regime (and higher), the bulk photovoltaic current of semimetallic BiTeI is dominated by the circular photogalvanic effect (CPGE),[37] but this is not true for semiconducting BiTeI.

Our *photovoltaic phase diagram* of BiTeI showcases the tight correlations between wave-function geometry and the bulk photovoltaic effect over a wide range of carrier densities and on both sides of the topological phase transition. We hypothesize that similar correlations hold for other topological materials, suggesting that the bulk photovoltaic effect is an unprecedented phenomenological framework in which to faithfully diagnose and sensitively characterize topological materials.

C. Transient versus steady photovoltaic current

Section II demonstrates that the steady, excitation-induced current $\boldsymbol{j}_{\mathrm{exc}}$ well approximates the transient current at the onset of radiation. By subtracting the transient current (measured at early times) from the steady current (measured at late times), would one obtain $\boldsymbol{j}_{\mathrm{rec}} + \boldsymbol{j}_{\mathrm{intra}}$?[38] Not quite, because the steady photovoltaic current includes not only the shift current but also the ballistic current.[39]

[37]The CPGE is larger than the LPGE by a factor $\omega\tau \gg 1$, with ω being the photon frequency and $\tau \gtrsim 100$ fs being the momentum relaxation time [57].

[38]We consider an experimental geometry where the photon-dragged current vanishes [1,2], and where the surface photovoltaic current is separable from the bulk photovoltaic current [62].

[39]The ballistic current results from a $(\boldsymbol{k} \to -\boldsymbol{k})$ asymmetry in the quasiparticle distribution [3,22,30,62]; such asymmetry can be induced by electron-impurity, electron-phonon [63], and electron-hole interactions [64]. Does a large phonon-induced ballistic current cor-

It may be argued that the transient current is insensitive to the temperature of a photoexcited semiconductor,[40] while the steady current includes $\boldsymbol{j}_{\text{intra}}$, which is sensitive to the temperature of the phonons. Suggestively, a substantial disparity in the temperature dependence (of the transient versus steady photocurrent) was observed for a ferroelectric charge-transfer complex, though the disparity was attributed by Nakamura *et al.* to the formation of a Schottky barrier at the sample-electrode interface [21]. It would be interesting to see if this disparity persists for a different electrode whose work function is identical to that of the sample.

Conversely, it has been reported that the Kraut-Baltz-Sipe-Shrekbtii formula [4,6,13] for the shift current adequately describes the photoconductivity measurements of *n*-GaP [65], suggesting that $\boldsymbol{j}_{\text{rec}}+\boldsymbol{j}_{\text{intra}}$ is small for this material; this hypothesis can be tested by an *ab initio* study of the intraband/interband Berry phase of GaP, which we would love to see.

D. The loop approach to shift currents

Our analysis of the direct-gap semiconductor relied on identifying a reduced set of quasiparticle transitions that concatenate into loops (cf. Fig. 4) and predominantly contribute to the steady shift current [Eq. (1)]. Appendix D shows how to exactly reformulate the BIS formula [Eq. (1)] as a sum over *loop currents*:

$$\boldsymbol{j}=-\frac{|e|}{\mathcal{V}}\sum_{B,B',m}\boldsymbol{S}^{m}_{B'\leftarrow B}\big(\mathcal{A}^{m}_{B'\leftarrow B}-\mathcal{E}^{m}_{B\leftarrow B'}\big)=\sum_{\text{loop}}\boldsymbol{j}_{\text{loop}},\tag{25}$$

with $\boldsymbol{j}_{\text{loop}}$ meaning the current contributed by a closed flow line (in energy-momentum space) of one-electron probability.

The loop formulation holistically treats excitation, relaxation, and recombination as inextricably linked processes; such linkage is epitomized by the shift loop in Eqs. (6)–(13). Using loops allows us to derive general properties of the steady shift current that do not necessarily apply to the transient shift current; in particular, a purported relation between the shift current and interband polarization differences [49] is shown in Appendix D 3 to apply to the transient current but not the steady current. A related advantage of loops is calculational: approximating the steady shift current by a reduced family of predominant loops [e.g., via Eq. (13)] requires far fewer computational resources than simulating a quasiparticle distribution f_B that sets the collisional integral to zero, and then inputting f_B into the BIS formula; cf. Appendix C 2.

Our loop current formulation is applicable beyond direct-gap semiconductors, with the caveat that the predominant loops may change depending on the context. For instance, recombination in indirect-gap semiconductors is intermediated by transitions between Bloch states and impurity-localized states; such transitions also contribute to the shift current [30]. In small-gap semiconductors or Dirac-Weyl semimetals, interband recombination may be contributed by electron-phonon scattering. It would be desirable to develop a theory of the steady shift current in Dirac-Weyl semimetals, for which the anomalous shift may potentially be large. It is hoped that photoconductivity measurements of TaAs [66] would benefit from a reinterpretation of what exactly is causing the shift.

E. The potential for solar-cell applications

A large short-circuit conductivity is not sufficient for solar-cell applications; also required is a large open-circuit photovoltage to generate sufficient electrical power. Though shift-current materials can have open-circuit photovoltages that greatly exceed the band gap [67–69], the product of the short-circuit current and open-circuit photovoltage may be limited depending on the architecture of the shift-driven solar cell.

Let us first consider Pusch *et al.*'s model [70] of a shift-driven solar cell, in which a homogeneous shift-current-carrying intrinsic semiconductor is connected via leads to an external load; henceforth, we refer to this as the *PRCE cell*. Assuming a few ideal conditions, namely that the contacts with the leads do not introduce additional resistance, and that temperature is sufficiently low ($k_BT\ll E_g$) to ignore the dark conductivity, the light-to-electrical energy conversion efficiency of a PRCE cell is calculated in Appendix K to be

$$\text{eff}=\frac{1}{4}\frac{m_r\mathfrak{s}^2/\tau_{\text{tr}}\tau_{\text{rec}}}{\hbar\omega}\tag{26}$$

under monochromatic illumination with frequency ω. Here, $m_r^{-1}=m_e^{-1}+m_h^{-1}$ is the reduced mass of an electron-hole pair in the parabolic-band approximation, τ_{tr} is the Drude-type transport lifetime for photoexcited carriers, and τ_{rec} is the recombination lifetime; cf. Sec. II. $\mathfrak{s}$ can be interpreted as the average shift per photoexcited electron-hole pair, and is expressible as a normalized affinity shift loop [cf. Eq. (10)]:

$$\mathfrak{s}=\frac{\text{ASL}_{\boldsymbol{\epsilon}_s,\omega}}{\langle f_{vc\boldsymbol{k}}|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2\rangle_\omega}.\tag{27}$$

Our formula for the efficiency is essentially identical to Eq. (11) in Ref. [70], except that our $\mathfrak{s}$ replaces their heuristically defined "average charge displacement R" with a precisely calculable formula; moreover, Eqs. (26) and (27) with Eqs. (6)–(10) clarify the oft-ignored roles of $\boldsymbol{j}_{\text{rec}}+\boldsymbol{j}_{\text{intra}}$ in the operation of a shift-driven solar cell.

Even if $\mathfrak{s}$ is of order the lattice period a, it may be seen that eff $\ll 1$ at solar frequencies and with typical values for $a\approx 5$ Å, $m_r\approx m_f$ (free-electron mass), $\tau_{\text{tr}}\approx 100$ fs, and $\tau_{\text{rec}}\approx 1$ ns for radiative recombination. More appreciable efficiencies are expected for heavy-fermion materials dominated by faster nonradiative recombination [71], and with extraordinarily large lattice periods that characterize (moiré) superlattices. Going beyond the PRCE model, inhomogeneous doping of the shift-current-carrying semiconductor would marry advantages of both the shift mechanism and conventional barrier layer photovoltaics; whether such a synergy is incremental or transformative remains to be seen.

relate with a large phonon-induced $\boldsymbol{j}_{\text{intra}}$? We leave this open question for future investigations.

[40] $\boldsymbol{j}_{\text{exc}}$ depends on temperature only through the photoexcitation transition rate, which is proportional to $(1-f_{c\boldsymbol{k}})f_{v\boldsymbol{k}}$, with f the Fermi-Dirac distribution. Assuming the band gap greatly exceeds k_BT_0, $(1-f_{c\boldsymbol{k}})f_{v\boldsymbol{k}}\approx 1$ is insensitive to T_0.

ACKNOWLEDGMENTS

We thank Boris Sturman for patient explanations of the BIS theory, Benjamin Fregoso for illuminating discussions of third-order optical responses, Michael Schuler for educating us on the electron-phonon interaction, and Pavlo Sukhachov for a detailed reading and commentary. An illuminating discussion with Andreas Pusch helped us formulate a more nuanced discussion of shift-driven solar cells. This research was supported in part by the National Science Foundation under Grant No. NSF PHY-1748958. In the final stages of this work, P.Z. received support from the Center for Emergent Materials, an NSF MRSEC, under Award No. DMR-2011876.

APPENDIX

The Appendixes contains several supplementary results for the specialized audience. Most of these results have been referenced and motivated in the main text. An organizational structure is presented to help with navigation:

Appendix A: This glossary collects many recurrent equations and symbols for easy reference. The equations include the Belinicher-Ivchenko-Sturman (BIS) formula for the shift current (Appendix A 1) and its threefold decomposition into contributions by excitation, recombination, and intraband transitions (Appendix A 4). The BIS formula inputs the quasiparticle, photonic, and phononic occupancies and outputs a current; certain assumptions about these occupancies are recorded in Appendix A 2. All explicit calculations of the shift current in this work have been based on a two-band, direct-gap semiconducting model; the terminology that surrounds this model is collected in Appendix A 3.

Appendix B: This Appendix summarizes several salient aspects of the steady nonequilibrium distribution of photoexcited carriers. Included is a review of the different relaxation mechanisms for a hot photoexcited carrier in a semiconductor (Appendix B 1) and a derivation of the associated kinetic model which applies in the regime of low carrier density (Appendix B 2). Simplified versions of this kinetic model are described if there is electron-hole symmetry (Appendix B 2 a) and if the collisional integral is constant along isoenergy surfaces (Appendix B 2 b).

Appendix C: The BIS formula of the shift current is pedagogically derived in Appendix C 1, numerically calculated in Appendix C 2, and compared to other formulas of the shift current in Appendix C 3. This comparison elucidates what is missing from the Kraut-Baltz-Sipe-Shkrebtii formula [4,6,13], as well as formulas derived from dissipative Floquet methods [17,19,20]. Appendix C 1 also demonstrates that to describe the steady state perturbatively, the zeroth-order state is emphatically *not* the thermal equilibrium state (in the absence of the light source).

Appendix D: The steady shift current is equivalently formulated in terms of loop currents. The basic ingredients of the loop formulation are presented in Appendix D 1, namely the loop decomposition of the probability flow network, the shift loop, and the loop current. A theorem derived in Appendix D 2 formalizes the equivalence between the BIS formula and a sum of loop currents. This theorem is applied to revise a purported relation [49] between the shift current and interband polarization differences (Appendix D 3), and to derive the geodesic approximation of the shift conductivity for 3D semiconductors (Appendix D 4) and quasi-2D semiconductors (Appendix D 6). The geodesic approximation is rigorously justified as the small-angle-scattering limit of the BIS formula in Appendix D 5.

Appendix E: The transient shift current is shown here to be well approximated by the excitation-induced component of the steady shift current. This Appendix rigorously elaborates an argument presented in Sec. II.

Appendix F: The shift conductivity is calculated for a model Hamiltonian with optical vortices to corroborate claims made in Sec. IV about the vortex-induced linear disparity of the shift conductivity.

Appendix G: A theorem relating Chern numbers to optical vorticity [cf. Eq. (18)] is proven here.

Appendix H: A few facts that support our case study on BiTeI are presented here, including an effective Hamiltonian of a massive Dirac fermion that holds near the topological phase transition, as well as the vanishing of the shift current at the band edge, for x- and y-polarized light.

Appendix I: An alternative derivation of the anomalous shift vector is provided.

Appendix J: A misconception about the electron-phonon scattering rate is exposed. The misconception traces back to a fallacious belief of a universally applicable gauge for the Bloch wave function.

Appendix K: The energy conversion efficiency is calculated for a solar cell based on an intrinsic, shift-current-carrying semiconductor.

APPENDIX A: GLOSSARY

1. The Belinicher-Ivchenko-Sturman formula

We use "Belinicher-Ivchenko-Sturman (BIS) formula" to refer to two sets of equations for the phonon-mediated and photon-mediated shift current. The former current is given by

$$\boldsymbol{j} = -\frac{|e|}{\mathcal{V}} \sum_{B,B',m} \boldsymbol{S}^{m}_{B'\leftarrow B}\left(\mathcal{A}^{m}_{B'\leftarrow B} - \mathcal{E}^{m}_{B\leftarrow B'}\right); \quad B = (b\boldsymbol{k}), \quad B' = (b'\boldsymbol{k}') \quad m = (\boldsymbol{q}p), \tag{A1}$$

with the phononic shift vector given by

$$\text{Phonon}: \quad \boldsymbol{S}^{m}_{B'\leftarrow B} = -(\nabla_{\boldsymbol{k}} + \nabla_{\boldsymbol{k}'}) \arg V^{m}_{B',B} + \boldsymbol{A}_{b'b'\boldsymbol{k}'} - \boldsymbol{A}_{bb\boldsymbol{k}} = -\boldsymbol{S}^{-m}_{B\leftarrow B'}, \tag{A2}$$

and with the difference in absorption and emission transition rates given by

$$\text{Phonon:}\quad \mathcal{A}^m_{B'\leftarrow B}-\mathcal{E}^m_{B\leftarrow B'}=\delta_{\boldsymbol{q},\boldsymbol{k}'-\boldsymbol{k}}\frac{2\pi}{\hbar}\left|V^m_{B'B}\right|^2\delta(E_{B'B}-\hbar\omega_m)\{(1-f_{B'})f_BN_m-(1-f_B)f_{B'}(N_m+1)\},\tag{A3}$$

with $E_{B'B}=E_{B'}-E_B$. Actually, the above equations are more general than those presented in Ref. [5], in that the above equations allow for interband phonon-mediated transitions, while the formula in the BIS paper does not. This being a minor generalization, we will anyway refer to our final formula as the BIS formula.

The BIS formula for the photon-mediated shift current combines Eq. (A1) with

$$\text{Photon:}\quad \boldsymbol{S}^m_{b'\boldsymbol{k}\leftarrow b\boldsymbol{k}}=-\nabla_{\boldsymbol{k}}\arg\boldsymbol{\epsilon}_m\cdot\boldsymbol{A}_{b'b\boldsymbol{k}}+\boldsymbol{A}_{b'b'\boldsymbol{k}}-\boldsymbol{A}_{bb\boldsymbol{k}}=-\boldsymbol{S}^{-m}_{b\boldsymbol{k}\leftarrow b'\boldsymbol{k}};\tag{A4}$$

$$\mathcal{A}^m_{B'\leftarrow B}-\mathcal{E}^m_{B\leftarrow B'}=\delta_{\boldsymbol{k},\boldsymbol{k}'}\frac{(2\pi e)^2\omega_m}{\mathcal{V}}\left|\boldsymbol{\epsilon}_m\cdot\boldsymbol{A}_{b'b\boldsymbol{k}}\right|^2\delta(E_{B'B}-\hbar\omega_m)\{(1-f_{B'})f_BN_m-(1-f_B)f_{B'}(N_m+1)\},\tag{A5}$$

in Gaussian units. In Ref. [5], $\mathcal{A}^m_{B'\leftarrow B}-\mathcal{E}^m_{B\leftarrow B'}$ appears with an additional multiplicative factor of $1/n^2$, with n the refractive index; this factor should not be there, according to our derivation in Appendix C 1.[41]

In the steady state, the quasiparticle distribution f_B satisfies a nondetailed balance condition that represents an invariance under simultaneous collisions with all bosons:

$$\text{For all } B,\ 0=\sum_m\sum_{B'}\left(\mathcal{A}^m_{B\leftarrow B'}+\mathcal{E}^m_{B\leftarrow B'}-\mathcal{A}^m_{B'\leftarrow B}-\mathcal{E}^m_{B'\leftarrow B}\right),\tag{A6}$$

with $\sum_m$ summing over all photonic and phononic modes. The right-hand side of the above equality may be viewed as the collisional integral evaluated to second order in the electron-boson coupling, i.e., the integral has a form expected from Fermi's golden rule. Equations (A1)–(A6) represent a closed set of equations to determine the shift current: one first determines f_B from Eq. (A6) and then inputs f_B into Eq. (A1).

2. Quasiparticle, photon, and phonon occupancies

The above equations show that the BIS current is a functional of the quasiparticle, photon, and phonon occupancies:

$$\boldsymbol{j}=\boldsymbol{j}\left[f_B,N^{\text{phot}}_m,N^{\text{phon}}_m\right],\tag{A7}$$

with the dependence on occupancies given solely by the transition rates $\mathcal{A}=\mathcal{A}[f_B,N^{\text{phot}}_m,N^{\text{phon}}_m]$ and $\mathcal{E}=\mathcal{E}[f_B,N^{\text{phot}}_m,N^{\text{phon}}_m]$.

Throughout this work, f symbolizes the occupancy of charged, fermionic quasiparticles that are long-lived in an insulator [72], though strictly speaking we do not account for the renormalization of the wave function [73]. When there is no T (for temperature) superscript on f_B, f_B should be understood as the nonequilibrium, steady distribution determined in a kinetic model (cf. Appendix B 2), hence f_B generically deviates from the thermal

$$\text{Fermi-Dirac distribution:}\quad f^T_B=\frac{1}{e^{(E_B-\mu)/k_BT}+1}.\tag{A8}$$

The phonons are assumed to thermalize with a lattice temperature T_l, meaning that the phonon occupancy is a Planck distribution:

$$\text{Phonon:}\quad N_m=N^{T_l}_m=\frac{1}{e^{\hbar\omega_m/k_BT_l}-1}.\tag{A9}$$

We will not always have the "phon" or "phot" superscript on N_m, so the meaning of m should be deduced from the local context. The minimal frequency for optical phonons is defined to be the optical phonon threshold:

$$\text{Optical phonon:}\ \min\{\omega_m\}=\Omega_o;\quad k_BT_l\ll\hbar\Omega_o\Rightarrow N^{T_l}_m\ll 1.\tag{A10}$$

$\hbar\Omega_o$ is typically comparable to k_B times room temperature. We assume in calculations of the shift conductivity that the lattice temperature is small compared to the optical phonon energy, hence the thermal occupancy of optical phonons is negligible.

The photon occupancy is assumed to be a sum of thermal and source-generated contributions:

$$\text{Photon:}\quad N_m=N^{T_l}_m+\Delta N_s\delta_{m,m_s};\quad \Delta N_s\gg 1.\tag{A11}$$

We have assumed that the source is bright ($\Delta N_s\gg 1$) and produces photons of a single, linearly polarized mode:

$$\text{Source:}\quad \text{mode}=m_s;\quad \text{frequency}=\omega_s=c||\boldsymbol{q}_s||>E_g/\hbar;\quad \text{polarization vector}=\boldsymbol{\epsilon}_s\in\mathbb{R}^3.\tag{A12}$$

[41] The $1/n^2$ factor is absent in the Sturman-Fridkin monograph [30], which followed after the BIS paper. A factor of $1/n$ appears only in the conversion of photon occupancies to the radiation intensity; cf. Appendix C 3.

In the classical approximation to the radiation field, ΔN_s can be be expressed in terms of the electric-wave amplitude according to[42]

$$\boldsymbol{\mathcal{E}} = \mathcal{E}_\omega \boldsymbol{\epsilon}_s e^{i(\boldsymbol{q}_s\cdot\boldsymbol{r}-\omega_s t)} + \text{c.c.}; \quad |\mathcal{E}_\omega|^2 = 2\pi\frac{\hbar\omega\Delta N_s}{\mathcal{V}}. \tag{A13}$$

The vector-valued *shift conductivity* is defined by the nonlinear current response:

$$\boldsymbol{j} = \boldsymbol{\sigma}_{\boldsymbol{\epsilon},\omega}|\mathcal{E}_\omega|^2, \tag{A14}$$

with $\boldsymbol{j}$ the shift current in Eq. (A1).

A basic property of the quasiparticle transition rates is that the absorption and emission rates cancel out if the fermions and bosons are thermalized with the same temperature:

$$\text{Detailed balance}: \quad \left(\mathcal{A}^m_{B'\leftarrow B} - \mathcal{E}^m_{B\leftarrow B'}\right)_{f_B^T, f_{B'}^T, N_m^T} = 0. \tag{A15}$$

This holds for both phonons and phonons, as one may verify by substituting the Fermi-Dirac and Planck distributions into Eqs. (A3)–(A5).

3. Direct-gap semiconducting model

All explicit calculations of the shift conductivity in this work are based on a model of a direct-gap intrinsic semiconductor with two bands (not counting spin):

$$\begin{aligned}&\text{Band indices}: \quad b = c \text{ (conduction)}; \quad b = v \text{ (valence)};\\ &\text{Bloch labels}: \quad C = (c\boldsymbol{k}); \quad V = (v\boldsymbol{k}).\end{aligned} \tag{A16}$$

$E_C - E_V$ is assumed to be minimized at a single wave vector, which we call the *extremal wave vector*:

$$\min\{E_{cv\boldsymbol{k}}\} = E_{cv\boldsymbol{k}_{\text{ext}}} = E_g, \quad E_{cv\boldsymbol{k}} = E_C - E_V. \tag{A17}$$

The band gap E_g is assumed to be large compared to the lattice temperature T_l (with source turned on) and the equilibrium temperature T_0 (with source turned off), hence the equilibrium Fermi-Dirac occupancies are close to being binary:

$$k_B T_0 \ll E_g \Rightarrow f_C^{T_0} \approx 0; \quad f_V^{T_0} \approx 1. \tag{A18}$$

Isoenergy surfaces of a band are defined to $\boldsymbol{k}$-surfaces in which the band dispersion is constant:

$$\text{Isoenergy surface of band } b \text{ and energy } E \;\equiv\; \text{all } \boldsymbol{k} \text{ satisfying} \quad E_B = E. \tag{A19}$$

The source photon frequency is assumed large enough that resonant absorption can occur across the band gap, and the excitation surface is defined as the $\boldsymbol{k}$-surface where resonant optical absorption can occur:

$$\text{ES} \equiv \text{ excitation surface } \equiv \text{ all } \boldsymbol{k} \text{ satisfying} \quad E_{cv\boldsymbol{k}} = \hbar\omega_s > E_g. \tag{A20}$$

We define $E_{b,\text{exc}}$ as the energies of b-band states on the excitation surface:

$$\text{Excitation energy}: \quad E_{c,\text{exc}} = \{E_C \mid \boldsymbol{k} \in \text{ES}\}; \quad E_{v,\text{exc}} = \{E_V \mid \boldsymbol{k} \in \text{ES}\}. \tag{A21}$$

We will encounter symmetric models where $E_{c,\text{exc}}$ is degenerate for all conduction-band states on the excitation surface, meaning the excitation surface is an isoenergy surface of energy $E_{c,\text{exc}}$.

For photonic modes that mediate resonant interband transitions, their thermal occupancy is negligible:

$$\text{Resonant photon}: \quad N_m^{T_l}\delta(E_{cv\boldsymbol{k}} - \hbar\omega_m) \ll \delta(E_{cv\boldsymbol{k}} - \hbar\omega_m) \Leftarrow k_B T_l \ll E_g. \tag{A22}$$

The passive and active regions of either band are defined with respect to the optical phonon threshold [Eq. (A10)]:

$$\begin{aligned}\text{Conduction: :} \quad & E_{c\boldsymbol{k}} - E_{c\boldsymbol{k}_{\text{ext}}} > \hbar\Omega_o \text{ (active)}; \quad 0 < E_{c\boldsymbol{k}} - E_{c\boldsymbol{k}_{\text{ext}}} < \hbar\Omega_o \text{ (passive)};\\ \text{Valence:} \quad & -(E_{v\boldsymbol{k}} - E_{v\boldsymbol{k}_{\text{ext}}}) > \hbar\Omega_o \text{ (active)}; \quad 0 < -(E_{v\boldsymbol{k}} - E_{v\boldsymbol{k}_{\text{ext}}}) < \hbar\Omega_o \text{ (passive)}.\end{aligned} \tag{A23}$$

4. Excitation, recombination, and intraband components of the shift current

As discussed in Sec. II, the steady shift current can be decomposed into contributions by excitation, recombination, and intraband relaxation processes:

$$\text{Threefold decomposition}: \quad \boldsymbol{j} = \boldsymbol{j}_{\text{exc}} + \boldsymbol{j}_{\text{intra}} + \boldsymbol{j}_{\text{rec}}. \tag{A24}$$

Here, we present the explicit expressions for each of the three components in Eqs. (A25)–(A30).

The *excitation-induced current* is defined to be the shift current contributed by interband, vertical transitions over the excitation surface [Eq. (A20)]:

$$\text{Excitation-induced current:} \quad \boldsymbol{j}_{\text{exc}} = \boldsymbol{j}[f]_{\boldsymbol{k}=\boldsymbol{k}'\in\text{ES}} \approx -2_{\uparrow\downarrow}\frac{|e|}{\mathcal{V}}\sum_{\boldsymbol{k}} \boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}} I^{m_s}_{\text{exc}\boldsymbol{k}}; \tag{A25}$$

$$\text{Excitation rate:} \quad I^{m_s}_{\text{exc}\boldsymbol{k}} = \frac{(2\pi e)^2\omega_s}{\mathcal{V}}|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2 f_{vc\boldsymbol{k}}\delta(E_{cv\boldsymbol{k}} - \hbar\omega_s)\Delta N_s. \tag{A26}$$

[42] This may be derived from the standard relation [74,75] between the classical electromagnetic vector potential and the photon number:

$$\boldsymbol{\mathcal{E}} = -\frac{1}{c}\frac{\partial \boldsymbol{A}^s_{\text{em}}}{\partial t}, \quad \boldsymbol{A}^s_{\text{em}} = 2c\sqrt{\frac{h\,\Delta N_s}{\omega_s\,\mathcal{V}}}\boldsymbol{\epsilon}_s\,\cos(\boldsymbol{q}_s\cdot\boldsymbol{r} - \omega_s t).$$

$2_{\uparrow\downarrow} = 2$, with the additional subscript reminding us that this two originates from the spin degree of freedom. $f_{vc\boldsymbol{k}}$ is defined to be $f_{v\boldsymbol{k}} - f_{c\boldsymbol{k}}$. Equation (A25) has been presented for the case in which a pair of spinless bands (labeled $b = v$ and c) are optically excited; if there are more than a pair, simply sum the right-hand side of Eq. (A25) over all pairs.

Equation (A25) is derived by restricting $\sum_{\boldsymbol{k}\boldsymbol{k}'}$ in Eq. (A1) with the condition that $\boldsymbol{k} = \boldsymbol{k}'$ lies on the excitation surface, and this is the meaning of $\boldsymbol{j}[f]_{\boldsymbol{k}=\boldsymbol{k}'\in\mathrm{ES}}$. In principle, $\sum_m$ in $\boldsymbol{j}[f]_{\boldsymbol{k}=\boldsymbol{k}'\in\mathrm{ES}}$ sums over all bosonic modes with the same frequency $\omega_m = \omega_s$ as the source-generated photons. This $\sum_m$ may be restricted to photonic modes, because the band gap is presumed to exceed the optical phonon energies. Photon-mediated vertical transitions over the excitation surface can be divided into five classes, according to Eqs. (A5) and (A11):

(i) Absorption of thermal photons with a rate $\propto N_m^{T_l} f_v(1 - f_c)$.

(ii) Stimulated emission of thermal photons with a rate $\propto N_m^{T_l} f_c(1 - f_v)$.

(iii) Absorption of source-generated photons with a rate $\propto \Delta N_s f_v(1 - f_c)$.

(iv) Stimulated emission of source-generated photons with a rate $\propto \Delta N_s f_c(1 - f_v)$.

(v) Spontaneous emission of photons with a rate $\propto f_c(1 - f_v)$.

In practice, only (iii) and (iv) are significant. Here is why. Since the timescale to spontaneously emit photons ($\sim$1 ns) greatly exceeds the timescale for scattering with phonons ($\sim$100 fs), the contribution of (v) to the shift current is negligible. By our assumptions that temperature is low and that carriers are resonantly excited, $k_B T_l \ll E_g \leqslant \hbar\omega_s = \hbar\omega_m$, hence the Planck occupancy $N_m^{T_l} \ll 1$. It follows that $\Delta N_s \gg 1 \gg N_m^{T_l}$, and we then assume (i) $\ll$ (iii), and (ii) $\ll$ (iv). Keeping only (iii) and (iv) leads to Eq. (A25).

The *intraband current* is defined to be the shift current contributed by intraband transitions:

$$\text{Intraband current:}\quad \boldsymbol{j}_{\mathrm{intra}} = \sum_{b=v,c} \boldsymbol{j}^b_{\mathrm{intra}};\quad \boldsymbol{j}^b_{\mathrm{intra}} = \boldsymbol{j}[f]_{b=b'} = -2_{\uparrow\downarrow}\frac{|e|}{\mathcal{V}|}\sum_m^{\text{phon}}\sum_{\boldsymbol{k},\boldsymbol{k}'} \boldsymbol{S}^m_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}}\left(\mathcal{A}^m_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}} - \mathcal{E}^m_{b\boldsymbol{k}\leftarrow b\boldsymbol{k}'}\right), \tag{A27}$$

with $\mathcal{A} - \mathcal{E}$ given in Eq. (A3). When bands do not overlap on the energy axis (as is true for our two-band semiconducting model), intraband transitions may be restricted to phononic modes, because the typical quasiparticle band velocity is much less than the speed of light.

It is useful to decompose the intraband current into contributions by acoustic and optical phonons: $\boldsymbol{j}^b_{\mathrm{intra}} = \boldsymbol{j}^{a,b}_{\mathrm{intra}} + \boldsymbol{j}^{o,b}_{\mathrm{intra}}$. Assuming (a) $k_B T_l \ll \hbar\Omega_o$, (b) the active region is much bigger than the passive region ($|E_{b,\mathrm{exc}} - E_{b,\boldsymbol{k}_{\mathrm{ext}}}| \gg \hbar\Omega_o$), and that (c) small-angle scattering predominates, the effect of acoustic phonons is substantially outweighed by that of optical phonons: $||\boldsymbol{j}^{a,b}_{\mathrm{intra}}|| \ll ||\boldsymbol{j}^{o,b}_{\mathrm{intra}}||$. Here is why. Assumption (c) allows us to employ the small-angle-limit of the phononic shift in Eq. (5); because this anomalous shift is proportional to $\delta k = ||\boldsymbol{k}' - \boldsymbol{k}||$, the net effect of transitions $\boldsymbol{k}' \leftarrow \boldsymbol{k}$ within the passive region is ignorable compared to transitions within the much larger active region; cf. a similar argument made in Sec. II. Within the active region, transitions mediated by optical phonons are predominantly that of spontaneous emission, because the thermal occupancies of optical phonons are small; cf. Eq. (A10). Transitions mediated by spontaneous emission of optical phonons predominantly result in a larger shift than transitions mediated by acoustic phonons; this is because the timescales for individual collisions are comparable to 100 fs for both types of phonons [31], but for optical phonons, δk has a lower bound given by Ω_o divided by the carrier group velocity, while the only lower bound for electron–acoustic-phonon scattering is the trivial bound $\delta k > 0$. Altogether, these considerations lead to the intraband shift current being dominated by

$$\boldsymbol{j}^{o,b}_{\mathrm{intra}} \approx 2_{\uparrow\downarrow}\frac{|e|}{\mathcal{V}|}\sum_m^{\text{optical phonons}}\sum_{\boldsymbol{k},\boldsymbol{k}'} \boldsymbol{S}^m_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}}\mathcal{E}^{sp,m}_{b\boldsymbol{k}\leftarrow b\boldsymbol{k}'}, \tag{A28}$$

$$\mathcal{E}^{sp,m}_{b\boldsymbol{k}\leftarrow b\boldsymbol{k}'} = \delta_{b,b'}\delta_{\boldsymbol{q},\boldsymbol{k}'-\boldsymbol{k}}\frac{2\pi}{\hbar}\left|V^m_{B'B}\right|^2\delta(E_{B'B} - \hbar\omega_m)(1 - f_B)f_{B'}. \tag{A29}$$

The *recombination-induced current* is defined to be the shift current contributed by vertical photon-mediated transitions ($c\boldsymbol{k} \leftrightarrow v\boldsymbol{k}$) for $\boldsymbol{k}$ outside the excitation surface:

$$\text{Recombination-induced current:}\quad \boldsymbol{j}_{\mathrm{rec}} = \boldsymbol{j}[f]_{\boldsymbol{k}=\boldsymbol{k}'\notin\mathrm{ES}} = 2_{\uparrow\downarrow}\frac{|e|}{\mathcal{V}}\sum_m^{\text{phot}}\sum_{\boldsymbol{k}} \boldsymbol{S}^m_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}} I^m_{\mathrm{rec}\boldsymbol{k}}, \tag{A30}$$

$$\text{Recombination rate:}\quad I^m_{\mathrm{rec}\boldsymbol{k}} = (1 - \delta_{\boldsymbol{k},\mathrm{ES}})\left(\mathcal{E}^m_{v\boldsymbol{k}\leftarrow c\boldsymbol{k}} - \mathcal{A}^m_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}\right), \tag{A31}$$

with $\mathcal{A} - \mathcal{E}$ given in Eq. (A5). We have introduced $\delta_{\boldsymbol{k},\mathrm{ES}}$ as the projector to the excitation surface, and $1 - \delta_{\boldsymbol{k},\mathrm{ES}}$ as the complementary projector. Equation (A30) may be simplified on the basis of two considerations:

(a) The $\sum_{\boldsymbol{k}}$ in Eq. (A30) may be further restricted to a small-$\boldsymbol{k}$ volume corresponding to the passive region, according to arguments presented in Sec. II and Appendix B 1.

(b) Because of the projection in Eq. (A30), $\boldsymbol{j}_{\mathrm{rec}}$ depends on the thermal photon occupancy $N_m^{T_l}$ but not the source-generated occupancy ΔN_s; since the thermal occupancies of resonant photons are small [cf. Eq. (A22)], one may as well retain only the transitions mediated by spontaneous emission.

Altogether, (a) and (b) imply

$$\boldsymbol{j}_{\rm rec} \approx 2_{\uparrow\downarrow} \frac{|e|}{\mathcal{V}} \sum_{m}^{\rm phot} \sum_{\boldsymbol{k}}^{\rm pass} \boldsymbol{S}^{m}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}} \mathcal{E}^{sp,m}_{v\boldsymbol{k}\leftarrow c\boldsymbol{k}}, \tag{A32}$$

$$\mathcal{E}^{sp,m}_{v\boldsymbol{k}\leftarrow c\boldsymbol{k}} = \delta_{\boldsymbol{k},\boldsymbol{k}'} \frac{(2\pi e)^2 \omega_m}{\mathcal{V}} \, |\boldsymbol{\epsilon}_m \cdot \boldsymbol{A}_{cv\boldsymbol{k}}|^2 \delta(E_{cv} - \hbar\omega_m)(1 - f_v) f_c. \tag{A33}$$

We take $\sum_{\boldsymbol{k}}^{\rm pass}$ to mean an integral over the passive $\boldsymbol{k}$-volume of either the conduction or valence band [cf. Eq. (A23)], whichever of the two volumes is smaller.

The threefold decomposition of the shift current in Eq. (A24) implies a threefold decomposition of the shift conductivity defined in Eq. (A14):

$$\boldsymbol{\sigma}_{\boldsymbol{\epsilon},\omega} = \boldsymbol{\sigma}^{\rm exc}_{\boldsymbol{\epsilon},\omega} + \boldsymbol{\sigma}^{\rm intra}_{\boldsymbol{\epsilon},\omega} + \boldsymbol{\sigma}^{\rm rec}_{\boldsymbol{\epsilon},\omega}, \quad \boldsymbol{\sigma}^{\rm exc}_{\boldsymbol{\epsilon},\omega} = \frac{\boldsymbol{j}_{\rm exc}}{|\mathcal{E}_\omega|^2},$$

$$\boldsymbol{\sigma}^{\rm intra}_{\boldsymbol{\epsilon},\omega} = \frac{\boldsymbol{j}_{\rm intra}}{|\mathcal{E}_\omega|^2}, \quad \boldsymbol{\sigma}^{\rm rec}_{\boldsymbol{\epsilon},\omega} = \frac{\boldsymbol{j}_{\rm rec}}{|\mathcal{E}_\omega|^2}. \tag{A34}$$

APPENDIX B: THE NONEQUILIBRIUM DISTRIBUTION OF PHOTOEXCITED CARRIERS

1. Relaxation mechanisms for photoexcited carriers

The steady shift current in a semiconductor cannot be calculated without understanding some basic aspects of the relaxation mechanisms and distribution of photoexcited carriers, which we review here. (We use "carrier" as a shorthand for hole and electron quasiparticles.) Much of this brief review derives from a more comprehensive review by Esipov and Levinson [27].

Which scattering process dominates the energy relaxation of carriers depends on (i) the radiation intensity J generated by a source, and (ii) the energy E of the carrier.

The dependence on J is because the rate of carrier-carrier scattering via the instantaneous Coulomb interaction increases with the density n of photoexcited carriers,[43] and n is roughly proportional to J.

The dependence on the carrier energy E is because the matrix elements for scattering depend on the initial and final states. An especially strong dependence develops for E near the optical phonon threshold $\hbar\Omega_o$, which is defined as the smallest optical phonon energy. Our convention is that E for an electron (hole) carrier is set to zero at the conduction-band minimum (valence-band maximum). For $E < \hbar\Omega_o$ (the passive region), a carrier is forbidden by energy conservation against emitting optical phonons, and energy relaxation is substantially suppressed relative to $E > \hbar\Omega_o$ (the active region); cf. Eq. (A23) and Figs. 3(b) and 3(c).

We define an upper bound n_h to the carrier density, such that if $n \lesssim n_h$ (meaning much less than or comparable in magnitude), scattering by optical phonons is the *primary*/dominant mechanism of energy relaxation for photoexcited carriers *in the active region*; if $n \gg n_h$, it would be carrier-carrier scattering that dominates energy relaxation in the active region. For instance, $n_h \sim 10^{18}$ cm^{-3} for GaAs [27]. We assume throughout this paper that optical phonons are the primary energy relaxers in the active region. Because the typical carrier–optical-phonon scattering time $\tau^o \sim 100$ fs [31,32], which is far smaller than the interband recombination time ($\tau_{\rm rec} \sim 1$ ns) [27,30], the majority of carriers would relax into the passive region where they await recombination.[44] In other words, the steady electron (hole) distribution in the passive region accounts for most of the electrons in the conduction band (holes in the valence band), as illustrated in Fig. 3(c).

It is also useful to identify the secondary/subdominant mechanism for energy relaxation in the active region; the two candidates are carrier-carrier scattering and carrier–acoustic-phonon scattering. We assume that the subdominant mechanism in the active region is also the dominant mechanism for energy relaxation in the passive region, where carrier–optical-phonon scattering "switches of" discontinuously. Let us define a second density $n_l \ll n_h$, such that in the intermediate density range: $n_l \ll n \lesssim n_h$ (e.g., $10^{13} \ll n \lesssim 10^{18}$ cm^{-3} for GaAs), carrier-carrier scattering is the subdominant relaxer in the active region; and in the low-density regime: $n \ll n_l$, carrier–acoustic-phonon scattering is the subdominant relaxer in the active region.

For at least a number of semiconductors, steady-state measurements of hot-carrier photoluminescence spectra[45] support the hypothesis that photoexcited electrons in the passive region largely follow a nondegenerate Maxwellian distribution with a source-dependent chemical potential μ_e and electron temperature T_e; likewise, the majority of photoexcited holes in the passive region are Maxwellian with parameters μ_h and T_h. μ_e and T_e are distinct from the equilibrium chemical potential and temperature: T_e simply equals the nonequilibrium lattice temperature T_l if electron–acoustic-phonon scattering is the dominant energy relaxer in the passive region ($n \ll n_l$); however, T_e may exceed T_l if electron-electron scattering is the dominant energy relaxer in the passive region ($n_l \ll n \lesssim n_h$). Typically, both $k_B T_e$ and $k_B T_h < \hbar\Omega_o$, so most of the photoexcited carriers occupy only a smaller fraction of the passive region, and recombination transitions predominantly occur between electrons (with energy $\lesssim k_B T_e$) and holes (with energy $\lesssim k_B T_h$). In large part, the theory that is presented in the main text is agnostic about fine-grained details of the carrier distribution within the passive region, meaning the theory is generally applicable whether or not a Maxwellian distribution develops in the passive region. However, if it does develop, then explicit kinetic models can be constructed that are based on the diffusive approximation for energy relaxation; cf. Appendixes B 2 b and E.

2. The kinetic model in the low-density regime

We will introduce a kinetic model that holds in the low-density regime ($n \ll n_l$) and forms the basis

[43] If the majority of photoexcited carriers follow a Maxwellian distribution, the rate of carrier-carrier scattering is simply proportional to n; cf. Eq. (2.3.6) in Ref. [27].

[44] Because $\hbar\Omega_o \sim 30$ meV, and a typical bandwidth ~1 eV, it takes at most 30 emissions of optical phonons for a hot carrier to relax into the active region. 30×100 fs is still much less than $\tau_{\rm rec}$.

[45] For instance, see Refs. [76,77]; more experiments are reviewed in Refs. [27,28].

for numerical simulations of the BIS formula in Appendix C 2.

The collisional integral for a quasiparticle in a two-band semiconductor may decomposed into vertical photon-mediated transitions and intraband phonon-mediated transitions:

$$I_{ck} = I_{\boldsymbol{k}}^{\text{phot}} + I_{ck}^{\text{phon}}, \quad I_{vk} = -I_{\boldsymbol{k}}^{\text{phot}} + I_{vk}^{\text{phon}}. \tag{B1}$$

The photon-mediated component can be further decomposed into an excitation rate [Eq. (A26)] and recombination rate [Eq. (A31)], depending on whether $\boldsymbol{k}$ lies on the excitation surface or not:

$$I_{\boldsymbol{k}}^{\text{phot}} = \delta_{\boldsymbol{k},\text{ES}} I_{\text{exc}\boldsymbol{k}}^{m_s} - (1-\delta_{\boldsymbol{k},\text{ES}}) \sum_m^{\text{phot}} I_{\text{rec}\boldsymbol{k}}^{m}. \tag{B2}$$

The phonon-mediated component can be decomposed into incoming transitions that increase the quasiparticle occupancy and outgoing transitions that decrease the quasiparticle occupancy:

$$I_{b\boldsymbol{k}}^{\text{phon}} = \sum_m^{\text{phon}} \sum_{\boldsymbol{k}'} \left(\mathcal{A}^m_{b\boldsymbol{k}\leftarrow b\boldsymbol{k}'} + \mathcal{E}^m_{b\boldsymbol{k}\leftarrow b\boldsymbol{k}'} - \mathcal{A}^m_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}} - \mathcal{E}^m_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}} \right), \tag{B3}$$

with $\mathcal{A}$ and $\mathcal{E}$ defined in Eq. (A3). Each of I_{exc}, I_{rec}, $\mathcal{A}$, and $\mathcal{E}$ depends on the quasiparticle distribution f_B; this distribution is is defined to be steady if it sets the collisional integral to zero:

$$\text{Steady distribution:} \quad I_{c\boldsymbol{k}}[f_B] = I_{v\boldsymbol{k}}[f_B] = 0 \quad \text{for all} \quad \boldsymbol{k}. \tag{B4}$$

Let us first address the contribution to $I_{b\boldsymbol{k}}^{\text{phon}}$ by carrier–optical-phonon scattering, which is assumed to be the dominant energy relaxation mechanism in the active region; cf. Appendix B 1. As justified in Appendix A 4, one may neglect the absorption and stimulated emission of optical phonons, retaining only the transition rate for spontaneous emission: $\mathcal{A}^m_{b\boldsymbol{k}\leftarrow b\boldsymbol{k}'} - \mathcal{E}^m_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}} \to -\mathcal{E}^{\text{sp},m}_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}}$, with $\mathcal{E}^{\text{sp}}$ defined in Eq. (A29). We assume that small-angle polarization scattering predominates over deformation scattering [35]. For simplicity in modeling, we focus on polarization scattering by a single branch of longitudinal optical phonons, in which case the phonon mode m is fully specified by a phonon wave vector $\boldsymbol{q}$; for small $\boldsymbol{q}$, the phonon frequency is assumed to be approximately a constant equal to Ω_o. The asymptotic expression for the collisional integral is then given by[46]

$$\text{Optical phonon:} \quad I^o_{b\boldsymbol{k}} = \sum_{\boldsymbol{q}\boldsymbol{k}'} \left(\mathcal{E}^{\text{sp},\boldsymbol{q}}_{b\boldsymbol{k}\leftarrow b\boldsymbol{k}'} - \mathcal{E}^{\text{sp},\boldsymbol{q}}_{b\boldsymbol{k}'\leftarrow b\boldsymbol{k}} \right), \tag{B5}$$

$$\mathcal{E}^{\text{sp},\boldsymbol{q}}_{b\boldsymbol{k}\leftarrow b\boldsymbol{k}'} = \delta_{\boldsymbol{k}'-\boldsymbol{k},\boldsymbol{q}} \frac{2\pi}{\hbar} \left| V^{\boldsymbol{q}}_{b\boldsymbol{k}'b\boldsymbol{k}} \right|^2 \delta(E_{b\boldsymbol{k}'\boldsymbol{k}} - \hbar\Omega_o)(1-f_{b\boldsymbol{k}}) f_{b\boldsymbol{k}'}; \tag{B6}$$

$$\left| V^{\boldsymbol{k}'-\boldsymbol{k}}_{b\boldsymbol{k}'b\boldsymbol{k}} \right|^2 = \left| V^{\boldsymbol{k}-\boldsymbol{k}'}_{b\boldsymbol{k}b\boldsymbol{k}'} \right|^2 \approx \frac{\hbar}{2\pi} \frac{\zeta a}{\mathcal{V}} \frac{|\langle u_{b\boldsymbol{k}'} | u_{b\boldsymbol{k}} \rangle_{\text{cell}}|^2}{|\boldsymbol{k}'-\boldsymbol{k}|^2}. \tag{B7}$$

ζ is a coupling parameter with dimensions of energy over time. The inner product of cell-periodic Bloch functions is related to the quantum metric tensor [79,80] as

$$\begin{aligned} |\langle u_{b\boldsymbol{k}'} | u_{b\boldsymbol{k}} \rangle_{\text{cell}}|^2 &= 1 - \delta k_i \delta k_j g^{ij}_{b\boldsymbol{k}} + O(\delta k^3); \\ g^{ij}_{b\boldsymbol{k}} &= \text{Re} \left\langle \nabla^i_{\boldsymbol{k}} u_b \middle| \nabla^j_{\boldsymbol{k}} u_b \right\rangle_{\text{cell}} - A^i_{b\boldsymbol{k}\boldsymbol{k}} A^j_{b\boldsymbol{k}\boldsymbol{k}}; \\ \delta\boldsymbol{k} &= \boldsymbol{k}' - \boldsymbol{k}, \end{aligned} \tag{B8}$$

with $\boldsymbol{A}_{b\boldsymbol{k}\boldsymbol{k}}$ the intraband Berry connection.[47] Below room temperature ($k_B T_l \ll \hbar\Omega_o$), spontaneous emission of optical phonons dominates over stimulated emission and absorption, meaning we drop all terms in Eq. (B7) that are proportional to the Planck occupancy: $N_{\boldsymbol{q}}^{T_l} \ll 1$.

Next we attend to the contribution to $I_{b\boldsymbol{k}}^{\text{phon}}$ by carrier–acoustic-phonon scattering, which has been assumed to be the subdominant energy relaxation mechanism in the active region; cf. Appendix B 1. Deformation scattering with acoustic phonons is typically outweighed by piezoacoustic scattering [35]. The precise expression of the transition rate/matrix element for piezoacoustic scattering will not be required, and because we will eventually employ a diffusive Fokker-Planck approximation to the collisional integral. For now, it is worth knowing that the matrix element depends on the quasiparticle band index only through[48]

$$\left| V^{\boldsymbol{k}'-\boldsymbol{k},p}_{b\boldsymbol{k}',b\boldsymbol{k}} \right|^2 \propto |\langle u_{b\boldsymbol{k}'} | u_{b\boldsymbol{k}} \rangle_{\text{cell}}|^2, \tag{B9}$$

just as for polarization scattering with optical phonons in Eq. (B7).

a. Electron-hole symmetric kinetic model

Because it is numerically intensive to simulate a steady distribution that satisfies $I_{c\boldsymbol{k}}[f_B] = I_{v\boldsymbol{k}}[f_B] = 0$ for all $\boldsymbol{k}$, we will resort to two model assumptions. The first is that band energies and electron-phonon-scattering matrix elements are

$$\begin{aligned} \text{Electron-hole symmetric:} \quad E_{c\boldsymbol{k}} &= -E_{v\boldsymbol{k}}, \\ \left| V^{m}_{c\boldsymbol{k}',c\boldsymbol{k}} \right|^2 &= \left| V^{-m}_{v\boldsymbol{k},v\boldsymbol{k}'} \right|^2, \end{aligned} \tag{B10}$$

with $-m$ being the momentum-inverted counterpart of m. This symmetry condition ensures for the phononic transition rates that

$$\begin{aligned} \mathcal{A}^m_{c\boldsymbol{k}\leftarrow c\boldsymbol{k}'}[1-f_v, 1-f_c] &= \mathcal{A}^{-m}_{v\boldsymbol{k}'\leftarrow v\boldsymbol{k}}[f_c, f_v], \\ \mathcal{E}^m_{c\boldsymbol{k}\leftarrow c\boldsymbol{k}'}[1-f_v, 1-f_c] &= \mathcal{E}^{-m}_{v\boldsymbol{k}'\leftarrow v\boldsymbol{k}}[f_c, f_v], \end{aligned} \tag{B11}$$

as may be verified by inspecting Eq. (A3); the meaning of $\mathcal{A}[1-f_v, 1-f_c]$ is to replace $f_{c\boldsymbol{k}} \to 1-f_{v\boldsymbol{k}}$ and $f_{v\boldsymbol{k}} \to 1-f_{c\boldsymbol{k}}$ for all terms in $\mathcal{A}$, and for all $\boldsymbol{k}$. The photonic transition rate satisfies $I_{\boldsymbol{k}}^{\text{phot}}[1-f_v, 1-f_c] = I_{\boldsymbol{k}}^{\text{phot}}[f_c, f_v]$, even without assuming electron-hole symmetry; cf. Eq. (B10).

[46] For general expressions, see Eq. (3.12) in Ref. [36] and the discussion in Sec. III. For the specific case of longitudinal optical phonons, Sec 1.3.E in Ref. [78] contains a concise derivation.

[47] Recently, the electron-phonon coupling has been related to an orbital-projected analog of the Fubini-Study metric [81].

[48] See Eqs. (3.15) and (3.16) in Ref. [36], bearing in mind a remark made in Sec. III about a missing factor.

Altogether,

$$I_{c\boldsymbol{k}}[1-f_v, 1-f_c] = I^{\text{phot}}_{\boldsymbol{k}}[f_c, f_v] + \sum_m^{\text{phon}} \sum_{\boldsymbol{k}'} \left(\mathcal{A}^{-m}_{v\boldsymbol{k}'\leftarrow v\boldsymbol{k}} + \mathcal{E}^{-m}_{v\boldsymbol{k}'\leftarrow v\boldsymbol{k}} - \mathcal{A}^{-m}_{v\boldsymbol{k}\leftarrow v\boldsymbol{k}'} - \mathcal{E}^{-m}_{v\boldsymbol{k}\leftarrow v\boldsymbol{k}'}\right)_{f_c,f_v} = -I_{v\boldsymbol{k}}[f_c, f_v]. \quad \text{(B12)}$$

Thus the steady-state condition $I_{c\boldsymbol{k}} = I_{v\boldsymbol{k}} = 0$ is solved by an electron-hole-symmetric distribution: $f_{c\boldsymbol{k}} = 1 - f_{v\boldsymbol{k}}$. In particular, if $f_{c\boldsymbol{k}}$ is found such that $I_{c\boldsymbol{k}}[f_c, 1-f_c] = 0$, then it is guaranteed that $I_{v\boldsymbol{k}}[f_c, 1-f_c] = 0$.

Let us check that our model for the quasiparticle Hamiltonian [$H(\boldsymbol{k})$ in Eq. (16)] and carrier-phonon scattering [Eqs. (B3)–(B7)] is electron-hole-symmetric in the sense of Eq. (B10). Suppose that the conduction- and valence-band wave functions are related by an antiunitary operation: $|u_{c\boldsymbol{k}}\rangle_{\text{cell}} = \hat{C}|u_{v\boldsymbol{k}}\rangle$, which implies $\langle u_{c\boldsymbol{k}}|u_{c\boldsymbol{k}'}\rangle = \langle u_{v\boldsymbol{k}'}|u_{v\boldsymbol{k}}\rangle$. Because the electron-phonon matrix element (for both polarization and piezoacoustic scatterings) only depends on the band index through $|\langle u_{b\boldsymbol{k}}|u_{b\boldsymbol{k}'}\rangle|^2$ [cf. Eq. (B9)], $|V^m_{c\boldsymbol{k},c\boldsymbol{k}'}|^2 = |V^m_{v\boldsymbol{k},v\boldsymbol{k}'}|^2$. Applying a general property of electron-phonon matrix elements: $V^m_{b\boldsymbol{k},b\boldsymbol{k}'} = \overline{V^{-m}_{b\boldsymbol{k}',b\boldsymbol{k}}}$ [cf. Eq. (C21)], one obtains the second equation in Eq. (B10). For the specific model Hamiltonian in Eq. (16), the antiunitary operation is simply $\hat{C} = \sigma_2 K$, with σ_2 the second Pauli matrix and K implementing complex conjugation. More generally, any $H(\boldsymbol{k})$ that is a sum of Pauli matrices satisfies

$$\sigma_2 \overline{H(\boldsymbol{k})} \sigma_2 = -H(\boldsymbol{k}), \quad E_{c\boldsymbol{k}} = -E_{v\boldsymbol{k}}. \quad \text{(B13)}$$

The last condition further implies that the excitation surface is an isoenergy surface:

$$0 = E_{cv\boldsymbol{k}} - \hbar\omega_s = 2E_{c\boldsymbol{k}} - \hbar\omega_s \quad \text{(B14)}$$

for any source radiation frequency ω_s. In other words, the set of excitation energies $E_{c,\text{exc}}$ defined in Eq. (A21) is just a single energy.

b. Isoenergy symmetric kinetic model

To recapitulate, we want to numerically simulate an electron-hole-symmetric distribution $f_{c\boldsymbol{k}}$ such that $I_{c\boldsymbol{k}}[f_c, 1-f_c] = 0$. Having reduced the problem to a single band by electron-hole symmetry, one may as well drop the band index on $f_{c\boldsymbol{k}} \to f_{\boldsymbol{k}}$, $E_{c\boldsymbol{k}} \to E_{\boldsymbol{k}}$, and $E_{c,\text{exc}} \to E_{\text{exc}}$. We further redefine $E = 0$ to be the minimal energy for the conduction band.

To simplify the simulation of $f_{\boldsymbol{k}}$, we further assume that $f_{\boldsymbol{k}}$ is approximately *isoenergy-symmetric*, meaning that $f_{\boldsymbol{k}}$ is approximately constant under variation of $\boldsymbol{k}$ within an isoenergy surface for $E_{\boldsymbol{k}}$:

$$f_{\boldsymbol{k}} \approx f_E, \quad f_E = \langle\langle f_{\boldsymbol{k}}\rangle\rangle_{kE},$$
$$\langle\langle \Xi(\boldsymbol{k})\rangle\rangle_{kE} = \sum_{\boldsymbol{k}} \delta_{kE}\, \Xi_{\boldsymbol{k}}, \quad \delta_{kE} = \frac{\delta(E_{\boldsymbol{k}} - E)}{\mathcal{V} g_E}, \quad \text{(B15)}$$

with g_E meaning the density of conduction-band states per unit volume ($\mathcal{V}$) and per spin orientation. We refer to $\langle\langle\cdots\rangle\rangle_{kE}$ as *isoenergy-averaging*, and f_E as the *isoenergy-averaged distribution*. For the purpose of computing the shift current, $f_{\boldsymbol{k}} \approx f_E$ is justified to the extent that the collisional integral in Eq. (B1) is isoenergy-symmetric: $I_{\boldsymbol{k}} \approx I_E = \langle\langle I_{\boldsymbol{k}}\rangle\rangle_{kE}$, which constrains the model Hamiltonians that we allow ourselves to numerically simulate.[49]

By averaging the kinetic equation $I_{\boldsymbol{k}}[f_c, 1-f_c] = 0$ over an isoenergy surface, one obtains[50]

$$E > E_{\text{cut}}: \quad g_E I_E[f] = G_\uparrow[f]\,\delta(E - E_{\text{exc}}) - \frac{g_E f_E}{\tau^o_E} + \frac{g_{E_+} f_{E_+}}{\tau^o_{E_+}} - \frac{g_E f_E}{\tau_{\text{rec}}} - \partial_E j^s_E = 0,$$
$$E_\pm = E \pm \hbar\Omega_o. \quad \text{(B16)}$$

We will explain the terms on the right-hand side in turn:

(i) Recalling the excitation energy E_{exc} [Eq. (A21)] to be the energy of conduction-band states on the excitation surface, $G_\uparrow\,\delta(E - E_{\text{exc}})$ is the rate of increase in the quasiparticle number density $g_E f_E$ due to the absorption of source-generated photons.[51] In other words, $G_\uparrow[f]$ is the rate at which source-generated photons are absorbed per unit volume and per spin orientation;

$$G_\uparrow[f] = \frac{\alpha_\uparrow[f]\mathcal{I}_{\text{rad}}}{\hbar\omega} \quad \text{(B17)}$$

can be expressed as a product of the single-spin absorption coefficient and the radiation intensity, divided by the source photon energy.

(ii) $-gf/\tau^o|_E$ ($+gf/\tau^o|_{E_+}$) in Eq. (B16) represents an outflow (inflow) of electrons due to spontaneous emission of optical phonons. τ^o_E is the average time for a quasiparticle with energy E to spontaneously emit an optical phonon; in the passive region ($E < \hbar\Omega_o$), spontaneous emission is forbidden by energy conservation, hence we set $\tau^o_E = \infty$. One may relate τ^o_E to the collisional integral $I^{\text{phon}}_{\boldsymbol{k}}$ by

$$\frac{g_{E_+} f_{E_+}}{\tau^o_{E_+}} - \frac{g_E f_E}{\tau^o_E} = \frac{\zeta a}{\mathcal{V}^2} \sum_{\boldsymbol{k}\boldsymbol{k}'}^{\text{cut}} \delta(E_{\boldsymbol{k}} - E) \frac{|\langle u_{\boldsymbol{k}'}|u_{\boldsymbol{k}}\rangle_{\text{cell}}|^2}{|\boldsymbol{k} - \boldsymbol{k}'|^2} \times \{f_{\boldsymbol{k}'}\delta(E_{\boldsymbol{k}'} - E_+) - f_{\boldsymbol{k}}\delta(E_{\boldsymbol{k}'} - E_-)\}. \quad \text{(B18)}$$

The right-hand side is obtained by applying $\sum_{\boldsymbol{k}} \delta(E_{\boldsymbol{k}} - E)/\mathcal{V}$ to the component of $I^{\text{phon}}_{\boldsymbol{k}}$ corresponding to spontaneous emission of optical phonons [cf. Eq. (B7)] and dropping all terms

[49] In general, it should be expected that the nonequilibrium distribution is isoenergy-asymmetric with respect to inverting $\boldsymbol{k}$: $f_{\boldsymbol{k}} \neq f_{-\boldsymbol{k}}$. This is possible because the continuous absorption of photons creates a nonequilibrium state with a direction for time. Consequently, the asymmetry ($f_{c\boldsymbol{k}} - f_{c,-\boldsymbol{k}}$) is proportional to the light intensity [30] and contributes to a "ballistic current" [3] but not the shift current.

[50] The following kinetic equation is very similar to one studied in Ref. [82]; however, we would rather not presume they adopted the same premises as we have adopted.

[51] $G_\uparrow\,\delta(E - E_{\text{exc}})$ is derived by applying $\sum_{\boldsymbol{k}} \delta(E_{\boldsymbol{k}} - E_{\text{exc}})/\mathcal{V}$ to $I^{\text{phot}}_{\boldsymbol{k}}$ [Eq. (B2)] and retaining terms that are proportional to the source photon number ΔN_s; cf. Eq. (A11).

that are nonlinear in the quasiparticle distribution: $f_{\boldsymbol{k}} f_{\boldsymbol{k}'} \ll f_{\boldsymbol{k}}$ and $f_{\boldsymbol{k}'}$; bear in mind that nondegenerate fermion statistics ($f_{\boldsymbol{k}} \ll 1$) apply to a wide range of continuous-wave laser experiments [27]. Because we are employing an asymptotic expression that is valid for small-angle scattering, we have introduced a cutoff in $\sum_{\boldsymbol{k}\boldsymbol{k}'}^{\text{cut}}$, so that $\delta k = |\boldsymbol{k} - \boldsymbol{k}'|$ is much less than the linear dimension of the Brillouin zone.

(iii) $-g_E f_E/\tau_{\text{rec}}$ represents the quasiparticle loss rate due to interband recombination by spontaneous emission of photons. The effects of absorption and stimulated emission of thermal photons are negligible, as was explained in Appendix A 4. In numerical simulations, we just take $\tau_{\text{rec}} \sim 1$ ns to be a typical, energy-averaged timescale for interband recombination [27,30]. In principle, one could refine the model by replacing $\tau_{\text{rec}} \to \tau_E^{\text{rec}}$, with τ^{rec} depending on E through the energy dependences of the dipole matrix element and f_E.[52] In practice, what matters to the shift current is the order-of-magnitude difference: $\tau_E^{\text{rec}} \gg\gg \tau_E^o$, which guarantees that recombination transitions predominantly occur in the passive region, independent of the precise energy dependence of τ^{rec}; cf. Sec. II and Eq. (A32).

(iv) $-\partial_E j_E^s$ is the rate of change of $g_E f_E$ induced by electron–acoustic-phonon scattering. A negative j_E^s represents a scattering-induced relaxation of the number density $g_E f_E$ toward decreasing energies, so we refer to j_E^s as the *energy-axis current*. In principle, this current should be an integral of f_E; however, the smallness of acoustic-phonon energies relative to typical electron energies allows us to employ the diffusive Fokker-Planck approximation [35,83]:

$$j_E^s = -\frac{g_E E}{\tau_E^s}(1 + k_B T_l \partial_E) f_E. \tag{B19}$$

E/τ_E^s is the dynamic friction coefficient [35], which is interpretable as minus the "drift speed" of a number-density-valued wave packet on the energy axis. The form of $(1 + k_B T_l \partial_E)$ encodes an Einstein relation between the dynamic friction coefficient and the diffusion coefficient.[53]

Our previous assumption that energy relaxation is dominated by optical phonons can now be expressed as a mathematical inequality, namely that the dynamic friction coefficient is much less than the energy relaxation rate due to spontaneous emission of optical phonons:

$$E > \hbar\Omega_o: \quad \eta_E = \frac{E}{\hbar\Omega_o}\frac{\tau_E^o}{\tau_E^s} \ll 1. \tag{B20}$$

The diffusive approximation is valid on the condition that the density of states is analytic and the collisions are quasielastic. The former condition rules out van Hove singularities [84]. The latter condition means precisely that the change in a quasiparticle's energy (due to a collision) is much less than the quasiparticle's initial energy [35]. This holds for most quasiparticle energies, since acoustic-phonon energies are a very small fraction of the quasiparticle bandwidth. We introduce a cutoff energy E_{cut} that is comparable to the typical acoustic phonon energy, such that the diffusive approximation holds for $E > E_{\text{cut}}$.

For energies less than the cutoff, we adopt the following kinetic equation:

$$0 < E < E_{\text{cut}}: \quad g_E I_E[f] = \frac{g_{E_+} f_{E_+}}{\tau^o_{E_+}} - \frac{g_E f_E}{\tau_{\text{rec}}} - \frac{j^s_{E_{\text{cut}}}}{E_{\text{cut}}} = 0. \tag{B21}$$

$-j^s_{E_{\text{cut}}}$ represents a (downward ≡ energy-relaxing) current of the number density $g_E f_E$ across the cutoff energy [cf. Eq. (B19)]; any density that relaxes across the cutoff is equally distributed between all conduction-band states below the cutoff.[54] This crude modeling of scattering below the cutoff can in principle be improved upon, but we remind the reader that the steady shift current is insensitive to fine details of the quasiparticle distribution within the passive region, due to arguments explained in Sec. II. Despite the crudeness of the model, it ensures that all phonon-mediated collisions conserve the total number of quasiparticles within the conduction band. In other words, if all the collisional terms in Eqs. (B16)–(B21), with the exception of terms involving $G_\uparrow$ and τ_{rec}, are collectively denoted as $g_E I_E^{\text{intra}}$, then $\int_0^\infty g_E I_E^{\text{intra}} dE = 0$.

APPENDIX C: BELINICHER-IVCHENKO-STURMAN FORMULA FOR THE SHIFT CURRENT

1. Derivation of the Belinicher-Ivchenko-Sturman formula

It has been expressed to the authors that the Belinicher-Ivchenko-Sturman theory [5] is difficult to penetrate. To our knowledge, no explicit derivation of the BIS formula yet exists in the literature. We will therefore derive their main formulas for pedagogy. More precisely, we mean to derive the form of the phonon-mediated (photon-mediated) shift current to be Eq. (A1), with the phononic (photonic) shift vector given in Eq. (A2) [Eq. (A4)], and with the difference in absorption and emission transition rates given by Eq. (A3) [Eq. (A5)].

Since the BIS formula encodes the spontaneous emission of photons, the derivation requires us to quantize the radiation field. If one were to quantize the radiation field but retain a first-quantized electron description, one would derive an analog of the BIS formula that is only applicable to nondegenerate Fermi statistics, i.e., one would miss a spontaneous-emission term that is nonlinear in the distribution function [cf. Eq. (A5) below]. To properly account for the Pauli exclusion principle in the presence of spontaneous emission, it is necessary to apply second quantization to the electron. It is fortuitous but misleading that terms that are

[52] $g_E f_E/\tau_E^{\text{rec}} = \sum_{\boldsymbol{k}} \delta(E_{\boldsymbol{k}} - E_{\text{exc}}) \mathcal{E}^{\text{sp},m}_{v\boldsymbol{k}\leftarrow c\boldsymbol{k}}/\mathcal{V}$, with the spontaneous emission rate defined in Eq. (A33).

[53] For electron–acoustic-phonon scattering, the Einstein relation is derived most directly from simplified expressions in Sec. 4.5 of Ref. [35], assuming that $k_B T_l \gg$ the acoustic phonon energy. In a subsequent discussion in Appendix E, we will also need an analogous Einstein relation for electron-electron scattering, which has been derived in Ref. [27].

[54] The $-j^s_{E_{\text{cut}}}/E_{\text{cut}}$ term can be viewed as a collisional term $g_E I_E^s[f]$ due to the secondary scattering mechanism. The crudeness in our approximation lies in assuming $g_E I_E^s[f]$ is independent of E for E below the cutoff. This amounts to assuming that I_E^s is independent of E, because the density of states is energy-independent for a quasi-2D parabolic band.

nonlinear in the distribution function cancel out if one considers only photon absorption and stimulated emission [Eq. (A5) with $N_m + 1 \approx N_m$]; thus it has been possible for theories (based on first quantization of the electron and a classical theory of radiation) to neglect the exclusion principle and yet derive correct formulas for the transient shift current, as will be elaborated in Appendix C 3.

Our derivation also manifests how a perturbation theory of the steady state differs dramatically from a perturbation theory of the transient state. Most practitioners who calculate nonlinear optical responses are calculating the transient response, and their zeroth-order state is the thermal equilibrium state in the absence of the light source. In steady-state perturbation theory, the zeroth-order state is emphatically *not* a thermal state; instead, Appendix C 1 e proves rigorously that if the state is steady, the zeroth-order quasiparticle distribution satisfies a nondetailed balance condition that represents an invariance under simultaneous collisions with all bosons. Moreover, we have no reason to believe that the zeroth-order state in steady perturbation theory is perturbatively connected to the thermal equilibrium state (in the absence of the light source).

The outline of the derivation is as follows:

(i) Appendix C 1 a sets up the problem and establishes the notation. We review salient properties of the independent-electron Hamiltonian, the crystal momentum representation, the independent-boson Hamiltonian, Fock space, and the electron-boson interaction. Finally, we express the shift current in terms of stationary density matrices, and we derive a perturbative expression for the stationary density matrix in the Lippmann-Schwinger scattering formalism.

(ii) The Lippmann-Schwinger formula for the stationary density matrix is expressed in terms of second-quantized matrix elements; these elements will be reduced to first-quantized matrix elements in Appendix C 1 b. The result of this reduction is an intermediate formula for the photonic and phononic shift current in Eqs. (C48) and (C57), respectively. These intermediate formulas are more formal than optimal: they are expressed in terms of an infinite number of band-off-diagonal matrix elements of the position operator.

(iii) Appendix C 1 c derives an optimal expression for the photonic shift current, with help from a sum rule derived from the first-quantized commutation relation between position and canonical momentum.

(iv) Appendix C 1 d derives an optimal expression for the phononic shift current, with help from a sum rule derived from the first-quantized commutation relation between position and the phonon-induced potential-energy field.

(v) Appendix C 1 e demonstrates that the zeroth-order density matrix is not thermal; instead, the zeroth-order quasiparticle distribution satisfies a nondetailed balance condition that represents an invariance under simultaneous collisions with all bosons.

a. Preliminaries

We decompose our Hamiltonian into two independent-particle terms and an electron-boson interaction:

$$H = H_0 + U, \quad H_0 = H_0^{\text{ele}} + H_0^{\text{bos}}. \tag{C1}$$

We will first explain the independent-particle terms:

Independent-electron Hamiltonian and the crystal momentum representation. H_0^{ele} is a mean-field Hamiltonian for independent electrons in a crystalline medium:

$$H_0^{\text{ele}} = \sum_B E_B c_B^\dagger c_B, \quad [c_B, c_{B'}^\dagger] = \delta_{B,B'},$$
$$B = (b, \boldsymbol{k}), \quad B' = (b', \boldsymbol{k}'), \tag{C2}$$

where $[x, y] = xy - yx$ (the commutator) and B is a collective index for both the band label and crystal wave vector. We assume throughout this work that spin-orbit coupling is negligible; to simplify notation, b should be understood as a spinless band label, and H as a Hamiltonian in one spin sector; only in the final steps will the current be multiplied by 2 to account for the spin degeneracy of bands.

c_B annihilates an electronic state with a wave function of the Bloch form: $e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{b\boldsymbol{k}}(\boldsymbol{r})/\sqrt{\mathcal{V}}$, with $u_{b\boldsymbol{k}}(\boldsymbol{r}) = u_{b\boldsymbol{k}}(\boldsymbol{r} + \boldsymbol{R})$ being periodic in Bravais-lattice translations, and $\mathcal{V}$ the volume of the medium. These cell-periodic functions are normalized as

$$\langle u_{b\boldsymbol{k}} | u_{b'\boldsymbol{k}} \rangle_{\text{cell}} = \delta_{b,b'}, \quad \langle X | Y \rangle_{\text{cell}} = \int \frac{d\boldsymbol{\tau}}{\mathcal{V}_{\text{cell}}} \overline{X(\boldsymbol{\tau})} Y(\boldsymbol{\tau}), \tag{C3}$$

with $\boldsymbol{\tau}$ the intracell coordinate, $\delta_{b,b'}$ a Kronecker delta function for the band labels, and $\mathcal{V}_{\text{cell}}$ the real-space volume of the primitive unit cell. The orthonormality and completeness of our basis of Bloch waves reads

$$\langle B | B' \rangle_1 = \delta_{B,B'} = \delta_{bb'} \delta_{\boldsymbol{k}\boldsymbol{k}'}, \quad I_1 = \sum_B |B\rangle\langle B|_1. \tag{C4}$$

I is the identity operator, and the superscript 1 in Eq. (C4) reminds us that we are dealing with a first-quantized, one-particle Hilbert space.

Our notation for $\boldsymbol{k}$ suggests misleadingly that $\boldsymbol{k}$ is a discrete wave vector: $\sum_B = \sum_b \sum_{\boldsymbol{k}}$ and $\delta_{B,B'} = \delta_{\boldsymbol{k},\boldsymbol{k}'} \delta_{b,b'}$. However, for the position operator to have a well-defined action on periodic Bloch states, one must take $\mathcal{V}$ to be infinite [85], hence $\delta_{\boldsymbol{k}\boldsymbol{k}'} = (2\pi)^3 \delta(\boldsymbol{k} - \boldsymbol{k}')/\mathcal{V}$ should be understood as a shorthand for a Dirac δ function, and we will be applying certain identities that apply to Dirac δ functions but not Kronecker δ functions:

$$\nabla_{\boldsymbol{k}} \delta_{\boldsymbol{k}\boldsymbol{k}'} = -\nabla_{\boldsymbol{k}'} \delta_{\boldsymbol{k}\boldsymbol{k}'}, \quad f_{\boldsymbol{k}'} \nabla_{\boldsymbol{k}} \delta_{\boldsymbol{k}\boldsymbol{k}'} - f_{\boldsymbol{k}} \nabla_{\boldsymbol{k}} \delta_{\boldsymbol{k}\boldsymbol{k}'} = \delta_{\boldsymbol{k}\boldsymbol{k}'} \nabla_{\boldsymbol{k}} f_{\boldsymbol{k}}. \tag{C5}$$

$\sum_{\boldsymbol{k}}$ should also be understood as an integral over the Brillouin zone: $\mathcal{V} \int_{\text{BZ}} d^3k/(2\pi)^3$. With these caveats in mind, we present the first-quantized position, canonical momentum, and velocity operators in the crystal momentum representation [85]:

$$\boldsymbol{r}_{BB'} = \langle B | \hat{\boldsymbol{r}} | B' \rangle_1 = i \delta_{bb'} \nabla_{\boldsymbol{k}} \delta_{\boldsymbol{k}\boldsymbol{k}'} + \delta_{\boldsymbol{k}\boldsymbol{k}'} \boldsymbol{A}_{bb'\boldsymbol{k}},$$
$$\boldsymbol{A}_{bb'\boldsymbol{k}} = \langle u_{b\boldsymbol{k}} | i \nabla_{\boldsymbol{k}} u_{b'} \rangle_{\text{cell}}, \tag{C6}$$

$$\boldsymbol{p}_{BB'} = \langle B | \hat{\boldsymbol{p}} | B' \rangle_1 = \delta_{\boldsymbol{k}\boldsymbol{k}'} \boldsymbol{P}_{bb'\boldsymbol{k}},$$
$$\boldsymbol{P}_{bb'\boldsymbol{k}} = \langle u_{b\boldsymbol{k}} | \hat{\boldsymbol{p}} | u_{b'\boldsymbol{k}} \rangle_{\text{cell}} = m_f \boldsymbol{v}_{bb'\boldsymbol{k}}, \tag{C7}$$

$$\boldsymbol{v}_{BB'} = \langle B | \hat{\boldsymbol{v}} | B' \rangle_1 = \delta_{\boldsymbol{k}\boldsymbol{k}'} \boldsymbol{v}_{bb'\boldsymbol{k}}, \quad \boldsymbol{v}_{bb'\boldsymbol{k}} = \langle u_{b\boldsymbol{k}} | \hat{\boldsymbol{v}} | u_{b'\boldsymbol{k}} \rangle_{\text{cell}}. \tag{C8}$$

We have assumed in the absence of spin-orbit coupling that $\hat{\boldsymbol{v}} = \hat{\boldsymbol{p}}/m_f$, with m_f the free-electron mass. It is also worth

defining the band-off-diagonal position operator as

$$\begin{aligned}\hat{\boldsymbol{r}}_{\rm off} &= \sum_{BB'} \boldsymbol{r}^{\rm off}_{BB'} |B\rangle\langle B'|_1,\\ \boldsymbol{r}^{\rm off}_{BB'} &= \delta_{\boldsymbol{k}\boldsymbol{k}'} \boldsymbol{A}^{\rm off}_{bb'\boldsymbol{k}},\\ \boldsymbol{A}^{\rm off}_{bb'\boldsymbol{k}} &= \boldsymbol{A}_{bb'\boldsymbol{k}}(1-\delta_{bb'}),\end{aligned} \tag{C9}$$

which is related to the band-off-diagonal elements of the velocity operator [85]:

$$\frac{\boldsymbol{v}^{\rm off}_{b'b\boldsymbol{k}}}{E_{bb'\boldsymbol{k}}} = -\frac{i}{\hbar}\boldsymbol{A}^{\rm off}_{b'b\boldsymbol{k}}, \quad E_{bb'\boldsymbol{k}} = E_{b\boldsymbol{k}} - E_{b'\boldsymbol{k}}. \tag{C10}$$

Independent-boson Hamiltonian. $H_0^{\rm bos}$ is the independent-boson Hamiltonian absent the zero-point energy:

$$\begin{aligned}H_0^{\rm bos} &= \sum_m \hbar\omega_m a_m^\dagger a_{m'},\\ a_m a^\dagger_{m'} + a^\dagger_{m'} a_m &= \delta_{m,m'}, \quad m = (\boldsymbol{q}, p),\end{aligned} \tag{C11}$$

where the index $m = (\boldsymbol{q}p)$ runs over both photonic and phononic modes.

We follow Fermi's prescription [86] in quantizing the transverse/solenoidal component of the electromagnetic vector potential in the Coulomb gauge [74,75]. For photons, $\boldsymbol{q}$ is a wave vector in $\mathbb{R}^3$ with a cutoff: $\hbar cq < E_{\rm cut}$; the cutoff energy may be taken as the largest energy difference between the Bloch bands, which are excited by the light source.[55] $p \in \{1, 2\}$ specifies one of the two possible transverse polarizations for a given $\boldsymbol{q}$; we adopt a linearly polarized basis, meaning the polarization vector is real-valued: $\boldsymbol{\epsilon}_m \equiv \boldsymbol{\epsilon}^{(p)}_{\hat{\boldsymbol{q}}} = \boldsymbol{\epsilon}^{(p)}_{-\hat{\boldsymbol{q}}} \in \mathbb{R}$. The photon frequency is polarization-independent: $\omega_m = cq$ with $q = ||\boldsymbol{q}||$.

For phonons, $\boldsymbol{q}$ is a wave vector in the BZ, and $p = 3, 4, \ldots, 3N_{\rm nuc} + 2$ is a label for a nondegenerate phonon band, with $N_{\rm nuc}$ being the number of nuclei per primitive unit cell. $\omega_{\boldsymbol{q}p} = \omega_{-\boldsymbol{q}p}$ is the renormalized phonon dispersion [87,88].

Altogether, $\sum_m = \sum_p \sum_{\boldsymbol{q}}$ with p running over $3N_{\rm nuc} + 2$ values, $\sum_{\boldsymbol{q}} = \mathcal{V}\int d^3q/(2\pi)^3$ with the integration domain depending on p, and $\delta_{m,m'}$ should be understood as $\delta_{\boldsymbol{q},\boldsymbol{q}'}\delta_{p,p'}$.

Fock space. Eigenstates of the independent-particle Hamiltonian are labeled by electronic occupancies $n_B \in \{0, 1\}$ and bosonic occupancies $N_m \in \{0, 1, 2, \ldots\}$:

$$\begin{aligned}(H_0 - E_\mu)|\mu\rangle = 0, \quad \mu &= \left(\left\{n_B^\mu\right\}_B, \left\{N_m^\mu\right\}_m\right),\\ E_\mu &= \sum_B E_B n_B^\mu + \sum_m \hbar\omega_m N_m^\mu.\end{aligned} \tag{C12}$$

Throughout this Appendix, Greek symbols (like μ) are used as a collective index for all electronic and bosonic occupancies. $\{n_B\}_B$ means a set of occupancies for all Bloch states, but we will often use the shorthand $\{n_B\}_B \to \{n\}$. Likewise for $\{N_m\}_m \to \{N\}$. We will refer to $|\mu\rangle$ as an *independent-particle state*. The set of independent-particle states forms an orthonormal basis ($\langle\mu|\nu\rangle = \delta_{\mu,\nu}$) for the combined-electron-boson Fock space. The resolution of identity is given by

$$\begin{aligned}I &= \sum_\mu |\mu\rangle\langle\mu| = \sum_{\{n\}}\sum_{\{N\}} |\{n\}\{N\}\rangle\langle\{n\}\{N\}|,\\ \sum_{\{n\}} &= \prod_B \sum_{n_B=0}^{1}, \quad \sum_{\{N\}} = \prod_m \sum_{N_m=0}^{\infty}.\end{aligned} \tag{C13}$$

Any operator O^e with an e superscript should be understood as acting only in the electronic Fock space, which is spanned by *independent-electron states* denoted as $|\{n\}\rangle_e$. (The existence or absence of subscripts distinguishes kets in different Hilbert spaces.) We will focus on bilinear electronic operators

$$O^e = \sum_{B,B'} O_{B,B'} c_B^\dagger c_{B'}, \quad O^e_{\mu\nu} = \langle\{n^\mu\}|O^e|\{n^\nu\}\rangle_e, \tag{C14}$$

with matrix elements denoted as $O^e_{\mu\nu}$; the commutator of two bilinear operators is expressible as

$$[G^e, O^e] = \sum_{BB'} [G, O]_{BB'} c_B^\dagger c_{B'}, \quad G^e = \sum_{B,B'} G_{B,B'} c_B^\dagger c_{B'}. \tag{C15}$$

Electron-boson interaction. We decompose U into a tensor product of operators acting in the electronic and photonic Fock spaces:

$$U = \sum_m U^e_m (a_m + a^\dagger_{-m}), \quad U^e_m = \sum_{B,B'} U^m_{B,B'} c_B^\dagger c_{B'} = \left(U^e_{-m}\right)^\dagger, \tag{C16}$$

with $-m = (-\boldsymbol{q}, p)$ the momentum-reversed partner of $m = (\boldsymbol{q}p)$. $U^e_m = (U^e_{-m})^\dagger$ ensures that U is self-adjoint. For m that is photonic (phononic), $U^m_{B,B'}$ is defined as the *electron-photon* (*electron-phonon*) *matrix element*:

$$U^{\boldsymbol{q}p}_{b\boldsymbol{k},b'\boldsymbol{k}'} = \begin{cases}\delta_{\boldsymbol{k},\boldsymbol{k}'} \check{W}^{\boldsymbol{q}p}_{bb'\boldsymbol{k}}, \quad \check{W}^{\boldsymbol{q}p}_{bb'\boldsymbol{k}} = \sqrt{\frac{he^2}{\omega_{\boldsymbol{q}p}\mathcal{V}}}\boldsymbol{\epsilon}^{(p)}_{\hat{\boldsymbol{q}}} \cdot \boldsymbol{v}_{bb'\boldsymbol{k}} & \text{(electron-photon)},\\ V^{\boldsymbol{q}p}_{b\boldsymbol{k},b'\boldsymbol{k}'} = \delta_{\boldsymbol{k},\boldsymbol{k}'+\boldsymbol{q}} \check{V}^{\boldsymbol{q}p}_{bb'\boldsymbol{k}}, \quad \check{V}^{\boldsymbol{q}p}_{bb'\boldsymbol{k}} = \mathcal{V}^{-1}\sum_{\boldsymbol{G}}^{\rm RL} \widetilde{PE}^{\boldsymbol{q}p}_{\boldsymbol{q}+\boldsymbol{G}} \langle u_{b\boldsymbol{k}}|e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}}|u_{b'\boldsymbol{k}-\boldsymbol{q}}\rangle_{\rm cell} & \text{(electron-phonon)}.\end{cases} \tag{C17}$$

We will describe each matrix element in turn.

[55] This cutoff is imposed for self-consistency: our use of the dipole approximation requires that q is much less than the linear dimension of the BZ.

The electron-photon matrix element is derived from the first-order term in the nonrelativistic minimal coupling: $|e|\boldsymbol{A}_\perp \cdot \boldsymbol{v}/c$, with $\boldsymbol{v}$ the second-quantized electron velocity operator and $\boldsymbol{A}_\perp$ the quantized electromagnetic vector potential satisfying $\boldsymbol{\nabla}\cdot\boldsymbol{A}_\perp = \boldsymbol{0}$ [74]. The photonic expression in Eq. (C17) is valid in the dipole approximation. Within this approximation, $U^e_m = U^e_{-m} = (U^e_m)^\dagger$ is self-adjoint. Minimal coupling also results in an electron-photon interaction proportional to e^2, but such a coupling does not contribute to the shift current because it cannot induce interband transitions within the dipole approximation [cf. Eq. (C23) below].

We adopt a simplified electron-phonon matrix element $\check{V}^{\boldsymbol{qp}}_{bb'\boldsymbol{k}}$ which is derived in the adiabatic approximation (where phonons are frozen from the electron's perspective) and by applying the Hartree approximation to electron-electron interactions [36,87]. In the expression for $\check{V}^{\boldsymbol{qp}}_{bb'\boldsymbol{k}}$, $\sum^{\mathrm{RL}}_{\boldsymbol{G}}$ sums over all reciprocal-lattice vectors, $\boldsymbol{q}$ is a wave vector in the Brillouin zone (BZ), and $\boldsymbol{q}+\boldsymbol{G}=\boldsymbol{Q}$ is a wave vector in $\mathbb{R}^3$. $\widetilde{\mathrm{PE}}^m_{\boldsymbol{Q}} = \int_{\mathbb{R}^3} e^{-i\boldsymbol{Q}\cdot\boldsymbol{r}}\mathrm{PE}^m_{\boldsymbol{r}}d\boldsymbol{r}$ is a Fourier transform of the one-electron potential energy $\mathrm{PE}^m_{\boldsymbol{r}}$ induced by annihilating a phonon of mode m. $\mathrm{PE}^m_{\boldsymbol{r}}$ is self-consistently [87] screened in a crystalline medium, and is linearly related to the bare potential energy $\mathrm{PE}^{m;0}_{\boldsymbol{r}}$:

$$\widetilde{\mathrm{PE}}^{\boldsymbol{qp}}_{\boldsymbol{q}+\boldsymbol{G}} = \sum_{\boldsymbol{G}'}^{\mathrm{RL}} \varepsilon^{-1}_{\boldsymbol{q}+\boldsymbol{G},\boldsymbol{q}+\boldsymbol{G}'}\widetilde{\mathrm{PE}}^{\boldsymbol{qp};0}_{\boldsymbol{q}+\boldsymbol{G}'} = \overline{\widetilde{\mathrm{PE}}^{-\boldsymbol{qp}}_{-\boldsymbol{q}-\boldsymbol{G}}}, \tag{C18}$$

with $\varepsilon^{-1}_{\boldsymbol{Q},\boldsymbol{Q}'} = \overline{\varepsilon^{-1}_{-\boldsymbol{Q},-\boldsymbol{Q}'}}$ the static, inverse dielectric function in the Hartree approximation.[56] The bare potential energy is expressible in terms of $\tilde{v}_{\boldsymbol{Q}} = 4\pi e^2/||\boldsymbol{Q}||^2$, the Fourier transform of the Coulomb interaction:

$$\begin{aligned}\widetilde{\mathrm{PE}}^{\boldsymbol{qp};0}_{\boldsymbol{q}+\boldsymbol{G}} &= i\tilde{v}_{\boldsymbol{q}+\boldsymbol{G}}\sum_j^{\text{nuclei}}\left(\frac{\hbar N_{\text{cell}}Z_j^2}{2\omega_{\boldsymbol{qp}}M_j}\right)^{1/2}(\boldsymbol{q}+\boldsymbol{G})\cdot\boldsymbol{\epsilon}^j_{\boldsymbol{qp}}e^{-i\boldsymbol{G}\cdot\boldsymbol{r}_j}\\ &= \overline{\widetilde{\mathrm{PE}}^{-\boldsymbol{qp};0}_{-\boldsymbol{q}-\boldsymbol{G}}},\end{aligned} \tag{C19}$$

with the caveat that $\tilde{v}_{\boldsymbol{0}} = 0$ to account for the electrical neutrality of the entire medium [89]. $N_{\text{cell}} = \mathcal{V}/\mathcal{V}_{\text{cell}}$ is the number of primitive unit cells; j labels the nuclei in one primitive unit cell; a nucleus labeled j has a charge $Z_j|e|$, mass M_j, and real-spatial coordinate $\boldsymbol{r}_j$; $\boldsymbol{\epsilon}^j_m = \overline{\boldsymbol{\epsilon}^j_{-m}}$ is the polarization vector of the jth nucleus.[57]

It is worth defining a first-quantized operator whose matrix elements (with respect to Bloch waves) are identical to the electron-phonon matrix element [Eq. (C17)]:

$$\hat{V}^{\boldsymbol{qp}} = \mathcal{V}^{-1}\sum_{\boldsymbol{G}}^{\mathrm{RL}}\widetilde{\mathrm{PE}}^{\boldsymbol{qp}}_{\boldsymbol{q}+\boldsymbol{G}}e^{i(\boldsymbol{q}+\boldsymbol{G})\cdot\hat{\boldsymbol{r}}}, \quad \langle B|\hat{V}^m|B'\rangle_1 = V^m_{BB'}. \tag{C20}$$

Because $\mathrm{PE}^m_{\boldsymbol{r}}$ is the one-body potential induced by a complex-valued wave (rather than a standing wave), the potential is not real-valued but satisfies $\mathrm{PE}^m_{\boldsymbol{r}} = \overline{\mathrm{PE}^{-m}_{\boldsymbol{r}}}$; moreover, $\hat{V}^m$ is not self-adjoint:

$$(\hat{V}^m)^\dagger = \hat{V}^{-m} = \hat{T}\hat{V}^m\hat{T}^{-1}, \quad \overline{V^m_{BB'}} = V^{-m}_{B'B}, \tag{C21}$$

with $\hat{T}$ being the first-quantized, time-reversal operator.

Shift current in terms of density matrices. We adopt the Schrödinger representation in which a_m, $a^\dagger_m$, U, and H are all time-independent, i.e., a_m is not accompanied with the multiplicative factor $e^{-i\omega_m t}$. This allows us to solve for the stationary density matrix

$$\partial_t\rho = -\frac{i}{\hbar}[H,\rho] = 0, \quad \rho = \rho^{(0)}+\rho^{(1)}+\rho^{(2)}+\cdots \tag{C22}$$

in time-independent perturbation theory, with $\rho^{(n)}$ proportional to the nth power of the perturbation U.

Because ρ is stationary, $-|e|\mathrm{Tr}[\boldsymbol{v}\rho]$ represents a direct current. $-|e|\mathrm{Tr}[\boldsymbol{v}\rho^{(0)}]$ represents the direct current in the absence of the light source, and vanishes by time-reversal symmetry. We will see in Appendix C 1 b that $\rho^{(1)}$ does not contribute to the direct current, but $\rho^{(2)}$ does. The shift current is the second-order direct current contributed by band-off-diagonal elements of the velocity matrix $\boldsymbol{v}_{bb'\boldsymbol{k}}$:

$$\begin{aligned}\boldsymbol{j} &= -\frac{|e|}{\mathcal{V}}\mathrm{Tr}[\boldsymbol{v}_{\text{off}}\rho^{(2)}],\\ \boldsymbol{v}_{\text{off}} &= \sum_{b,b',\boldsymbol{k}}\boldsymbol{v}^{\text{off}}_{bb'\boldsymbol{k}}c^\dagger_{b\boldsymbol{k}}c_{b'\boldsymbol{k}},\\ \boldsymbol{v}^{\text{off}}_{bb'\boldsymbol{k}} &= \boldsymbol{v}_{bb'\boldsymbol{k}}(1-\delta_{b,b'}).\end{aligned} \tag{C23}$$

Band-diagonal elements contribute to the "ballistic current" [3], which we do not touch upon in this work.

Stationary density matrix from the Lippmann-Schwinger formalism. We will derive $\rho^{(n)}$ based on the Lippmann-Schwinger scattering formalism [90,91], which we briefly review.

For any independent-particle state $|\mu\rangle$ with energy E_μ, one can construct an "in" state $|\mu_+\rangle$ that is an eigenstate of the full Hamiltonian with the same energy:

$$\begin{aligned}&(H-E_\mu)|\mu_+\rangle = 0,\\ &|\mu_+\rangle = |\mu\rangle + G^+_{E_\mu}U|\mu\rangle,\\ &G^+_E = \frac{1}{E-H+i0^+},\end{aligned} \tag{C24}$$

with G^+ the retarded Green's function and 0^+ a positive infinitesimal. An "in" state has the same normalization as its independent-particle counterpart [91]. Since the set of independent-particle states forms an orthonormal basis, so then does the set of all "in" states: $\langle\mu_+|\nu_+\rangle = \delta_{\mu,\nu}$.

Let us motivate the imaginary infinitesimal by a wave-packet interpretation proposed in Ref. [91]. The above correspondence between $|\mu\rangle$ and $|\mu_+\rangle$ allows us to parametrize $|\mu_+\rangle$ by the one-particle wave vectors $(\boldsymbol{k}_1,\boldsymbol{k}_2,\ldots,\boldsymbol{q}_1,\boldsymbol{q}_2,\ldots)$ of electrons and bosons that make up $|\mu\rangle$. Thus it is possible to form a wave packet by smoothly linearly combining $|\mu_+\rangle$ with slightly different values for the one-particle wave vectors. The $i0^+$ guarantees that such a wave packet behaves essentially as a superposition of independent particles in the far past: $t\to-\infty$.[58] The use of "in" states thus simulates a scattering

[56] An explicit expression can be found in Eq. (12.16) of Ref. [87].

[57] The above expressions are obtained from equations (2.9)–(2.11) in Ref. [36].

process in which localized wave packets of electrons and bosons are initially separated (in real space) but subsequently approach each other, and in so doing they evolve into an entangled, polaritonic/polaronic state with a nontrivial current.

Let us construct a density matrix by summing over outer products of "in" states weighted by probability coefficients F_μ:

$$\rho = \sum_\mu F_\mu |\mu_+\rangle\langle\mu_+|, \quad 1 = \sum_\mu F_\mu. \tag{C25}$$

Because the "in" state is an eigenstate of H, ρ satisfies the stationary condition in Eq. (C22). By iteratively expanding the Green's function in a perturbative series

$$\begin{aligned} G^+ &= G_0^+ + G_0^+ U G_0^+ + G_0^+ U G_0^+ U G_0^+ + \cdots, \\ G^+_{0;E} &= \frac{1}{E - H_0 + i0^+}, \end{aligned} \tag{C26}$$

one obtains a perturbative series for the density matrix:

$$\begin{aligned} \rho &= \sum_\mu F_\mu |\mu_+\rangle\langle\mu_+| = \rho^{(0)} + \rho^{(1)} + \rho^{(2)} + \cdots, \\ \rho^{(0)} &= \sum_\mu F_\mu |\mu\rangle\langle\mu|. \end{aligned} \tag{C27}$$

Because the zeroth-order component $\rho^{(0)}$ is stationary with respect to the noninteracting Hamiltonian H_0, one may as well take F_μ to be a product of one-particle probabilities p_{n_B} and P_{N_m}:

$$F_\mu = p_{\{n^\mu\}} P_{\{N^\mu\}}, \quad p_{\{n\}} = \prod_B p_{n_B}, \quad P_{\{N\}} = \prod_m P_{N_m}. \tag{C28}$$

The sense in which p and P are one-particle probabilities is that

$$\begin{aligned} 1 &= \sum_{n_B=0}^{1} p_{n_B} = \sum_{N_m=0}^{\infty} P_{N_m}, \\ \langle n_B\rangle &= \sum_{n_B=0}^{1} p_{n_B} n_B = \sum_\mu F_\mu n_B^\mu, \\ \langle N_m\rangle &= \sum_{N_m=0}^{\infty} P_{N_m} N_m = \sum_\mu F_\mu N_m^\mu, \end{aligned} \tag{C29}$$

with $\langle n_B\rangle$ and $\langle N_m\rangle$ being the average number of electrons and bosons with the one-particle labels B and m, respectively. In a generic, nonequilibrium state, p_{n_B} does not have the Fermi-Dirac form, and instead satisfies a nondetailed balance condition that represents an invariance under simultaneous collisions with all bosons, as detailed in Appendix C 1 e.

It is convenient to introduce the shorthand

$$F_{\mu\nu} = F_\mu - F_\nu, \quad E_{\mu\nu} = E_\mu - E_\nu, \tag{C30}$$

and express $\rho^{(1)}$ and $\rho^{(2)}$ in terms of their matrix elements in the independent-particle basis:

$$\rho^{(1)}_{\mu\nu} = \langle\mu|\rho^{(1)}|\nu\rangle = \frac{F_{\mu\nu} U_{\mu\nu}}{E_{\mu\nu} - i0^+}, \tag{C31}$$

$$\begin{aligned} \rho^{(2)}_{\mu\nu} &= \sum_\lambda U_{\mu\lambda} U_{\lambda\nu} \left[\frac{F_\lambda}{(E_{\lambda\mu} + i0^+)(E_{\lambda\nu} - i0^+)} + \frac{F_\nu}{(E_{\nu\mu} + i0^+)(E_{\nu\lambda} + i0^+)} + \frac{F_\mu}{(E_{\mu\lambda} - i0^+)(E_{\mu\nu} - i0^+)} \right] \\ &= \sum_\lambda \frac{U_{\mu\lambda} U_{\lambda\nu}}{E_{\mu\nu} - i0^+} \left[\frac{F_{\mu\lambda}}{E_{\mu\lambda} - i0^+} + \frac{F_{\nu\lambda}}{E_{\lambda\nu} - i0^+} \right]. \end{aligned} \tag{C32}$$

b. From second-quantized matrix elements to first-quantized matrix elements

We need only concern ourselves with matrix elements $\rho^{(n)}_{\mu\nu}$ with $|\mu\rangle$ and $|\nu\rangle$ having identical occupations numbers for all bosonic modes. After all, for any operator $O = O^e \otimes (identity)$ that acts trivially in the bosonic Fock space,

$$\mathrm{Tr}[O\rho] = \sum_{\mu\nu} O_{\nu\mu} \rho_{\mu\nu} \delta_{\{N^\mu\},\{N^\nu\}}. \tag{C33}$$

In particular, Eq. (C33) holds for O being the electronic velocity operator $\boldsymbol{v}$. An immediate implication is that $\rho^{(1)}$ does not contribute to the direct current: $\mathrm{Tr}[\boldsymbol{v}\rho^{(1)}] = 0$, because $\rho^{(1)}_{\mu\nu} \propto U_{\mu\nu}$ [cf. Eq. (C31)] and U necessarily changes the boson number; cf. Eq. (C16).

Let us apply Eq. (C33) to the shift current [Eq. (C23)] with $\rho^{(2)} = \sum_\lambda U_{\mu\lambda} U_{\lambda\nu} \ldots$ given in Eq. (C32). If $U_{\lambda\nu}$ represents the creation (annihilation) of a boson of mode m, then $U_{\mu\lambda}$ must represent the annihilation (creation) of a boson of the same mode. Thus,

$$\rho^{(2)}_{\mu\nu} \delta_{\{N^\mu\},\{N^\nu\}} = \sum_{\lambda,m} \frac{\langle\mu|U^e_m a_m|\lambda\rangle \langle\lambda|U^e_{-m} a^\dagger_m|\nu\rangle + \langle\mu|U^e_{-m} a^\dagger_m|\lambda\rangle \langle\lambda|U^e_m a_m|\nu\rangle}{E_{\mu\nu} - i0^+} \left[\frac{F_{\mu\lambda}}{E_{\mu\lambda} - i0^+} + \frac{F_{\nu\lambda}}{E_{\lambda\nu} - i0^+} \right] \delta_{\{N^\mu\},\{N^\nu\}}. \tag{C34}$$

[58] One can construct "out" states by flipping the sign of $i0^+$, such that the wave packet becomes essentially noninteracting in the far future. This wave-packet interpretation is elaborated in Chap. 3 of Ref. [91]. In other derivations of the conductivity [15,92], $i0^+$ appears as a result of an adiabatic turn-on process in accordance with Kubo tradition [93], yet no such adiabatic process exists in the typical experiment, e.g., with lasers.

In particular, Eq. (C33) implies there are no "cross terms" proportional to $\langle\cdots U^e_m\cdots\rangle\langle\cdots U^e_{m'}\cdots\rangle$ with m photonic and m' phononic.

Equation (C34) manifests two classes of intermediate states $|\lambda\rangle$—one with $\{N^\lambda\}$ differing from $\{N^\mu\}$ only in that $N^\lambda_m = N^\mu_m + 1$, and another with $\{N^\lambda\}$ differing from $\{N^\mu\}$ only in that $N^\lambda_m = N^\mu_m - 1$. We distinguish the two classes by the notation $\{N^\lambda\} = \{\ldots, N^\mu_m \pm 1, \ldots\}$, which allows us to express

$$\rho^{(2)}_{\mu\nu}\delta_{\{N^\mu\},\{N^\nu\}} = \sum_{\{n^\lambda\}}\sum_m \left\{ \frac{\left(U^e_m\right)_{\mu\lambda}\left(U^e_{-m}\right)_{\lambda\nu}}{E^e_{\mu\nu} - i0^+}\left(N^\mu_m + 1\right)\left[\frac{p_{\{n^\mu\}}P_{\{N^\mu\}} - p_{\{n^\lambda\}}P_{\{\ldots,N^\mu_m+1,\ldots\}}}{E^e_{\mu\lambda} - \hbar\omega_m - i0^+} + \frac{p_{\{n^\nu\}}P_{\{N^\mu\}} - p_{\{n^\lambda\}}P_{\{\ldots,N^\mu_m+1,\ldots\}}}{E^e_{\lambda\nu} + \hbar\omega_m - i0^+}\right]\right.$$
$$\left. + \frac{\left(U^e_{-m}\right)_{\mu\lambda}\left(U^e_m\right)_{\lambda\nu}}{E^e_{\mu\nu} - i0^+}N^\mu_m\left[\frac{p_{\{n^\mu\}}P_{\{N^\mu\}} - p_{\{n^\lambda\}}P_{\{\ldots,N^\mu_m-1,\ldots\}}}{E^e_{\mu\lambda} + \hbar\omega_m - i0^+} + \frac{p_{\{n^\nu\}}P_{\{N^\mu\}} - p_{\{n^\lambda\}}P_{\{\ldots,N^\mu_m-1,\ldots\}}}{E^e_{\lambda\nu} - \hbar\omega_m - i0^+}\right]\right\}\delta_{\{N^\mu\},\{N^\nu\}}. \tag{C35}$$

The above expression utilizes the definition of O^e in Eq. (C14) and a new definition for the electronic component of the total energy:

$$E^e_\mu = \sum_B n^\mu_B E_B, \quad E^e_{\mu\nu} = E^e_\mu - E^e_\nu. \tag{C36}$$

The factors of N^μ_m and $N^\mu_m + 1$ in Eq. (C35) are obtained from the standard matrix elements for bosonic creation and annihilation.

In evaluating $\mathrm{Tr}[\boldsymbol{v}_{\mathrm{off}}\rho^{(2)}]$, we first perform a partial trace by summing over the bosonic occupancies. In this manner, one converts expressions involving N^μ_m to expressions involving average occupancies:

$$\mathrm{Tr}[\boldsymbol{v}_{\mathrm{off}}\rho^{(2)}] = \sum_{\{n^\mu,n^\nu,n^\lambda\}}\sum_m \frac{\left(\boldsymbol{v}^e_{\mathrm{off}}\right)_{\nu\mu}}{E^e_{\mu\nu} - i0^+}\left\{\left(U^e_m\right)_{\mu\lambda}\left(U^e_{-m}\right)_{\lambda\nu}\left[\frac{p_{\{n^\mu\}}\langle N_m+1\rangle - p_{\{n^\lambda\}}\langle N_m\rangle}{E^e_{\mu\lambda} - \hbar\omega_m - i0^+} + \frac{p_{\{n^\nu\}}\langle N_m+1\rangle - p_{\{n^\lambda\}}\langle N_m\rangle}{E^e_{\lambda\nu} + \hbar\omega_m - i0^+}\right]\right.$$
$$\left. + \left(U^e_{-m}\right)_{\mu\lambda}\left(U^e_m\right)_{\lambda\nu}\left[\frac{p_{\{n^\mu\}}\langle N_m\rangle - p_{\{n^\lambda\}}\langle N_m+1\rangle}{E^e_{\mu\lambda} + \hbar\omega_m - i0^+} + \frac{p_{\{n^\nu\}}\langle N_m\rangle - p_{\{n^\lambda\}}\langle N_m+1\rangle}{E^e_{\lambda\nu} - \hbar\omega_m - i0^+}\right]\right\}. \tag{C37}$$

Let us apply the relation between band-off-diagonal elements of the velocity operator and band-off-diagonal elements of the position operator [Eq. (C10)], which translates to the following identity in second quantization:

$$\frac{\left(\boldsymbol{v}^e_{\mathrm{off}}\right)_{\nu\mu}}{E^e_{\mu\nu} - i0^+} = -\frac{i}{\hbar}\sum_{\boldsymbol{k}bb'}\boldsymbol{A}^{\mathrm{off}}_{b'b\boldsymbol{k}}\langle\{n^\nu\}|c^\dagger_{b'\boldsymbol{k}}c_{b\boldsymbol{k}}|\{n^\mu\}\rangle_e \equiv -\frac{i}{\hbar}\left(\boldsymbol{A}^e_{\mathrm{off}}\right)_{\nu\mu}. \tag{C38}$$

In dropping the $i0^+$, we have assumed that $E^e_{\mu\nu} = E_{bb'\boldsymbol{k}}$ (for some $b \neq b'$) is nonzero for the bands and wave vectors of interest; it is worth recalling that b does not include the spin label, hence one should not expect an energy degeneracy due to spin. By plugging Eq. (C38) into Eq. (C37) and recognizing that two of the four terms are complex conjugates of the other two,

$$\mathrm{Tr}\left[\boldsymbol{v}_{\mathrm{off}}\rho^{(2)}\right] = -\frac{i}{\hbar}\sum_{\{n^\mu,n^\nu,n^\lambda\}}\sum_m \left(\boldsymbol{A}^e_{\mathrm{off}}\right)_{\nu\mu}\left[\left(U^e_m\right)_{\mu\lambda}\left(U^e_{-m}\right)_{\lambda\nu}\frac{p_{\{n^\mu\}}\langle N_m+1\rangle - p_{\{n^\lambda\}}\langle N_m\rangle}{E^e_{\mu\lambda} - \hbar\omega_m - i0^+}\right.$$
$$\left. + \left(U^e_{-m}\right)_{\mu\lambda}\left(U^e_m\right)_{\lambda\nu}\frac{p_{\{n^\nu\}}\langle N_m\rangle - p_{\{n^\lambda\}}\langle N_m+1\rangle}{E^e_{\lambda\nu} - \hbar\omega_m - i0^+}\right] + \mathrm{c.c.} \tag{C39}$$

Let us interchange variables $\{n^\mu\} \leftrightarrow \{n^\nu\}$ for the first term and $\{n^\lambda\} \leftrightarrow \{n^\nu\}$ for the second, and then apply the resolution of identity within the electronic Fock space: $\sum_{\{n^\mu\}}|\{n^\mu\}\rangle\langle\{n^\mu\}| = I^e$,

$$\mathrm{Tr}[\boldsymbol{v}_{\mathrm{off}}\rho^{(2)}] = \frac{i}{\hbar}\sum_{\{n^\nu,n^\lambda\}}\sum_m \frac{\left(U^e_m\right)_{\nu\lambda}\left[\boldsymbol{A}^e_{\mathrm{off}}, U^e_{-m}\right]_{\lambda\nu}}{E^e_{\nu\lambda} - \hbar\omega_m - i0^+}\left(\langle N_m\rangle\left(p_{\{n^\nu\}} - p_{\{n^\lambda\}}\right) + p_{\{n^\nu\}}\right) + \mathrm{c.c.}, \tag{C40}$$

with $[\boldsymbol{A}^e_{\mathrm{off}}, U^e_m]_{\lambda\nu}$ meaning a matrix element of the commutator of two electronic operators, as defined in Eqs. (C14) and (C15). By splitting $\sum_m = \sum^{\mathrm{photon}}_m + \sum^{\mathrm{phonon}}_m$ in Eq. (C40), one decomposes $\mathrm{Tr}[\boldsymbol{v}_{\mathrm{off}}\rho^{(2)}] = \mathrm{Tr}[\boldsymbol{v}_{\mathrm{off}}\rho^{(2)}_{\mathrm{phot}}] + \mathrm{Tr}[\boldsymbol{v}_{\mathrm{off}}\rho^{(2)}_{\mathrm{phon}}]$, which we tackle separately.

Evaluating $Tr[\boldsymbol{v}_{off}\rho^{(2)}_{\mathrm{phot}}]$. Recalling the definitions of U^e_m, $\boldsymbol{A}^{\mathrm{off}}_{\boldsymbol{k}}$, and $\boldsymbol{A}^e_{\mathrm{off}}$ in Eqs. (C17), (C10), and (C38), and that $U^e_m = U^e_{-m}$ within the dipole approximation,

$$\left(U^e_m\right)_{\nu\lambda}\left[\boldsymbol{A}^e_{\mathrm{off}}, U^e_{-m}\right]_{\lambda\nu} = \frac{he^2}{\omega_m\mathcal{V}}\sum_{\boldsymbol{k}'\boldsymbol{k}}\sum_{aa'bb'}\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}}\left[\boldsymbol{A}^{\mathrm{off}}_{\boldsymbol{k}'}, \boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{\boldsymbol{k}'}\right]_{aa'}\langle\{n^\nu\}|c^\dagger_{b\boldsymbol{k}}c_{b'\boldsymbol{k}}|\{n^\lambda\}\rangle_e\langle\{n^\lambda\}|c^\dagger_{a\boldsymbol{k}'}c_{a'\boldsymbol{k}'}|\{n^\nu\}\rangle_e, \tag{C41}$$

with a, a', b, and b' being band labels, and $[\boldsymbol{A}^{\mathrm{off}}_{\boldsymbol{k}'}, \boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{\boldsymbol{k}'}]$ being a commutator of two matrices in the band indices.

The product $\langle c^\dagger_{b\boldsymbol{k}} c_{b'\boldsymbol{k}}\rangle_e \langle c^\dagger_{a\boldsymbol{k}'} c_{a'\boldsymbol{k}'}\rangle_e$ is given by

$$\text{(i)}\ \delta_{AA'}\delta_{BB_2} n^\nu_{b\boldsymbol{k}} n^\nu_{a\boldsymbol{k}'} \delta_{\{n^\nu\},\{n^\lambda\}} + \text{(ii)}\ \big(1-n^\lambda_B\big) n^\lambda_{B_2} n^\nu_B \big(1-n^\nu_{B_2}\big)\delta_{AB_2}\delta_{A'B}\delta_{\{n^\nu\},\{n^\lambda\}-B_2+B}, \tag{C42}$$

with $B=(b\boldsymbol{k})$, $B_2=(b'\boldsymbol{k})$, $A=(a\boldsymbol{k}')$, and $A'=(a'\boldsymbol{k}')$, and $\{n^\lambda\}-B_2+B$ labels an electronic Fock basis state that differs from $\{n^\lambda\}$ only in having one-particle state B_2 be unoccupied and B be occupied. Equation (C42) implies two additive contributions to $(U^e_m)_{\nu\lambda}[\boldsymbol{A}^e_{\rm off}, U^e_{-m}]_{\lambda\nu} = (\mathrm{i}')+(\mathrm{ii}')$, namely

$$\begin{aligned}
(\mathrm{i}') &= \frac{he^2}{\omega_m \mathcal{V}} \left\{ \sum_{b\boldsymbol{k}} \boldsymbol{\epsilon}_m \cdot \boldsymbol{v}_{bb\boldsymbol{k}} n^\nu_{b\boldsymbol{k}} \right\} \left\{ \sum_{b'\boldsymbol{k}'} \big[\boldsymbol{A}^{\rm off}_{\boldsymbol{k}'}, \boldsymbol{\epsilon}_m \cdot \boldsymbol{v}_{\boldsymbol{k}'}\big]_{b'b'} n^\nu_{b'\boldsymbol{k}'} \right\} \delta_{\{n^\nu\},\{n^\lambda\}}, \\
(\mathrm{ii}') &= \frac{he^2}{\omega_m \mathcal{V}} \sum_{\boldsymbol{k}} \sum_{bb'} \boldsymbol{\epsilon}_m \cdot \boldsymbol{v}_{bb'\boldsymbol{k}} \big[\boldsymbol{A}^{\rm off}_{\boldsymbol{k}}, \boldsymbol{\epsilon}_m \cdot \boldsymbol{v}_{\boldsymbol{k}}\big]_{b'b} \big(1-n^\lambda_B\big) n^\lambda_{B_2} n^\nu_B \big(1-n^\nu_{B_2}\big) \delta_{\{n^\nu\},\{n^\lambda\}-B_2+B},
\end{aligned} \tag{C43}$$

where we replace the dummy index a by b' in (i'). Plugging $(\mathrm{i}')+(\mathrm{ii}')$ into Eq. (C40) leads to two additive contributions to $\mathrm{Tr}[\boldsymbol{v}_{\rm off}\rho^{(2)}_{\rm phot}] = (\mathrm{i}'')+(\mathrm{ii}'')$.

It should be seen that (i'') is at least fourth order in the electron charge and therefore does not contribute to the second-order $\mathrm{Tr}[\boldsymbol{v}_{\rm off}\rho^{(2)}_{\rm phot}]$. This follows from

$$\sum_{\{n^\nu,n^\lambda\}} \frac{\langle N_m\rangle (p_{\{n^\nu\}} - p_{\{n^\lambda\}}) + p_{\{n^\nu\}}}{E^e_{\mu\lambda} - \hbar\omega_m - i0^+} n^\nu_{b\boldsymbol{k}} n^\nu_{b'\boldsymbol{k}'} \delta_{\{n^\nu\},\{n^\lambda\}} = \sum_{\{n\}} \frac{p_{\{n\}}}{-\hbar\omega_m} n_{b\boldsymbol{k}} n_{b'\boldsymbol{k}'} = -\frac{\langle n_{b\boldsymbol{k}}\rangle \langle n_{b'\boldsymbol{k}'}\rangle}{\hbar\omega_m} \tag{C44}$$

and

$$(\mathrm{i}'')\ \propto e^2 \sum_{\boldsymbol{k}} \boldsymbol{\epsilon}_m \cdot \boldsymbol{v}_{bb\boldsymbol{k}} \langle n_{b\boldsymbol{k}}\rangle = e^2 \sum_{\boldsymbol{k}} \boldsymbol{\epsilon}_m \cdot \boldsymbol{v}_{bb\boldsymbol{k}} \frac{\langle n_{b\boldsymbol{k}}\rangle - \langle n_{b,-\boldsymbol{k}}\rangle}{2} \propto e^4. \tag{C45}$$

Due to time-reversal symmetry, $\boldsymbol{v}_{bb\boldsymbol{k}}$ is odd under $\boldsymbol{k}\to-\boldsymbol{k}$. The same symmetry would constrain $\langle n_{b\boldsymbol{k}}\rangle$ to be an even function if the average were taken in a state of thermal equilibrium. However, optical excitation creates a nonequilibrium state that breaks time-reversal symmetry, which is reflected in a nonzero $(\langle n_{b\boldsymbol{k}}\rangle - \langle n_{b,-\boldsymbol{k}}\rangle)$ that is proportional to the source intensity, i.e., to e^2.[59]

What remains of $\mathrm{Tr}[\boldsymbol{v}_{\rm off}\rho^{(2)}_{\rm phot}]$ is (ii''). To evaluate (ii''), we point out that the energy denominator in Eq. (C40) reduces to

$$\frac{\big(1-n^\lambda_B\big) n^\lambda_{B_2} n^\nu_B \big(1-n^\nu_{B_2}\big)\delta_{\{n^\nu\},\{n^\lambda\}-B_2+B}}{E^e_{\nu\lambda} - \hbar\omega_m - i0^+} = \frac{\big(1-n^\lambda_B\big) n^\lambda_{B_2} n^\nu_B \big(1-n^\nu_{B_2}\big)\delta_{\{n^\nu\},\{n^\lambda\}-B_2+B}}{E_{bb'\boldsymbol{k}} - \hbar\omega_m - i0^+}. \tag{C46}$$

We need two more identities that follow from $p_{\{n\}}$ being a probability function for independent particles [cf. Eq. (C28)]:

$$\begin{aligned}
\sum_{\{n^\nu,n^\lambda\}} p_{\{n^\nu\}} \big(1-n^\lambda_B\big) n^\lambda_{B_2} n^\nu_B \big(1-n^\nu_{B_2}\big)\delta_{\{n^\nu\},\{n^\lambda\}-B_2+B} &= \sum_{n^\lambda_B, n^\lambda_{B_2}} \big(1-n^\lambda_B\big) n^\lambda_{B_2} \sum_{n^\nu_B} p_{n^\nu_B} n^\nu_B \sum_{n^\nu_{B_2}} p_{n^\nu_{B_2}} \big(1-n^\nu_{B_2}\big) = \langle n_B\rangle \big\langle 1-n_{B_2}\big\rangle, \\
\sum_{\{n^\nu,n^\lambda\}} p_{\{n^\lambda\}} \big(1-n^\lambda_B\big) n^\lambda_{B_2} n^\nu_B \big(1-n^\nu_{B_2}\big)\delta_{\{n^\nu\},\{n^\lambda\}-B_2+B} &= \big\langle n_{B_2}\big\rangle \langle 1-n_B\rangle.
\end{aligned} \tag{C47}$$

Altogether, the photonic shift current is expressible as

$$\boldsymbol{j}_{\rm phot} = -\frac{2\pi i|e|^3}{\omega_m \mathcal{V}^2} \sum_m^{\rm photon} \sum_{bb'\boldsymbol{k}} \{N_m f_{bb'\boldsymbol{k}} - f_{b'\boldsymbol{k}}(1-f_{b\boldsymbol{k}})\} \frac{\boldsymbol{\epsilon}_m \cdot \boldsymbol{v}_{b'b\boldsymbol{k}} \big[\boldsymbol{A}^{\rm off}_{\boldsymbol{k}}, \boldsymbol{\epsilon}_m \cdot \boldsymbol{v}_{\boldsymbol{k}}\big]_{bb'}}{E_{bb'\boldsymbol{k}} + \hbar\omega_m + i0^+} + \text{c.c.}, \quad f_{bb'\boldsymbol{k}} = f_{b\boldsymbol{k}} - f_{b'\boldsymbol{k}}. \tag{C48}$$

In this last step, we interchanged $b\leftrightarrow b'$ and simplified our notation as $\langle N_m\rangle \to N_m$ and $\langle n_B\rangle \to f_B$, to be consistent with the rest of the paper.

To go from Eq. (C48) to the final expression for the photonic shift current [Eqs. (A1), (A4), and (A5)] involves a sum rule derived from the first-quantized commutation relation: $[\hat{r}^n, \hat{p}^{n'}] = i\hbar\delta_{n,n'}$, with n and n' denoting the components of three-vectors. We follow this through in Appendix C 1 c.

Evaluating $Tr[\boldsymbol{v}_{off}\rho^{(2)}_{phon}]$. Recalling the definitions of U^e_m, $\check{V}^m$, $\boldsymbol{A}^{\rm off}_{\boldsymbol{k}}$, and $\boldsymbol{A}^e_{\rm off}$ in Eqs. (C17), (C10), and (C38),

$$\big(U^e_m\big)_{\nu\lambda} \big[\boldsymbol{A}^e_{\rm off}, U^e_{-m}\big]_{\lambda\nu} = \sum_{\boldsymbol{k}\boldsymbol{k}'} \sum_{aa'bb'} \check{V}^m_{bb'\boldsymbol{k}} \big(\boldsymbol{A}^{\rm off}_{\boldsymbol{k}'} \check{V}^{-m}_{\boldsymbol{k}'} - \check{V}^{-m}_{\boldsymbol{k}'} \boldsymbol{A}^{\rm off}_{\boldsymbol{k}'+\boldsymbol{q}}\big)_{aa'} \langle\{n^\nu\}|c^\dagger_{b\boldsymbol{k}} c_{b'\boldsymbol{k}-\boldsymbol{q}}|\{n^\lambda\}\rangle_e \langle\{n^\lambda\}|c^\dagger_{a\boldsymbol{k}'} c_{a'\boldsymbol{k}'+\boldsymbol{q}}|\{n^\nu\}\rangle_e, \tag{C49}$$

with $m=(\boldsymbol{q}, p)$, where a, a', b, and b' are band labels, and $\boldsymbol{A}^{\rm off}_{\boldsymbol{k}'} \check{V}^{-m}_{\boldsymbol{k}'}$ is a product of two matrices indexed by band labels.

[59] The "ballistic current" is essentially $-|e|/2\mathcal{V}\sum_{b\boldsymbol{k}} \boldsymbol{v}_{bb\boldsymbol{k}}(\langle n_{b\boldsymbol{k}}\rangle - \langle n_{b,-\boldsymbol{k}}\rangle)$ [3].

Imitating Eqs. (C42) and (C43), we find two additive contributions to $(U^e_m)_{\nu\lambda}[\boldsymbol{A}^e_{\rm off}, U^e_{-m}]_{\lambda\nu} = (\mathrm{i}') + (\mathrm{ii}')$, the first of which is nontrivial only if the phonon wave vector vanishes:

$$
\begin{aligned}
(\mathrm{i}') &= \delta_{\boldsymbol{q},\boldsymbol{0}} \left\{ \sum_{b\boldsymbol{k}} \check{V}^m_{bb\boldsymbol{k}} n^\nu_{b\boldsymbol{k}} \right\} \left\{ \sum_{b'\boldsymbol{k}'} \left[\boldsymbol{A}^{\rm off}_{\boldsymbol{k}'}, \check{V}^{-m}_{\boldsymbol{k}'} \right]_{b'b'} n^\nu_{b'\boldsymbol{k}'} \right\} \delta_{\{n^\nu\},\{n^\lambda\}}, \\
(\mathrm{ii}') &= \sum_{\boldsymbol{k}} \sum_{bb'} \check{V}^m_{bb'\boldsymbol{k}} \left(\boldsymbol{A}^{\rm off}_{\boldsymbol{k}-\boldsymbol{q}} \check{V}^{-m}_{\boldsymbol{k}-\boldsymbol{q}} - \check{V}^{-m}_{\boldsymbol{k}-\boldsymbol{q}} \boldsymbol{A}^{\rm off}_{\boldsymbol{k}} \right)_{b'b} \left(1 - n^\lambda_B\right) n^\lambda_{B_3} n^\nu_B \left(1 - n^\nu_{B_3}\right) \delta_{\{n^\nu\},\{n^\lambda\}-B_3+B},
\end{aligned}
\tag{C50}
$$

with $B = (b\boldsymbol{k})$ and $B_3 = (b', \boldsymbol{k} - \boldsymbol{q})$. The contribution to (i') is only by zero-wave-vector optical phonons, since zero-wave-vector acoustic phonons do not admit quantization.[60] Plugging $(\mathrm{i}') + (\mathrm{ii}')$ into Eq. (C40) leads to two additive contributions to $\mathrm{Tr}[\boldsymbol{v}_{\rm off}\rho^{(2)}_{\rm phon}] = (\mathrm{i}'') + (\mathrm{ii}'')$.

It should be seen that (i'') is at least fourth order in the electron-boson coupling and therefore does not contribute to the second-order $\mathrm{Tr}[\boldsymbol{v}_{\rm off}\rho^{(2)}_{\rm phon}]$. To appreciate this, apply Eq. (C44) once again, noting that the $1/\omega_m$ factor in Eq. (C44) is well-defined for optical phonons as $\boldsymbol{q} \to \boldsymbol{0}$. Then,

$$
(\mathrm{i}'') \propto \left(\sum_{\ldots} \check{V}^m_{\ldots} \cdots \right) \sum_{b\boldsymbol{k}} \left[\boldsymbol{A}^{\rm off}_{\boldsymbol{k}}, \check{V}^{-m}_{\boldsymbol{k}} \right]_{b,b} \langle n_{b\boldsymbol{k}} \rangle = \left(\sum_{\ldots} \check{V}^m_{\ldots} \cdots \right) \sum_{b\boldsymbol{k}} \left[\boldsymbol{A}^{\rm off}_{\boldsymbol{k}}, \check{V}^{-m}_{\boldsymbol{k}} \right]_{b,b} \frac{\langle n_{b\boldsymbol{k}} \rangle - \langle n_{b,-\boldsymbol{k}} \rangle}{2}.
\tag{C51}
$$

Because $(\langle n_{b\boldsymbol{k}} \rangle - \langle n_{b,-\boldsymbol{k}} \rangle)$ is proportional to e^2, altogether (i'') is quadratic in both the electron-photon and electron-phonon couplings.

To arrive at the last line in Eq. (C51), we had applied that $[\boldsymbol{A}^{\rm off}_{\boldsymbol{k}}, \check{V}^{\boldsymbol{0}p}_{\boldsymbol{k}}]_{b,b}$ is odd under $\boldsymbol{k}$-inversion, due to time-reversal symmetry. Indeed, the antiunitary nature of time reversal,

$$
\hat{T}\,|u_{b,-\boldsymbol{k}}\rangle_{\rm cell} = e^{i\phi_{b\boldsymbol{k}}} |u_{b\boldsymbol{k}}\rangle_{\rm cell}, \quad \langle u_B | \hat{T}\, u_{B'} \rangle_{\rm cell} = \langle u_{B'} | \hat{T}^{-1} u_B \rangle_{\rm cell},
\tag{C52}
$$

results in a transposition of the band labels for matrix elements:

$$
\boldsymbol{A}^{\rm off}_{bb',-\boldsymbol{k}} = e^{i(\phi_{b\boldsymbol{k}} - \phi_{b'\boldsymbol{k}})} \boldsymbol{A}^{\rm off}_{b'b\boldsymbol{k}}, \quad \check{V}^{\boldsymbol{0}p}_{b'b,-\boldsymbol{k}} = e^{i(\phi_{b'\boldsymbol{k}} - \phi_{b\boldsymbol{k}})} \check{V}^{\boldsymbol{0}p}_{bb',\boldsymbol{k}} \Rightarrow \left(\boldsymbol{A}^{\rm off}_{-\boldsymbol{k}} \check{V}^{\boldsymbol{0}p}_{-\boldsymbol{k}} \right)_{b,b} = \left(\check{V}^{\boldsymbol{0}p}_{\boldsymbol{k}} \boldsymbol{A}^{\rm off}_{\boldsymbol{k}} \right)_{b,b}.
\tag{C53}
$$

To elaborate on the middle equality, we utilize our general expression for the self-consistently-screened electron-phonon matrix element [Eq. (C17)] and massage the matrix element as

$$
\langle u_{b',-\boldsymbol{k}} | e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}} | u_{b,-\boldsymbol{k}} \rangle = \langle u_{b',-\boldsymbol{k}} | \hat{T}^{-1} e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}} \hat{T} | u_{b,-\boldsymbol{k}} \rangle = \overline{\langle \hat{T}\, u_{b',-\boldsymbol{k}} | e^{-i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}} | \hat{T}\, u_{b,-\boldsymbol{k}} \rangle} = e^{i(\phi_{b'\boldsymbol{k}} - \phi_{b\boldsymbol{k}})} \langle u_{b\boldsymbol{k}} | e^{i\boldsymbol{G}\cdot\hat{\boldsymbol{r}}} | u_{b'\boldsymbol{k}} \rangle,
\tag{C54}
$$

omitting the *cell* superscript in the above equation.

What remains of $\mathrm{Tr}[\boldsymbol{v}_{\rm off}\rho^{(2)}_{\rm phon}]$ is (ii''). To evaluate (ii''), we follow steps closely analogous to Eqs. (C46) and (C47), replacing the Bloch label $B_2 \to B_3$. This leads to the following expression for the phononic shift current:

$$
\boldsymbol{j}_{\rm phon} = -\frac{i|e|}{\hbar\mathcal{V}} \sum_m^{\rm phonon} \sum_{bb'\boldsymbol{k}} \left\{ N_m \left(f_{B_3} - f_B \right) - f_B \left(1 - f_{B_3} \right) \right\} \frac{\left(\boldsymbol{A}^{\rm off}_{\boldsymbol{k}-\boldsymbol{q}} \check{V}^{-m}_{\boldsymbol{k}-\boldsymbol{q}} - \check{V}^{-m}_{\boldsymbol{k}-\boldsymbol{q}} \boldsymbol{A}^{\rm off}_{\boldsymbol{k}} \right)_{b'b} \check{V}^m_{bb'\boldsymbol{k}}}{E_{B_3} - E_B + \hbar\omega_m + i0^+} + \mathrm{c.c.},
\tag{C55}
$$

with $m = (\boldsymbol{q}, p)$, $B = (b\boldsymbol{k})$, and $B_3 = (b', \boldsymbol{k} - \boldsymbol{q})$. Utilizing our definition of the band-off-diagonal position operator [Eq. (C9)] and the first-quantized electron-phonon operator [Eq. (C20)],

$$
[\hat{\boldsymbol{r}}_{\rm off}, \hat{V}^{-m}]_{B'B} = \langle B' | [\hat{\boldsymbol{r}}_{\rm off}, \hat{V}^{-m}] | B \rangle_1, \quad \sum_{\boldsymbol{k}'} [\hat{\boldsymbol{r}}_{\rm off}, \hat{V}^{-m}]_{B'B} V^m_{BB'} = \left(\boldsymbol{A}^{\rm off}_{\boldsymbol{k}-\boldsymbol{q}} \check{V}^{-m}_{\boldsymbol{k}-\boldsymbol{q}} - \check{V}^{-m}_{\boldsymbol{k}-\boldsymbol{q}} \boldsymbol{A}^{\rm off}_{\boldsymbol{k}} \right)_{b'b} \check{V}^m_{bb'\boldsymbol{k}},
\tag{C56}
$$

with $B' = (b'\boldsymbol{k}')$. This identity can be inserted into Eq. (C55) to obtain an equivalent expression for the phononic shift current:

$$
\boldsymbol{j}_{\rm phon} = -\frac{i|e|}{\hbar\mathcal{V}} \sum_m^{\rm phonon} \sum_{BB'} \{ N_m (f_{B'} - f_B) - f_B (1 - f_{B'}) \} \frac{[\hat{\boldsymbol{r}}_{\rm off}, \hat{V}^{-m}]_{B'B} V^m_{BB'}}{E_{B'} - E_B + \hbar\omega_m + i0^+} + \mathrm{c.c.}
\tag{C57}
$$

To go from Eq. (C57) to the final expression for the phononic shift current [Eqs. (A1)–(A3)] involves a sum rule derived from $[\hat{\boldsymbol{r}}, \hat{V}^{-m}] = 0$. The zero is because $\hat{V}^{-m}$ is defined in terms of the position operator but not the momentum operator [Eq. (C20)]. We follow this through in Appendix C 1 d.

[60] One way to see this is that in the quantization of the displacement field, the prefactor in front of a_m is inversely proportional to $\sqrt{\omega_m}$ [35].

c. Sum rule for the photonic shift current

The first-quantized commutation relation

$$i\hbar\delta_{n,n'}\delta_{BB'} = \langle B'|[\hat{r}^n, \hat{p}^{n'}]|B\rangle_1 = \sum_{B''}\left(r^n_{BB''}p^{n'}_{B''B'} - p^{n'}_{BB''}r^n_{B''B'}\right) \tag{C58}$$

will be used to prove

$$[A^{\mathrm{off}n}, P^{n'}]_{bb'} = i\hbar\delta_{n,n'}\delta_{b,b'} + \left[-i\nabla^n_{\boldsymbol{k}} + \left(A^n_{b'b'\boldsymbol{k}} - A^n_{bb\boldsymbol{k}}\right)\right]P^{n'}_{bb'}, \tag{C59}$$

with all $\boldsymbol{k}$-dependent quantities evaluated at the same $\boldsymbol{k}$. By inserting Eqs. (C6) and (C7) into the right-hand side of Eq. (C58) and carrying out $\sum_{B''}$,

$$i\hbar\delta_{n,n'}\delta_{BB'} = iP^{n'}_{bb'\boldsymbol{k}'}\nabla^n_{\boldsymbol{k}}\delta_{\boldsymbol{k}\boldsymbol{k}'} - iP^{n'}_{bb'\boldsymbol{k}}\nabla^n_{\boldsymbol{k}}\delta_{\boldsymbol{k}\boldsymbol{k}'} + \delta_{\boldsymbol{k}\boldsymbol{k}'}[A^n, P^{n'}]_{bb'}. \tag{C60}$$

By applying the second Dirac-δ identity [Eq. (C5)] and separating diagonal and off-diagonal components of A^n, one derives Eq. (C59).

Let us plug Eq. (C59) into our expression for the shift current [Eq. (C48)]. It should be remarked that the $i\hbar\delta_{n,n'}\delta_{b,b'}$ term in Eq. (C59) does not contribute to the current because the band-diagonal velocity $\boldsymbol{v}_{bb\boldsymbol{k}}$ is an odd function of $\boldsymbol{k}$ and the rest of the integrand may be taken as even.[61] What remains is

$$\boldsymbol{j}_{\mathrm{phot}} = \mathrm{Im}\sum_m \frac{4\pi|e|^3}{\omega_m\mathcal{V}^2}\sum_{bb'\boldsymbol{k}}\{N_m f_{bb'\boldsymbol{k}} - f_{b'\boldsymbol{k}}(1-f_{b\boldsymbol{k}})\}\frac{\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{b'b\boldsymbol{k}}[-i\nabla_{\boldsymbol{k}} + (\boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})]\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}}}{E_{bb'\boldsymbol{k}} + \hbar\omega_m + i0^+}, \tag{C61}$$

with all $\boldsymbol{k}$ subscripts omitted for simplicity. By applying the Sokhotski-Plemelj theorem: $1/(x+i0^+) = \mathrm{CPV}[1/x] - i\pi\delta(x)$, with CPV meaning Cauchy's principal value, one can decompose $\boldsymbol{j} = (\mathrm{a}) + (\mathrm{b})$, with

$$(\mathrm{a}) \propto \sum_{\boldsymbol{k}}\{N_m f_{bb'\boldsymbol{k}} - f_{b'\boldsymbol{k}}(1-f_{b\boldsymbol{k}})\}\mathrm{CPV}\frac{\mathrm{Im}\,\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{b'b\boldsymbol{k}}[-i\nabla_{\boldsymbol{k}} + (\boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})]\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}}}{E_{bb'\boldsymbol{k}} + \hbar\omega_m},$$

$$(\mathrm{b}) = -\sum_m \frac{4\pi^2|e|^3}{\omega_m\mathcal{V}^2}\sum_{bb'\boldsymbol{k}}\{N_m f_{bb'\boldsymbol{k}} - f_{b'\boldsymbol{k}}(1-f_{b\boldsymbol{k}})\}\delta(E_{b'b\boldsymbol{k}} - \hbar\omega_m)\mathrm{Re}\,\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{b'b\boldsymbol{k}}[-i\nabla_{\boldsymbol{k}} + (\boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})]\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}}. \tag{C62}$$

(a) vanishes by time-reversal symmetry, which imposes that

$$\mathrm{Im}\,\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{b'b\boldsymbol{k}}[-i\nabla_{\boldsymbol{k}} + (\boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})]\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}} = -|\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}}|\nabla_{\boldsymbol{k}}|\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}}| \tag{C63}$$

is an odd function of $\boldsymbol{k}$. To appreciate this, apply that $\boldsymbol{\epsilon}_m$ is real, the velocity operator inverts sign under time reversal, and the time-reversal symmetry of cell-periodic wave functions [Eq. (C52)]

$$\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb',-\boldsymbol{k}} = -e^{i\phi_{b\boldsymbol{k}} - i\phi_{b'\boldsymbol{k}}}\overline{\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}}}. \tag{C64}$$

(b) is related to the photonic shift vector [Eq. (A4)] by the following identity:

$$\mathrm{Re}\,\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{b'b\boldsymbol{k}}[-i\nabla_{\boldsymbol{k}} + (\boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})]\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}} = |\boldsymbol{\epsilon}_m\cdot\boldsymbol{v}_{bb'\boldsymbol{k}}|^2\boldsymbol{S}^m_{b'\boldsymbol{k}\leftarrow b\boldsymbol{k}}. \tag{C65}$$

Plugging the above equation and Eq. (C10) into Eq. (C62), one finally derives Eq. (A1) with Eqs. (A4) and (A5).

d. Sum rule for the phononic shift current

Substituting Eqs. (C6) and (C7) into the right-hand side of

$$0 = \langle B'|[\hat{\boldsymbol{r}}, \hat{V}^{-m}]|B\rangle_1 = \sum_{B''}\left(\boldsymbol{r}_{B'B''}V^{-m}_{B''B} - V^{-m}_{B'B''}\hat{\boldsymbol{r}}_{B''B}\right), \tag{C66}$$

applying the standard identity $f(x,x')\partial_x\delta(x-x') = \delta(x-x')\partial_{x'}f(x,x')$, and separating the band-diagonal and band-off-diagonal matrix elements of the position operator, one obtains

$$0 = (i\nabla_{\boldsymbol{k}'} + i\nabla_{\boldsymbol{k}} + \boldsymbol{A}_{b'b'\boldsymbol{k}'} - \boldsymbol{A}_{bb\boldsymbol{k}})V^{-m}_{B'B} + \delta_{\boldsymbol{k}',\boldsymbol{k}-\boldsymbol{q}}\left(\boldsymbol{A}^{\mathrm{off}}_{\boldsymbol{k}-\boldsymbol{q}}\check{V}^{-m}_{\boldsymbol{k}-\boldsymbol{q}} - \check{V}^{-m}_{\boldsymbol{k}-\boldsymbol{q}}\boldsymbol{A}^{\mathrm{off}}_{\boldsymbol{k}}\right)_{b'b}. \tag{C67}$$

Plugging this into our expression for the phononic shift current [Eq. (C57)],

$$\boldsymbol{j}_{\mathrm{phon}} = -\,\mathrm{Im}\,\frac{2|e|}{\hbar\mathcal{V}}\sum_m^{\mathrm{phonon}}\sum_{BB'}\{N_m f_{B'B} - f_B(1-f_{B'})\}\frac{V^m_{BB'}(i\nabla_{\boldsymbol{k}'} + i\nabla_{\boldsymbol{k}} + \boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})V^{-m}_{B'B}}{E_{B'B} + \hbar\omega_m + i0^+}, \tag{C68}$$

[61] An argument can be constructed that is analogous to the one used in Eq. (C45).

with $B = (b\boldsymbol{k})$, $B' = (b\boldsymbol{k}')$, $f_{B'B} = f_{B'} - f_B$, and $E_{B'B} = E_{B'} - E_B$. By applying the Sokhotski-Plemelj theorem, one can decompose $\boldsymbol{j}_{\text{phon}} = \text{(a)} + \text{(b)}$, with

$$\text{(a)} \propto \sum_{\boldsymbol{k}\boldsymbol{k}'\boldsymbol{q}} \{N_m f_{B'B} - f_B(1-f_{B'})\}\text{CPV}\frac{\text{Im}\, V^m_{BB'}(i\nabla_{\boldsymbol{k}'} + i\nabla_{\boldsymbol{k}} + \boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})V^{-m}_{B'B}}{E_{B'B} + \hbar\omega_m}, \tag{C69}$$

$$\text{(b)} = \frac{2\pi|e|}{\hbar\mathcal{V}}\sum_m^{\text{phonon}}\sum_{BB'}\{N_m f_{B'B} - f_B(1-f_{B'})\}\delta(E_{BB'} - \hbar\omega_m)\text{Re}\, V^m_{BB'}(i\nabla_{\boldsymbol{k}'} + i\nabla_{\boldsymbol{k}} + \boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})V^{-m}_{B'B}. \tag{C70}$$

To simplify the above expressions, it is worth recalling $\overline{V^m_{BB'}} = V^{-m}_{B'B}$ from Eq. (C21).

(a) vanishes by time-reversal symmetry, which imposes that

$$\text{Im}\, V^m_{BB'}(i\nabla_{\boldsymbol{k}'} + i\nabla_{\boldsymbol{k}} + \boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})V^{-m}_{B'B} = \left|V^m_{BB'}\right|(\nabla_{\boldsymbol{k}'} + \nabla_{\boldsymbol{k}})\left|V^m_{BB'}\right| \tag{C71}$$

is odd under simultaneously inverting $(\boldsymbol{k}, \boldsymbol{k}', \boldsymbol{q}) \to (-\boldsymbol{k}, -\boldsymbol{k}', -\boldsymbol{q})$, and the rest of the integrand in Eq. (C69) is even. (Certainly all energies are even functions, and we have argued for $f_B \approx f_{-B}$; we suppose further that $N_m \approx N_{-m}$, i.e., that any time-reversal breaking of the phonon occupations is proportional to the light intensity, and does not affect the second-order shift current.) To prove oddness of Eq. (C71), it suffices to show that $|V^m_{BB'}|$ is even, i.e., $|V^m_{BB'}| = |V^{-m}_{-B,-B'}|$ with the minus signs denoting a reversal in wave vectors. Recalling how time reversal acts on $\hat{V}^m$ [Eq. (C21)] and on Bloch waves [Eq. (C52)],

$$V^m_{-B,-B'} = e^{i(\phi_B - \phi_{B'})}V^m_{B'B} \Rightarrow \left|V^{-m}_{-B,-B'}\right|^2 = V^{-m}_{-B,-B'}V^m_{-B',-B} = V^{-m}_{B'B}V^m_{BB'} = \left|V^m_{BB'}\right|^2. \tag{C72}$$

Plugging

$$\text{Re}\, V^m_{BB'}(i\nabla_{\boldsymbol{k}'} + i\nabla_{\boldsymbol{k}} + \boldsymbol{A}_{b'b'\boldsymbol{k}} - \boldsymbol{A}_{bb\boldsymbol{k}})V^{-m}_{B'B} = -\left|V^m_{BB'}\right|^2\left\{-(\nabla_{\boldsymbol{k}'} + \nabla_{\boldsymbol{k}})\arg V^m_{BB'} + \boldsymbol{A}_{bb\boldsymbol{k}} - \boldsymbol{A}_{b'b'\boldsymbol{k}}\right\} \tag{C73}$$

into Eq. (C70) and interchanging $B \leftrightarrow B'$, one finally derives Eqs. (A1)–(A3).

It is worth justifying our interpretation of Eq. (A3) as a difference between absorption and emission rates:

(i) Suppose a Bloch state transits from $B \to B'$ while absorbing a phonon of mode m; this is implemented by the electron-phonon interaction $U^e_m(a_m + a^\dagger_{-m})$ [cf. Eq. (C16)], or more specifically by $V^m_{B'B}c^\dagger_{B'}c_B a_m$ [cf. Eq. (C17)]. Thus one expects the associated shift vector for this process to be $-\nabla_{\boldsymbol{k}}\arg V^m_{B'B} + \boldsymbol{A}_{b'b'\boldsymbol{k}'} - \boldsymbol{A}_{bb\boldsymbol{k}} = \boldsymbol{S}^m_{B'\leftarrow B}$ [cf. Eq. (A2)]. By the Golden Rule, one expects a transition probability that is proportional to $|V^m_{B'B}|^2$ and given by the first term in Eq. (A3), namely $\mathcal{A}^m_{B'\leftarrow B}$. The associated contribution to the current is then $-(|e|/\mathcal{V})\,\boldsymbol{S}^m_{B'\leftarrow B}\,\mathcal{A}^m_{B'\leftarrow B}$, which is the first term in Eq. (A1).

(ii) Suppose a Bloch state transits from $B' \to B$ while emitting a phonon of mode m; this is implemented by the electron-phonon interaction $U^e_{-m}(a_{-m} + a^\dagger_m)$ [cf. Eq. (C16)], or more specifically by $V^{-m}_{BB'}c^\dagger_B c_{B'}a^\dagger_m$ [cf. Eq. (C17)]. Thus one expects the associated shift vector for this process to be $-\nabla_{\boldsymbol{k}}\arg V^{-m}_{BB'} + \boldsymbol{A}_{bb\boldsymbol{k}} - \boldsymbol{A}_{b'b'\boldsymbol{k}'} = \boldsymbol{S}^{-m}_{B\leftarrow B'}$ [cf. Eq. (A2)]. By the Golden Rule, one expects a transition probability that is proportional to $|V^{-m}_{BB'}|^2 = |V^m_{B'B}|^2$ [cf. Eq. (C21)] and given by (the negative of) the second term in Eq. (A3), namely $\mathcal{E}^m_{B\leftarrow B'}$. Why the minus sign in Eq. (A3); equivalently, why the minus sign in Eq. (A1)? The reason is that the current contributed by this transition is

$$-\frac{|e|}{\mathcal{V}}\boldsymbol{S}^{-m}_{B\leftarrow B'}\mathcal{E}^m_{B\leftarrow B'} = -\frac{|e|}{\mathcal{V}}\big(-\boldsymbol{S}^m_{B'\leftarrow B}\big)\mathcal{E}^m_{B\leftarrow B'}, \tag{C74}$$

which is the second term in Eq. (A1). Note that $\boldsymbol{S}^{-m}_{B\leftarrow B'} = -\boldsymbol{S}^m_{B'\leftarrow B}$ follows from $\overline{V^m_{BB'}} = V^{-m}_{B'B}$ [cf. Eq. (C21)].

e. The zeroth-order quasiparticle distribution is not thermal

Let us define the nonperturbative quasiparticle distribution as

$$f^{\text{stat}}_B = \text{Tr}\big[n^e_B\rho\big], \quad n^e_B = c^\dagger_B c_B. \tag{C75}$$

In the Schrödinger representation (indicated by $\overset{S}{=}$ below), density matrices can be time-dependent, but operators (such as n^e_B) are time-independent:

$$\partial_t f^{\text{stat}}_B \overset{S}{=} \text{Tr}\big[n^e_B\partial_t\rho\big]. \tag{C76}$$

Because ρ is stationary, the nonperturbative quasiparticle distribution is steady:

$$0 = \partial_t\rho = -\frac{i}{\hbar}[H, \rho] \Rightarrow 0 = \partial_t f^{\text{stat}}_B. \tag{C77}$$

In the Heisenberg representation, density matrices are generally time-independent, but operators (like n^e_B) satisfy Heisenberg's equation of motion:

$$\partial_t f^{\text{stat}}_B = \text{Tr}\big[\big(\partial_t n^e_B\big)_H\rho_H\big] = \frac{i}{\hbar}\text{Tr}\big\{\big[U_H, \big(n^e_B\big)_H\big]\rho_H\big\}, \tag{C78}$$

with O_H denoting an operator O in the Heisenberg representation;[62] here, it should be recalled that $H = H_H = (H_0)_H + U_H$ and $[H_0, n_B^e] = 0 \Rightarrow [(H_0)_H, (n_B^e)_H] = 0$. Since traces are independent of the representation,

$$\mathrm{Tr}\{[U_H, (n_B^e)_H]\rho_H\} = \mathrm{Tr}\{[U, n_B^e]\rho\}, \tag{C79}$$

and we may insert the perturbative expansion for ρ in Eq. (C25). A term in this perturbative expansion that is even in powers of U has a vanishing contribution to $\mathrm{Tr}\{[U, n_B^e]\rho\}$, because one traces over an odd multiple of the bosonic creation/annihilation operator. In particular, $\mathrm{Tr}\{[U, n_B^e]\rho^{(0)}\} = 0$ because $\rho^{(0)} = \sum_\mu F_\mu |\mu\rangle\langle\mu|$ [cf. Eq. (C25)] and $\langle\mu|a_m + a_{-m}^\dagger|\mu\rangle = (a_m + a_{-m}^\dagger)_{\mu\mu} = 0$. Let us therefore evaluate $\mathrm{Tr}\{[U, n_B^e]\rho^{(1)}\}$, using our expression for $\rho^{(1)}$ in Eq. (C31):

$$-i\hbar\partial_t f_B^{\mathrm{stat}} = \sum_{m\mu\nu} \{[U_m^e, n_B^e](a_m + a_{-m}^\dagger)\}_{\nu\mu} \frac{F_{\mu\nu}U_{\mu\nu}}{E_{\mu\nu} - i0^+} + O(U^4). \tag{C80}$$

Each photon/phonon that is created must be subsequently annihilated, and vice versa:

$$-i\hbar\partial_t f_B^{\mathrm{stat}} = \sum_{m\mu\nu} [U_m^e, n_B^e]_{\nu\mu} (U_{-m}^e)_{\mu\nu} \frac{F_{\mu\nu}U_{\mu\nu}}{E_{\mu\nu} - i0^+} \{(a_m)_{\nu\mu}(a_m^\dagger)_{\mu\nu} + (a_{-m}^\dagger)_{\nu\mu}(a_{-m})_{\mu\nu}\}. \tag{C81}$$

Switching $m \to -m$ in the second term, and applying the standard matrix elements for bosonic operators,

$$-i\hbar\partial_t f_B^{\mathrm{stat}} = \sum_{m\mu\nu} [U_m^e, n_B^e]_{\nu\mu} (U_{-m}^e)_{\mu\nu} \frac{F_{\mu\nu}}{E_{\mu\nu} - i0^+} N_m^\mu \delta_{N^\mu, N^\nu + m} + \sum_{m\mu\nu} [U_{-m}^e, n_B^e]_{\nu\mu} (U_m^e)_{\mu\nu} \frac{F_{\mu\nu}}{E_{\mu\nu} - i0^+} (N_m^\mu + 1)\delta_{N^\mu + m, N^\nu}. \tag{C82}$$

$\delta_{N^\mu, N^\nu + m}$ is a Kronecker delta function enforcing $N_{m'}^\mu = N_{m'}^\nu$ for all m', except for $N_m^\nu + 1 = N_m^\mu$; for $\delta_{N^\mu + m, N^\nu}$, it is $N_m^\mu + 1 = N_m^\nu$, which is the exception. We use this δ function to kill the summation over N^ν:

$$\begin{aligned} -i\hbar\partial_t f_B^{\mathrm{stat}} = &\sum_{mn^\mu n^\nu N^\mu} [U_m^e, n_B^e]_{\nu\mu} (U_{-m}^e)_{\mu\nu} \frac{p_{n^\mu} P_{N^\mu} - p_{n^\nu} P_{\cdots N_m^\mu - 1 \cdots}}{E_{\mu\nu}^e + \hbar\omega_m - i0^+} N_m^\mu \\ &+ \sum_{mn^\mu n^\nu N^\mu} [U_{-m}^e, n_B^e]_{\nu\mu} (U_m^e)_{\mu\nu} \frac{p_{n^\mu} P_{N^\mu} - p_{n^\nu} P_{\cdots N_m^\mu + 1 \cdots}}{E_{\mu\nu}^e - \hbar\omega_m - i0^+} (N_m^\mu + 1). \end{aligned} \tag{C83}$$

E^e is the electronic component of E. Carrying out the sum over N^μ,

$$-i\hbar\partial_t f_B^{\mathrm{stat}} = \sum_{mn^\mu n^\nu} [U_m^e, n_B^e]_{\nu\mu} (U_{-m}^e)_{\mu\nu} \frac{p_{n^\mu}\langle N_m\rangle - p_{n^\nu}\langle N_m + 1\rangle}{E_{\mu\nu}^e + \hbar\omega_m - i0^+} + \sum_{mn^\mu n^\nu} [U_{-m}^e, n_B^e]_{\nu\mu} (U_m^e)_{\mu\nu} \frac{p_{n^\mu}\langle N_m + 1\rangle - p_{n^\nu}\langle N_m\rangle}{E_{\mu\nu}^e - \hbar\omega_m - i0^+}. \tag{C84}$$

Interchanging summation variables $n^\mu \leftrightarrow n^\nu$ for the second term,

$$-i\hbar\partial_t f_B^{\mathrm{stat}} = \sum_{mn^\mu n^\nu} \left(- \frac{[n_B^e, U_m^e]_{\nu\mu} (U_{-m}^e)_{\mu\nu}}{E_{\mu\nu}^e + \hbar\omega_m - i0^+} + \frac{[U_{-m}^e, n_B^e]_{\mu\nu} (U_m^e)_{\nu\mu}}{E_{\mu\nu}^e + \hbar\omega_m + i0^+} \right) (p_{n^\mu}\langle N_m\rangle - p_{n^\nu}\langle N_m + 1\rangle). \tag{C85}$$

By applying that n_B^e is self-adjoint and $U_m^e = (U_{-m}^e)^\dagger$ [cf. Eq. (C16)], one recognizes one fraction to be the complex conjugate of the other:

$$-i\hbar\partial_t f_B^{\mathrm{stat}} = \sum_{mn^\mu n^\nu} \left(2i\,\mathrm{Im}\, \frac{[U_{-m}^e, n_B^e]_{\mu\nu} (U_m^e)_{\nu\mu}}{E_{\mu\nu}^e + \hbar\omega_m + i0^+} \right) (p_{n^\mu}\langle N_m\rangle - p_{n^\nu}\langle N_m + 1\rangle). \tag{C86}$$

At this point we split the photonic and phononic contributions:

$$\partial_t f_B^{\mathrm{stat}} = (\partial_t f_B^{\mathrm{stat}})^{\mathrm{phot}} + (\partial_t f_B^{\mathrm{stat}})^{\mathrm{phon}} \tag{C87}$$

by splitting the sum over the bosonic modes: $\sum_m = \sum_m^{\mathrm{phot}} + \sum_m^{\mathrm{phon}}$. Focusing first on the photonic contribution, we evaluate the numerator in Eq. (C86) with help from Eq. (C15), Eq. (C17), and $\epsilon_m = \epsilon_{-m} \in \mathbb{R}$,

$$[U_{-m}^e, n_B^e] = \sum_{b'} \sharp\epsilon_{-m} \cdot (v_{b'\boldsymbol{b}\boldsymbol{k}}^{\mathrm{off}} c_{b'\boldsymbol{k}}^\dagger c_{b\boldsymbol{k}} - (b \leftrightarrow b')); \quad \sharp = \sqrt{\frac{he^2}{\omega_m \mathcal{V}}}, \tag{C88}$$

$$[U_{-m}^e, n_B^e]_{\mu\nu} (U_m^e)_{\nu\mu} = \sum_{b'} |\sharp\epsilon_m \cdot v_{bb'\boldsymbol{k}}^{\mathrm{off}}|^2 \{(1 - n_{B'}^\nu) n_B^\nu (1 - n_B^\mu) n_{B'}^\mu \delta_{n^\mu, n^\nu - B + B'} - (b \leftrightarrow b')\}. \tag{C89}$$

[62] This may be verified by substituting $\rho = e^{-iHt/\hbar}\rho_H e^{iHt/\hbar}$ and $n_B^e = e^{iHt/\hbar}(n_B^e)_H e^{-iHt/\hbar}$ into Eq. (C76).

Since the numerator in Eq. (C86) is manifestly real, it suffices to evaluate the imaginary part of the denominator:

$$\delta_{n^\mu,n^\nu-B+B'}\mathrm{Im}\,\frac{1}{E^e_{\mu\nu}+\hbar\omega_m+i0^+}=-\pi\,\delta_{n^\mu,n^\nu-B+B'}\delta\big(E^e_{\mu\nu}+\hbar\omega_m\big)=-\pi\,\delta_{n^\mu,n^\nu-B+B'}\delta(\varepsilon_{B'B}+\hbar\omega_m), \tag{C90}$$

with ε_B being a one-electron energy. Summing over electron occupancies,

$$\sum_{n^\mu n^\nu} p^{n^\mu}\big(1-n^\nu_{B'}\big)n^\nu_B\big(1-n^\mu_B\big)n^\mu_{B'}\delta_{n^\mu,n^\nu-B+B'}=(1-f_B)f_{B'},\quad f_B=\langle n_B\rangle. \tag{C91}$$

Combining it all, we arrive at a steady-state condition on the quasiparticle occupancies:

$$\begin{aligned}
0=\partial_t f^{\mathrm{stat}}_B&=\big(\partial_t f^{\mathrm{stat}}_B\big)^{\mathrm{phot}}_{\mathrm{gain}}-\big(\partial_t f^{\mathrm{stat}}_B\big)^{\mathrm{phot}}_{\mathrm{loss}}+\big(\partial_t f^{\mathrm{stat}}_B\big)^{\mathrm{phon}}+O(U^4),\\
\big(\partial_t f^{\mathrm{stat}}_B\big)^{\mathrm{phot}}_{\mathrm{gain}}&=\frac{2\pi}{\hbar}\sum_{mb'}\big|\sharp\epsilon_m\cdot v^{\mathrm{off}}_{bb'\boldsymbol{k}}\big|^2(1-f_B)f_{B'}\{\langle N_m\rangle\delta(\varepsilon_{BB'}-\hbar\omega_m)+\langle N_m+1\rangle\delta(\varepsilon_{BB'}+\hbar\omega_m)\},\\
\big(\partial_t f^{\mathrm{stat}}_B\big)^{\mathrm{phot}}_{\mathrm{loss}}&=\frac{2\pi}{\hbar}\sum_{mb'}\big|\sharp\epsilon_m\cdot v^{\mathrm{off}}_{bb'\boldsymbol{k}}\big|^2(1-f_{B'})f_B\{\langle N_m\rangle\delta(\varepsilon_{B'B}-\hbar\omega_m)+\langle N_m+1\rangle\delta(\varepsilon_{B'B}+\hbar\omega_m)\}.
\end{aligned} \tag{C92}$$

It may be seen that the gain and loss rates are of the form expected from Dirac's time-dependent perturbation theory, i.e., Fermi's Golden Rule. The phononic contribution may be evaluated analogously and also has the form expected from Dirac's time-dependent perturbation theory.

In conclusion, for the nonperturbative quasiparticle distribution f^{stat}_B to be steady (up to U^4 corrections), the zeroth-order quasiparticle distribution f_B is the steady solution of $I_{\mathrm{coll}}[f_B]=0$, where I_{coll} is the collisional integral (evaluated by Fermi's Golden Rule) in the presence of the light source. In particular, f_B is *not* the thermal quasiparticle distribution in the absence of the light source, contrary to the way in which most authors approach perturbation theory in nonlinear optical response.

2. Numerical implementation of the BIS shift-current formula

This section explains how to simulate an isoenergy-averaged quasiparticle distribution f_E that is a steady solution to the kinetic equation derived in Appendix B 2, and how f_E is subsequently input to the BIS formula [Eq. (A1)] to determine the shift conductivity and its threefold decomposition. The conductivity will be determined for the model Hamiltonian [Eq. (16) with $\tilde{Q}=1$ and $\tilde{P}=4$] that is characterized by large time-reversal-symmetric Berry curvature; in particular, we would like the reader to be able to reproduce the conductivity plot in Fig. 5(e).

In Appendix B, we have motivated the momentum-resolved collisional integral in Eqs. (B1)–(B7), and we derived the corresponding isoenergy-averaged collisional integral in Eqs. (B16)–(B21), having assumed that the quasiparticle distribution is isoenergy-symmetric: $f_{\boldsymbol{k}}\approx f_E$; cf. Eq. (B15). This assumption is justified to the extent that the collisional integral is isoenergy-symmetric, meaning that Eqs. (B1)–(B7) are well-approximated by Eqs. (B16)–(B21). Whether this is a good approximation depends on the parameters chosen in our model Hamiltonian [Eq. (16)] as well as the source radiation frequency ω_s. We have checked that the e-isotropy condition approximately holds with our chosen parameters ($\tilde{Q}=1$ and $\tilde{P}=4$) in the frequency range $\hbar\omega/E_o\in[0.8,1.5]$.[63] The isoenergy-symmetric assumption was made to save computational simulation time, but one may do without this assumption if one is numerically sophisticated.

There remains some work in fixing the parameters in both sets of collisional integrals, chief among them being the electron–optical-phonon coupling constant ζ in Eq. (B7), as well as the timescale τ^o_E for spontaneous emission of optical phonons in Eq. (B16). The two parameters are related through Eq. (B18), which can be simplified as

$$\zeta\,\frac{a}{\mathcal{V}^2 g_E}\sum_{\mathbf{k}\mathbf{k}'}^{\mathrm{cut}}\frac{|\langle u_{\boldsymbol{k}'}|u_{\boldsymbol{k}}\rangle_{\mathrm{cell}}|^2}{|\mathbf{k}-\mathbf{k}'|^2}\delta(E_{c\boldsymbol{k}}-E)\delta(E_{c\mathbf{k}\mathbf{k}'}-\hbar\Omega_o)=\frac{1}{\tau^o_E}. \tag{C93}$$

The summation is restricted by the condition $\delta k=|\mathbf{k}-\mathbf{k}'|\leqslant\mathscr{G}/10$, with $\mathscr{G}=2\pi/a$ being the reciprocal-lattice period. A typical scale for τ^o_E is 0.1 ps [31,32] hence we set $\tau^o_{E^*}=0.1$ ps for a reference energy $E^*=0.413\,75E_0$ in the active region; this fixes $\zeta=6.329\frac{E_0}{0.1\,\mathrm{ps}}$ and causes τ^o_E to vary from 33.3 to 165.2 fs in the active region, as illustrated in Fig. 11(a). To be clear, all plotted energies are defined to equal zero in the middle of the gap, in contrast to the carrier energies defined with respect to the band extrema. The other parameters in the kinetic model are fixed to be $\tau_{\mathrm{rec}}=1$ ns (a typical interband recombination time [30]); $\tau^s_E=1$ ns for all E (a typical energy relaxation time due to spontaneous emission of acoustic phonons [28]);[64] and $a=5$ Å (a typical lattice period). All calculation in this Appendix are presented for a linearly polarized source: $\boldsymbol{\epsilon}_s=\vec{x}$.

Our first step is to simulate f_E, which sets the isoenergy-averaged collisional integral [Eqs. (B16)–(B21)] to zero. We begin by discretizing the energy: $\ldots,E_j,E_{j+1},E_{j+2},\ldots$ such that adjacent energy levels are separated by $\Delta E=E_{j+1}-E_j$. Conduction-band Bloch states are binned according to the following rule: if $E_j-\Delta E/2\leqslant E_{c\boldsymbol{k}}<E_j+\Delta E/2$, then the

[63] The dipole matrix element becomes isoenergy-asymmetric at higher frequencies, as explained in Sec. IV.

[64] $\tau^s_E\gg\tau^o_E$ and the optical phonon scattering explicitly dominates in the active region.

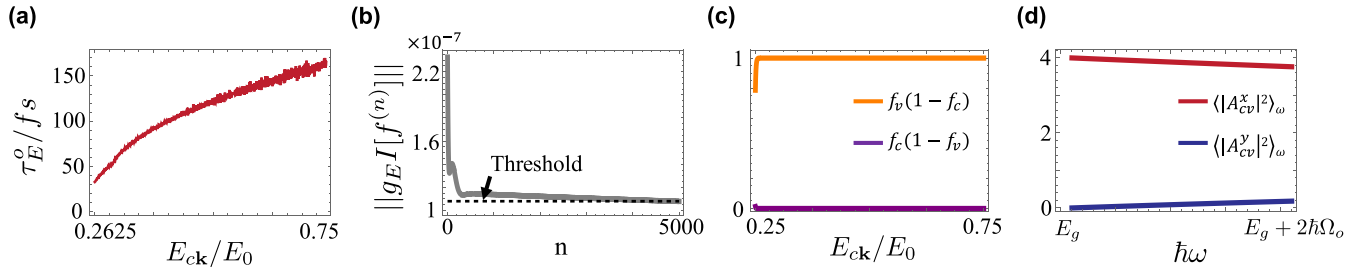

FIG. 11. Part (a) shows the plot of τ_E^o vs E in the active region. Part (b) shows the plot of $||g_E I[f^{(n)}]||$ vs the evolution step n. Part (c) compares $f_c(1-f_v)$ and $f_v(1-f_c)$. Part (d) compares $\langle|A^x_{cv}|^2\rangle_\omega$ and $\langle|A^y_{cv}|^2\rangle_\omega$ in the passive region, i.e., $\hbar\omega \in [E_g, E_g+2\hbar\Omega_o]$.

Bloch state $(c\boldsymbol{k})$ belongs in the jth bin. By choosing $\hbar\omega = n_\omega \Delta E$ and $\hbar\Omega_o = n_\Omega \Delta E$ to be integer multiples of ΔE, one can translate Dirac δ functions to Kronecker δ functions: $\delta(E_{c\mathbf{k}\mathbf{k}'} - \hbar\Omega_o) \to \delta_{E_{c\mathbf{k}\mathbf{k}'},\hbar\Omega_o}/\Delta E$ and $\delta(E_{cv\mathbf{k}} - \hbar\omega) \to \delta_{E_{cv\mathbf{k}},\hbar\omega}/\Delta E$. For instance, $\delta_{E_{c\mathbf{k}\mathbf{k}'},\hbar\Omega_o} = 1$ if and only if $E_{c\boldsymbol{k}} \to E_j$ and $E_{c\boldsymbol{k}'} \to E_{j-n_\Omega}$ for some bin index j. Then the photon-absorption term is discretized as

$$G_\uparrow[f]\,\delta(E - E_{c,\mathrm{exc}}) \to g_E \sum_{m_s} I^{m_s}_{\mathrm{exc},E},$$

$$I^m_{\mathrm{exc},E} = 2\pi^2 \alpha_{\mathrm{fs}} c \frac{\Delta N_m}{\mathcal{V}} 2E\,(1-2f_E) \times \langle|\epsilon_m \cdot \mathbf{A}_{cv\mathbf{k}}|^2\rangle_{\omega_m} \frac{\delta_{2E,\hbar\omega_m}}{\Delta E}, \tag{C94}$$

with $\alpha_{\mathrm{fs}} = |e|^2/(\hbar c)$ being the fine-structure constant, and $I^m_{\mathrm{exc},E}$ being the isoenergy average of $I^{\omega\epsilon}_{\mathrm{exc}\boldsymbol{k}}$; cf. Eq. (A26). In practice, we have chosen $\Delta E = E_0/1600$ and $\hbar\Omega_o = 20\Delta E$. To avoid certain artifacts of our energy discretization scheme, we introduced a small frequency bandwidth $(4\Delta E/\hbar)$ for the source-generated photons; this means that the source produces an equal number of photons in each of four modes $(m_s = 1, 2, 3, 4)$, with differing frequencies $\omega_s - 2\Delta E/\hbar$, $\omega_s - \Delta E/\hbar$, ω_s, $\omega_s + \Delta E/\hbar$ but identical polarization $\boldsymbol{\epsilon}_s$.

We initialize the distribution as a Boltzmann-Maxwell distribution: $f_E^{(0)} = f^T_{E,\mathrm{BM}} = \exp(-(E-\mu_e)/k_BT)$, which is the steady distribution favored by the Fokker-Planck term: $(1 + k_BT\,\partial_E) f_E^T = 0$.[65] μ_e is generically not the chemical potential in thermal equilibrium; instead, it is determined by balancing recombination and excitation rates for the conduction band as a whole: $\sum_E g_E f_E^T \Delta E/\tau_{\mathrm{rec}} = G_\uparrow$, with $\sum_E \Xi(E)$ our shorthand for $\sum_j \Xi(E_j)$.

Beginning from our ansatz Maxwellian distribution, we evolve the system over a discrete time interval δt to obtain a new distribution:

$$f_E^{(n+1)} = f_E^{(n)} + I_E[f^{(n)}]\delta t \tag{C95}$$

for $n = 0, 1, 2, \ldots$, with the collisional integral $I_E[f]$ defined in Eqs. (B16)–(B21). This is a numerical procedure to obtain a steady state, and in no way reflects the actual time evolution of quasiparticle distribution in an experiment. We stop this iterative process when the norm

$$||g_E I[f^{(n)}]|| = \sqrt{\sum_E (g_E I[f^{(n)}(E)])^2} \tag{C96}$$

decays below a certain threshold, i.e., 0.05% of $\sum_{E,m_s} g_E I^{m_s}_{\mathrm{exc}}$. Supposing the threshold is crossed when $n = n_0$, then we say $f^{(n_0)}$ is a *numerically steady* solution of the kinetic equation.

For illustration, Fig. 3(c) represents a numerically steady distribution calculated using the above scheme, with $\Delta N_{m_s}/\mathcal{V} = (1\ \mathrm{ns})^{-1} \frac{\Delta E \mathscr{G}^2}{440\pi^2 c\alpha_{\mathrm{fs}} E_0} \approx 10^{10}\ \mathrm{cm}^{-3}$ for each of the four source modes, $E_{\mathrm{exc}} = 5\hbar\Omega_o$, $n_0 = 5000$ steps, and a time step $\delta t = 1$ fs. Figure 11(b) illustrates a decay of $||g_E\, I[f_n]||$ below our threshold of $(5\times10^{-4})\sum_E g_E I_{\mathrm{exc}} \approx 1.08\times10^{-7}$.

To calculate the shift current, we input the numerically steady $f_E^{(n_0)}$ to the threefold-decomposed current formulas in Eqs. (A25), (A28), and (A32). The discrete analogs of these formulas are

$$\mathbf{j}_{\mathrm{exc}} = -2_{\uparrow\downarrow} \frac{|e|}{\mathcal{V}} \sum_{m_s} \sum_{\mathbf{k}} S^{\epsilon_s}_{c\mathbf{k}\leftarrow v\mathbf{k}} I^m_{\mathrm{exc},E_k}, \tag{C97}$$

$$\mathbf{j}_{\mathrm{intra}} = 2_{\uparrow\downarrow} 2_{cv} \frac{|e|}{\mathcal{V}} \sum_{\mathbf{k}\mathbf{k}'}^{\mathrm{cut}} \frac{\zeta a}{\mathcal{V}} \frac{1}{|\mathbf{k}'-\mathbf{k}|^2} f_{c\boldsymbol{k}'}(1-f_{c\boldsymbol{k}}) \times \frac{\delta_{E_{\mathbf{k}\mathbf{k}'},\hbar\Omega_o}}{\Delta E} \boldsymbol{\Omega}_{c,(\boldsymbol{k}+\boldsymbol{k}')/2} \times (\boldsymbol{k}'-\boldsymbol{k}), \tag{C98}$$

$$\mathbf{j}_{\mathrm{rec}} = 2_{\uparrow\downarrow} \frac{|e|}{\mathcal{V}} \sum_{\mathbf{k}\in\mathrm{pass}} \mathbf{S}^{\hat{x}}_{c\boldsymbol{k}_{\mathrm{ext}}\leftarrow v\boldsymbol{k}_{\mathrm{ext}}} \frac{f_{c\mathbf{k}}}{\tau_{\mathrm{rec}}}. \tag{C99}$$

We will explain each equation in turn:

(i) *Excitation*. I^m_{exc} was defined in Eq. (C94) and $\sum_{m_s}$ sums over the aforementioned source modes.

(ii) *Intra*. Equation (C98) is derived by substituting the electron-phonon matrix element [Eq. (B7)] and the anomalous shift vector [Eq. (5)] into Eq. (A28), and then summing over both conduction and valence bands. In this sum, each band contributes equally due to the presumed electron-hole symmetry ($f_{c\boldsymbol{k}} = 1 - f_{v\boldsymbol{k}}$; cf. Appendix B 2 a), hence the factor of $2_{cv} = 2$ in Eq. (C98). To see why, note for any two-band model that $\boldsymbol{\Omega}_{c\boldsymbol{k}} = -\boldsymbol{\Omega}_{v\boldsymbol{k}}$, hence $\boldsymbol{S}^{\mathrm{ano}}_{c;\boldsymbol{k}'\leftarrow\boldsymbol{k}} = -\boldsymbol{S}^{\mathrm{ano}}_{v;\boldsymbol{k}'\leftarrow\boldsymbol{k}}$ and $f_{c\boldsymbol{k}'}(1-f_{c\boldsymbol{k}})\boldsymbol{S}^{\mathrm{ano}}_{c;\boldsymbol{k}'\leftarrow\boldsymbol{k}} = (1-f_{v\boldsymbol{k}'})f_{v\boldsymbol{k}}(-\boldsymbol{S}^{\mathrm{ano}}_{v;\boldsymbol{k}'\leftarrow\boldsymbol{k}})$. Recognizing from Eq. (5) that $\boldsymbol{S}^{\mathrm{ano}}_{v;\boldsymbol{k}'\leftarrow\boldsymbol{k}} = -\boldsymbol{S}^{\mathrm{ano}}_{v;\boldsymbol{k}\leftarrow\boldsymbol{k}'}$, we find that Eq. (A28) is identical for valence and conduction bands.

[65] Hot-carrier photoluminescence spectra support the hypothesis that most photoexcited carriers are distributed in the manner of Maxwell-Boltzmann; cf. Appendix B 1.

(iii) *Recombination*. $\sum_{\mathbf{k}\in\text{pass}}$ in Eq. (C99) integrates over the passive $\boldsymbol{k}$-volume, based on a previous argument (cf. Sec. II) that the majority of photoexcited carriers are steadily distributed within the passive region; this argument is corroborated by our numerical simulation in Fig. 3(c), bearing in mind that g_E is constant in our quasi-2D model. Because a dipole selection rule fixes $A^y_{cv\boldsymbol{k}}=0$ for $k_x=0$, and $A^y_{cv\boldsymbol{k}}$ cannot vary substantially in the small passive region (assuming the band gap is not anomalously small), it may be deduced that $|A^y_{cv,\mathbf{k}}|^2 \ll |A^x_{cv,\mathbf{k}}|^2$ everywhere in the passive region; cf. Fig. 11(d). Therefore, one may as well approximate all recombination transitions as being mediated by x-polarized photons, with the spontaneous emission rate $\mathcal{E}^{\text{sp},\hat{x}}_{v\mathbf{k}\leftarrow c\mathbf{k}}=f_{c\mathbf{k}}/\tau_{\text{rec}}$. The corresponding photonic shift vector $\mathbf{S}^{\hat{x}}_{v\mathbf{k}\leftarrow c\mathbf{k}}$ is also approximated as $\mathbf{S}^{\hat{x}}_{v\boldsymbol{k}_{\text{ext}}\leftarrow c\boldsymbol{k}_{\text{ext}}}$, because the variation of the photonic shift vector within the passive region is small.

The threefold-decomposed conductivities are obtained by diving each of $\boldsymbol{j}_{\text{exc}}$, $\boldsymbol{j}_{\text{intra}}$, and $\boldsymbol{j}_{\text{rec}}$ by $|\mathcal{E}_\omega|^2$; cf. Eq. (A34). It is advantageous to express the squared electric amplitude $|\mathcal{E}_\omega|^2$ in terms of the discrete I^m_{exc} [Eq. (C94)]:

$$|\mathcal{E}_\omega|^2=\left(\sum_{E,m_s} g_E\Delta E I^{m_s}_{\text{exc},E}\right)\Big/\Big[2\pi c\alpha_{fs}(1-2f_E)\times\left\langle\left|\epsilon_{m_s}\cdot\mathbf{A}_{cv\mathbf{k}}\right|^2\right\rangle_\omega \text{JDOS}_\uparrow\Big], \tag{C100}$$

in accordance with $\sum_{m_s}\Delta N_{m_s}\hbar\omega_s/\mathcal{V}=|\mathcal{E}_\omega|^2/(2\pi)$; cf. Eq. (A13). For the conductivity plot in Fig. 5(e), we had chosen $\sum_{m_s}\Delta N_{m_s}/\mathcal{V}\approx 10^{10}\ \text{cm}^{-3}$. For comparison, in a typical argon-ion-laser experiment with a radiation intensity of 40 W cm^{-2} [94], the number density of source photons is approximately $1/3\times 10^{10}\ \text{cm}^{-3}$.

3. Comparison with the Kraut-Baltz-Sipe-Shkrebtii formula and dissipative Floquet methods

The Kraut-Baltz-Sipe-Shkrebtii formula (KBSS) for the shift current is [4,6,13]

$$\boldsymbol{j}_{\text{KBSS}}=\boldsymbol{\sigma}^{\text{KBSS}}_{\boldsymbol{\epsilon},\omega}|\mathcal{E}_\omega|^2,$$
$$\boldsymbol{\sigma}^{\text{KBSS}}_{\boldsymbol{\epsilon},\omega}=-2\pi\frac{|e|^3}{\hbar}\sum_{bb'}\int\frac{d^3k}{(2\pi)^3}f^T_{bb'\boldsymbol{k}}|\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{b'b\boldsymbol{k}}|^2\boldsymbol{S}_{b'\boldsymbol{k}\leftarrow b\boldsymbol{k}}\times\delta(E_{b'b\boldsymbol{k}}-\hbar\omega), \tag{C101}$$

with $E_{b'b\boldsymbol{k}}=E_{b'\boldsymbol{k}}-E_{b\boldsymbol{k}}$ and $f^T_{bb'\boldsymbol{k}}=f^T_{b\boldsymbol{k}}-f^T_{b'\boldsymbol{k}}$. One can convert Eq. (C101) to a proportionality relation with the radiation intensity (within the dielectric medium) by $\mathcal{I}_{\text{rad}}=(c/2\pi)n_\omega|\mathcal{E}_\omega|^2$, assuming the medium is nonmagnetic with a frequency-dependent refractive index n_ω that is spatially uniform and isotropic.[66]

The KBSS formula has been derived in a variety of models and methods [4,6,7,12,13,15,17,19,20,96], which may have created an impression that it is universally truthful. The actual reason for the universality is a largely unjustifiable and often implicit assumption shared by all these models, namely that the electronic quasiparticle distribution retains its equilibrium value under continuous-wave irradiation. It is an experimental fact that this assumption does not hold, as is most vividly demonstrated by hot-carrier photoluminescence spectroscopy [27,28].

This formula was originally derived by Kraut and Baltz [4,13] and subsequently rederived by Sipe and Shkrebtii [6] using more-or-less standard perturbation theory. In the Kraut-Baltz derivation, relaxation was accounted for in a crude relaxation-time approximation, with the relaxation time eventually taken to be arbitrarily small compared to the Rabi oscillation period at resonance; in other words, relaxation to equilibrium is assumed to be such a strong effect (relative to the optical excitation) that the electronic quasiparticle distribution never deviates from the equilibrium value. (Similar perturbative derivations [12,15,96] have proposed without rigorous justification to view the imaginary infinitesimals in the energy denominator as an inverse relaxation time.) In the Sipe-Shkrebtii derivation [6] (and similar diagrammatic methods [7]), relaxation was omitted entirely; because their method is based on perturbing an equilibrium state in the lowest orders for the electric field, it is not surprising that their final formula is expressed in terms of the equilibrium quasiparticle distribution. The KBSS formula has been alternatively derived from dissipative Floquet methods [17,19,20] in the regime of strong dissipation: relaxation rate $\gg$ Rabi frequency. This is another model where relaxation to equilibrium is assumed to be overwhelmingly strong.

The rest of this Appendix will be used to demonstrate that the BIS formula *also* reduces to the KBSS formula if the electronic quasiparticle distribution is thermal. On the one hand, this planned demonstration can be viewed as a consistency check of the BIS formula. On the other hand, the BIS-to-KBSS reduction crystallizes what is missing from the KBSS formula: namely, the missed photocurrent can be precisely attributed to the deviation of the steady quasiparticle distribution from its equilibrium value, given a realistic model of relaxation in which relaxation also causes shifts.

Without further ado, the KBSS formula in Eq. (C101) is related to the BIS formula in Eq. (A1) by

$$\boldsymbol{j}_{\text{KBSS}}=\boldsymbol{j}\left[f^T_B, N^{T;\text{phot}}_m+\Delta N_s\delta_{m,m_s}, N^{T;\text{phon}}_m\right]. \tag{C102}$$

That the BIS formula is a functional of the quasiparticle, photon, and phonon occupancies has been explained in Appendix A 2. The KBSS formula is thus the BIS formula with a very specific input for occupancies: f^T_B is a Fermi-Dirac distribution [Eq. (A8)], $N^{T;\text{phon}}_m$ is a Planck distribution [Eq. (A9)] with the same temperature, and the photon occupancy is a sum of thermal and nonthermal contributions; the nonthermal photons are generated by a monomodal source with mode index m_s.

[66]The time-averaged Poynting vector (within the dielectric medium) has the form $\mathcal{I}_{\text{rad}}\hat{\boldsymbol{q}}$, with $\mathcal{I}_{\text{rad}}=(c/2\pi)n_\omega|\mathcal{E}_\omega|^2$ having dimensions of energy per unit area per unit time, and $\hat{\boldsymbol{q}}$ being the unit directional vector of the electromagnetic wave propagation. We adopt the same, real-valued definition of the refractive index as in Ref. [95]. In an absorptive medium, $\mathcal{I}_{\text{rad}}$ should be multiplied by a coordinate-dependent, exponential damping factor [95]; however, this factor is negligible if the attenuation length greatly exceeds the thickness of the medium.

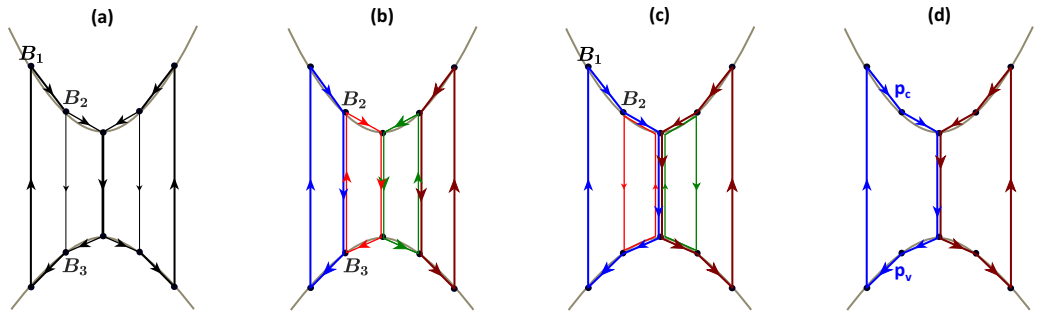

FIG. 12. (a) Caricature of a probability flow network. Panels (b) and (c) illustrate two distinct loop decompositions of the network in panel (a). Panel (d) illustrates a geodesic approximation of the network in (a). Part of the approximation amounts to neglecting the radiative recombination transition between the two Bloch states labeled B_2 and B_3, as justified in Sec. III.

All bosonic modes with a thermal occupancy cannot contribute to the shift current, due to detailed balance; cf. Eq. (A15). For the source mode m_s, the net transition rate [Eq. (A5)] can be decomposed as

$$\begin{aligned}&\left(\mathcal{A}^{m_s}_{C\leftarrow V}-\mathcal{E}^{m_s}_{V\leftarrow C}\right)_{f_B^T,N_{m_s}^T+\Delta N_s}\\&\quad=\left(\mathcal{A}^{m_s}_{C\leftarrow V}-\mathcal{E}^{m_s}_{V\leftarrow C}\right)_{f_B^T,N_{m_s}^T}\\&\qquad+\frac{(2\pi e)^2\omega_s}{\mathcal{V}}\,|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2\delta(E_{cv\boldsymbol{k}}-\hbar\omega_s)f^T_{vc\boldsymbol{k}}\Delta N_s,\end{aligned}\tag{C103}$$

with $C=(c\boldsymbol{k})$ and $V=(v\boldsymbol{k})$. The first term on the right-hand side of Eq. (C103) vanishes by detailed balance [Eq. (A15)], hence the right-hand side of Eq. (C102) reduces to Eq. (C101).

We will say a few words about what is missed from the KBSS formula, how the BIS formula does better, and why dissipative Floquet models (in their present formulation) do not. As explained in Sec. II and elaborated on in Appendix E, the KBSS current is approximately the transient photocurrent, or equivalently the excitation component of the steady photocurrent:

$$\boldsymbol{\sigma}^{\rm KBSS}_{\boldsymbol{\epsilon},\omega}\approx\boldsymbol{\sigma}^{\rm exc}_{\boldsymbol{\epsilon},\omega},\quad \boldsymbol{\sigma}_{\boldsymbol{\epsilon},\omega}=\boldsymbol{\sigma}^{\rm exc}_{\boldsymbol{\epsilon},\omega}+\boldsymbol{\sigma}^{\rm intra}_{\boldsymbol{\epsilon},\omega}+\boldsymbol{\sigma}^{\rm rec}_{\boldsymbol{\epsilon},\omega}.\tag{C104}$$

As defined through the BIS formula, the shift conductivity $\boldsymbol{\sigma}_{\boldsymbol{\epsilon},\omega}$ [Eq. (A14)] has a threefold decomposition explained in Eq. (14); apparently, the KBSS formula misses out on current contributions by intraband relaxation and interband recombination.

Because the KBSS formula is derived by dissipative Floquet methods in the strongly dissipative regime [17,19,20], it is evident that these methods also miss out on the effects of intraband relaxation and interband recombination. The present formulation of Floquet methods is inadequate for the following reasons: (a) The premise of time-periodic Hamiltonians relies on a classical approximation of the radiation field, and precludes the quantum effect of radiative recombination by spontaneous emission. (b) In Refs. [17] and [19], the use of experimentally unrealizable "fermionic baths" as a relaxation mechanism precludes the phonon-induced shift [Eq. (5)] responsible for $\boldsymbol{\sigma}^{\rm intra}_{\boldsymbol{\epsilon},\omega}$. (c) In Ref. [20], Barik and Sau considered electron-phonon scattering as a relaxation mechanism; however, they also missed the phonon-induced shift [Eq. (A2)] due to an unjustifiable assumption that the electron-phonon matrix element is momentum-independent.

APPENDIX D: LOOP FORMULATION OF THE STEADY SHIFT CURRENT

We present an equivalent formulation of the steady shift current, namely that the BIS formula in Eq. (A1) is equivalent to a sum of loop currents:

$$\text{Loop current theorem:}\quad \boldsymbol{j}=-\frac{|e|}{\mathcal{V}}\sum_{B,B',m}\boldsymbol{S}^m_{B'\leftarrow B}\left(\mathcal{A}^m_{B'\leftarrow B}-\mathcal{E}^m_{B\leftarrow B'}\right)=\sum_{\rm loop}\boldsymbol{j}_{\rm loop},\tag{D1}$$

with $\boldsymbol{j}_{\rm loop}$ being the current contributed by a closed flow line (in energy-momentum space) of one-electron probability, as illustrated in Figs. 12(b) and 12(c). The precise definition of $\boldsymbol{j}_{\rm loop}$ is given in Eq. (D15) after some preliminary preparations.

As a first step to reformulating the shift current in terms of loop currents, Appendix D 1 shows how to interpret the flow of one-electron probabilities in energy-momentum space as an oriented graph with nodes corresponding to Bloch states, as illustrated in Fig. 12(a); it will be shown that this graph can be decomposed into loops, and for each loop one can associate a net shift vector [Eq. (D14)] and a current [Eq. (D15)]. We will then prove the loop current theorem in Appendix D 2, and subsequently discuss two applications:

(i) The loop-current formula manifests that the intraband-Berry connection terms $(\boldsymbol{A}_{b'b'\boldsymbol{k}'}-\boldsymbol{A}_{bb\boldsymbol{k}})$ in the shift vector [Eqs. (2) and (3)] always cancel out when all transitions in the steady state are accounted for. From this follows a revision of a purported relation between the shift current and interband polarization differences [49], as discussed in Appendix D 3.

(ii) The loop formulation naturally leads to equitable approximations of the shift current, which treat excitation, relaxation, and recombination on an equal footing. The approximation lies in identifying a reduced family of loops

that contribute most substantially to the shift current. Once a reduced family of loops is identified, calculating the shift current via Eq. (D1) requires far fewer computational resources than a direct calculation of the BIS formula (cf. Appendix C 2). This work focuses on the geodesic loops [Fig. 12(d)] that predominate the shift current in direct-gap semiconductors. For 3D semiconductors, the geodesic approximation to the shift conductivity [Eq. (13)] is derived from the loop-current formula [Eq. (D1)] in Appendix D 4, and as a small-angle-scattering limit of the BIS formula [Eq. (A1)] in Appendix D 5. Finally, the geodesic approximation is extended to quasi-2D semiconductors in Appendix D 6.

1. The shift loop and the loop current

Let us define a *link* as a pair of Bloch labels. A link is said to be ordered if the band energy of the first label is larger than the band energy of the second:

$$\text{ordered link} \equiv (B'{>}B), \quad E_{B'B} = E_{B'} - E_B > 0. \tag{D2}$$

A *general link* written as (B', B) admits any possible ordering of $E_{B'}$ and E_B. For instance, given the three Bloch labels in Fig. 12(a), one may write $(B_1 > B_2)$ and (B_3, B_2) but not $(B_3 > B_2)$.

For every ordered link, we define the *ordered transition rate* as the sum of one-electron transition rates over all possible bosonic modes indexed by m:

$$\text{ordered transition rate } = R_{(B'>B)} = \sum_m \left(\mathcal{A}^m_{B'\leftarrow B} - \mathcal{E}^m_{B\leftarrow B'}\right), \tag{D3}$$

and the *ordered shift vector* as a weighted average of the shift vector [cf. Eqs. (A2)–(A4)] over all bosonic modes:

$$\text{ordered shift vector } \boldsymbol{S}_{(B'>B)} = \sum_m \boldsymbol{S}^m_{B'\leftarrow B} \frac{\mathcal{A}^m_{B'\leftarrow B} - \mathcal{E}^m_{B\leftarrow B'}}{R_{(B'>B)}}. \tag{D4}$$

As a reminder, $\mathcal{A}$ is the absorption rate and $\mathcal{E}$ is the emission rate defined in Eqs. (A3)–(A5). Because $\mathcal{A}^m_{B'\leftarrow B} - \mathcal{E}^m_{B\leftarrow B'} \propto \delta(E_{B'B} - \hbar\omega_m)$ with a bosonic energy $\hbar\omega_m$ that is strictly positive,[67] we wrote $(B'{>}B)$ in Eq. (D4) rather than $(B > B')$. Let us discuss two classes of ordered shift vectors:

Ex. 1: Phononic ordered shift vector

If B' and B differ in electronic wave vectors, then, within the dipole approximation for the electron-photon coupling, one can restrict $\sum_m$ in Eqs. (D3) and (D4) to phononic modes.

(a) If the difference in wave vectors ($\boldsymbol{k}$ and $\boldsymbol{k}'$) is small and the transition is intraband ($b = b'$), $\boldsymbol{S}^{qp}_{B'\leftarrow B}$ in Eq. (D4) is well-approximated by the anomalous shift $\boldsymbol{S}^{\text{ano}}_{b;\boldsymbol{k}'\leftarrow\boldsymbol{k}}$ [Eq. (5)], which does not depend on the phonon branch p. It should also be recalled from Eq. (A3) that $\mathcal{A}^m_{B'\leftarrow B} - \mathcal{E}^m_{B\leftarrow B'} \propto \delta_{\boldsymbol{q},\boldsymbol{k}'-\boldsymbol{k}}$. Altogether, these imply that Eq. (D4) reduces to $\boldsymbol{S}_{(b\boldsymbol{k}'>b\boldsymbol{k})} = \boldsymbol{S}^{ano}_{b;\boldsymbol{k}'\leftarrow\boldsymbol{k}}$.

(b) If the difference in wave vectors ($\boldsymbol{k}$ and $\boldsymbol{k}'$) is not necessarily small, but the phonon energy $\hbar\omega_m = E_{B'B}$ is nondegenerate, then $\sum_m$ in Eqs. (D3) and (D4) is restricted to one phonon branch (say m), and $\boldsymbol{S}_{(B'>B)} = \boldsymbol{S}^m_{B'\leftarrow B}$ as defined in Eq. (A2).

Ex. 2: Photonic ordered shift vector

If B' and B are identical in electron wave vectors ($\boldsymbol{k} = \boldsymbol{k}'$), with $E_{b'b\boldsymbol{k}} = E_{b'\boldsymbol{k}} - E_{b\boldsymbol{k}}$ exceeding the optical phonon energies, than one may restrict $\sum_m$ in Eqs. (D3) and (D4) to photonic modes.

(a) If $\boldsymbol{k}$ does not lie on the excitation surface, $\sum_m$ in Eqs. (D3) and (D4) is restricted (by energy conservation) to photonic modes whose occupations are thermal, i.e., $N_m = N^{T_l}_m$ has the Planck form and does not depend on the orientation $\hat{\boldsymbol{q}}$ of the photon wave vector. In fact, the only quantities in Eq. (D4) that depend on $\hat{\boldsymbol{q}}$ are the photonic shift vector [Eq. (A4)] and the square of the interband Berry connection [Eq. (A5)]. Thus Eq. (D4) simplifies to

$$\boldsymbol{S}_{(b'\boldsymbol{k}>b\boldsymbol{k})} = \frac{\int d\lambda_{\hat{\boldsymbol{q}}} \sum_{p=1}^2 |\boldsymbol{\epsilon}_{\boldsymbol{q}p}\cdot\boldsymbol{A}_{b'b\boldsymbol{k}}|^2 \boldsymbol{S}^{qp}_{b'\boldsymbol{k}\leftarrow b\boldsymbol{k}}}{\int d\lambda_{\hat{\boldsymbol{q}}} \sum_{p=1}^2 |\boldsymbol{\epsilon}_{\boldsymbol{q}p}\cdot\boldsymbol{A}_{b'b\boldsymbol{k}}|^2}, \tag{D5}$$

where we integrate over $\hat{\boldsymbol{q}}$ (parametrized by solid angle $\lambda_{\hat{\boldsymbol{q}}}$) and sum over both transverse polarizations.

(b) If $\boldsymbol{k}$ lies on the excitation surface, $\sum_m$ in Eqs. (D3) and (D4) sum over all photonic modes with the same frequency ω_s as the source-generated photons. For a bright source, an argument in Appendix A 4 conveys that $\sum_m$ in Eqs. (D3) and (D4) may as well be restricted to the single source mode m_s, so that Eq. (D4) simplifies to $\boldsymbol{S}_{(b'\boldsymbol{k}>b\boldsymbol{k})} = \boldsymbol{S}^{m_s}_{b'\boldsymbol{k}\leftarrow b\boldsymbol{k}}$ [Eq. (A4)].

It would also be useful to discuss the net transition rate for $B' \leftarrow B$, with $E_{B'}$ not necessarily greater than E_B. For this purpose, we define

$$\text{oriented transition rate } R_{B'\leftarrow B} = \text{sgn}[E_{B'B}] R_{(B'>B)} = -R_{B\leftarrow B'}, \tag{D6}$$

[67] As remarked earlier in Appendix C 1 b, quantized phonons/photons are not well-defined for zero ω_m.

such that $R_{B'\leftarrow B} > 0$ represents a net probability flow from B to B', independent of the ordering of band energies.

We may draw a cartoon to visualize the flow of probability in energy-momentum space. In Fig. 12(a), we represent every link by an arrow; the thickness of the arrow shaft is proportional to $|R_{(B'>B)}|$; the arrowhead points from $B' \leftarrow B$ if $R_{(B'>B)} > 0$, and vice versa. Our cartoon is thus an oriented graph/network, with each node/vertex corresponding to a Bloch state, and with each link/edge oriented according to the direction of the probability flow. We will use node=B interchangeably.

By comparing the BIS formula [Eq. (A1)] with the definitions of $R_{(B'>B)}$ and $\boldsymbol{S}_{(B'>B)}$ in Eqs. (D3) and (D4), one deduces that the shift current is essentially the sum of $R_{(B'>B)}\boldsymbol{S}_{(B'>B)}$ over all ordered links in the probability-flow network:

$$\boldsymbol{j} = -\frac{|e|}{\mathcal{V}} \sum_{(B'>B)} \boldsymbol{S}_{(B'>B)} R_{(B'>B)}. \tag{D7}$$

In the steady state, the time independence of the occupancy of each Bloch state implies that for each node (say, B) in the graph, incoming transition rates must exactly balance outgoing transition rates: $\sum_{B'} R_{B'\leftarrow B} = 0$. The probability-flow network can therefore be viewed as a discrete analog of a divergence-free/solenoidal vector field. This discrete solenoidal condition allows us to decompose the probability-flow network into loops, as illustrated in Fig. 12(b).[68] Each loop represents the closed flow line of an electron's probability in energy-momentum space, with the perspective that forward-moving holes are backward-moving electrons.

More precisely, here are three defining properties of a loop:

(a) The first property of a loop is that it is a closed concatenation of general links:

$$\begin{aligned}\text{loop with } N \text{ links } &= (B_N, B_{N-1})(B_{N-1}, B_{N-2})\cdots \\ &\quad (B_2, B_1)(B_1, B_N).\end{aligned} \tag{D8}$$

If (B', B) is one of the N links appearing above, then we say the link is contained in the loop: $(B', B) \in$ loop; if $E_{B'B} > 0$ (<0), we would further say that $(B' > B) \in$ loop [$(B > B') \in$ loop].

(b) To each loop, we associate a positive-valued *loop rate* $|\delta R_{\text{loop}}|$, which is the magnitude of the probability flow rate along the loop.

(c) Each loop has a $\mathbb{Z}_2$-valued orientation (Or_{loop}) that determines the direction of probability flow:

$$Or_{\text{loop}} = +1: \quad B_1 \to B_2 \to \cdots \to B_N \to B_1, \tag{D9}$$

$$Or_{\text{loop}} = -1: \quad B_1 \leftarrow B_2 \leftarrow \cdots \leftarrow B_N \leftarrow B_1. \tag{D10}$$

It follows from (a)–(c) that one can assign an *oriented loop rate* to each link in the loop:

$$\delta R^{\text{loop}}_{B_{n+1}\leftarrow B_n} = -\delta R^{\text{loop}}_{B_n\leftarrow B_{n+1}} = Or_{\text{loop}}|\delta R_{\text{loop}}| \text{ with } B_{N+1} \equiv B_1. \tag{D11}$$

The sense in which the probability-flow network is decomposed to loops is that for each link in the network,

$$R_{B'\leftarrow B} = \sum_{\text{loop}\ni(B',B)} \delta R^{\text{loop}}_{B'\leftarrow B}, \tag{D12}$$

where the summation is over all loops that contain the link (B', B); $R_{B'\leftarrow B}$ is given by Fermi's Golden Rule [cf. Eqs. (A3) and (A5)] and depends on the carrier distribution. Equivalently, for every ordered link in the network,

$$R_{(B'>B)} = \sum_{\text{loop}\ni(B'>B)} \delta R^{\text{loop}}_{B'\leftarrow B}. \tag{D13}$$

Consider the cartoon of Fig. 12(b) for illustration: (B_3, B_2) is contained in two loops colored red and blue, hence $R_{B_3\leftarrow B_2}$ is given by a sum of two δR's. The loop decomposition is not unique, meaning that a different set of loops may satisfy Eq. (D13) for the same network, as illustrated in Fig. 12(c).

For each loop, the shift loop is defined by summing the ordered shift vector over all ordered links in the loop, weighted by a sign that encodes the direction of probability flow in that loop:

$$\text{Shift loop} \quad \boldsymbol{S}_{\text{loop}} = \sum_{(B'>B)\in\text{loop}} \text{sgn}\big[\delta R^{\text{loop}}_{B'\leftarrow B}\big]\, \boldsymbol{S}_{(B'>B)}. \tag{D14}$$

Because the summation is over ordered links, $E_{B'B} > 0$, and $\text{sgn}[\delta R^{\text{loop}}_{B'\leftarrow B}] = +1$ (-1) if the loop-decomposed probability flow is toward increasing band energies (decreasing band energies).

To motivate this sign factor, consider an example that elaborates on case (b) of *Ex. 1* [cf. the box under Eq. (D4)]. For the conduction-band link $(B_1 > B_2)$ illustrated in Fig. 12(c), the probability flow is toward decreasing band energies, which reflects the predominance of phonon emission over absorption. Then the link's contribution to $\boldsymbol{S}_{\text{loop}}$ is simply $\text{sgn}[\delta R^{\text{loop}}_{B_1\leftarrow B_2}]\, \boldsymbol{S}_{(B_1>B_2)} = -\boldsymbol{S}^{\boldsymbol{q}p}_{B_1\leftarrow B_2}$, which equals $\boldsymbol{S}^{-\boldsymbol{q},p}_{B_2\leftarrow B_1}$ by the inversion symmetry of the phonon shift vector [Eq. (A2)]. As explained at the end of Appendix C 1 d, $\boldsymbol{S}^{-m}_{B_2\leftarrow B_1}$ is precisely the shift vector associated with emitting a phonon of mode m.

With $\boldsymbol{S}_{\text{loop}}$ and $|\delta R_{\text{loop}}|$ in hand, we can now define the *loop current*

$$\boldsymbol{j}_{\text{loop}} = -\frac{|e|}{\mathcal{V}} \boldsymbol{S}_{\text{loop}} |\delta R_{\text{loop}}|, \tag{D15}$$

[68] Analogously, a divergence-free vector field can be approximated by a superposition of elementary solenoids, which includes the case of finite-length loops [97].

which enters our loop current theorem in Eq. (D1). It may be seen that the loop current depends implicitly on the carrier population through $|\delta R_{\rm loop}|$, as per Eq. (D12) and with identifying $R_{B'\leftarrow B}$ as the Golden-Rule transition rates in Eqs. (A3) and (A5).

2. Derivation of loop current theorem

Beginning from the right-hand side of Eq. (D1), we input the definitions of the loop current in Eq. (D15) and the shift loop in Eq. (D14),

$$\sum_{\rm loop} \boldsymbol{j}_{\rm loop} = -\frac{|e|}{\mathcal{V}} \sum_{\rm loop} \boldsymbol{S}_{\rm loop} |\delta R_{\rm loop}| = -\frac{|e|}{\mathcal{V}} \sum_{\rm loop} \sum_{(B'>B)\in {\rm loop}} {\rm sgn}\left[\delta R^{\rm loop}_{B'\leftarrow B}\right] |\delta R_{\rm loop}|\, \boldsymbol{S}_{(B'>B)}. \tag{D16}$$

Utilizing the definition of the oriented loop rate in Eq. (D11),

$$\sum_{\rm loop} \boldsymbol{j}_{\rm loop} = -\frac{|e|}{\mathcal{V}} \sum_{\rm loop} \sum_{(B'>B)\in {\rm loop}} \delta R^{\rm loop}_{B'\leftarrow B}\, \boldsymbol{S}_{(B'>B)} = -\frac{|e|}{\mathcal{V}} \sum_{(B'>B)} \sum_{{\rm loop}\ni(B'>B)} \delta R^{\rm loop}_{B'\leftarrow B}\, \boldsymbol{S}_{(B'>B)}. \tag{D17}$$

In the last step, we have applied that summing over all ordered links in a given loop and subsequently summing over all loops is equivalent to summing over all loops that contain a given ordered link and subsequently summing over all ordered links. Carrying out the restricted summation over loops on the right-hand side of Eq. (D17) and utilizing the rate decomposition condition in Eq. (D13), we obtain the BIS formula [Eq. (A1)], which completes the proof.

3. Gauge invariance of the reduced shift loop

The theorem allows us to simply derive general properties of the steady shift current. We focus on one such property, namely that the $(\boldsymbol{A}_{b'b'\boldsymbol{k}'} - \boldsymbol{A}_{bb\boldsymbol{k}})$ terms in both phononic and photonic shift vectors [Eqs. (A2)–(A4)] cancel out when all transitions are accounted for. This cancellation was pointed out by BIS without an explicit demonstration [5], but it is a simple consequence of the loop current theorem.

Recall that the shift vector (in either the photonic or phononic case) may be decomposed into a term that depends on the bosonic mode and terms that do not:

$$\boldsymbol{S}^m_{B'\leftarrow B} = \delta \boldsymbol{S}^m_{B'\leftarrow B} + \boldsymbol{A}_{B'} - \boldsymbol{A}_B, \quad \boldsymbol{A}_{B'} = \boldsymbol{A}_{b'b'\boldsymbol{k}'}, \quad \boldsymbol{A}_B = \boldsymbol{A}_{bb\boldsymbol{k}}. \tag{D18}$$

The mode-dependent term is the negative gradient of an argument of a certain transition matrix element [Eqs. (A2)–(A4)]; we will refer to $\delta \boldsymbol{S}^m_{B'\leftarrow B}$ as the *reduced shift vector*. It follows that the ordered shift vector [Eq. (D4)] decomposes similarly as

$$\boldsymbol{S}_{(B'>B)} = \boldsymbol{A}_{B'} - \boldsymbol{A}_B + \sum_m \delta \boldsymbol{S}^m_{B'\leftarrow B} \frac{\mathcal{A}^m_{B'\leftarrow B} - \mathcal{E}^m_{B\leftarrow B'}}{R_{(B'>B)}}. \tag{D19}$$

One may verify that the intraband connection terms cancel out in the shift loop $\boldsymbol{S}_{\rm loop}$ for any loop. Indeed, in the case of the first orientation in Eq. (D9), the shift loop decomposes as

$$\boldsymbol{S}_{\rm loop} = \delta \boldsymbol{S}_{\rm loop} + \left(\boldsymbol{A}_{B_2} - \boldsymbol{A}_{B_1}\right) + \left(\boldsymbol{A}_{B_3} - \boldsymbol{A}_{B_2}\right) + \cdots + \left(\boldsymbol{A}_{B_N} - \boldsymbol{A}_{B_{N-1}}\right) + \left(\boldsymbol{A}_{B_1} - \boldsymbol{A}_{B_N}\right) = \delta \boldsymbol{S}_{\rm loop}. \tag{D20}$$

$\delta \boldsymbol{S}_{\rm loop}$, the *reduced shift loop*, is defined by replacing all shift vectors by reduced shift vectors [cf. Eq. (D18)] in Eq. (D14). Thus it follows that each loop current, being proportional to $\boldsymbol{S}_{\rm loop}$, is invariant if the intraband connection terms are dropped. Finally, the steady shift current, being a sum of loop currents, also satisfies the same invariance property.

Equation (D20) implies that the reduced shift loop is a well-defined, gauge-invariant quantity, despite the fact that the reduced shift vector [Eq. (D18)] of a single transition is not gauge-invariant. By "gauge-invariance," we mean being invariant under redefining one-electron Bloch wave functions by a Bloch-label-dependent phase ϕ_B that is differentiable with respect to $\boldsymbol{k}$: $|u_B\rangle_{\rm cell} \to e^{i\phi_B}|u_B\rangle_{\rm cell}$.

The cancellation in Eq. (D20) calls into question a claim made by Fregoso-Morimoto-Moore [49], namely that large polarization differences $[-|e|\int_{\rm BZ}(\boldsymbol{A}_{cc\boldsymbol{k}} - \boldsymbol{A}_{vv\boldsymbol{k}})d^3k/(2\pi)^3]$ between the conduction and valence bands imply a large shift current in the absence of optical vortices. At best, the Fregoso-Morimoto-Moore claim holds for the frequency-integrated transient shift conductivity (Appendix E), but not the steady shift conductivity.

4. Geodesic approximation of loop currents for 3D semiconductors

The steady shift current is well approximated by keeping the most relevant loops in Eq. (D1). This section focuses on the reduced family of *geodesic loops*, which predominate the shift current in an intrinsic, direct-gap semiconductor with a single minimum for $E_{c\boldsymbol{k}} - E_{v\boldsymbol{k}}$, namely $E_{c\boldsymbol{k}_{\rm ext}} - E_{v\boldsymbol{k}_{\rm ext}} = E_g$, and conditioned on (i) carrier–optical-phonon scattering being the dominant mechanism for energy relaxation in the active region, (ii) small optical phonon energies (relative to E_g and the largest energy of a photoexcited carrier), and (iii) low temperature $k_BT_l \ll E_g, \hbar\Omega_o$ (Ω_o is the optical phonon threshold frequency). The goal of this section is to derive the geodesic approximation to the shift conductivity [Eq. (13)] from the loop-current formula [Eq. (D1)].

To motivate the geodesic loop, let us first consider a pair of Bloch states with Bloch labels $V = (v, \boldsymbol{k}_{\rm exc})$ and $C = (c, \boldsymbol{k}_{\rm exc})$; $\boldsymbol{k}_{\rm exc}$ lies on the optical surface, v denotes the valence band, and c is the conduction band. The oriented transition rate $R_{C\leftarrow V}$ [cf. Eq. (D6)] is assumed to be dominated by the absorption of nonthermal, source-created photons. The probability-flow subgraph that includes the link (C, V) is caricatured in Fig. 13. We will not repeat the arguments (detailed in Secs. II and III) that explain why such a subgraph is predominant; our goal here is to explain how such a subgraph can be approximated by a geodesic loop.

Granted some poetic license, one may view the subgraph as a cyclic probability river that rises in elevation, then splits into tributaries that eventually merge into a waterfall. The splitting reflects the multiple possible intraband relaxation pathways in the conduction band; the merging reflects the existence of a band-energy extremum that causes relaxation pathways to

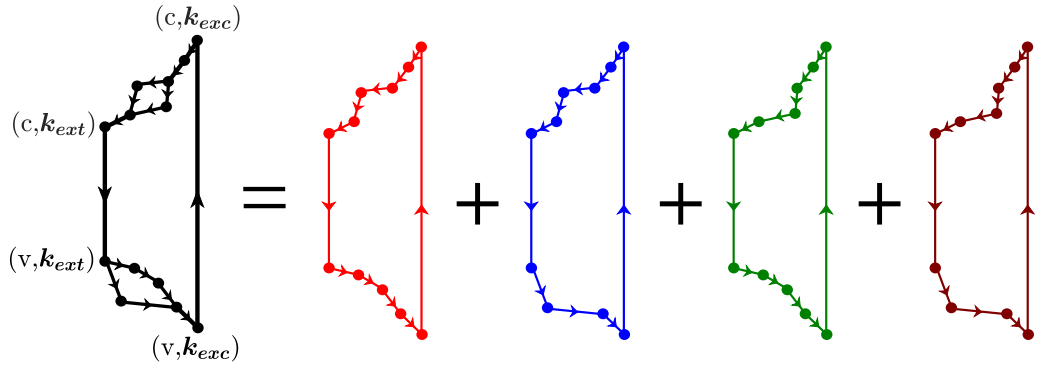

FIG. 13. Loop decomposition of a probability flow subgraph. A similar subgraph was considered in Fig. 4(c).

converge toward the extremal wave vector $\boldsymbol{k}_{\rm ext}$. The cyclic river may be decomposed into N cyclic streams, such that for each stream, the flow rate is constant along the stream. ($N=4$ for our caricature in Fig. 13.) This constant flow rate is identified with $|\delta R_{{\rm loop}_n}|$, with $\{{\rm loop}_n\}_{n=1}^N$ being labels for the N streams. All streams merge at (C,V), such that the sum of the stream flow rates [Eq. (D11)] equals the river flow rate:

$$\sum_{n=1}^{N}\left|\delta R_{{\rm loop}_n}\right| = R_{C\leftarrow V}, \quad C=(c\boldsymbol{k}_{\rm exc}), \quad V=(v\boldsymbol{k}_{\rm exc}). \tag{D21}$$

We have chosen a stream decomposition such that all the streams flow with the same orientation as the river, and this is always possible to choose. In principle, one may choose a stream decomposition in which some of the streams flow against the river along (C,V); then for those counterflowing streams, one would replace $|\delta R_{{\rm loop}_n}| \to -|\delta R_{{\rm loop}_n}|$ in Eq. (D21).

We proceed without further use of metaphors. The contribution of the above subgraph to the shift current[69] is a sum of N loop currents:

$$\boldsymbol{j}_{{\rm subgraph}[\boldsymbol{k}_{\rm exc}]} = -\frac{|e|}{\mathcal{V}}\sum_{n=1}^{N}\boldsymbol{S}_{{\rm loop}_n}\left|\delta R_{{\rm loop}_n}\right|. \tag{D22}$$

As argued in Sec. III, the predominant intraband-relaxation pathways do not deviate far from geodesic paths connecting $\boldsymbol{k}_{\rm exc}$ to $\boldsymbol{k}_{\rm ext}$; we remind the reader that the geodesic path is orthogonal to all isoenergy contours. Let us define the *geodesic loop* as combining an excitation transition at $\boldsymbol{k}_{\rm exc}$, geodesic-path relaxation to $\boldsymbol{k}_{\rm ext}$ through the conduction band, recombination at $\boldsymbol{k}_{\rm ext}$, and geodesic-path relaxation back to $\boldsymbol{k}_{\rm exc}$ through the valence band, as caricatured in Fig. 12(d). We denote the geodesic, oriented $\boldsymbol{k}$-paths by p_c and p_v, respectively, and the geodesic loop by loop$[\boldsymbol{k}_{\rm exc}]$; the associated shift loop is denoted $\boldsymbol{S}_{{\rm loop}[\boldsymbol{k}_{\rm exc}]}$, with $\boldsymbol{S}_{\rm loop}$ generally defined in Eq. (D14).

It is of interest to show how $\boldsymbol{S}_{{\rm loop}[\boldsymbol{k}_{\rm exc}]}$ simplifies to an expression for the shift loop [Eqs. (6)–(8)] that we have used in the main text:

(i) For the recombination transition associated with $(B'>B)=(c\boldsymbol{k}_{\rm ext}>v\boldsymbol{k}_{\rm ext})$ (leftmost link in Fig. 13), one applies Eq. (D5) to show that ${\rm sgn}[\delta R^{\rm loop}_{B'\leftarrow B}]\,\boldsymbol{S}_{(B'>B)}=\boldsymbol{S}_{\rm rec}$, as defined in Eq. (8). ${\rm sgn}[\delta R^{\rm loop}_{B'\leftarrow B}]$ being -1 accounts for the reversed orientation in a recombination transition, but this minus sign can be absorbed by $-\boldsymbol{S}^m_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}=\boldsymbol{S}^m_{v\boldsymbol{k}\leftarrow c\boldsymbol{k}}$.

(ii) For the excitation transition associated with $(B'>B)=(c\boldsymbol{k}_{\rm exc}>v\boldsymbol{k}_{\rm exc})$, ${\rm sgn}[\delta R^{\rm loop}_{B'\leftarrow B}]\,\boldsymbol{S}_{(B'>B)}=\boldsymbol{S}^{m_s}_{c\boldsymbol{k}_{\rm exc}\leftarrow v\boldsymbol{k}_{\rm exc}}$ assuming that the source is monomodal and bright (cf. argument in Appendix A 4).

(iii) For an intraband transition associated with $(B'>B)=(c\boldsymbol{k}'>c\boldsymbol{k})$, it is assumed small-angle scattering predominates ($||\boldsymbol{k}'-\boldsymbol{k}||\ll$ Brillouin-zone dimension), such that ${\rm sgn}[\delta R^{\rm loop}_{B'\leftarrow B}]\,\boldsymbol{S}_{(B'>B)}$ reduces to the asymptotic expression $-\boldsymbol{\Omega}_{c\boldsymbol{k}_{\rm ave}}\times\delta\boldsymbol{k}$ [cf. case (a) in *Ex. 1* of Appendix D 1]. By approximating a discrete sum over intraband links as a line integral, one obtains the first line integral in Eq. (6). The second line integral is obtained in an analogous manner.

Because ${\rm loop}_n$ does not deviate far from loop$[\boldsymbol{k}_{\rm exc}]$, we approximate $\boldsymbol{S}_{{\rm loop}_n}\approx\boldsymbol{S}_{{\rm loop}[\boldsymbol{k}_{\rm exc}]}$ for all loops that make up the subgraph; this is the geodesic approximation. The approximation is justified to the extent that small-angle scattering predominates over large-angle scattering, as elaborated on in Appendix D 5. Applying the geodesic approximation to Eqs. (D21) and (D22),

$$\boldsymbol{j}_{{\rm subgraph}[\boldsymbol{k}_{\rm exc}]}\approx-\frac{|e|}{\mathcal{V}}\boldsymbol{S}_{{\rm loop}[\boldsymbol{k}_{\rm exc}]}R_{C\leftarrow V}. \tag{D23}$$

The shift current is a sum of loop currents over loops that constitute the full probability-flow network, and not just the subgraph containing (C,V). [For the one-dimensional caricature in Fig. 12(d), the full network is composed of two subgraphs.] In other words, one should sum $\boldsymbol{j}_{{\rm subgraph}[\boldsymbol{k}_{\rm exc}]}$ over all $\boldsymbol{k}_{\rm exc}$ on the excitation surface:

$$\boldsymbol{j}\approx-2_{\uparrow\downarrow}\frac{|e|}{\mathcal{V}}\sum_{\boldsymbol{k}}\boldsymbol{S}_{{\rm loop}[\boldsymbol{k}]}R_{(c\boldsymbol{k})\leftarrow(v\boldsymbol{k})}, \tag{D24}$$

with the understanding that $R_{(c\boldsymbol{k})\leftarrow(v\boldsymbol{k})}\propto\delta(E_{cv\boldsymbol{k}}-\hbar\omega)$ [Eq. (A5)] constrains $\sum_{\boldsymbol{k}}$ to the excitation surface; we have also included a factor of 2 to account for spin. Assuming a bright, monomodal light source, we may follow the argument in Appendix A 4 to derive that $R_{(c\boldsymbol{k})\leftarrow(v\boldsymbol{k})}\approx I^{\epsilon\omega}_{{\rm exc}\boldsymbol{k}}$, with $I_{\rm exc}$ defined in Eq. (A26). Converting the source mode occupancy ΔN_s to an electric-field amplitude $\mathcal{E}_\omega$ through Eq. (A13), one obtains

$$\boldsymbol{j}\approx-2\pi\frac{|e|^3}{\hbar}\langle f_{vc\boldsymbol{k}}|\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2\boldsymbol{S}_{{\rm loop}[\boldsymbol{k}]}\rangle_\omega(2{\rm JDOS}_\uparrow)|\mathcal{E}_\omega|^2, \tag{D25}$$

[69] Bear in mind that "current" has nothing to do with the metaphorical river current.

with $f_{vc} = f_{v\boldsymbol{k}} - f_{c\boldsymbol{k}}$ a difference in the steady-state quasiparticle distribution functions, and $\langle\cdots\rangle_\omega$ and $\mathrm{JDOS}_\uparrow$ defined in Eq. (11). When expressed in terms of a nonlinear conductivity, $\boldsymbol{j} = \boldsymbol{\sigma}_{\boldsymbol{\epsilon},\omega}|\mathcal{E}_\omega|^2$, Eq. (D25) is equivalent to Eq. (13).

5. Geodesic approximation as a small-angle-scattering limit of the BIS formula

Because the geodesic approximation of the shift conductivity [Eq. (13)] has been used in all model calculations, it is of interest to clarify the regime of validity of the approximation. Here, we will demonstrate that Eq. (13) derives as a *small-angle-scattering limit* of the BIS formula Eq. (A1): $\lim_{s\to\infty} \boldsymbol{\sigma}_{\mathrm{BIS}} = \boldsymbol{\sigma}_{\mathrm{geo}}$, with s a parameter that controls the angle of scattering. The choice of s is not unique. One possible choice is to increase the power in the square of the electron-phonon matrix element: $|V^m_{b\boldsymbol{k}',b\boldsymbol{k}}|^2 \propto 1/||\boldsymbol{k}' - \boldsymbol{k}||^s$ [Eq. (C17)], bearing in mind that this is a theoretical exercise to elucidate the essence of the geodesic approximation; the physical value of s is 2 for polarization scattering with optical phonons; cf. Sec. III.

Implementing the threefold decomposition of both $\boldsymbol{\sigma}_{\mathrm{geo}}$ [Eq. (14)] and $\boldsymbol{\sigma}_{\mathrm{BIS}}$ (Appendix A 4), one can straightforwardly verify that the excitation components match exactly, while the recombination components match to a good approximation, bearing in mind that recombination transitions predominantly occur at $\boldsymbol{k}$ near the extremal wave vector.[70] This section will demonstrate for the intraband components that $\lim_{s\to\infty} \boldsymbol{\sigma}^{\mathrm{intra}}_{\mathrm{BIS}} = \boldsymbol{\sigma}^{\mathrm{intra}}_{\mathrm{geo}}$.

Assuming only two bands are optically excited, the intraband conductivity decomposes into contributions by individual bands: $\boldsymbol{\sigma}^{\mathrm{intra}} = \boldsymbol{\sigma}^{\mathrm{intra}}_c + \boldsymbol{\sigma}^{\mathrm{intra}}_v$, and we will prove for the conduction band that

$$\text{BIS-geodesic reduction:}\quad \lim_{s\to\infty} \boldsymbol{\sigma}^{\mathrm{intra}}_{c,\mathrm{BIS}} = \boldsymbol{\sigma}^{\mathrm{intra}}_{c,\mathrm{geo}} + O_r\left(\sqrt{\frac{\hbar\Omega_o}{E_{\mathrm{exc}}}}, \sqrt{\frac{\hbar\Omega_o}{E_g}}\right), \tag{D26}$$

with Ω_o the optical phonon threshold [Eq. (A10)] and E_{exc} the excitation energy measured from the conduction-band minimum [Eq. (A21)]; the meaning of O_r is relative error, i.e., $a + O_r(b,c)$ means $O_r(b,c)$ has a magnitude less than or comparable to maximum$\{|ba|, |ca|\}$, assuming b and c are dimensionless. The BIS-geodesic reduction for the valence band [Eq. (D26) with $c \to v$] also holds true, but is a straightforward extension requiring no further substantiation.

To clarify, $\boldsymbol{\sigma}^{\mathrm{intra}}_{c,\mathrm{geo}}$ is given by Eqs. (6)–(13) with the shift loop reduced to the line integral over the geodesic path $p_{c\boldsymbol{k}}$ connecting $\boldsymbol{k}$ to $\boldsymbol{k}_{\mathrm{ext}}$:

$$\boldsymbol{\sigma}^{\mathrm{intra}}_{c,\mathrm{geo}} = -2\pi\frac{|e|^3}{\hbar}\left\langle f_{vc\boldsymbol{k}}|\boldsymbol{\epsilon}\cdot A_{cv\boldsymbol{k}}|^2 \int_{p_{c\boldsymbol{k}}} \boldsymbol{\Omega}_c \times d\boldsymbol{k}\right\rangle_\omega 2_{\uparrow\downarrow}\mathrm{JDOS}_\uparrow, \tag{D27}$$

while $\boldsymbol{\sigma}^{\mathrm{intra}}_{c,\mathrm{BIS}}$ is taken from Eqs. (A28)–(A34):

$$\boldsymbol{\sigma}^{\mathrm{intra}}_{c,\mathrm{BIS}} = -2_{\uparrow\downarrow}\frac{|e|}{\mathcal{V}|\mathcal{E}_\omega|^2}\sum_{\boldsymbol{k},\boldsymbol{k}'} \boldsymbol{S}_{c\boldsymbol{k}\leftarrow c\boldsymbol{k}'}\mathcal{E}^{\mathrm{sp}}_{c\boldsymbol{k}\leftarrow c\boldsymbol{k}'}. \tag{D28}$$

We have omitted the phonon mode $m = (\boldsymbol{q}p)$ superscript on the phononic shift $\boldsymbol{S}^m_{c\boldsymbol{k}'\leftarrow c\boldsymbol{k}} = -\boldsymbol{S}^{-m}_{c\boldsymbol{k}\leftarrow c\boldsymbol{k}'}$ and the spontaneous emission rate $\mathcal{E}^{\mathrm{sp},m}$, with the understanding that p is fixed to a single branch of optical phonons, and $\boldsymbol{q} = \boldsymbol{k} - \boldsymbol{k}'$ is fully determined by momentum conservation; cf. Eq. (A3). Henceforth, we will simplify the notation by omitting the c subscript on all quantities, except in instances where such an omission may lead to confusion.

In addition to certain assumptions that justify the predominance of geodesic loops (summarized in the beginning of Appendix D 4), we will make additional model assumptions that simplify the demonstration of the BIS-geodesic reduction [Eq. (D26)], though we do not believe these additional assumptions are ultimately necessary for the reduction:

(i) The optical phonon frequency is roughly a constant equal to $\hbar\Omega_o$ for the small phonon wave vectors we consider.

(ii) Both conduction and valence bands have isotropic dispersions, i.e., $E_{c\boldsymbol{k}}$ and $E_{v\boldsymbol{k}}$ depend on $\boldsymbol{k}$ through $||\boldsymbol{k}||$, as may be expected near band extrema with cubic symmetry.

(iii) In the active region, electron–optical-phonon scattering overwhelmingly dominates over electron–acoustic-phonon scattering as the primary mechanism for energy relaxation. One way to formalize this is to take η_E defined in Eq. (B20) to zero.

Some implications of (i)–(iii) will hereby be elucidated in preparation to prove the BIS-geodesic reduction [Eq. (D26)].

Excitation rate. (i) and (ii) imply that the excitation energy E_{exc} [cf. Eq. (A21)] of conduction-band states is degenerate, i.e., the excitation rate [Eq. (A26)] is nonzero only if $E_{\boldsymbol{k}} = E_{\mathrm{exc}}$:

$$I_{\mathrm{exc}\boldsymbol{k}} = \tilde{I}_{\mathrm{exc}\boldsymbol{k}}\delta_{\boldsymbol{k},E_{\mathrm{exc}}},$$
$$\tilde{I}_{\mathrm{exc}\boldsymbol{k}} = \frac{2\pi|e|^2}{\hbar} f_{vc\boldsymbol{k}}|\epsilon\cdot\mathbf{A}_{cv\boldsymbol{k}}|^2\frac{|\nabla_{\boldsymbol{k}}E_c|}{|\nabla_{\boldsymbol{k}}E_{cv}|}\mathcal{V}g_{E_{\mathrm{exc}}}|\mathcal{E}_\omega|^2. \tag{D29}$$

We collect here a few useful properties of *surface projectors*:

$$\delta_{\boldsymbol{k},E} = \frac{\delta(E_{\boldsymbol{k}} - E)}{\mathcal{V}g_E}, \quad \sum_{\mathbf{k}}\delta_{\mathbf{k},E} = \sum_E \delta_{\mathbf{k},E} = 1,$$
$$\delta_{\boldsymbol{k},E}\delta_{\boldsymbol{k},E} = \delta_{\boldsymbol{k},E}, \quad \delta_{\boldsymbol{k},E_j}\delta_{\boldsymbol{k},E_{j'}} = \delta_{\boldsymbol{k},E_j}\delta_{jj'}, \tag{D30}$$

[70] As described in Sec. II and elaborated here, $\sigma^{\mathrm{rec}}_{\mathrm{BIS}}$ reduces to $\sigma^{\mathrm{rec}}_{\mathrm{geo}}$ if the photonic shift vector $\boldsymbol{S}^m_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$ [in Eq. (A32), with $\boldsymbol{k}$ in the passive $\boldsymbol{k}$-volume] is approximated to be $\boldsymbol{S}^m_{c\boldsymbol{k}_{\mathrm{ext}}\leftarrow v\boldsymbol{k}_{\mathrm{ext}}}$. This approximation leads to a relative error of order $\hbar\Omega_o/E_g$, assuming that the band gap E_g is the energy scale for significant variation of the shift vector. If the photoexcited carriers within the passive region follow a Maxwellian distribution, with electron temperature $k_BT_e < \hbar\Omega_o$ and hole temperature $k_BT_h < \hbar\Omega_o$ (cf. Appendix B 1), then the relative error is reducible to $\sigma^{\mathrm{rec}}_{\mathrm{BIS}} = \sigma^{\mathrm{rec}}_{\mathrm{geo}} + O_r(k_BT/E_g)$, with T being the smaller of $\{T_e, T_h\}$.

which encode their completeness (with $\sum_E$ meaning $\int \mathcal{V} g_E dE$), idempotence, and orthogonality.[71] Integrating a surface-projected test function is equivalent to averaging the test function over a two-sphere parametrized by the solid angle λ:

$$\langle\langle \Xi(\boldsymbol{k})\rangle\rangle_{kE} = \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k},E}\ \Xi(\boldsymbol{k}) = \int \frac{d\lambda}{4\pi}\ \Xi(\boldsymbol{k})\Big|_{\boldsymbol{k}=(k_E,\lambda)},$$
$$\sum_{\boldsymbol{k}} = \int_0^{\text{cutoff}} \frac{\mathcal{V}k^2 dk}{(2\pi)^3} \int d\lambda, \tag{D32}$$

with k_E being the inverse of the isotropic band dispersion E_k. $\langle\langle \Xi(\boldsymbol{k})\rangle\rangle_{kE}$ is referred to as the *isoenergy average* of $\Xi(\boldsymbol{k})$.

Quasiparticle distribution. (i)–(iii) imply that the nonequilibrium quasiparticle distribution within the active region is singularly peaked at periodic intervals [27,82]:

$$f_{\boldsymbol{k}} = \sum_{j=0}^{j_{\max}} \tilde{f}_{\boldsymbol{k},j}\delta_{\boldsymbol{k},E_j}, \quad E_j = E_{\text{exc}} - j\hbar\Omega_o. \tag{D33}$$

$f_{\boldsymbol{k}}$ thus has a ladderlike structure, with the top rung of the ladder corresponding to the excitation energy ($E_0 = E_{\text{exc}}$), and the lowest rung $E_{j_{\max}}$ lying just above the passive region. The singular nature of $f_{\boldsymbol{k}}$ originates from the source being monochromatic and the predominant phonons being dispersionless. One may verify that the regular function f_E in Eq. (E9) becomes proportional to a Dirac δ function as $\eta_E \to 0$.

Spontaneous emission rate. A related implication of (i)–(iii) is that the spontaneous emission rate is a sum of terms that connect adjacent rungs of the ladder:

$$\mathcal{E}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'} = \sum_{j=0}^{j_{\max}-1} \frac{\tilde{\mathcal{E}}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}}{||\boldsymbol{k}-\boldsymbol{k}'||^s}\delta_{\boldsymbol{k},E_{j+1}}\delta_{\boldsymbol{k}',E_j}. \tag{D34}$$

We have extracted $1/||\boldsymbol{k}-\boldsymbol{k}'||^s$ and the singular δ functions such that $\tilde{\mathcal{E}}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}$ is regular as $\boldsymbol{k}$ approaches $\boldsymbol{k}'$. To derive the surface projector $\delta_{\boldsymbol{k}',E_j}$ in Eq. (D34), apply that $\mathcal{E}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}$ [Eq. (A29)] is proportional to the singular distribution $f_{\boldsymbol{k}'}$ [Eq. (D33)]; the second surface projector $\delta_{\boldsymbol{k},E_{j+1}}$ in Eq. (D34) originates from energy conservation: $E_{\boldsymbol{k}'\boldsymbol{k}} = \hbar\Omega_o$ [Eq. (A29)]. The ladder structure in Eq. (D34) implies that the operator $\sum_{j=0}^{j_{\max}-1}\delta_{\boldsymbol{k}',E_j}$ acts trivially on the emission rate:

$$\sum_{j=0}^{j_{\max}-1} \delta_{\boldsymbol{k}',E_j}\mathcal{E}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'} = \mathcal{E}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}, \tag{D35}$$

due to the idempotence of surface projectors; cf. Eq. (D30).

[71] If the reader is bothered by $(\delta_{\boldsymbol{k},E})^2$ being a product of two Dirac δ functions, one may regularize the surface projector as

$$\delta^{\mathcal{V}}_{\boldsymbol{k},E} = \begin{cases} 1 & |E_{\boldsymbol{k}} - E| < 1/2\mathcal{V}g_E, \\ 0 & \text{otherwise}, \end{cases} \tag{D31}$$

multiply two regularized projectors, and then subsequently take $\mathcal{V} \to \infty$.

Kinetic equation. The kinetic equation for the steady quasiparticle distribution [Eqs. (B1)–(B7)] simplifies to

$$I_{\text{exc},\boldsymbol{k}'} - \sum_{\boldsymbol{k}} \mathcal{E}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'} + \sum_{\boldsymbol{k}''} \mathcal{E}^{\text{sp}}_{\boldsymbol{k}'\leftarrow\boldsymbol{k}''} = 0 \tag{D36}$$

for $\boldsymbol{k}'$ in the active region [Eq. (A23)]; $I_{\text{exc},\boldsymbol{k}'}$ is given in Eq. (D29) and $\mathcal{E}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}$ in Eq. (D34). We have dropped the recombination component [Eq. (B2)] of the kinetic equation because the loss rate due to spontaneous emission of optical phonons greatly outweighs the loss rate due to interband recombination; cf. the discussion under Eq. (E8).

The last preparation for the BIS-geodesic reduction [Eq. (D26)] will be to relate the excitation and spontaneous emission rates as

$$\tilde{I}_{\text{exc},(k_0,\lambda)} = \tilde{\tilde{\mathcal{E}}}^{\text{sp}}_{(k_{j+1},\lambda)\leftarrow(k_j,\lambda)} \lim_{s\to\infty} \langle 1/q^s\rangle_j^{j+1},$$
$$\langle 1/q^s\rangle_j^{j+1} \equiv \int \frac{d\lambda'}{4\pi} \frac{1}{||\boldsymbol{k}-\boldsymbol{k}'||^s}\Big|_{\mathbf{k}=(k_{j+1},\lambda');\boldsymbol{k}'=(k_j,\lambda)}. \tag{D37}$$

k_j is short for k_{E_j}, meaning it is the radius of the spherical isoenergy surface with energy E_j. The term on the right-hand side of $\tilde{I}_{\text{exc}}$ can be interpreted as the rate at which a quasiparticle on the jth isoenergy surface drops to the $(j+1)$th surface by spontaneously emitting an optical phonon.

Equation (D37) is ultimately a consequence of the conservation of probability flow in energy-momentum space. Proving Eq. (D37) takes three steps: (A) we first relate the excitation rate to the rate of phonon-mediated transitions between the zeroth/excitation surface to the first isoenergy surface. (B) We then relate the rate of phonon-mediated transitions between the $(j-1)$th and jth surfaces to the rate of phonon-mediated transitions between the jth and $(j+1)$th surfaces. (C) Combining our relations from (A) and (B) and taking the small-angle-scattering limit gives us Eq. (D37).

Step (A): Projecting the kinetic equation [Eq. (D36)] onto the excitation surface tells us

$$0 = \delta_{\boldsymbol{k}',E_0}\left(I_{\text{exc},\boldsymbol{k}'} - \sum_{\boldsymbol{k}} \mathcal{E}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'} + \sum_{\boldsymbol{k}''} \mathcal{E}^{sp}_{\boldsymbol{k}'\leftarrow\boldsymbol{k}''}\right)$$
$$= I_{\text{exc},\boldsymbol{k}'} - \delta_{\boldsymbol{k}',E_0}\sum_{\boldsymbol{k}} \mathcal{E}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}, \tag{D38}$$

with the last term dropping out because there are no quasiparticles with energies exceeding E_{exc} that can drop to the excitation surface by emitting a phonon; cf. Eq. (D34). Let us substitute the ladder formula for the emission rate [Eq. (D34)] into Eq. (D38) and apply the orthogonality of surface projectors [Eq. (D30)] to reduce $\sum_j$ to the $j=0$ term. We then convert $\sum_{\boldsymbol{k}}\delta_{\boldsymbol{k}E_1}$ to a solid-angular integral via Eq. (D32) to obtain

$$\delta_{\boldsymbol{k}',E_0}\tilde{I}_{\text{exc},\boldsymbol{k}'} = \delta_{\boldsymbol{k}',E_0}\int \frac{d\lambda}{4\pi} \frac{\tilde{\tilde{\mathcal{E}}}^{\text{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}}{||\boldsymbol{k}-\boldsymbol{k}'||^s}\Big|_{\mathbf{k}=(k_1,\lambda)}. \tag{D39}$$

Step (B): If we project the kinetic equation [Eq. (D36)] to the jth isoenergy surface with $j \neq 0$ and $\neq j_{\max}$, then it is the

excitation term that drops out:

$$\delta_{\boldsymbol{k}',E_j}\left(\sum_{\boldsymbol{k}''}\mathcal{E}^{\mathrm{sp}}_{\boldsymbol{k}'\leftarrow\boldsymbol{k}''}-\sum_{\boldsymbol{k}}\mathcal{E}^{\mathrm{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}\right)=0. \tag{D40}$$

Like how we derived the right-hand side of Eq. (D39), Eq. (D40) can be massaged to the form

$$\int\frac{d\lambda''}{4\pi}\left.\frac{\tilde{\mathcal{E}}^{\mathrm{sp}}_{\boldsymbol{k}'\leftarrow\boldsymbol{k}''}}{||\boldsymbol{k}'-\boldsymbol{k}''||^s}\right|_{\boldsymbol{k}''=(k_{j-1},\lambda'')}=\int\frac{d\lambda}{4\pi}\left.\frac{\tilde{\mathcal{E}}^{\mathrm{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}}{||\boldsymbol{k}-\boldsymbol{k}'||^s}\right|_{\mathbf{k}=(k_{j+1},\lambda)} \tag{D41}$$

for any $\boldsymbol{k}'$ on the jth isoenergy surface. With $\boldsymbol{k}'$ as a reference point, Eq. (D41) encodes that the incoming probability flow from the $(j-1)$th surface matches the outgoing probability flow to the $(j+1)$th surface.

Step (C): Both Eqs. (D39) and (D41) involve solid-angular integrals that simplify in the small-angle-scattering limit: fixing $\boldsymbol{k}'=(k_j,\lambda')$,

$$\begin{aligned}\lim_{s\to\infty}\int\frac{d\lambda}{4\pi}\left.\frac{\tilde{\mathcal{E}}^{\mathrm{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}}{||\boldsymbol{k}-\boldsymbol{k}'||^s}\right|_{\mathbf{k}=(k_{j+1},\lambda)}&\\ =\tilde{\mathcal{E}}^{\mathrm{sp}}_{(k_{j+1},\lambda')\leftarrow\boldsymbol{k}'}\lim_{s\to\infty}\langle 1/q^s\rangle_j^{j+1},&\end{aligned} \tag{D42}$$

with $\langle 1/q^s\rangle$ defined in Eq. (D37). The crucial step taken here is to replace $\tilde{\tilde{\mathcal{E}}}^{\mathrm{sp}}_{\boldsymbol{k}\leftarrow\boldsymbol{k}'}$ in the integral by its value when $||\boldsymbol{k}-\boldsymbol{k}'||^{-s}$ is maximized, or equivalently when $||\boldsymbol{k}-\boldsymbol{k}'||$ is minimized. This replacement is justified asymptotically as $s\to\infty$, and it may be seen as an application of Laplace's method [98]. To manifest the usual form of the integral seen in Laplace's method, we momentarily adopt spherical-angular coordinates $\lambda=(\cos\theta,\phi)$ such that $\boldsymbol{k}=(k\sin\theta\cos\phi,k\sin\theta\sin\phi,k\cos\theta)$ and $\boldsymbol{k}'=(0,0,k')$; then for any smooth function $f_{\boldsymbol{k}}=f(k,\lambda)$,

$$\begin{aligned}L_s&=\int\frac{d\lambda}{4\pi}\frac{f(k,\lambda)}{||\boldsymbol{k}-\boldsymbol{k}'||^s}\\&=\int_{-1}^{1}e^{sR(x)}\left[\int_0^{2\pi}\frac{f(k,x,\phi)}{4\pi}d\phi\right]dx,\\ R(x)&=-\frac{1}{2}\ln\left[k^2+k'^2-2kk'x\right],\end{aligned} \tag{D43}$$

with $x=\cos\theta$. $R(x)$ has a unique global maximum at $x=1$, which is an end point of the interval of integration. Applying a standard formula from asymptotic analysis [98],

$$\begin{aligned}\lim_{s\to\infty}L_s&=\frac{f(k,1,\phi)}{2}\frac{e^{sR(1)}}{sR'(1)}+O_r(s^{-1})\\&=f_{0,0,k}\frac{1}{2skk'|k-k'|^{s-2}}+O_r(s^{-1}),\end{aligned} \tag{D44}$$

with $R'=dR/dx$. In our application, $f_{0,0,k}$ corresponds to $\tilde{\tilde{\mathcal{E}}}^{\mathrm{sp}}_{(k_{j+1},\lambda')\leftarrow(k_j,\lambda')}$ in Eq. (D42).

Substituting Eq. (D42) into Eq. (D41), we relate the transition rates between two adjacent pairs of isoenergy surfaces as

$$\frac{\tilde{\tilde{\mathcal{E}}}^{\mathrm{sp}}_{(k_j,\lambda)\leftarrow(k_{j-1},\lambda)}}{\tilde{\tilde{\mathcal{E}}}^{\mathrm{sp}}_{(k_{j+1},\lambda)\leftarrow(k_j,\lambda)}}=\lim_{s\to\infty}\frac{\langle 1/q^s\rangle_j^{j+1}}{\langle 1/q^s\rangle_j^{j-1}}. \tag{D45}$$

Combining Eqs. (D39) and (D45), we relate the excitation rate to the transition rate between a pair of isoenergy surfaces:

$$\tilde{I}_{\mathrm{exc},(k_0,\lambda)}=\tilde{\tilde{\mathcal{E}}}^{\mathrm{sp}}_{(k_{j+1},\lambda)\leftarrow(k_j,\lambda)}\lim_{s\to\infty}\langle 1/q^s\rangle_0^1\prod_{i=1}^{j}\frac{\langle 1/q^s\rangle_i^{i+1}}{\langle 1/q^s\rangle_i^{i-1}}. \tag{D46}$$

The solid-angular integral in Eq. (D37) is evaluated to be

$$\begin{aligned}\langle 1/q^s\rangle_j^{j+1}&=\frac{1}{2(s-2)k_jk_{j+1}}\\&\quad\times\left[\frac{1}{|k_j-k_{j+1}|^{s-2}}-\frac{1}{(k_j+k_{j+1})^{s-2}}\right],\end{aligned} \tag{D47}$$

which manifests that $\langle 1/q^s\rangle_j^{j+1}$ is symmetric under interchanging j and $j+1$, hence Eq. (D46) simplifies to Eq. (D37), as desired.

Proof of BIS-geodesic reduction [Eq. (D26)]. Let us begin the proof by demonstrating that the ratio between Eqs. (D27) and (D28) reduces to

$$\frac{\boldsymbol{\sigma}^{\mathrm{intra}}_{\mathrm{geo}}}{\boldsymbol{\sigma}^{\mathrm{intra}}_{\mathrm{BIS}}}=\frac{\left\langle\left\langle\tilde{I}_{\mathrm{exc},\mathbf{k}}\int_{p_k}\boldsymbol{\Omega}\times d\mathbf{k}\right\rangle\right\rangle_{kE_0}}{\sum_{j=0}^{j_{\max}-1}\left\langle\left\langle\sum_{\mathbf{k}}\boldsymbol{S}_{\mathbf{k}\leftarrow\mathbf{k}'}\mathcal{E}^{\mathrm{sp}}_{\mathbf{k}\leftarrow\mathbf{k}'}\right\rangle\right\rangle_{\boldsymbol{k}'E_j}}. \tag{D48}$$

Beginning with the geodesic expression in Eq. (D27), we insert the integral expression of $\langle\cdots\rangle_\omega$ from Eq. (11) and decompose the excitation rate $I_{\mathrm{exc}\boldsymbol{k}}$ according to Eq. (D29):

$$-\frac{\mathcal{V}|\mathcal{E}_\omega|^2}{2_{\uparrow\downarrow}|e|}\boldsymbol{\sigma}^{\mathrm{intra}}_{\mathrm{geo}}=\sum_{\boldsymbol{k}}\delta_{\boldsymbol{k},E_{\mathrm{exc}}}\tilde{I}_{\mathrm{exc}\boldsymbol{k}}\int_{p_k}\boldsymbol{\Omega}\times d\boldsymbol{k}. \tag{D49}$$

The right-hand side of the above equation is simply the numerator of Eq. (D48), per our definition of isoenergy averaging in Eq. (D32). Working now on the BIS formula [Eq. (D28)], we insert the trivial operator $\sum_{j=0}^{j_{\max}-1}\delta_{\boldsymbol{k}',E_j}$ [Eq. (D35)] and apply again the definition of isoaveraging in Eq. (D32):

$$-\frac{\mathcal{V}|\mathcal{E}_\omega|^2}{2_{\uparrow\downarrow}|e|}\boldsymbol{\sigma}^{\mathrm{intra}}_{\mathrm{BIS}}=\sum_{j=0}^{j_{\max}-1}\left\langle\left\langle\sum_{\mathbf{k}}\boldsymbol{S}_{\mathbf{k}\leftarrow\mathbf{k}'}\mathcal{E}^{\mathrm{sp}}_{\mathbf{k}\leftarrow\mathbf{k}'}\right\rangle\right\rangle_{\boldsymbol{k}'E_j}. \tag{D50}$$

Taking the ratio of Eqs. (D49) and (D50) gives Eq. (D48), as desired.

Focusing on a summand of fixed j and taking the small-angle-scattering limit,

$$\begin{aligned}S\mathcal{E}_j&=\lim_{s\to\infty}\left\langle\left\langle\sum_{\mathbf{k}}\boldsymbol{S}_{\mathbf{k}\leftarrow\mathbf{k}'}\mathcal{E}^{\mathrm{sp}}_{\mathbf{k}\leftarrow\mathbf{k}'}\right\rangle\right\rangle_{\boldsymbol{k}'E_j}\\&=\int\frac{d\lambda}{4\pi}\sum_{\boldsymbol{k}}\boldsymbol{S}_{\mathbf{k}\leftarrow(k_j,\lambda)}\lim_{s\to\infty}\mathcal{E}^{\mathrm{sp}}_{\mathbf{k}\leftarrow(k_j,\lambda)}.\end{aligned} \tag{D51}$$

We then apply Laplace's method [Eqs. (D43) and (D44)] and replace $\boldsymbol{S}_{\mathbf{k}\leftarrow\mathbf{k}'}\tilde{\tilde{\mathcal{E}}}^{\mathrm{sp}}_{\mathbf{k}\leftarrow\mathbf{k}'}$ by its value when $||\boldsymbol{k}-\boldsymbol{k}'||$ is minimized:

$$\begin{aligned}S\mathcal{E}_j&=\int\frac{d\lambda}{4\pi}\boldsymbol{S}_{(k_{j+1},\lambda)\leftarrow(k_j,\lambda)}\tilde{\tilde{\mathcal{E}}}^{\mathrm{sp}}_{(k_{j+1},\lambda)\leftarrow(k_j,\lambda)}\lim_{s\to\infty}\langle 1/q^s\rangle_j^{j+1}\\&=\int\frac{d\lambda}{4\pi}\boldsymbol{\Omega}_{\boldsymbol{k}^j_{\mathrm{ave}}}\times\delta\boldsymbol{k}^j\,\tilde{I}_{\mathrm{exc},(k_0,\lambda)},\end{aligned} \tag{D52}$$

with $\boldsymbol{k}^j_{\mathrm{ave}}=(\boldsymbol{k}^j+\boldsymbol{k}'^j)/2$, $\delta\boldsymbol{k}^j=\boldsymbol{k}'^j-\boldsymbol{k}^j$, $\boldsymbol{k}'^j=(k_{j+1},\lambda)$, and $\boldsymbol{k}^j=(k_j,\lambda)$. In the last step, we substituted the spontaneous emission rate with the excitation rate in accordance with

Eq. (D37), and we replaced the phonon-mediated shift vector with its asymptotic small-angle limit [Eq. (5)]. The sum of $S\mathcal{E}_j$ over j may be regarded as a Riemann sum, which approximates a line integral over the geodesic path:

$$\sum_j \boldsymbol{\Omega}_{\boldsymbol{k}^j_{\rm ave}} \times \delta\boldsymbol{k}^j = \int_{p_{\boldsymbol{k}}} \boldsymbol{\Omega} \times d\boldsymbol{k} + O_r\left(\sqrt{\frac{\hbar\Omega_o}{E_{\rm exc}}}, \sqrt{\frac{\hbar\Omega_o}{E_g}}\right). \tag{D53}$$

Indeed, it may be seen that the discrete transitions between isoenergy surfaces,

$$(k_{j_{\rm max}}, \lambda) \leftarrow (k_{j_{\rm max}-1}, \lambda) \leftarrow \cdots \leftarrow (k_1, \lambda) \leftarrow (k_0, \lambda) = \boldsymbol{k}, \tag{D54}$$

concatenate into a straight path $p'_{\boldsymbol{k}}$ of fixed solid-angular orientation; the geodesic path $p_{\boldsymbol{k}}$ similarly connects $\boldsymbol{k}_{\rm ext} \leftarrow \boldsymbol{k}$ in a straight path. One caveat is that $(k_{j_{\rm max}}, \lambda)$ is not $\boldsymbol{k}_{\rm ext}$ and lies just outside the passive $\boldsymbol{k}$-volume [Eq. (A23)], thus $|p'_{\boldsymbol{k}}|$ is shorter than $|p_{\boldsymbol{k}}|$ by about $k_{\hbar\Omega_o}$. In the parabolic-band approximation, $k_{\hbar\Omega_o}/k_{E_{\rm exc}} = (\hbar\Omega_0/E_{\rm exc})^{1/2}$, which is the reason for the relative error in Eq. (D53). This estimate presumes the band gap E_g is comparable to $E_{\rm exc}$. For semiconductors with anomalously small band gaps, the Berry curvature may be concentrated in an energy interval comparable to E_g, hence the relative error is modified to $(\hbar\Omega_0/E_g)^{1/2}$. Altogether,

$$\begin{aligned}\lim_{s\to\infty} \sum_{j=0}^{j_{\rm max}-1} \left\langle\!\!\left\langle \sum_{\mathbf{k}} \boldsymbol{S}_{\mathbf{k}\leftarrow\mathbf{k}'} \mathcal{E}^{sp}_{\mathbf{k}\leftarrow\mathbf{k}'} \right\rangle\!\!\right\rangle_{k'E_j} \\ = \left\langle\!\!\left\langle \tilde{I}_{{\rm exc},\mathbf{k}} \int_{p_{\boldsymbol{k}}} \boldsymbol{\Omega} \times d\mathbf{k} \right\rangle\!\!\right\rangle_{kE_0} \\ + O_r\left(\sqrt{\frac{\hbar\Omega_o}{E_{\rm exc}}}, \sqrt{\frac{\hbar\Omega_o}{E_g}}\right),\end{aligned} \tag{D55}$$

which combines with Eq. (D48) to give the BIS-geodesic reduction [Eq. (D26)].

For finite $s = 2$, which is appropriate to polarization scattering, one should expect the ratio $\boldsymbol{\sigma}^{\rm intra}_{\rm geo}/\boldsymbol{\sigma}^{\rm intra}_{\rm BIS}$ to deviate from unity. In practice, we find this deviation to be small: for the model calculation in Sec. III, the ratio turns out to be 1.08 for a source photon energy of $\hbar\Omega = 0.8E_0$; cf. Fig. 5(e).

6. Geodesic approximation for quasi-2D semiconductors

Having formulated the geodesic approximation for 3D direct-gap semiconductors, we would like to extend the notion to quasi-2D direct-gap semiconductors, as exemplified by the model Hamiltonian in Eq. (16).

By quasi-2D, we mean that the electronic band energies $E_{b\boldsymbol{k}}$ and cell-periodic wave functions $|u_{b\boldsymbol{k}}\rangle_{\rm cell}$ are approximately independent of one wave-vector coordinate, say, k_z. The former condition implies that the band gap is minimized not at a single $\boldsymbol{k}$-point but along a $\boldsymbol{k}$-line. In our model [Eq. (16)], this $\boldsymbol{k}$-line is parametrized by $\boldsymbol{k} = (0, 0, k_z)$, as illustrated by the purple line in Fig. 14. The latter condition on the wave function implies that the intraband Berry curvature vector is collinear with the z unit directional vector: $\boldsymbol{\Omega}_{b\boldsymbol{k}} = \Omega^z_{b\boldsymbol{k}}\vec{z}$, and that the shift current vanishes in the z direction. Indeed, a nonzero z-component of the photonic/phononic shift vector requires that $|u_{b\boldsymbol{k}}\rangle_{\rm cell}$ depends nontrivially on k_z, as deducible from Eqs. (A2)–(A4) and Eq. (C17).

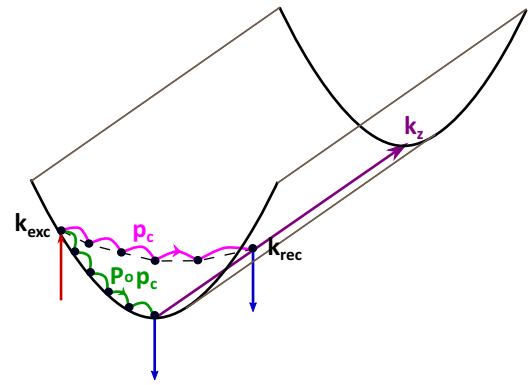

FIG. 14. Quasi-two-dimensional conduction band plotted over (E, k_x, k_z), with E parametrizing an implicit vertical axis.

Let us then consider the shift current $\boldsymbol{j}_\perp$ orthogonal to $\vec{z}$. We would like to demonstrate that $\boldsymbol{j}_\perp$ is well approximated by Eq. (D25), with loop[$\boldsymbol{k}$] reinterpreted as a *planar geodesic loop*: a geodesic loop confined to the k_x-k_y plane that contains $\boldsymbol{k}$. Precisely, we mean that all nodes in loop[$\boldsymbol{k}$] have identical values for k_z, and $\boldsymbol{k}$ is connected by a geodesic path to the extremal wave vector that lies closest to $\boldsymbol{k}$, as illustrated by the green trajectory in Fig. 14. A consequence of loop[$\boldsymbol{k}$] being planar is that the affinity shift loop in Eq. (D25) simplifies to a planar integral:

$$\begin{aligned}&\langle f_{vc\boldsymbol{k}} |\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2 \boldsymbol{S}_{{\rm loop}[\boldsymbol{k}]}\rangle_\omega \\ &\quad = \int \frac{dk_x dk_y}{(2\pi)^2 a_z} \frac{\delta(E_{cv\boldsymbol{k}} - \hbar\omega)}{{\rm JDOS}_\uparrow} f_{vc\boldsymbol{k}} |\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2 \boldsymbol{S}_{{\rm loop}[\boldsymbol{k}]}.\end{aligned} \tag{D56}$$

We have introduced a lattice constant a_z such that $\int dk_z = 2\pi/a_z$. The z component of $\boldsymbol{k}$ in the integrand can be arbitrarily chosen, and the integrand only depends on band energies and wave functions within the arbitrarily chosen k_x-k_y plane. This justifies our use of the planar model Hamiltonian in Eq. (16), which explicitly depends on k_x and k_y but not k_z.

To recapitulate, being quasi-two-dimensional allows us to simplify the loop analysis to planar loops, as if the problem were strictly two-dimensional. Such a simplification is not *a priori* obvious, since a hot photoexcited electron with initial wave vector $\boldsymbol{k}_{\rm exc}$ (on the excitation surface) may relax to any point along the conduction-band minimum, including points that differ from $\boldsymbol{k}_{\rm exc}$ in the z component (cf. the pink trajectory in Fig. 14). If subgraph[$\boldsymbol{k}_{\rm exc}$], the probability-flow subgraph that includes the link $(c\boldsymbol{k}_{\rm exc} > v\boldsymbol{k}_{\rm exc})$ (cf. Fig. 13), is decomposed into loops, one expects to find loops that are extended in the k_z direction.

Let us denote the pink-colored trajectory by p_c; the lower-energy boundary point of p_c corresponds to the wave vector $\boldsymbol{k}_{\rm rec}$ of recombination; it is assumed that the electron traces a path p_v from Bloch label $(v, \boldsymbol{k}_{\rm exc})$ to $(v, \boldsymbol{k}_{\rm rec})$, which is not illustrated in Fig. 14. Altogether, p_c, p_v and the vertical links at $\boldsymbol{k}_{\rm exc}$ and $\boldsymbol{k}_{\rm rec}$ combine to form loop$(\boldsymbol{k}_{\rm exc}, \boldsymbol{k}_{\rm rec})$; its associated shift loop $\boldsymbol{S}_{{\rm loop}(\boldsymbol{k}_{\rm exc}, \boldsymbol{k}_{\rm rec})}$ is defined through Eq. (D14). We define $P \circ {\rm loop}(\boldsymbol{k}_{\rm exc}, \boldsymbol{k}_{\rm rec})$ as the projection of loop$(\boldsymbol{k}_{\rm exc}, \boldsymbol{k}_{\rm rec})$ onto the k_x-k_y plane containing $\boldsymbol{k}_{\rm exc}$. If $k^z_{\rm exc} = 0$, then this projection

amounts to setting $k_z = 0$ for all nodes along the loop, such that the pink trajectory collapses to the green trajectory in Fig. 14.

We would prove that the shift loop is invariant under such a projection: $\boldsymbol{S}_{\mathrm{loop}(\boldsymbol{k}_{\mathrm{exc}},\boldsymbol{k}_{\mathrm{rec}})} = \boldsymbol{S}_{P\circ\mathrm{loop}(\boldsymbol{k}_{\mathrm{exc}},\boldsymbol{k}_{\mathrm{rec}})}$. Points (i)–(iii) in Appendix D 4 can be used to show that $\boldsymbol{S}_{\mathrm{loop}(\boldsymbol{k}_{\mathrm{exc}},\boldsymbol{k}_{\mathrm{rec}})}$ has the same form as the right-hand side of Eq. (6), with $\boldsymbol{S}_{\mathrm{rec}}$ defined as in Eq. (8) but with $\boldsymbol{k}_{\mathrm{ext}}$ replaced by $\boldsymbol{k}_{\mathrm{rec}}$. Because the photonic shift vector [Eq. (A4)] and the interband Berry connection are purely a function of $|u_{b\boldsymbol{k}}\rangle$, which is k_z-independent, the two photonic terms in Eq. (6) are invariant under changing the z component of $\boldsymbol{k}_{\mathrm{rec}}$. What remains is to demonstrate a similar invariance for the anomalous component of the shift loop, which is given by a sum of the two line integrals in Eq. (6). p_c differs from $P \circ p_c$ only in that the $\boldsymbol{k}$-path of integration is extended in the k_z direction (Fig. 14). Since $\boldsymbol{\Omega}_{c\boldsymbol{k}} = \Omega^z_{c\boldsymbol{k}}\hat{z}$ is k_z-independent, it follows that $\boldsymbol{\Omega}_{c\boldsymbol{k}} \times d\boldsymbol{k} = \boldsymbol{\Omega}_{c,P\circ\boldsymbol{k}} \times P \circ d\boldsymbol{k}$, meaning $\int_{p_c} \boldsymbol{\Omega}_{c\boldsymbol{k}} \times d\boldsymbol{k}$ is invariant under projecting $p_c \to P \circ p_c$. The same argument and conclusion hold for $v \to c$. This completes our proof of invariance for the shift loop.

It is a straightforward generalization to demonstrate that the invariance property $\boldsymbol{S}_{\mathrm{loop}} = \boldsymbol{S}_{P\circ\mathrm{loop}}$ holds for any loop, not just the simple loop we considered above. Thus for the purpose of evaluating the loop current contribution by subgraph[$\boldsymbol{k}_{\mathrm{exc}}$] [Eq. (D22)], one may as well project the entire subgraph to the k_x-k_y plane containing $\boldsymbol{k}_{\mathrm{exc}}$.

At this point, one may apply essentially the same arguments that led to approximating Eq. (D22) by Eq. (D25), with the only modification being that all loops are now planar, and in particular, loop($\boldsymbol{k}$) in Eqs. (D23) and (D24) is a planar geodesic loop. This completes the proof of Eq. (D56).

APPENDIX E: THE TRANSIENT CURRENT APPROXIMATES THE EXCITATION-INDUCED CURRENT

Appendix B establishes concepts and notations that are prerequisite to understanding this Appendix.

We focus on the photoexcited carrier density regime: $n \lesssim n_h$, where energy relaxation in the active region is dominated by optical phonons. Assuming that the excitation energy [Eq. (A21)] of photoexcited carriers lies in the active region, we would demonstrate that the transient current $\boldsymbol{j}_{\mathrm{tran}}$ is well approximated by the excitation-induced component $\boldsymbol{j}_{\mathrm{exc}}$ [cf. Eq. (A25)] of the steady current.

Before tackling the transient and nonequilibrium currents, let us take a step back to consider an equilibrated mix of electrons, photons, and phonons in the absence of the light source. The quasiparticle occupancy then follows the Fermi-Dirac distribution: $f^{T_0}_{b\boldsymbol{k}}$ [Eq. (A8)], while the occupancy of photons and phonons follows the Planck distribution: $N^{T_0}_m$ [Eq. (A9)] with the same equilibrium temperature. The shift current, viewed as a functional of the quasiparticle, photon, and phonon occupancies [Eq. (A7)], vanishes:

$$\text{Equilibrium:} \quad \boldsymbol{j}\big[f^{T_0}_B, N^{T_0;\mathrm{phot}}_m, N^{T_0;\mathrm{phon}}_m\big] = 0, \tag{E1}$$

due to detailed balance; cf. Eq. (A15).

At the onset of turning on a light source (with frequency ω_s, mode m_s, polarization $\boldsymbol{\epsilon}_s$), the quasiparticles and phonons retain their equilibrium distributions, but the photon occupancy is modified to $N^{T_0;\mathrm{phot}}_m + \Delta N_s \delta_{m,m_s}$. We define the transient current as the current at the onset of radiation:

$$\text{Onset:} \quad \boldsymbol{j}_{\mathrm{tran}} = \boldsymbol{j}\big[f^{T_0}_B, N^{T_0;\mathrm{phot}}_m + \Delta N_s \delta_{m,m_s}, N^{T_0;\mathrm{phon}}_m\big]. \tag{E2}$$

All bosonic modes with a thermal occupancy cannot contribute to the shift current, due to detailed balance; cf. Eq. (A15). For the source mode m_s, the net transition rate [Eq. (A5)] can be decomposed just as in Eq. (C103), with T replaced by T_0. Because the first term on the right-hand side of Eq. (C103) vanishes by detailed balance [Eq. (A15)], the transient current is simply proportional to the source-generated photon occupancy:

$$\boldsymbol{j}_{\mathrm{tran}} = -\frac{|e|}{\mathcal{V}} \sum_{\boldsymbol{k}} \boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{C\leftarrow V} \frac{(2\pi e)^2 \omega_s}{\mathcal{V}} |\boldsymbol{\epsilon}_s \cdot \boldsymbol{A}_{cv\boldsymbol{k}}|^2 \times \delta(E_{cv\boldsymbol{k}} - \hbar\omega) f^{T_0}_{cv\boldsymbol{k}} \Delta N_s. \tag{E3}$$

The formula here assumes a two-band semiconducting model (Appendix A 3), but more generally one would just sum over contributions from all resonant interband transitions. It should be borne in mind that $f^{T_0}_C$ is exponentially suppressed with exponent $E_g/k_B T_e \gg 1$ for an intrinsic semiconductor:

$$f^{T_0}_C \ll 1 \quad \text{and} \quad 1 - f^{T_0}_V \ll 1. \tag{E4}$$

Equation (E3) manifests that $\boldsymbol{j}_{\mathrm{tran}} \neq 0$ must originate solely from the disruption of detailed balance between pairs of Bloch states that are resonantly coupled by the light source, i.e., pairs labeled $(c\boldsymbol{k})$ and $(v\boldsymbol{k})$, with $\boldsymbol{k}$ on the excitation surface ES; cf. Eq. (A20). It follows that in evaluating $\boldsymbol{j}_{\mathrm{tran}} = \boldsymbol{j}[f^{T_0}, \ldots]$, one may as well restrict the wave-vector summations $\sum_{\boldsymbol{k}\boldsymbol{k}'}$ in Eq. (A1) with the condition $\boldsymbol{k} = \boldsymbol{k}' \in \mathrm{ES}$:

$$\boldsymbol{j}_{\mathrm{tran}} = \boldsymbol{j}\big[f^{T_0}_B, N^{T_0;\mathrm{phot}}_m + \Delta N_s \delta_{m,m_s}, N^{T_0;\mathrm{phon}}_m\big]_{\boldsymbol{k}=\boldsymbol{k}'\in\mathrm{ES}}. \tag{E5}$$

As derived in Appendix A 4, the excitation-induced component of the steady shift current [cf. Eq. (A25)] differs from Eq. (E3) only in that $f^{T_0}_{cv\boldsymbol{k}}$ is replaced by the nonequilibrium $f_{cv\boldsymbol{k}} = f_C - f_V$. If one accepts that the nonequilibrium quasiparticle distribution over the excitation surface satisfies

$$\forall \boldsymbol{k} \in \mathrm{ES}: \quad f_C \ll 1 \quad \text{and} \quad 1 - f_V \ll 1, \tag{E6}$$

then $f_{cv\boldsymbol{k}} \approx f^{T_0}_{cv\boldsymbol{k}}$ (on the excitation surface), and therefore the excitation-induced current approximates the transient current:

$$\boldsymbol{j}_{\mathrm{tran}} = \boldsymbol{j}[f^{T_0}]|_{\boldsymbol{k}=\boldsymbol{k}'\in\mathrm{ES}} \approx \boldsymbol{j}[f]|_{\boldsymbol{k}=\boldsymbol{k}'\in\mathrm{ES}} = \boldsymbol{j}_{\mathrm{exc}}, \tag{E7}$$

in accordance with Eqs. (E3)–(E5).

For $n \ll n_h$, we believe the inequalities in Eq. (E6) hold generally, due to an argument presented in the main text and reproduced here: the smallness of f_C and $(1 - f_V)$ originates from the slowness in optical excitations compared to the fastness of energy relaxation by carrier-carrier and carrier-phonon scatterings.

We will flesh out this argument by deriving an explicit expression of f_C for the kinetic model set up in Appendix B 2. This model encodes certain assumptions that caricature reality, as detailed in Appendixes B 2 a and B 2 b. Thus our explicit expression for f_C should be understood as an order-of-magnitude estimate for more realistic distributions; this is fine because the advertised inequality [Eq. (E6)] is a statement about orders of magnitude.

With this caveat in mind, let us reproduce from Eq. (B16) the kinetic equation for the isoenergy-averaged quasiparticle distribution f_E in the conduction band:

$$E > \hbar\Omega_o: \quad G_\uparrow \delta(E - E_{c,\mathrm{exc}}) - \frac{g_E f_E}{\tau_E^o} + \frac{g_{E_+} f_{E_+}}{\tau_{E_+}^o} + \partial_E \left[\frac{g_E E}{\tau_E^s} (1 + k_B T_e \partial_E) f_E \right] = 0,$$
$$E_+ = E + \hbar\Omega_o. \tag{E8}$$

We assume the reader has read the discussion leading to Eq. (B16), and we will not repeat the definitions and descriptions of each term in the kinetic equation. However, we will mention two slight differences between the above equation and Eq. (B16):

(i) We have dropped the interband recombination term $(-g_E f_E/\tau_{\mathrm{rec}})$ that was present in Eq. (B16). This is alright for $E > \hbar\Omega_o$ (the active region), because electron–optical-phonon scattering results in a substantially larger loss rate: $-g_E f_E/\tau_E^o$, given that $\tau_{\mathrm{rec}} \sim 1$ ns and $\tau_E^o \sim 100$ fs [30–32].

(ii) The diffusive Fokker-Planck term in Eq. (E8) carries a more general meaning than the corresponding term in Eq. (B16):

(ii-a) For $n \ll n_l$, the diffusive term encodes electron–acoustic-phonon scattering, and $\tau^s \equiv \tau_A$ is the energy relaxation time due to spontaneous emission of acoustic phonons; a typical value is $\tau_A \sim 1$ ns [27,28].

(ii-b) For $n_h \gg n \gg n_l$, the diffusive term encodes electron-electron scattering, and $\tau^s = \tau_{ee}$ is the time taken for a hot "test electron" (with initial energy $\gg k_B T_e$) to cool down to an energy comparable to $k_B T_e$ [27]. By assumption for this density regime, electron-electron scattering is more efficient in relaxing an electron's energy than electron–acoustic-phonon scattering, meaning $\tau^{ee} \ll \tau^A \sim 1$ ns. It is also possible for electron-electron collisions to establish an electron temperature T_e that exceeds the lattice temperature T_l [27].

The solution to the differential equation [Eq. (E8)] has been derived in Refs. [82] and [27]. Here, we extract a few salient facts from these references that help to prove Eq. (E6): In the absence of the secondary scattering process ($\tau_E^s \to \infty$), the distribution is a sum of Dirac-δ functions centered at $E_k := E_{c;\mathrm{exc}} - k\hbar\Omega_o$. The effects of the secondary scatterers are that each peak shifts as $E_k \to E_k - \eta_{E_k}(k+1)\hbar\Omega_o$, as well as broadens to a regular function. Assuming $\hbar\Omega_o/k_B T_e \gtrsim 1$ and $k \sim 1$, the width of each peak remains small compared to $\hbar\Omega_o$. The highest peak ($k = 0$) has the functional form

$$f_E = \frac{G_\uparrow \tau^o}{g} \frac{\varrho - 1}{4 k_B T_e \sqrt{\varrho}} \exp[-x - \sqrt{\varrho}|x|]|_{x=(E-E_{c;\mathrm{exc}})/2k_B T_e},$$
$$\varrho = 1 + 4\frac{k_B T_e}{\hbar\Omega_o}\eta^{-1}, \tag{E9}$$

with g, τ^o, and η evaluated at $E_{c;\mathrm{exc}}$.[72] In particular,

$$f_{E_{c;\mathrm{exc}}} = \frac{G_\uparrow \tau^o}{g\hbar\Omega_o \eta\sqrt{\varrho}} \overset{\eta\hbar\Omega_o/k_B T_e \ll 1}{\approx} \frac{1}{2}\frac{G_\uparrow}{g}\sqrt{\frac{\tau^s \tau^o}{E_{c;\mathrm{exc}} k_B T_e}}. \tag{E10}$$

Let us estimate $G_\uparrow/g$ under realistic experimental conditions. Recalling $G_\uparrow = \alpha_\uparrow \mathcal{I}_{\mathrm{rad}}/\hbar\omega$ from Eq. (B17), and assuming typical values for the lattice period $a \sim 5$ Å, density of states $g \sim eV/a^3$, absorption coefficient $\alpha_\uparrow \sim 10^3$ cm^{-1} [99], and continuous-wave laser intensity $\mathcal{I}_{\mathrm{rad}} \sim 40$ W cm^{-2} [69], one finds a modest value for $G_\uparrow/g \sim 10$ eV/s.

$f_{E_{c;\mathrm{exc}}}$ is the product of $G_\uparrow/g$ with a quantity that has dimensions of time over energy. This quantity encodes the microscopic energy relaxation processes, which occur at much shorter timescales than 1 s: as a reminder, $\tau^o \sim 100$ fs and $\tau^s \lesssim 1$ ns [27,28,31,32]. Thus, $f_{E_{c;\mathrm{exc}}} \ll 1$ even at the low temperature of $T_e \sim 1$ K. Given that $f_C \ll 1$ for $\boldsymbol{k} \in$ ES, $1 - f_V = f_C \ll 1$ immediately follows from the electron-hole symmetry of our model; cf. Appendix B 2 a. This completes our demonstration of Eq. (E6).

APPENDIX F: MODEL CALCULATIONS WITH OPTICAL VORTICES

This Appendix details the model calculations that support certain claims stated in Sec. IV, which we reproduce here for easy reference:

(I) $\boldsymbol{\sigma}_{\vec{x},\omega}$ is dominated by the recombination-induced current.

(II) $\boldsymbol{\sigma}_{\vec{y},\omega}$ is dominated by the excitation-induced and intraband currents.

(III) The signs of $\boldsymbol{\sigma}^y_{\vec{x},\omega}$ and $\boldsymbol{\sigma}^y_{\vec{y},\omega}$ differ over a broad range of frequencies.

(IV) The linear disparity in the conductivity is large: $|\boldsymbol{\sigma}^y_{\vec{x},\omega} - \boldsymbol{\sigma}^y_{\vec{y},\omega}| \sim$ mA V^{-2}.

(V) The current response to unpolarized light is given by $|\boldsymbol{\sigma}^y_{\vec{x},\omega} + \boldsymbol{\sigma}^y_{\vec{y},\omega}|/2 \sim 0.1$ mA V^{-2}.

Some aspects of the following demonstration will be a more quantitative elaboration of qualitative arguments made in Sec. IV.

The form of our model Hamiltonian is identical to the one studied in the context of the anomalous shift; cf. Eq. (16). Having studied the case of $\tilde{Q} = 1$, we now tune $\tilde{Q}$ from positive to negative values. The conduction and valence bands touch (at $\boldsymbol{k} = \boldsymbol{0}$) when $\tilde{Q} = 0$ and subsequently untouch for negative $\tilde{Q}$. This untouching is accompanied by the nucleation of two time-reversal-related $\vec{x}$-vortex lines at $(\tilde{k}_x, \tilde{k}_y) \approx (\pm\sqrt{-\tilde{Q}/(1-\tilde{Q}/2)}, 0)$, as illustrated in Fig. 15(a); there are no $\vec{y}$-vortices in this model [Fig. 15(b)]. Henceforth, we fix $\tilde{Q} = -1$.

First, let us consider the case of an $\vec{x}$-polarized light source and make the case that the excitation-induced current is outweighed by the recombination-induced current: $||\boldsymbol{j}_{\mathrm{exc}}[\vec{x}]|| \ll ||\boldsymbol{j}_{\mathrm{rec}}[\vec{x}]||$, due to the vortex-induced orientational disorder of the photonic shift vector field. This inequality simplifies to $||\boldsymbol{\sigma}^{\mathrm{exc},y}_{\vec{x},\omega}|| \ll ||\boldsymbol{\sigma}^{\mathrm{rec},y}_{\vec{x},\omega}||$ for the y-component of the shift conductivity [Eqs. (6)–(14)], because a mirror symmetry ($x \to -x$) of the model Hamiltonian[73] constrains the x component of the shift current to vanish, while the z component vanishes due to the quasi-two-dimensionality of the model; cf. Appendix D 6. Because the shift conductivity is essentially the product of the

[72] The solution presented in Ref. [27] is missing a factor of $1/k_B T_e$, which we presume is a minor typographical oversight.

[73] $M_x H(\boldsymbol{k}) M_x^{-1} = H(-k_x, k_y)$ with $M_x = \boldsymbol{\sigma}_3$.

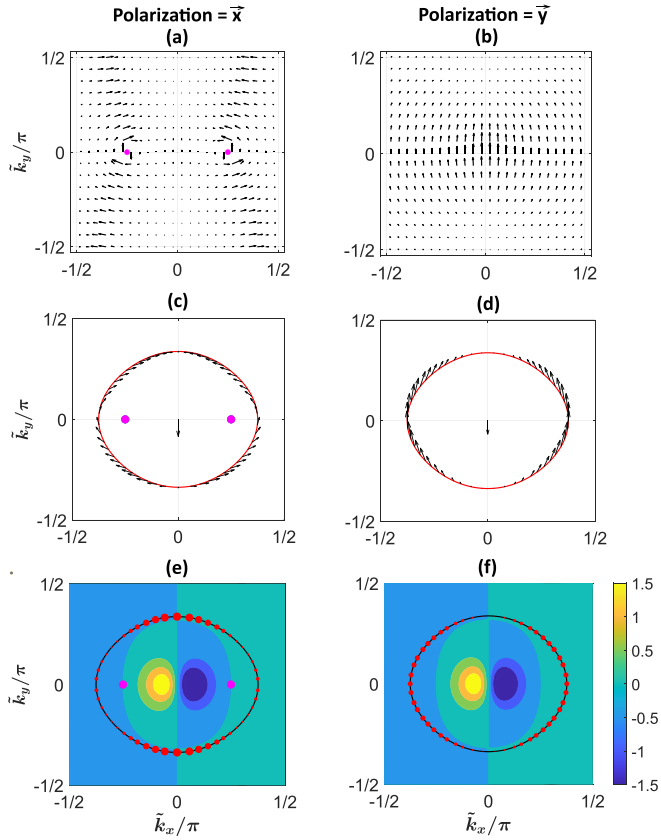

FIG. 15. Characterization of the model Hamiltonian in Eq. (16) with $\tilde{Q} = -1$. Panels (a), (c), and (e) are characterizations for a light source with linear polarization vector $\boldsymbol{\epsilon}_s = \vec{x}$, and (b), (d), and (f) for $\boldsymbol{\epsilon}_s = \vec{y}$. The pink dots in (a), (c), and (e) represent the $\boldsymbol{k}$-locations of $\vec{x}$-vortices. Panels (a) and (b) depict the photonic shift vector field $\boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$, with $\boldsymbol{\epsilon}_s = \vec{x}$ and $\vec{y}$, respectively. For panels (c) and (d), the red ellipse represents the excitation surface for a photon frequency $\hbar\omega = 4.5E_0$; arrows on the ellipse represent the vectors $|\boldsymbol{\epsilon}_s \cdot \boldsymbol{A}_{cv\boldsymbol{k}}|^2 \boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$ for $\boldsymbol{k}$ on the excitation surface; the central arrow represents the recombination component of the affinity shift loop: $\mathrm{ASL}^{\mathrm{rec}}_{\boldsymbol{\epsilon},4.5E_0}$; cf. Eq. (F1). For panels (e) and (f), the size of the red dots indicates the magnitude of $|\boldsymbol{\epsilon}_s \cdot \boldsymbol{A}_{cv\boldsymbol{k}}|^2$ for $\boldsymbol{k}$ on the excitation surface; the colored background represents the Berry curvature scalar field $\Omega^z_{c\boldsymbol{k}}$ in units of $a^2 = (\mathcal{V}_{\mathrm{cell}})^{2/3}$, with a color legend on the right.

joint density of states ($\mathrm{JDOS}_\uparrow$) [cf. Eq. (12)] with the affinity shift loop [cf. Eqs. (10)–(13)], one may as well compare the excitation and recombination components of the affinity shift loop:

$$\mathrm{ASL}^{\mathrm{exc}}_{\boldsymbol{\epsilon},\omega} \equiv \langle |\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2 S^{\boldsymbol{\epsilon}}_{y,c\boldsymbol{k}\leftarrow v\boldsymbol{k}}\rangle_\omega \quad \text{vs}$$
$$\mathrm{ASL}^{\mathrm{rec}}_{\boldsymbol{\epsilon},\omega} \equiv \langle |\boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2\rangle_\omega S_{y,\mathrm{rec}} \qquad \text{(F1)}$$

for $\boldsymbol{\epsilon} = \vec{x}$, and $\langle\cdots\rangle_\omega$ denoting an average over the excitation surface; cf. Eq. (11). S_y means the y component of $\boldsymbol{S}$, and $S^{\boldsymbol{\epsilon}}_{y,c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$ is the photonic shift vector defined in Eq. (A4). The recombination shift $\boldsymbol{S}_{\mathrm{rec}}$ is defined in Eq. (8) but simplifies in the present context to $\boldsymbol{S}^{\vec{x}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}}$, due to a mirror-symmetry-imposed dipole selection rule.[74] A numerical calculation of Eq. (F1) reveals for a wide range of photon frequencies that $\mathrm{ASL}^{\mathrm{exc}}_{\vec{x}}$ and $\mathrm{ASL}^{\mathrm{rec}}_{\vec{x}}$ have opposite signs, and that $|\mathrm{ASL}^{\mathrm{exc}}_{\vec{x}}| \ll |\mathrm{ASL}^{\mathrm{rec}}_{\vec{x}}|$ by a multiplicative factor ranging from 1/5 to 1/8, as illustrated in Fig. 16(a).

To rationalize this multiplicative factor, we illustrate $|A^x_{cv\boldsymbol{k}}|^2 \boldsymbol{S}^{\vec{x}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$ as arrows in Fig. 15(c) for $\boldsymbol{k}$ along a representative excitation surface encircling the $\vec{x}$-vortices. The central arrow in Fig. 15(c) represents $\langle |A^x_{cv\boldsymbol{k}}|^2\rangle_\omega \boldsymbol{S}^{\vec{x}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}}$. All arrows are drawn with a common scale to allow for mutual comparison. It is evident that proximity to the $\vec{x}$-vortex causes the direction of $|A^x_{cv\boldsymbol{k}}|^2 \boldsymbol{S}^{\vec{x}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$ to rotate along the excitation surface; the average of $|A^x_{cv\boldsymbol{k}}|^2 \boldsymbol{S}^{\vec{x}}_{c\boldsymbol{k}\leftarrow v\boldsymbol{k}}$ over the excitation surface is therefore diminished; this average just equals $\mathrm{ASL}^{\mathrm{exc}}_{\boldsymbol{\epsilon},\omega}$; cf. Eq. (F1). In contrast, recombination occurs in the vicinity of the extremal wave vector $\boldsymbol{k}_{\mathrm{ext}}$, where the photonic shift vector is roughly constant. Thus follows a general principle: *ceteris paribus*, the orientational disorder induced by $\vec{x}$-vorticity reduces $\boldsymbol{j}_{\mathrm{exc}}[\vec{x}]$ relative to $\boldsymbol{j}_{\mathrm{rec}}[\vec{x}]$, for an $\vec{x}$-polarized source.

Ceteris paribus, the same orientational disorder reduces $\boldsymbol{j}_{\mathrm{exc}}[\vec{x}]$ relative to $\boldsymbol{j}_{\mathrm{exc}}[\vec{y}]$, for reasons explained in Sec. IV. This implies a linear disparity of the excitation-induced current $\boldsymbol{j}_{\mathrm{exc}}$, which applies to a broad range of photon frequencies; compare red curves of Figs. 16(a) and 16(b).

To understand the linear disparity of the intraband current $\boldsymbol{j}_{\mathrm{intra}}$, we have indicated the $\boldsymbol{k}$-dependent magnitude of $|A^x_{cv}|^2$ ($|A^y_{cv}|^2$) by the size of dots imprinted over the excitation surface in Fig. 15(e) [Fig. 15(f)]; in both figures, the same Berry curvature scalar field ($\Omega^z_{c\boldsymbol{k}}$) is represented by a color plot. It may be seen that $|A^x_{cv}|^2$ and $|A^y_{cv}|^2$ are both anisotropic over the excitation surface, but each favors a different segment of the excitation surface for reasons explained in Sec. IV. We deduce for the $\vec{y}$-polarized source that the predominant relaxation pathways are roughly parallel to k_x [cf. Fig. 7(f)] and intersect the Berry-curvature hot spots, leading to a larger anomalous shift than the case of the $\vec{x}$-polarized source. Once again, this effect is not limited to a fine-tuned photon frequency; compare green curves of Figs. 16(a) and 16(b).

Altogether, the linear disparity of $\boldsymbol{j}_{\mathrm{exc}}$ and $\boldsymbol{j}_{\mathrm{intra}}$ results in the net shift current being dominated by $\boldsymbol{j}_{\mathrm{exc}} + \boldsymbol{j}_{\mathrm{intra}}$ for a $\vec{y}$-polarized source [cf. the black curve in Fig. 16(b) and claim (II)], and by $\boldsymbol{j}_{\mathrm{rec}}$ for a $\vec{x}$-polarized source [black curve in Fig. 16(a) and claim (I)]; the net current changes sign if the polarization is flipped [claim (III)]. The linear disparity of the affinity shift loop [i.e., the difference of the two black curves in Fig. 16(a) vs 16(b)] is comparable to -1 (in units of $\mathcal{V}_{\mathrm{cell}}$, the real-space volume of the primitive unit cell) over a broad range of frequencies; this corresponds to a linear disparity of the conductivity: $\boldsymbol{\sigma}^y_{\vec{x},\omega} - \boldsymbol{\sigma}^y_{\vec{y},\omega} \approx 2\ \mathrm{mA\,V^{-2}}$ [cf. Eq. (13), Fig. 2(d), and claim (IV)], assuming a generic value for $\mathrm{JDOS}_\uparrow \approx (\mathcal{V}_{\mathrm{cell}} eV)^{-1}$.[75]

[74] Conduction- and valence-band states with $k_x = 0$ transform under different representations of mirror symmetry M_x, hence $A^y_{cv\boldsymbol{k}_{\mathrm{ext}}} = A^z_{cv\boldsymbol{k}_{\mathrm{ext}}} = 0$. This implies for any $\boldsymbol{\epsilon}$ that is not orthogonal to $\vec{x}$ that $\nabla_{\boldsymbol{k}} \arg \boldsymbol{\epsilon}\cdot\boldsymbol{A}_{cv\boldsymbol{k}_{\mathrm{ext}}} = \nabla_{\boldsymbol{k}} \arg A^x_{cv\boldsymbol{k}_{\mathrm{ext}}}$ and $\boldsymbol{S}^{\boldsymbol{\epsilon}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}} = \boldsymbol{S}^{\vec{x}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}}$.

[75] We choose $\tilde{Q} = -1$ and $\tilde{P} = 12$ such that $\mathrm{JDOS}_\uparrow \approx (\mathcal{V}_{\mathrm{cell}} eV)^{-1}$.

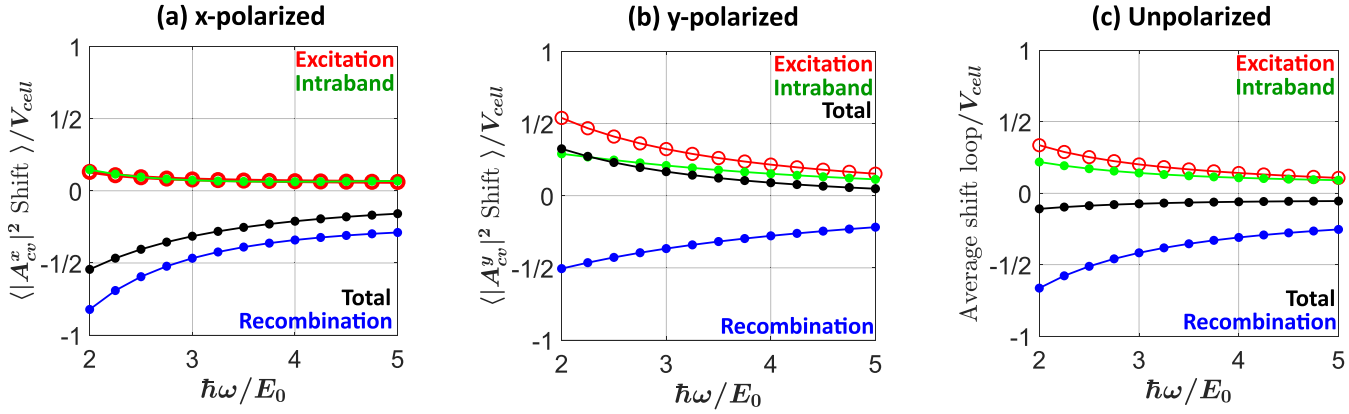

FIG. 16. The black curves in panels (a) and (b) represent the affinity shift loop $\langle|\boldsymbol{\epsilon}_s \cdot \boldsymbol{A}_{cv}|^2 \boldsymbol{S}^{\boldsymbol{\epsilon}_s}_{\text{y,loop}}\rangle_\omega$ (in units of $\mathcal{V}_{\text{cell}}$) vs the photon frequency ω, for polarization $\boldsymbol{\epsilon}_s = \vec{x}$ and $\vec{y}$, respectively. Nonblack curves represent the three components of the affinity shift loop: excitation (red curve), intraband relaxation (green curve), and recombination (blue curve). Panel (c) averages the affinity shift loop over two orthogonal light polarizations.

The response to an unpolarized light source is given by $(\boldsymbol{\sigma}^y_{\vec{x},\omega} + \boldsymbol{\sigma}^y_{\vec{y},\omega})/2$, which $\approx$0.2 mA V^{-2} over a broad range of frequencies [cf. the black curve in Fig. 16(c) and claim (V)].

We end this Appendix with a caveat: the calculated values of $\boldsymbol{\sigma}$ should be taken with a grain of salt. A reliable calculation of $\boldsymbol{\sigma}$ should also account for the dependence of Bloch wave functions over continuous space [48], but such a dependence is discarded when the Hilbert space is reduced to a two-dimensional vector space at each $\boldsymbol{k}$ point, as was done for all model Hamiltonians in this work. A more realistic model would incorporate *ab initio*–derived wave functions as additional model parameters [11,100]. Reassuringly, our qualitative arguments for vortex-induced shifts do not rely on the two-band approximation and are equally applicable to realistic, continuous-space Hamiltonians.

APPENDIX G: CHERN-VORTICITY THEOREM

The Chern-vorticity theorem in Eq. (18) relates the Chern numbers (C_v, C_c) of the valence and conduction states (over any closed 2D $\boldsymbol{k}$-manifold $\boldsymbol{\Sigma}$) to the net optical vorticity (Vort) within $\boldsymbol{\Sigma}$.

To prove the theorem, we first recall that if C_v $(C_v) \neq 0$, the wave function cannot be made continuous and periodic over $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{A}_{vvk}$ $(\boldsymbol{A}_{cck})$ must be singular somewhere on $\boldsymbol{\Sigma}$. To be concrete, supposing $\boldsymbol{\Sigma}$ were a two-torus; Fig. 17 illustrates how $\boldsymbol{\Sigma}$ is decomposed into two patches, such that the wave function in the interior of each patch is analytic in $\boldsymbol{k}$, but $\boldsymbol{A}_{vvk}$ $(\boldsymbol{A}_{cck})$ is singular at the patch boundary:

$$\lim_{\delta\to 0} \frac{1}{2\pi}\left(\int_{\mathcal{L}_1} + \int_{\mathcal{L}_2}\right) \boldsymbol{A}_{bbk} \cdot d\boldsymbol{k} = C_b \quad \text{with } b = v, c. \tag{G1}$$

Here, C_b is the Chern number of the band labeled by b, and δ is an infinitesimal parameter illustrated in Fig. 17.

Performing the same line integral with the Berry connection replaced by the photonic shift vector gives zero for any linear polarization vector $\boldsymbol{\epsilon}$:

$$\lim_{\delta\to 0} \frac{1}{2\pi}\left(\int_{\mathcal{L}_1} + \int_{\mathcal{L}_2}\right) \boldsymbol{S}^{\boldsymbol{\epsilon}}_{ck\leftarrow vk} \cdot d\boldsymbol{k} = \boldsymbol{0}, \tag{G2}$$

because the shift vector is gauge-invariant and smoothly defined except at optical vortices, and one can always choose the patch boundary to avoid those vortex points.

Comparing the last two equations with the definition of the photonic shift vector in Eq. (3), one infers that there must be nonzero circulations in $\arg[\boldsymbol{\epsilon} \cdot \boldsymbol{A}_{cvk}]$ to compensate for the singularity of $\boldsymbol{A}_{vvk}$ $(\boldsymbol{A}_{cck})$. Specifically,

$$\lim_{\delta\to 0} \frac{1}{2\pi}\left(\int_{\mathcal{L}_1} + \int_{\mathcal{L}_2}\right) \nabla_{\boldsymbol{k}} \arg[\boldsymbol{\epsilon} \cdot \boldsymbol{A}_{cvk}] \cdot d\boldsymbol{k} = C_c - C_v, \tag{G3}$$

which means that the net phase vorticity of $\boldsymbol{\epsilon} \cdot \boldsymbol{A}_{cvk}$ over $\boldsymbol{\Sigma}$ is $C_c - C_v$. The line integral over $\mathcal{L}_1$ $(\mathcal{L}_2)$ equals the winding number of $\arg[\boldsymbol{\epsilon} \cdot \boldsymbol{A}_{cvk}]$ in patch 1 (patch 2), and is thus topologically invariant upon contracting the $\mathcal{L}_1$ $(\mathcal{L}_2)$ to infinitesimally encircle any optical vortex in patch 1 (patch 2); this contraction is illustrated in Fig. 17. This invariance implies that Eq. (G3) is equivalent to Eq. (18).

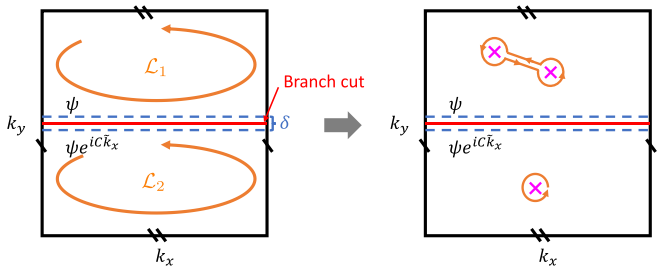

FIG. 17. Illustration of the two-patch decomposition of $\boldsymbol{\Sigma}$, as well as the paths for the line integrals in Eqs. (G1) and (G3). The magenta crosses represent the optical vortices.

APPENDIX H: SUPPORTING OUR CASE STUDY OF BiTeI

Our case study of BiTeI is based on a four-band Hamiltonian $H_{\text{BiTeI}}(\boldsymbol{k})$ [cf. Eq. (19)] with energies ordered as $E_1 \leqslant E_2 < E_3 \leqslant E_4$. We focus on photon frequencies that resonantly excite quasiparticles from the highest-energy valence band to the lowest-energy conduction band: $b' = 3$ and $b = 2$. Minimizing $E_{3\boldsymbol{k}}$ with respect to $\boldsymbol{k}$ defines a circular ring contained in the zero-k_z plane; maximizing $E_{2\boldsymbol{k}}$ with respect to

$\boldsymbol{k}$ also defines a circular ring contained in the zero-k_z plane; actually, the two rings coincide, as suggested pictorially in Fig. 9. This coincidence may be rationalized: the O(2) symmetry about the z axis, combined with time-reversal symmetry (T), imply the existence of a $C_{2z}T$ symmetry (twofold rotation composed with time reversal) which maps $\boldsymbol{k} \to (k_x, k_y, -k_z)$; thus if $E_{3\boldsymbol{k}}$ is minimized on a single O(2)-symmetric ring, this ring must lie on the $C_{2z}T$-symmetric plane with $k_z = 0$. Within this plane, the Hamiltonian has a chiral symmetry that relates positive to negative energies:

$$\tau_2\sigma_3 H_{\mathrm{BiTeI}}(k_x, k_y, 0)\tau_2\sigma_3 = -H_{\mathrm{BiTeI}}(k_x, k_y, 0), \tag{H1}$$

which implies that $E_{2k_xk_y0}$ is maximized wherever $E_{3k_xk_y0}$ is minimized. We will refer to this ring as the *band-edge ring*.

1. Effective description by a massive Dirac fermion

Near the topological phase transition between a trivial insulator and a $\mathbb{Z}_2$ topological insulator, the two bands that touch are effectively described by a massive Dirac fermion in two momentum dimensions. Here, we provide a detailed derivation of the massive Dirac Hamiltonian in the $k_{x^+} - k_z$ half-plane [cf. Eq. (21)].

We start by restricting H_{BiTeI} [Eq. (19)] to the $k_{x^+} - k_z$ half-plane and Taylor-expanding the Hamiltonian around $\mathbf{k}_0 = (\lambda, 0, 0)$:

$$\begin{aligned} H_{\mathrm{BiTeI}} &= H_0 + H_1; \\ H_0 &= \hbar v\lambda(\tau_1\sigma_3 - \tau_2\sigma_1), \\ H_1 &= (m' - 2A\lambda q_x)\tau_3\sigma_0 - \hbar v(q_x\tau_2\sigma_1 + q_z\tau_2\sigma_3) + O(q^2). \end{aligned} \tag{H2}$$

We have introduced wave numbers q_x and q_z, which are the deviations from $\mathbf{k}_0$ in the $k_{x^+} - k_z$ half-plane; $\tau_i\sigma_j$ is the Kronecker product of τ_i and σ_j, i.e., $\tau_i \otimes \sigma_j$ [101]. $\mathbf{k}_0$ is a point where the bands touch during the topological phase transition; by construction, H_1 vanishes at $\boldsymbol{k}_0$ when $m' = 0$, and the touching bands correspond to the two zero-energy eigenstates of H_0, which we label as $|1\rangle = (i, 1, 0, 0)^T/\sqrt{2}$ and $|2\rangle = (0, 0, -i, 1)^T/\sqrt{2}$.

In the low-energy subspace spanned by $|1\rangle$ and $|2\rangle$, the effective Hamiltonian is given by degenerate perturbation theory as

$$H'_{i,j} = \langle i|H_1|j\rangle, \quad i, j = 1, 2. \tag{H3}$$

Given that

$$\begin{aligned} \langle i|\tau_3\sigma_0|j\rangle &= (\gamma_3)_{ij}, \\ \langle i|\tau_2\sigma_1|j\rangle &= -(\gamma_1)_{ij}, \\ \langle i|\tau_2\sigma_3|j\rangle &= -(\gamma_2)_{ij}\ i, j = 1, 2, \end{aligned} \tag{H4}$$

with $\gamma_{1,2,3}$ being Pauli matrices of the Hilbert space spanned by $|1\rangle$ and $|2\rangle$, Eq. (H4) directly gives Eq. (21) in the main text.

2. Vanishing shift at the band edge, for x- and y-polarized light

This section aims to explain why $\vec{z}\cdot\boldsymbol{j}_{\mathrm{exc}}$ dominates over $\vec{z}\cdot\boldsymbol{j}_{\mathrm{rec}}$ in the low-frequency regime of Fig. 8(c). This reduces to explaining the smallness of the recombination shift vector $\vec{z}\cdot\boldsymbol{S}_{\mathrm{rec}}$ [Eq. (8)] relative to the excitation shift vector [Eq. (7)], according to the average-shift-loop formula in Eqs. (6)–(13). Given that $\vec{z}\cdot\boldsymbol{S}_{\mathrm{rec}}$ is an affinity-weighted average of $\vec{z}\cdot\boldsymbol{S}^{\boldsymbol{\epsilon}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}}$ over all polarization vectors of the spontaneously emitted photon [Eq. (8)], it may be argued that $\vec{z}\cdot\boldsymbol{S}_{\mathrm{rec}}$ is small because of the vanishing of the band-edge shift vectors,

$$\text{On the band-edge ring:}\quad \vec{z}\cdot\boldsymbol{S}^{\vec{x}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}} = \vec{z}\cdot\boldsymbol{S}^{\vec{y}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}} = 0. \tag{H5}$$

This is equivalent to the vanishing of the band-edge shift connections,

$$\text{On the band-edge ring:}\quad \vec{z}\cdot\boldsymbol{S}^{\vec{x}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}}\left|\vec{x}\cdot\boldsymbol{A}_{cv\boldsymbol{k}_{\mathrm{ext}}}\right|^2 = \vec{z}\cdot\boldsymbol{S}^{\vec{y}}_{v\boldsymbol{k}_{\mathrm{ext}}\leftarrow c\boldsymbol{k}_{\mathrm{ext}}}\left|\vec{y}\cdot\boldsymbol{A}_{cv\boldsymbol{k}_{\mathrm{ext}}}\right|^2 = 0, \tag{H6}$$

because the optical affinity is nonvanishing throughout the band-edge ring; after all, there are no optical vortex loops intersecting the band-edge ring, as illustrated in Fig. 9.

For any tight-binding Hamiltonian $H(\boldsymbol{k})$, the photonic shift connection can be expressed as [44]

$$\vec{z}\cdot\boldsymbol{S}^{\vec{x}}_{b'\leftarrow b}|\vec{x}\cdot\boldsymbol{A}_{b'b}|^2 = \mathrm{Im}\left\{\frac{\overline{v^x_{bb'}}}{(\omega_{bb'})^2}\left[\langle u_{b\boldsymbol{k}}|\partial_z\partial_x H|u_{b'\boldsymbol{k}}\rangle - \frac{v^z_{bb'}\Delta^x_{bb'} + v^x_{bb'}\Delta^z_{bb'}}{\omega_{bb'}} + \sum_{b''\neq b,b'}\left(\frac{v^z_{bb''}v^x_{b''b'}}{\omega_{bb''}} - \frac{v^x_{bb''}v^z_{b''b'}}{\omega_{b''b'}}\right)\right]\right\}. \tag{H7}$$

Here, $\partial_z \equiv \partial_{k_z}$, $v^z_{bb'} = \langle u_{b\boldsymbol{k}}|\frac{1}{\hbar}\partial_z H|u_{b'\boldsymbol{k}}\rangle$, $\Delta^z_{bb'} = \partial_z E_b - \partial_z E_{b'}$, $\omega_{bb'} = (E_b - E_{b'})/\hbar$, and $\sum_{b''\neq b,b'}$ means to sum over all band indices b'' that are neither b nor b'.

Let us show that Eq. (H7) vanishes for $H = H_{\mathrm{BiTeI}}$ throughout the band-edge ring:

(i) The first term in the square brackets of Eq. (H7) vanishes, because H_{BiTeI} depends quadratically on $\boldsymbol{k}$ as $k_x^2 + k_y^2 + k_z^2$.

(ii) The second term in the square brackets vanishes, because band energy functions are extremized at the band edge: $\partial_z E_3 = \partial_z E_2 = 0$.

(iii) The third term also vanishes, but the argument is longer: first, observe from Eq. (19) that $\partial_z H_{\mathrm{BiTeI}}|_{k_z=0} = \tau_2\sigma_3$ is simply the chirality operator in Eq. (H1), meaning that $\partial_z H_{\mathrm{BiTeI}}$ maps between energy eigenstate with inverted energies. This implies $\hbar v^z_{bb''} = \langle u_{b\boldsymbol{k}}|\partial_z H_{\mathrm{BiTeI}}|u_{b''\boldsymbol{k}}\rangle|_{k_z=0}$ is only nonzero if $E_{b''} = -E_b$, but $E_{b''} = -E_b$ cannot be satisfied because of the constraint $\sum_{b''\neq b,b'}$ in Eq. (H7). A similar argument proves that $v^z_{b''b'} = 0$, hence altogether the third term in the square bracket vanishes.

The above demonstration holds if one replaces $x \to y$, meaning that the z-component of the shift connection also

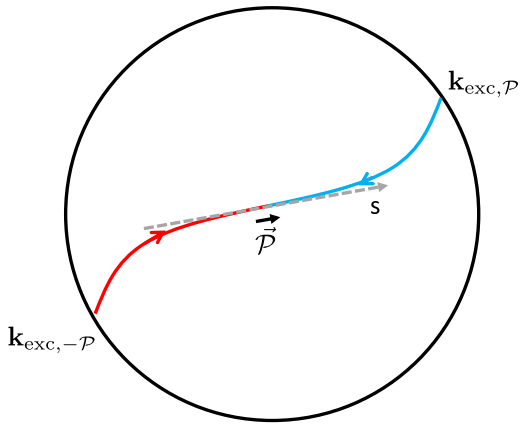

FIG. 18. Representative pair of diametrically opposite geodesic paths, with $-\mathcal{P}$ in red and $\mathcal{P}$ in blue. The precise meaning of "diametrically opposite" is that both paths approach the band-extremal wave vector ($\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}$) along the same tangent line, which we illustrate as a gray dashed line with tangent vector $\vec{\mathcal{P}}$.

vanishes for $\vec{y}$-polarized light, throughout the band-edge ring. This completes the proof of Eq. (H6).

3. Asymptotic behavior of anomalous-shift integrals

One result in Sec. V was the $(1/E_g)$-divergence of the intraband shift conductivity across the topological phase transition, with $|E_g|$ the band gap and $\mathrm{sgn}(E_g) = -1$ on the trivial side of the transition. This divergence relied on an inequality between two anomalous-shift integrals:

$$\left|\left(\int_{-\mathcal{P}} + \int_{\mathcal{P}}\right)\vec{z}\cdot\boldsymbol{\Omega}_{c\boldsymbol{k}}\times d\boldsymbol{k}\right| \ll \left|\left(\int_{-\mathcal{P}} - \int_{\mathcal{P}}\right)\vec{z}\cdot\boldsymbol{\Omega}_{c\boldsymbol{k}}\times d\boldsymbol{k}\right|, \tag{H8}$$

which is asymptotically valid as $|E_g|$ approaches zero; $\pm\mathcal{P}$ are any pair of diametrically opposite geodesic paths, as representatively illustrated in Fig. 18.

We have demonstrated in Sec. V that the right-hand side of Eq. (H8) diverges as $1/E_g$, thus to prove Eq. (H8) it suffices to show that the magnitude of the left integral is decreasing as $|E_g| \to 0$. This is the main result of this subsection.

To begin, consider the Berry curvature within one cross-section of the torus enclosed by the excitation surface, as exemplified by the k_x^+-k_x half-plane. As $|E_g| \to 0$, $\vec{z}\cdot\boldsymbol{\Omega}_{c\boldsymbol{k}}\times d\boldsymbol{k}/|d\boldsymbol{k}| = \Omega^y_{c\boldsymbol{k}}$ becomes localized to a "hot spot" centered at the band-extremal wave vector $\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}$, with a spot width comparable to $|m'|/\hbar v \propto |E_g|$. Indeed, writing the massive Dirac Hamiltonian as a dot product of three-vectors:

$$H' = \boldsymbol{d}\cdot\boldsymbol{\gamma}, \quad \boldsymbol{d} = (d_1,d_2,d_3) = (\hbar v q_x, \hbar v q_z, m' - 2A\lambda q_x),$$
$$\boldsymbol{\gamma} = (\gamma_1,\gamma_2,\gamma_3), \tag{H9}$$

the conduction-band Berry curvature can be expressed as

$$\Omega^h_{c\boldsymbol{k}} = -\frac{1}{4d^3}\epsilon_{hij}\boldsymbol{d}\cdot\nabla_{k_i}\boldsymbol{d}\times\nabla_{k_j}\boldsymbol{d} \Rightarrow \Omega^y_{c\boldsymbol{k}} = \frac{1}{2}\frac{m'}{D^{3/2}},$$
$$D = (d_1)^2 + (d_2)^1 + (d_3)^2, \quad d = \sqrt{D}. \tag{H10}$$

The $\boldsymbol{k}$-location of the Berry-curvature maximum can be identified by

$$0 = \partial_s\Omega^y_{c\boldsymbol{k}} = \frac{3}{4}m'\frac{\partial_s D}{D^{5/2}}, \tag{H11}$$

with $(\partial_s)^m$ being the mth-order derivative in the direction that is tangential to $\mathcal{P}$ at the band extremum. The Berry-curvature maximum (of the hot spot) coincides (in $\boldsymbol{k}$-location) with the band-extremal wave vector $\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}$; this is because the energy spectrum of H' has an $E \to -E$ symmetry at each $\boldsymbol{k}$, which implies that extremizing the $\boldsymbol{k}$-dependent energy gap ($E_{c\boldsymbol{k}} - E_{v\boldsymbol{k}}$) is equivalent to extremizing the conduction-band energy:

$$E_{c\boldsymbol{k}} = \sqrt{D_{\boldsymbol{k}}} \Rightarrow 0 = \partial_s E_c|_{\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}} = \left.\frac{\partial_s D}{\sqrt{D}}\right|_{\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}}. \tag{H12}$$

The band gap is defined as the extremal value of ($E_{c\boldsymbol{k}} - E_{v\boldsymbol{k}}$):

$$|E_g| = 2d|_{\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}} = \frac{m'}{\sqrt{1+u^2}}, \quad u = \frac{2A\lambda}{\hbar v}, \tag{H13}$$

and the extremal value of the Berry curvature can be expressed in terms of the signed band gap ($E_g = \mathrm{sgn}[m']|E_g|$) as

$$\Omega^y_{c\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}} = \left.\frac{1}{2}\frac{m'}{D^{3/2}}\right|_{\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}} = 4\sqrt{1+u^2}\frac{E_g}{|E_g|^3}. \tag{H14}$$

That the width of the Berry-curvature hot spot is of order $|m'|/\hbar v$ can be deduced from dimensional analysis of Eq. (H10), assuming that $|A\lambda|$ is less than or comparable to $|\hbar v|$.

The localization of Berry curvature in momentum space allows us to express the anomalous-shift integral as

$$\left(\int_{-\mathcal{P}} + \int_{\mathcal{P}}\right)\vec{z}\cdot\boldsymbol{\Omega}_{c\boldsymbol{k}}\times d\boldsymbol{k} = \int_{-\Lambda}^{\Lambda}\Omega_s^{(0)}\Theta_s ds + \text{correction}, \tag{H15}$$

with a cutoff $\Lambda > 0$ for the integration variable s along the gray dashed tangent line in Fig. 18; we have introduced $\Omega_s^{(0)} \equiv \Omega^y_{c\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}+s\vec{\mathcal{P}}}$, with $\vec{\mathcal{P}}$ being the unit-norm vector parallel to the tangent line. Θ_s is the symmetric step function that equals -1 for positive s, and $+1$ for negative s. This step function arises because $\pm\mathcal{P}$ are oriented paths beginning on diametrically opposed points on the excitation surface and ending at the same point: $\boldsymbol{k}_{\mathrm{ext},\mathcal{P}}$. For fixed Λ, it is evident that the magnitude of the correction in Eq. (H15) decreases as $E_g \to 0$, due to the increasing localization of the hot spot (which has a width $|m'|/\hbar v \propto |E_g|$).

To prove the main result of this section, what remains is to show that $\int_{-\Lambda}^{\Lambda}\Omega_s^{(0)}\Theta_s ds$ is also decreasing; actually, we will prove a stronger statement that this integral just vanishes. Indeed, so long as the massive-Dirac Hamiltonian has an energy gap (i.e., $m' \neq 0$), $\boldsymbol{\Omega}_{c\boldsymbol{k}}$ is an analytic function of $\boldsymbol{k}$ and hence $\Omega_s^{(0)}$ is an analytic function of s, meaning it admits a convergent Taylor expansion $\Omega_s^{(0)} = \sum_{n\in\mathbb{N}}\Omega_0^{(n)}s^n/n!$, with the nth-order derivative $\Omega_0^{(n)}$ to be $(\partial_s)^n\Omega_s^{(0)}$ evaluated at the band extremum ($s = 0$). Only the odd-order derivatives contribute to Eq. (H15), due to the symmetric step function Θ_s

being an odd function of s, hence

$$\left(\int_{-\mathcal{P}}+\int_{\mathcal{P}}\right)\vec{z}\cdot\boldsymbol{\Omega}_{c\boldsymbol{k}}\times d\boldsymbol{k}\approx 2\sum_{n\in 2\mathbb{N}+1}\frac{\Omega_0^{(n)}}{n!}\int_0^{\Lambda}s^n ds. \tag{H16}$$

The following discussion proves that $\Omega_0^{(n)}$ vanishes for any odd n. It follows from a dimensional analysis of Eq. (H10) that the nth-order derivative can be expanded as

$$\Omega_s^{(n)}=\frac{1}{D^{3/2+n}}\sum_{m_1=0}^{2}\cdots\sum_{m_n=0}^{2}c_{m_1\cdots m_n}\left(\partial_{s_1}\right)^{m_1}\cdots \left(\partial_{s_n}\right)^{m_n}D_{s_1}\cdots D_{s_n}\Big|_{s_j\to s}^{m_1+\cdots m_n=n}, \tag{H17}$$

with linear coefficients $c_{m_1\cdots m_n}$ that depend on m_j. The meaning of the subscript $s_j\to s$ is that after performing all the differentiations $[(\partial_{s_1})^{m_1}\cdots(\partial_{s_n})^{m_n}]$, the resultant function of $(s_1,\ldots,s_n)$ is to be replaced by a function of $(s,\ldots,s)$. The summations over m_j are restricted such that $m_1+m_2+\cdots+m_n=n$, and each $\sum_{m_j=0}^{2}$ is capped at two, because D is a second-order polynomial of momenta variables, which follows from the linearization of the massive-Dirac Hamiltonian. If $\Omega_s^{(n)}$ is evaluated at the band extremum ($s=0$), then one can further drop $m_j=1$ (in each of the summations over m_j) because $\partial_s D|_{s=0}=0$ [cf. Eq. (H12)]. It becomes apparent that the condition $m_1+\cdots m_n=n$ cannot be satisfied for odd n and $m_j\in\{0,2\}$, implying that $\Omega_0^{(n)}=0$ for odd n. This completes the proof.

APPENDIX I: ALTERNATIVE DERIVATION OF THE ANOMALOUS SHIFT

We provide an alternative derivation of the anomalous shift vector [Eq. (5)] that aims to demystify the appearance of the Berry curvature. Beginning from an expression for the intraband phononic shift that was derived in Sec. III and is valid for small momentum transfer:

$$\text{Phonon:}\quad \boldsymbol{S}^{m}_{\boldsymbol{k}'\leftarrow\boldsymbol{k}}\approx-\nabla_{\boldsymbol{k}_{\text{ave}}}\left(\boldsymbol{A}_{\boldsymbol{k}_{\text{ave}}}\cdot\delta\boldsymbol{k}\right)+\boldsymbol{A}_{\boldsymbol{k}'}-\boldsymbol{A}_{\boldsymbol{k}}, \tag{I1}$$

with $\boldsymbol{A}_{\boldsymbol{k}}$ the intraband Berry connection, $\boldsymbol{k}_{\text{ave}}=(\boldsymbol{k}+\boldsymbol{k}')/2$, and $\delta\boldsymbol{k}=\boldsymbol{k}'-\boldsymbol{k}$. We have omitted the band index to simplify notation. We consider the y-component of the above shift vector, and we express the derivative as the limiting value of a difference:

$$\vec{y}\cdot\nabla_{\boldsymbol{k}_{\text{ave}}}\left(\boldsymbol{A}_{\boldsymbol{k}_{\text{ave}}}\cdot\delta\boldsymbol{k}\right)=\lim_{\epsilon\to 0}\frac{\left(\boldsymbol{A}_{\boldsymbol{k}_{\text{ave}}+\epsilon\vec{y}/2}-\boldsymbol{A}_{\boldsymbol{k}_{\text{ave}}-\epsilon\vec{y}/2}\right)\cdot\delta\boldsymbol{k}}{\epsilon}, \tag{I2}$$

such that the shift vector component can be expressed as a line integral of the connection:

$$\left(\boldsymbol{S}^{m}_{\boldsymbol{k}'\leftarrow\boldsymbol{k}}\right)^{y}\approx\lim_{\epsilon\to 0}\frac{1}{\epsilon}\oint\boldsymbol{A}_{\boldsymbol{k}}\cdot d\boldsymbol{k}\approx\left(\boldsymbol{\Omega}_{\boldsymbol{k}_{\text{ave}}}\times\delta\boldsymbol{k}\right)^{y} \tag{I3}$$

along an infinitesimally thin parallelogram drawn in Fig. 19. Finally, one converts the line integral to an area integral of the curvature by Stokes' theorem. This proof is easily generalized for the x and z components.

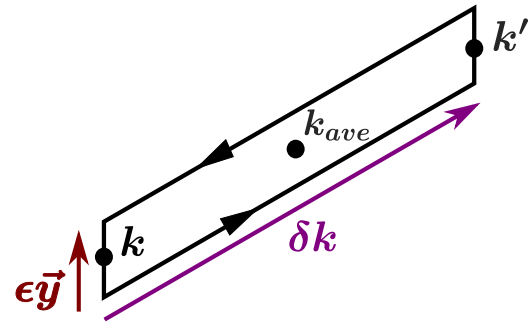

FIG. 19. Infinitesimally thin parallelogram.

APPENDIX J: DIFFICULTIES OF THE PARALLEL-TRANSPORT GAUGE

It has been claimed in the literature that $\langle u_{\boldsymbol{k}p}|u_{\boldsymbol{k}}\rangle_{\text{cell}}=1+O(\delta k^2)$ can be chosen as a gauge choice for the wave function [36,102]. This gauge corresponds to a parallel transport condition [$\delta\boldsymbol{k}\cdot\boldsymbol{A}_{\boldsymbol{k}\boldsymbol{k}}=\boldsymbol{0}$] in the direction of $\delta\boldsymbol{k}=\boldsymbol{k}'-\boldsymbol{k}$ [103,104]. It is not uncommon to find textbooks that ignore the wave-function dependence of the electron-phonon scattering rate [31]. All Berry-curvature effects (including the anomalous shift [Eq. (5)]) are missed if one blithely adopts the parallel-transport gauge.

It is therefore of interest to expose the fallacies inherent in $\langle u_{\boldsymbol{k}'}|u_{\boldsymbol{k}}\rangle_{\text{cell}}=1+O(\delta k^2)$,[76] of which there are two related kinds:

(i) For fixed $\delta\boldsymbol{k}$, it is generically impossible to set $\delta\boldsymbol{k}\cdot\boldsymbol{A}_{\boldsymbol{k}\boldsymbol{k}}=\boldsymbol{0}$ for all $\boldsymbol{k}$ in the Brillouin zone; this is tantamount to assuming that the single-band Berry phase vanishes for all momentum loops parallel to $\delta\boldsymbol{k}$. This assumption may hold if $\delta\boldsymbol{k}$ is orthogonal to a mirror plane, in which case the Berry phase of a single spinless band (in the absence of spin-orbit coupling) is indeed quantized to 0 or π, but one cannot rule out the case of π *a priori*.

(ii) For fixed $\delta\boldsymbol{k}$, it *is* possible to impose $\delta\boldsymbol{k}\cdot\boldsymbol{A}_{\boldsymbol{k}\boldsymbol{k}}=\boldsymbol{0}$ for all $\boldsymbol{k}$ in a ball-shaped subregion of the Brillouin zone. However, it is generically impossible to simultaneously impose $\delta\boldsymbol{k}'\cdot\boldsymbol{A}_{\boldsymbol{k}\boldsymbol{k}}=0$ within the same ball for $\delta\boldsymbol{k}'$ that is not collinear with $\delta\boldsymbol{k}$. The simultaneous imposition is equivalent to assuming a vanishing Berry phase for an infinitesimal loop encircling $\boldsymbol{k}$, i.e., that the Berry curvature $\delta\boldsymbol{k}\times\delta\boldsymbol{k}'\cdot\boldsymbol{\Omega}_b$ vanishes at $\boldsymbol{k}$. Certainly, one must allow for phonons of all possible wave vectors ($\delta\boldsymbol{k},\delta\boldsymbol{k}'\in\text{BZ}$) to completely describe the electron-phonon interaction. Without fine-tuning, the Berry curvature $\boldsymbol{\Omega}_b$ vanishes at a generic $\boldsymbol{k}$-point only in PT-symmetric materials with negligible spin-orbit coupling [105]. PT is certainly not a symmetry in the present case study of noncentric (meaning no P), nonmagnetic (meaning T-symmetric) materials.

APPENDIX K: ENERGY CONVERSION EFFICIENCY

We will derive an ideal expression for the energy conversion efficiency for Pusch *et al.*'s model [70] of a shift-current

[76] There is, of course, no controversy in the claim that $|\langle u_{\boldsymbol{k}'}|u_{\boldsymbol{k}}\rangle_{\text{cell}}|^2=1+O(\delta k^2)$; cf. Eq. (B8).

photovoltaic cell. Our derivation closely follows that in Sec. II of Ref. [70], which we recommend as prerequisite reading. However, our final expression for the efficiency [Eq. (K9)] is less heuristic than Eq. (11) of Ref. [70] in that ours is wholly expressed in terms of kinetic and band-structure parameters, which can be extracted from *ab initio* calculations.

We adopt the same device geometry that is illustrated in Fig. 1 of Ref. [70]: light falls onto a semiconductor facet with illuminated area $A_{\rm illum} = dw$; d is the separation between two electrodes, and w is the width of each electrode. For concreteness, we will fix the facet's normal vector to be parallel to the unit directional vector $\vec{z}$; the photovoltaic current flows between the electrodes in the x direction, and w is the linear dimension of the electrode in the y direction. Assuming that the radiation falls onto the facet with normal incidence, the Poynting vector within the semiconductor decays exponentially as

$$\text{Poynting vector } = \mathcal{I}_{\rm rad}(z)\vec{z}, \quad \mathcal{I}_{\rm rad}(z) = \mathcal{I}_{\rm rad}(0)e^{-\alpha_{\rm abs}z}, \tag{K1}$$

with an attenuation length given by the inverse of the absorption coefficient:

$$\alpha_{\rm abs} = \frac{4\pi^2}{\alpha_{fs}}\frac{\hbar\omega}{n_\omega}\langle f_{vc\boldsymbol{k}}|\boldsymbol{\epsilon}_s\cdot\boldsymbol{A}_{cv\boldsymbol{k}}|^2\rangle_\omega \,\mathrm{JDOS}_{\uparrow\downarrow}. \tag{K2}$$

Here, $\alpha_{fs} \approx 137$ is the fine structure, n_ω is the refractive index, $f_{vc\boldsymbol{k}} = f_V - f_C$ is a difference of the steady quasiparticle distributions, and $\mathrm{JDOS}_{\uparrow\downarrow}$ is the spin-doubled joint density of states. Our semiclassical expression for the absorption coefficient presumes that $\alpha_{\rm abs}^{-1}$ greatly exceeds the lattice period; the same type of semiclassical approximation implies that the shift current density has the same exponential decay due to being proportional to $\mathcal{I}_{\rm rad}(z)$:

$$j_x(z) = -|e|\mathfrak{s}\,\mathrm{Abs}(z), \quad \mathrm{Abs}(z) = \frac{\alpha_{\rm abs}\mathcal{I}_{\rm rad}(z)}{\hbar\omega}. \tag{K3}$$

$\mathfrak{s}$, the average shift per photoexcited electron-hole pair, has been defined in Eq. (27); $\mathrm{Abs}(z)$ is understood as the photon absorption rate per unit volume, at a distance z from the illuminated facet. Equation (K3) is equivalent to $j_x = \sigma_{\boldsymbol{\epsilon}_s,\omega}|\mathcal{E}_\omega|^2$, with $\sigma_{\boldsymbol{\epsilon}_s,\omega}$ the geodesic-approximated shift conductivity in Eq. (13); to derive the equivalence, revert to Gaussian units and replace $|\mathcal{E}_\omega(z)|^2 = 2\pi\mathcal{I}_{\rm rad}(z)/cn_\omega$ (cf. footnote 66 in Appendix C 3) and $e^2/\hbar c \approx 1/137$.

Assuming ideally that the contacts with the electrodes do not introduce additional resistance, the energy conversion efficiency is given by

$$\mathrm{eff} = \frac{1}{4}\frac{V_{\rm oc}I_{\rm sc}}{\mathcal{I}_{\rm rad}A_{\rm illum}}, \tag{K4}$$

with $V_{\rm oc}$ the open-circuit photovoltage and $I_{\rm sc}$ the short-circuit shift current. The latter quantity is obtained by integrating

$$I_{\rm sc} = w\int_0^t j_x(z)dz \approx w\,j_x(0)/\alpha_{\rm abs}, \tag{K5}$$

with t the thickness of the semiconductor in the z direction. In the last step of Eq. (K5), we assumed $t \gg \alpha_{\rm abs}^{-1}$. The open-circuit photovoltage is determined by the condition that the shift and drift currents cancel out at each z:

$$j_x(z) = \sigma_{\rm ph}(z)\frac{V_{\rm oc}}{d}. \tag{K6}$$

Assuming ideally that the temperature is sufficiently low ($k_BT \ll E_g$) for the dark conductivity to be negligible,[77] the drift current is simply proportional to the linear conductivity of photoexcited carriers; this conductivity is assumed to have the Drude form

$$\sigma_{\rm ph}(z) = e^2\tau_{\rm tr}\left(\frac{n(z)}{m_e} + \frac{p(z)}{m_h}\right), \tag{K7}$$

with $\tau_{\rm tr}$ the transport lifetime, n (p) the photoexcited electron density (hole density), and m_e (m_h) the effective mass for electrons (holes). Assuming that the semiconductor is intrinsic,

$$n(z) = p(z) = \tau_{\rm rec}\mathrm{Abs}(z), \quad \sigma_{\rm ph}(z) = \frac{e^2\tau_{\rm tr}\tau_{\rm rec}}{m_r}\mathrm{Abs}(z), \tag{K8}$$

with $\tau_{\rm rec}$ the recombination time [cf. Eq. (B16)] and $m_r^{-1} = m_e^{-1} + m_h^{-1}$. Combining all the above equations,

$$\mathrm{eff} = \frac{1}{4}\frac{.511\ \mathrm{MeV}}{\hbar\omega}\frac{m_r}{m_f}\frac{\mathfrak{s}^2/\tau_{\rm tr}\tau_{\rm rec}}{c^2}, \tag{K9}$$

which is equivalent to Eq. (26).

[77] See Sec. V of Ref. [70].

[1] A. M. Danishevskii, A. A. Kastal'skii, S. M. Ryvkin, and I. D. Yaroshetskii, Dragging of free carriers by photons in direct interband transitions in semiconductors, JETP **31**, 292 (1970).

[2] A. A. Grinberg, Theory of the photoelectric and photomagnetic effects produce by light pressure, JETP **31**, 531 (1970).

[3] V. I. Belinicher and B. I. Sturman, The photogalvanic effect in media lacking a center of symmetry, Sov. Phys. Usp. **23**, 199 (1980).

[4] R. von Baltz and W. Kraut, Theory of the bulk photovoltaic effect in pure crystals, Phys. Rev. B **23**, 5590 (1981).

[5] V. Belinicher, E. Ivchenko, and B. Sturman, Kinetic theory of the displacement photovoltaic effect in piezoelectric, Sov. Phys. JETP **56**, 359 (1982).

[6] J. E. Sipe and A. I. Shkrebtii, Second-order optical response in semiconductors, Phys. Rev. B **61**, 5337 (2000).

[7] D. E. Parker, T. Morimoto, J. Orenstein, and J. E. Moore, Diagrammatic approach to nonlinear optical response with application to weyl semimetals, Phys. Rev. B **99**, 045121 (2019).

[8] S. M. Young and A. M. Rappe, First principles calculation of the shift current photovoltaic effect in ferroelectrics, Phys. Rev. Lett. **109**, 116601 (2012).

[9] 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).
[10] 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).
[11] J. Ibañez-Azpiroz, S. S. Tsirkin, and I. Souza, *Ab initio* calculation of the shift photocurrent by wannier interpolation, Phys. Rev. B **97**, 245143 (2018).
[12] J. Ahn, G.-Y. Guo, N. Nagaosa, and A. Vishwanath, Riemannian geometry of resonant optical responses, Nat. Phys. **18**, 290 (2021).
[13] W. Kraut and R. von Baltz, Anomalous bulk photovoltaic effect in ferroelectrics: A quadratic response theory, Phys. Rev. B **19**, 1548 (1979).
[14] O. Matsyshyn and I. Sodemann, Nonlinear Hall acceleration and the quantum rectification sum rule, Phys. Rev. Lett. **123**, 246602 (2019).
[15] T. Holder, D. Kaplan, and B. Yan, Consequences of time-reversal-symmetry breaking in the light-matter interaction: Berry curvature, quantum metric, and diabatic motion, Phys. Rev. Res. **2**, 033100 (2020).
[16] D. Hornung and R. von Baltz, Quantum kinetics of the magnetophotogalvanic effect, Phys. Rev. B **103**, 195203 (2021).
[17] T. Morimoto and N. Nagaosa, Topological nature of nonlinear optical effects in solids, Sci. Adv. **2**, 10.1126/sciadv.1501524 (2016).
[18] T. Morimoto and N. Nagaosa, Topological aspects of nonlinear excitonic processes in noncentrosymmetric crystals, Phys. Rev. B **94**, 035117 (2016).
[19] O. Matsyshyn, F. Piazza, R. Moessner, and I. Sodemann, Rabi regime of current rectification in solids, Phys. Rev. Lett. **127**, 126604 (2021).
[20] T. Barik and J. D. Sau, Nonequilibrium nature of nonlinear optical response: Application to the bulk photovoltaic effect, Phys. Rev. B **101**, 045201 (2020).
[21] M. Nakamura, S. Horiuchi, F. Kagawa, N. Ogawa, T. Kurumaji, Y. Tokura, and M. Kawasaki, Shift current photovoltaic effect in a ferroelectric charge-transfer complex, Nat. Commun. **8**, 281 (2017).
[22] B. I. Sturman, Ballistic and shift currents in the bulk photovoltaic effect theory, Phys. Usp. **63**, 407 (2020).
[23] L. Z. Tan and A. M. Rappe, Enhancement of the bulk photovoltaic effect in topological insulators, Phys. Rev. Lett. **116**, 237402 (2016).
[24] J. Ahn, G.-Y. Guo, and N. Nagaosa, Low-frequency divergence and quantum geometry of the bulk photovoltaic effect in topological semimetals, Phys. Rev. X **10**, 041041 (2020).
[25] C.-K. Chan, N. H. Lindner, G. Refael, and P. A. Lee, Photocurrents in weyl semimetals, Phys. Rev. B **95**, 041104(R) (2017).
[26] X. Yang, K. Burch, and Y. Ran, Divergent bulk photovoltaic effect in weyl semimetals, arXiv:1712.09363 [cond-mat.mes-hall].
[27] S. Esipov and Y. Levinson, The temperature and energy distribution of photoexcited hot electrons, Adv. Phys. **36**, 331 (1987).
[28] B. P. Zakharchenya, D. N. Mirlin, V. I. Perel', and I. I. Reshina, Spectrum and polarization of hot-electron photoluminescence in semiconductors, Sov. Phys. Usp. **25**, 143 (1982).
[29] P. A. M. Dirac, The quantum theory of the emission and absorption of radiation, Proc. R. Soc. London A **114**, 243 (1927).
[30] B. I. Sturman and V. M. Fridkin, *The Photovoltaic and Photorefractive Effects in Noncentrosymmetric Materials* (Gordon and Breach, London, 1992).
[31] M. Lundstrom, *Fundamentals of Carrier Transport* (Cambridge University Press, Cambridge, UK, 2010).
[32] M. X. Na, A. K. Mills, F. Boschini, M. Michiardi, B. Nosarzewski, R. P. Day, E. Razzoli, A. Sheyerman, M. Schneider, G. Levy, S. Zhdanovich, T. P. Devereaux, A. F. Kemper, D. J. Jones, and A. Damascelli, Direct determination of mode-projected electron-phonon coupling in the time domain, Science **366**, 1231 (2019).
[33] M.-C. Chang and Q. Niu, Berry phase, hyperorbits, and the hofstadter spectrum: Semiclassical dynamics in magnetic bloch bands, Phys. Rev. B **53**, 7010 (1996).
[34] G. Sundaram and Q. Niu, Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and Berry-phase effects, Phys. Rev. B **59**, 14915 (1999).
[35] V. F. Gantmakher and Y. Levinson, *Carrier Scattering in Metals and Semiconductors* (North Holland, Amsterdam, 1987).
[36] P. Vogl, Microscopic theory of electron-phonon interaction in insulators or semiconductors, Phys. Rev. B **13**, 694 (1976).
[37] H. Fröhlich, Theory of electrical breakdown in ionic crystals, Proc. R. Soc. London A **160**, 230 (1937).
[38] H. Fröhlich and N. F. Mott, The mean free path of electrons in polar crystals, Proc. R. Soc. London A **171**, 496 (1939).
[39] H. Meijer and D. Polder, Note on polar scattering of conduction electrons in regular crystals, Physica **19**, 255 (1953).
[40] J. Bardeen and W. Shockley, Deformation potentials and mobilities in non-polar crystals, Phys. Rev. **80**, 72 (1950).
[41] L.-k. Shi, D. Zhang, K. Chang, and J. C. W. Song, Geometric photon-drag effect and nonlinear shift current in centrosymmetric crystals, Phys. Rev. Lett. **126**, 197402 (2021).
[42] H. Wang, X. Tang, H. Xu, J. Li, and X. Qian, Generalized wilson loop method for nonlinear light-matter interaction, npj Quantum Mater. **7**, 61 (2022).
[43] A. Alexandradinata, C. Fang, M. J. Gilbert, and B. A. Bernevig, Spin-orbit-free topological insulators without time-reversal symmetry, Phys. Rev. Lett. **113**, 116403 (2014).
[44] 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).
[45] A. Alexandradinata, A. Nelson, and A. A. Soluyanov, Teleportation of Berry curvature on the surface of a Hopf insulator, Phys. Rev. B **103**, 045107 (2021).
[46] A. Nelson, T. Neupert, T. Bzdušek, and A. Alexandradinata, Multicellularity of delicate topological insulators, Phys. Rev. Lett. **126**, 216404 (2021).
[47] A. Nelson, T. Neupert, A. Alexandradinata, and T. Bzdušek, Delicate topology protected by rotation symmetry: Crystalline hopf insulators and beyond, Phys. Rev. B **106**, 075124 (2022).
[48] A. Alexandradinata, Quantization of intraband and interband Berry phases in the shift current, Phys. Rev. B **110**, 075159 (2024).
[49] 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).

[50] A. Shevelkov, E. Dikarev, R. Shpanchenko, and B. Popovkin, Crystal structures of bismuth tellurohalides BiTeX (X= Cl, Br, I) from x-ray powder diffraction data, J. Solid State Chem. **114**, 379 (1995).

[51] K. Ishizaka, M. Bahramy, H. Murakawa, M. Sakano, T. Shimojima, T. Sonobe, K. Koizumi, S. Shin, H. Miyahara, A. Kimura *et al.*, Giant Rashba-type spin splitting in Bulk bitei, Nat. Mater. **10**, 521 (2011).

[52] T. Das and A. V. Balatsky, Engineering three-dimensional topological insulators in rashba-type spin-orbit coupled heterostructures, Nat. Commun. **4**, 1972 (2013).

[53] Y. Qi, W. Shi, P. G. Naumov, N. Kumar, R. Sankar, W. Schnelle, C. Shekhar, F.-C. Chou, C. Felser, B. Yan, and S. A. Medvedev, Topological quantum phase transition and superconductivity induced by pressure in the bismuth tellurohalide BiTeI, Adv. Mater. **29**, 1605965 (2017).

[54] T. Ideue, J. G. Checkelsky, M. S. Bahramy, H. Murakawa, Y. Kaneko, N. Nagaosa, and Y. Tokura, Pressure variation of rashba spin splitting toward topological transition in the polar semiconductor bitei, Phys. Rev. B **90**, 161107(R) (2014).

[55] B. A. B. with T. L. Hughes, *Topological Insulators and Topological Superconductors* (Princeton University Press, Princeton, NJ, 2013).

[56] N. A. Sinitsyn, Q. Niu, and A. H. MacDonald, Coordinate shift in the semiclassical boltzmann equation and the anomalous hall effect, Phys. Rev. B **73**, 075318 (2006).

[57] I. Sodemann and L. Fu, Quantum nonlinear hall effect induced by berry curvature dipole in time-reversal invariant materials, Phys. Rev. Lett. **115**, 216806 (2015).

[58] J. I. Facio, D. Efremov, K. Koepernik, J.-S. You, I. Sodemann, and J. van den Brink, Strongly enhanced berry dipole at topological phase transitions in bitei, Phys. Rev. Lett. **121**, 246403 (2018).

[59] A. Tomokiyo, T. Okada, and S. Kawano, Phase diagram of system (Bi_2Te_3)-(BiI_3) and crystal structure of BiTeI, Jpn. J. Appl. Phys. **16**, 291 (1977).

[60] P. Lošťák, J. Horák, A. Vaško, and N. t. Dich, Optical properties of BiTeI crystals, Phys. Status Solidi A **59**, 311 (1980).

[61] C.-R. Wang, J.-C. Tung, R. Sankar, C.-T. Hsieh, Y.-Y. Chien, G.-Y. Guo, F. C. Chou, and W.-L. Lee, Magnetotransport in copper-doped noncentrosymmetric BiTeI, Phys. Rev. B **88**, 081104(R) (2013).

[62] V. L. Alperovich, V. I. Belinicher, V. N. Novikov, and A. S. Terekhov, Photogalvanic effects investigation in gallium arsenide, Ferroelectrics **45**, 1 (1982).

[63] Z. Dai, A. M. Schankler, L. Gao, L. Z. Tan, and A. M. Rappe, Phonon-assisted ballistic current from first-principles calculations, Phys. Rev. Lett. **126**, 177403 (2021).

[64] Z. Dai and A. M. Rappe, First-principles calculation of ballistic current from electron-hole interaction, Phys. Rev. B **104**, 235203 (2021).

[65] D. Hornung, R. von Baltz, and U. Rössler, Band structure investigation of the bulk photovoltaic effect in n-gap, Solid State Commun. **48**, 225 (1983).

[66] 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, Colossal mid-infrared bulk photovoltaic effect in a type-i weyl semimetal, Nat. Mater. **18**, 471 (2019).

[67] P. Brody, Large polarization-dependent photovoltages in ceramic $BaTiO_3$ + 5wt.% $CaTiO_3$, Solid State Commun. **12**, 673 (1973).

[68] A. M. Glass, D. von der Linde, and T. J. Negran, High-voltage bulk photovoltaic effect and the photorefractive process in $LiNbO_3$, Appl. Phys. Lett. **25**, 233 (1974).

[69] W. Koch, R. Munser, W. Ruppel, and P. Würfel, Bulk photovoltaic effect in $BaTiO_3$, Solid State Commun. **17**, 847 (1975).

[70] A. Pusch, U. Römer, D. Culcer, and N. J. Ekins-Daukes, Energy conversion efficiency of the bulk photovoltaic effect, PRX Energy **2**, 013006 (2023).

[71] J. I. Pankove, *Optical Processes in Semiconductors* (Dover, New York, 1975).

[72] W. Kohn, Effective mass theory in solids from a many-particle standpoint, Phys. Rev. **105**, 509 (1957).

[73] L. J. Sham, Theory of the shallow impurity states in semiconductors, Phys. Rev. **150**, 720 (1966).

[74] W. Heitler, *The Quantum Theory of Radiation* (Dover, New York, 1954).

[75] J. J. Sakurai, *Advanced Quantum Mechanics* (Addison-Wesley, Reading, MA, 1967).

[76] J. Shah and R. C. C. Leite, Radiative recombination from photoexcited hot carriers in gaas, Phys. Rev. Lett. **22**, 1304 (1969).

[77] E. Meneses, N. Jannuzzi, and R. Leite, Dependence of hot carriers temperature on lattice temperature in CdS, Solid State Commun. **13**, 245 (1973).

[78] G. D. Mahan, *Many-Particle Physics*, 2nd ed. (Plenum, New York, 1990).

[79] J. P. Provost and G. Vallee, Riemannian structure on manifolds of quantum states, Commun. Math. Phys. **76**, 289 (1980).

[80] R. Resta, The insulating state of matter: a geometrical theory, Eur. Phys. J. B **79**, 121 (2011).

[81] J. Yu, C. J. Ciccarino, R. Bianco, I. Errea, P. Narang, and B. A. Bernevig, Non-trivial quantum geometry and the strength of electron–phonon coupling, Nat. Phys. **20**, 1262 (2024).

[82] S. Esipov and Y. Levinson, novel type of oscillatory effects for hot photoelectrons, JETP **59**, 1113 (1984).

[83] E. Lifshitz and L. P. Pitaevskii, *Physical Kinetics* (Butterworth Heinemann, Oxford, 1979).

[84] L. Van Hove, The occurrence of singularities in the elastic frequency distribution of a crystal, Phys. Rev. **89**, 1189 (1953).

[85] E. I. Blount, Formalisms of band theory, in *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic, New York, 1962), Vol. 13.

[86] E. Fermi, Quantum theory of radiation, Rev. Mod. Phys. **4**, 87 (1932).

[87] L. Sham and J. Ziman, *The electron-phonon interaction* (Academic, Cambridge, MA, US, 1963), pp. 221–298.

[88] P. N. Keating, Dielectric screening and the phonon spectra of metallic and nonmetallic crystals, Phys. Rev. **175**, 1171 (1968).

[89] J. Bardeen and D. Pines, Electron-phonon interaction in metals, Phys. Rev. **99**, 1140 (1955).

[90] B. A. Lippmann and J. Schwinger, Variational principles for scattering processes. i, Phys. Rev. **79**, 469 (1950).

[91] S. Weinberg, *The Quantum Theory of Fields, Volume 1: Foundations* (Cambridge University Press, Cambridge, 2005).

[92] L. Gao, Z. Addison, E. J. Mele, and A. M. Rappe, Intrinsic fermi-surface contribution to the bulk photovoltaic effect, Phys. Rev. Res. **3**, 033169 (2021).

[93] R. Kubo, Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems, J. Phys. Soc. Jpn. **12**, 570 (1957).

[94] W. T. H. Koch, R. Munser, W. Ruppel, and P. Würfel, Anomalous photovoltage in $BaTiO_3$, Ferroelectrics **13**, 305 (1976).

[95] L. D. Landau, E. M. Lifshitz, and L. P. Pitaevskii, *Electrodynamics of Continuous Media*, 2nd ed. (Elsevier, Singapore, 1960).

[96] H. Watanabe and Y. Yanase, Chiral photocurrent in parity-violating magnet and enhanced response in topological antiferromagnet, Phys. Rev. X **11**, 011001 (2021).

[97] S. K. Smirnov, Decomposition of solenoidal vector charges into elementary solenoids and the structure of normal one-dimensional flows, St. Petersburg Math. J. **5**, 841 (1994).

[98] F. W. Olver, *Asymptotics and Special Functions* (Academic, San Diego, 1974).

[99] M. DiDomenico and S. H. Wemple, Optical properties of perovskite oxides in their paraelectric and ferroelectric phases, Phys. Rev. **166**, 565 (1968).

[100] J. Ibañez-Azpiroz, F. de Juan, and I. Souza, Assessing the role of interatomic position matrix elements in tight-binding calculations of optical properties, SciPost Phys. **12**, 070 (2022).

[101] Wikipedia contributors, Kronecker product—Wikipedia, the free encyclopedia (2024), [Online; accessed 12-August-2024].

[102] E. Antoncík and P. T. Landsberg, Overlap integrals for bloch electrons, Proc. Phys. Soc. **82**, 337 (1963).

[103] A. A. Soluyanov and D. Vanderbilt, Smooth gauge for topological insulators, Phys. Rev. B **85**, 115415 (2012).

[104] A. Alexandradinata, Z. Wang, B. A. Bernevig, and M. Zaletel, Glide-resolved photoemission spectroscopy: Measuring topological invariants in nonsymmorphic space groups, Phys. Rev. B **101**, 235166 (2020).

[105] A. Alexandradinata and L. Glazman, Semiclassical theory of landau levels and magnetic breakdown in topological metals, Phys. Rev. B **97**, 144422 (2018).