Editors' Suggestion

Semiclassical theory of Landau levels and magnetic breakdown in topological metals

A. Alexandradinata* and Leonid Glazman
Department of Physics, Yale University, New Haven, Connecticut 06520, USA

(Received 9 February 2018; published 26 April 2018)

The Bohr-Sommerfeld quantization rule lies at the heart of the semiclassical theory of a Bloch electron in a magnetic field. This rule is predictive of Landau levels and de Haas–van Alphen oscillations for conventional metals, as well as for a host of topological metals which have emerged in the recent intercourse between band theory, crystalline symmetries, and topology. The essential ingredients in any quantization rule are connection formulas that match the semiclassical (WKB) wave function across regions of strong quantum fluctuations. Here, we propose (a) a multicomponent WKB wave function that describes transport within degenerate-band subspaces, and (b) the requisite connection formulas for saddle points and type-II Dirac points, where tunneling respectively occurs within the same band, and between distinct bands. (a) and (b) extend previous works by incorporating phase corrections that are subleading in powers of the field; these corrections include the geometric Berry phase, and account for the orbital magnetic moment and the Zeeman coupling. A comprehensive symmetry analysis is performed for such phase corrections occurring in closed orbits, which is applicable to solids in any (magnetic) space group. We have further formulated a graph-theoretic description of semiclassical orbits. This allows us to systematize the construction of quantization rules for a large class of closed orbits (with or without tunneling), as well as to formulate the notion of a topological invariant in semiclassical magnetotransport—as a quantity that is invariant under continuous deformations of the graph. Landau levels in the presence of tunneling are generically quasirandom, i.e., disordered on the scale of nearest-neighbor level spacings but having longer-ranged correlations; we develop a perturbative theory to determine Landau levels in such quasirandom spectra.

DOI: 10.1103/PhysRevB.97.144422

I. INTRODUCTION

The Peierls-Onsager-Lifshitz semiclassical theory of Bloch electrons in a weak magnetic field is the bridge that connects experimentally accessible, field-induced oscillations to properties of a metal at zero field [1–5]. This theory underlies the phenomenological construction of the Fermi surface of normal metals [6,7] and superconductors [8]—from measuring the oscillatory period of the magnetization [9] or resistivity [10]. These de Haas–van Alphen (dHvA) oscillations are generically disrupted by field-induced quantum tunneling between semiclassical orbits. Such tunneling, known as magnetic breakdown, occurs wherever semiclassical orbits intersect at saddle points (in the energy-momentum dispersion) [11] or at band-touching points [12]. The experimental discovery of breakdown in magnesium [6] sparked an extension of the semiclassical theory to incorporate tunneling [11–19].

The semiclassical theory has been further extended to incorporate two modern concepts: a wave packet that orbits around the Fermi surface accumulates a geometric Berry phase [20,21] as well as a second phase associated with the orbital magnetic moment of a wave packet around its center of mass [22]. Both phases were first derived from the effective-Hamiltonian theory pioneered in the 1960s [14,23–26]; analogs of these phases appear ubiquitously in the asymptotic theory of coupled-wave equations [27–30], which apply in a much wider variety of physical contexts than the present study. Only more recently have the physical consequences of the geometric phase and the orbital magnetic moment been explored for solids [25,31,32]—especially in the complementary theory of wave packets [22,33–35].

Both of the above phases are evaluated on semiclassical orbits uniquely determined by Hamilton's equation. On the other hand, semiclassical orbits are no longer unique in the presence of breakdown. The challenge is to resolve this tension. Recently, we have synthesized the geometric phase, the orbital moment, and tunneling—into a single, generalized Bohr-Sommerfeld quantization rule [36]. This rule is not only predictive of Landau levels and de Haas–van Alphen oscillations for conventional metals, but it is also critically relevant to describe a host of topological metals which have emerged in the recent intercourse between band theory, crystalline symmetries, and topology. Such topological metals have intrinsically unremovable geometric phase, owing to the presence of Dirac-Weyl points where conically dispersing bands touch [37–44]; their Fermi surfaces are twisted into unusual topologies [45–47] such that breakdown is also unavoidable.

The essential ingredients in any quantization rule are connection formulas that match the semiclassical (WKB) wave function across regions of strong quantum fluctuations. The main subject of this work is the derivation of these ingredients and the systematic construction of quantization rules for a large class of closed orbits—with and without breakdown. Our results are summarized in the following section.

II. SUMMARY AND ORGANIZATION OF RESULTS

The semiclassical theory is a method to approximate the wave functions and energy levels of a Bloch electron in a

*Corresponding author: aris.alex@yale.edu

magnetic field. These approximations become increasingly accurate in the limit where a classical action function—which characterizes the solid at zero field—is much larger than a parameter characteristic of the magnetic field. These quantities are simplest to exemplify for a semiclassical orbit on a Fermi surface having the topology of a sphere; an orbiting wave packet evolves according to Hamilton's equation of motion:

$$\hbar\dot{\boldsymbol{k}} = -\frac{|e|}{c\hbar}\nabla_{\boldsymbol{k}}\varepsilon \times \boldsymbol{B}. \tag{1}$$

The semiclassical approximation is valid where the the area $|S|$ (in $\boldsymbol{k}$ space) of this orbit is much larger than $1/l^2$, with the magnetic length defined as

$$l = \sqrt{\frac{\hbar c}{|e|B}}. \tag{2}$$

To simplify notation, we henceforth adopt a coordinate system where the magnetic field $\boldsymbol{B} = -B\hat{z}$, and the semiclassical orbit is therefore a band contour at fixed energy E (and wave vector k_z, for 3D solids), equipped with an orientation from Eq. (1). The Bohr-Sommerfeld quantization rules of the semiclassical theory are derived in the effective-Hamiltonian formalism [1,5,14,23,24,48–50], which we briefly review in Sec. IV. Certain notations used in this review, and throughout the text, are collected in Sec. III for easy reference.

We begin properly in Sec. V by deriving the quantization rules for closed orbits in the absence of breakdown; these rules are summarized in Eqs. (68)–(74), and their consequences for Landau levels and de Haas–van Alphen oscillations are discussed in Sec. V D. These rules are equivalent to a continuity condition of the WKB solution to the above-reviewed effective Hamiltonian. The single-component WKB wave function [51,52] (as applied to a nondegenerate band) is reviewed in Sec. V A 1, and we generalize this to a multicomponent wave function in Sec. V A 2 (as applied to bands of arbitrary degeneracy D). As this wave function is continued around a closed orbit, it accumulates a phase proportional to $1/B$ and the oriented area of the orbit; the subleading-in-B variation of the wave function is described by a $D \times D$ unitary propagator $\mathcal{A}$ [cf. Eq. (74)], which is generated by a one-form that includes the Berry connection (non-Abelian for $D > 1$), the orbital magnetic moment, and the Zeeman coupling. Each eigenphase (i.e., phase of each eigenvalue) of this propagator enters the quantization rule as an $O(1)$ phase correction. In addition, there is a subleading Maslov correction originating from turning points on the orbit where the WKB solution is invalid. A general method to determine Maslov corrections is described in Sec. V B, which is applicable to twisted Fermi surfaces whose orbits intersect at points. We emphasize that there are no further $O(1)$ corrections to the quantization rule.

Section VI is an exposition of the effects of symmetry (in any space or magnetic space group) in the quantization condition. By a symmetry analysis of the propagator $\mathcal{A}$ [cf. Eq. (74)], we ascertain how symmetry constrains the degeneracy and energetic offsets of the Landau levels, as well as phase offsets in the de Haas–van Alphen oscillations. In addition, we provide a general symmetry analysis of the orbital magnetic moment and Zeeman coupling in Sec. VI B; this may be applied to $\boldsymbol{k}$-resolved measurements of the orbital magnetic moment, e.g., through circular dichroism in photoemission [53].

In Sec. VII, we describe quantization rules which are applicable to orbits which intersect at saddle points in the energy-momenta dispersion. Saddle points are the nuclei of Lifshitz transitions in the Fermi-surface topology, as exemplified by the surface states of topological crystalline insulators [45,54]. In the vicinity of saddle points lie regions of strong intraband tunneling, where the WKB solutions lose their validity. WKB wave functions away from the saddle point are patched together by a connection formula that we derive in Sec. VII A. The generalized quantization rule is equivalent to the continuity of patched-up WKB wave functions over the intersecting orbit; the general algorithm for constructing such rules is presented in Sec. VII B. This algorithm is then applied to two case studies: a Weyl metal near a metal-insulator phase transition (cf. Sec. VII C), and the surface states in the SnTe class of topological crystalline insulators [55] (cf. Sec. VII D).

A qualitatively distinct type of breakdown occurs where orbits intersect at a touching point between two bands which are otherwise nondegenerate at generic wave vectors. In this work, we focus on touching points for which the nearby band dispersion is conical. From a general classification of Fermi surfaces near conical touching points [56], we identify the orbit intersection as a type-II Dirac point (in short, a II-Dirac point) [57–59], which might be viewed as an over-tilted version of the conventional, rotationally symmetric Dirac point [60]. Due to the discontinuity of the Bloch wave function across the II-Dirac point [61], the effective Hamiltonian that was reviewed in Sec. IV is not applicable. What we require is a different representation for the effective Hamiltonian, where the basis functions evolve smoothly across the II-Dirac point. Inspired by the basis functions proposed by Slutskin [12], we formulate such an effective Hamiltonian in Sec. VIII, as summarized in Eqs. (230)–(237). This effective Hamiltonian extends previous formulations [12,17] by (i) being applicable to any band-touching point (of any degeneracy and dispersion, e.g., Weyl [37–39], multi-Weyl [62], and spin-1 Weyl points [40], fourfold-degenerate Dirac points [41–43], and charge-2 Dirac points [44]), and by (ii) accounting for subleading-in-B corrections, which includes the multiband orbital magnetic moment at the band-touching point.

The solution of the above effective Hamiltonian—particularized to a II-Dirac point—affords us a connection formula presented in Sec. IX C. This rule is a crucial ingredient in quantization rules for orbits that intersect at a II-Dirac point. We demonstrate how to construct such rules for orbits surrounding an isolated, over-tilted Weyl point in Sec. IX D. The Landau-level spectrum in the presence of interband breakdown (and also intraband breakdown in low-symmetry metals) is generically quasirandom, i.e., disordered on the scale of nearest-neighbor level spacings but having longer-ranged correlations. A perturbative theory to determine Landau levels in quasirandom spectra is presented in Sec. IX E and applied to our case study.

Throughout the text, we will employ a graph-theoretic description of orbits that is summarized in Sec. III F for easy reference. Such a description is not only useful in systematizing the construction of quantization rules (with or without breakdown); it allows us to define an equivalence class of Fermi surfaces—through the homotopy equivalence of their corresponding graphs. This allows us to precisely define a

topological invariant in semiclassical magnetotransport: as a quantity that is invariant under continuous deformations of the Hamiltonian that preserves the homotopy class of the graph. The generalization to symmetry-protected topological invariants is simply described in Sec. III F. Examples of such topological invariants have been presented in our previous companion works [32,36].

III. PRELIMINARIES

We review the exact Hamiltonian of a Bloch electron, with and without a magnetic field, to establish notation that will be used throughout this paper.

A. Bloch Hamiltonian in the crystal momentum representation

In materials with light elements and consequently weak spin-orbit coupling, we will apply the field-free Schrödinger Hamiltonian, defined as

$$\hat{H}_0 = \frac{\boldsymbol{p}^2}{2m} + V(\boldsymbol{r}); \tag{3}$$

otherwise, we apply the Pauli Hamiltonian:

$$\hat{H}_0 = \frac{1}{2m}\left(\boldsymbol{p} + \frac{\hbar}{4mc^2}\boldsymbol{\sigma}\times\boldsymbol{\nabla}V\right)^2 + V(\boldsymbol{r}), \tag{4}$$

which is accurate to order E/mc^2. We use the same symbol $\hat{H}_0$ for both Hamiltonians, and unless otherwise stated in the context, we assume that expressions with $\hat{H}_0$ apply to both types of Hamiltonians. Each eigenstate of $\hat{H}_0$ may be expressed as a Bloch function

$$\psi_{n\boldsymbol{k}} = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{n\boldsymbol{k}}, \tag{5}$$

where $u_{n\boldsymbol{k}} = u_{n\boldsymbol{k}}(\boldsymbol{r})$ in the Schrödinger case, and $= u_{n\boldsymbol{k}}(\boldsymbol{r},s)$ with additional spin index s in the Pauli case. In both cases, $u_{n\boldsymbol{k}}$ is periodic with respect to Bravais-lattice translations $\boldsymbol{r}\to\boldsymbol{r}+\boldsymbol{R}$, and shall henceforth be referred to as cell-periodic functions. It is convenient to define the cell-periodic position coordinate $\boldsymbol{\tau}$ with the equivalence $\boldsymbol{\tau}\sim\boldsymbol{\tau}+\boldsymbol{R}$, as well as the variable α, which is a flexible shorthand for $\boldsymbol{\tau}$ in the Schrödinger case, and for $(\boldsymbol{\tau},s)$ in the Pauli case. It is well known that cell-periodic functions form an orthonormal set which is complete with respect to the space of α:

$$\sum_\alpha u^*_{m\boldsymbol{k}}(\alpha)u_{n\boldsymbol{k}}(\alpha) = \delta_{mn}, \quad \text{and}$$
$$\sum_m u_{m\boldsymbol{k}}(\alpha)u^*_{m\boldsymbol{k}}(\beta) = \delta_{\alpha\beta}. \tag{6}$$

Here and henceforth, we employ the Dirac notation: $\langle u|v\rangle = \sum_\alpha u^*(\alpha)v(\alpha)$, where $\sum_\alpha$ should be interpreted as an integration of $\boldsymbol{\tau}$ over the unit cell (normalized multiplicatively by the volume of the Brillouin torus), and possibly also a sum over the spin indices. Analogously, $\delta_{\alpha\beta}$ denotes the Dirac delta function $\delta(\boldsymbol{\tau}-\boldsymbol{\tau}')$, possibly multiplied with a Kronecker delta function in spin space. When there is no topological obstruction to constructing Wannier functions (W_n), we will find it useful to expand the cell-periodic function in terms of Wannier functions as

$$u_{n\boldsymbol{k}}(\boldsymbol{\tau},s) = \frac{1}{\sqrt{N}}\sum_{\boldsymbol{R}} e^{-i\boldsymbol{k}\cdot(\boldsymbol{\tau}-\boldsymbol{R})}W_n(\boldsymbol{\tau}-\boldsymbol{R},s). \tag{7}$$

The Hamiltonian acts on cell-periodic functions as

$$\hat{H}_0(\boldsymbol{k}) = e^{-i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}\hat{H}_0 e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}; \tag{8}$$

we will refer to $\hat{H}_0$ as the Hamiltonian, and $\hat{H}_0(\boldsymbol{k})$ as the Bloch Hamiltonian. The velocity operator is defined by

$$\hat{\boldsymbol{\Pi}} = -\frac{i}{\hbar}[\hat{\boldsymbol{r}},\hat{H}_0] = \boldsymbol{\nabla}_{\boldsymbol{p}}\hat{H}_0, \tag{9}$$

and it acts on cell-periodic functions as

$$\hat{\boldsymbol{\Pi}}(\boldsymbol{k}) := e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\,\hat{\boldsymbol{\Pi}}\,e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = \hat{\boldsymbol{\Pi}} + \frac{\hbar\boldsymbol{k}}{m} = \boldsymbol{\nabla}_{\boldsymbol{p}}\hat{H}_0(\boldsymbol{k})$$
$$= \boldsymbol{\nabla}_{\boldsymbol{k}}\hat{H}_0(\boldsymbol{k}) = \begin{cases} \frac{\hat{\boldsymbol{p}}+\hbar\boldsymbol{k}}{m}, \\ \frac{\hat{\boldsymbol{p}}+\hbar\boldsymbol{k}}{m} - \frac{\mu_B}{2emc}\boldsymbol{\sigma}\times\boldsymbol{\nabla}V, \end{cases} \tag{10}$$

with the Bohr magneton $\mu_B = |e|\hbar/2mc$. The Bloch Hamiltonian may always be expanded around a chosen wave vector $\boldsymbol{k}_0$ as

$$\hat{H}_0(\boldsymbol{k}) = \hat{H}_0(\boldsymbol{k}_0) + \hbar(\boldsymbol{k}-\boldsymbol{k}_0)\cdot\hat{\boldsymbol{\Pi}}(\boldsymbol{k}_0) + \frac{\hbar^2(\boldsymbol{k}-\boldsymbol{k}_0)^2}{2m}. \tag{11}$$

Any operator which acts on functions of $\boldsymbol{r}$ (and possibly also on spin index s) are denoted with a hat, as exemplified in Eqs. (3)–(10); the same operator in the basis of cell-periodic functions is a matrix denoted by the same symbol with a tilde. Unless specified otherwise, we will usually employ a basis of cell-periodic functions which correspond to energy bands, i.e., for which the Hamiltonian matrix is diagonal,

$$\tilde{H}_0(\boldsymbol{k})_{mn} = \langle u_{m\boldsymbol{k}}|\hat{H}_0(\boldsymbol{k})|u_{n\boldsymbol{k}}\rangle = \varepsilon_{n\boldsymbol{k}}\delta_{mn}. \tag{12}$$

Another example is the velocity matrix

$$\tilde{\boldsymbol{\Pi}}(\boldsymbol{k})_{mn} = \langle u_{m\boldsymbol{k}}|\hat{\boldsymbol{\Pi}}(\boldsymbol{k})|u_{n\boldsymbol{k}}\rangle, \tag{13}$$

which may be identified, in the basis of energy eigenstates, as

$$\hbar\tilde{\boldsymbol{\Pi}}(\boldsymbol{k})_{mn} = \boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon_n\delta_{mn} + i\tilde{\mathfrak{X}}(\boldsymbol{k})_{mn}(\varepsilon_{m\boldsymbol{k}} - \varepsilon_{n\boldsymbol{k}}). \tag{14}$$

Here, we have introduced

$$\tilde{\mathfrak{X}}(\boldsymbol{k})_{mn} = i\langle u_{m\boldsymbol{k}}|\boldsymbol{\nabla}_{\boldsymbol{k}}u_{n\boldsymbol{k}}\rangle, \tag{15}$$

which occurs as part of the matrix elements of the position operator in the crystal-momentum representation [63]. It is also useful to define the diagonal component of the velocity matrix as

$$\tilde{\boldsymbol{v}}(\boldsymbol{k})_{mn} = \frac{1}{\hbar}\boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon_n\delta_{mn}, \tag{16}$$

as well as the spin-half matrix

$$\frac{\hbar}{2}\tilde{\boldsymbol{\sigma}}(\boldsymbol{k})_{mn} = \frac{\hbar}{2}\langle u_{m\boldsymbol{k}}|\hat{\boldsymbol{\sigma}}|u_{n\boldsymbol{k}}\rangle. \tag{17}$$

While these matrices are formally infinite-dimensional, we are often interested in the physics of a finite number of (possibly degenerate) bands, projected by

$$P(\boldsymbol{k}) = \sum_{n=1}^{D}|u_{n\boldsymbol{k}}\rangle\langle u_{n\boldsymbol{k}}|, \tag{18}$$

with D the dimension of said subspace at each wave vector. Bands not in P are henceforth labeled with an extra bar: $\bar{m},\bar{n}$,

and their corresponding projection is

$$Q(\boldsymbol{k}) = \sum_{\bar{n}} |u_{\bar{n}\boldsymbol{k}}\rangle\langle u_{\bar{n}\boldsymbol{k}}| = I - P(\boldsymbol{k}). \tag{19}$$

Let us then define the restriction of any matrix in the infinite basis to the subspace projected by P as

$$P:\ \{\tilde{\boldsymbol{\Pi}},\tilde{\mathfrak{X}},\tilde{\boldsymbol{v}},\tilde{\boldsymbol{\sigma}}\} \to \{\boldsymbol{\Pi},\mathfrak{X},\boldsymbol{v},\boldsymbol{\sigma}\}. \tag{20}$$

These finite-dimensional matrices are distinguished notationally by having no accents. A case in point is $\mathfrak{X}$, which is the Berry connection [20] for solids [64]. This connection manifests whenever one differentiates operators represented in the D-dimensional cell-periodic basis: for any $\hat{O}(\boldsymbol{k})$,

$$\begin{aligned}\boldsymbol{\nabla}_{\boldsymbol{k}} O(\boldsymbol{k})_{mn} &= \boldsymbol{\nabla}_{\boldsymbol{k}}\langle u_{m\boldsymbol{k}}|\hat{O}(\boldsymbol{k})|u_{n\boldsymbol{k}}\rangle \\ &= \langle u_{m\boldsymbol{k}}|\boldsymbol{\nabla}_{\boldsymbol{k}}\hat{O}|u_{n\boldsymbol{k}}\rangle + i[\mathfrak{X}(\boldsymbol{k}),O(\boldsymbol{k})]_{mn}.\end{aligned} \tag{21}$$

B. Gauge transformations in band theory

We will often deal with $U(D)$ basis transformations in the cell-periodic functions in P:

$$|u_{n\boldsymbol{k}}\rangle \to \sum_{m=1}^{D} |u_{m\boldsymbol{k}}\rangle V_{mn}(\boldsymbol{k}), \quad V^{-1} = V^{\dagger}. \tag{22}$$

We will refer to this as a gauge transformation within P. With respect to this transformation, certain objects are invariant (such as the projection P itself); other objects transform covariantly; i.e., they change only in being conjugated by the unitary V (e.g., the just-defined spin matrix $\boldsymbol{\sigma} \to V^{-1}\boldsymbol{\sigma}V$); other objects have a more complicated transformation rule; e.g., the non-Abelian Berry connection transforms as [65]

$$\mathfrak{X} \to V^{-1}\mathfrak{X}V + iV^{-1}\boldsymbol{\nabla}_{\boldsymbol{k}}V. \tag{23}$$

C. Review of symmetry in Bloch Hamiltonians

Let g denote a symmetry in the (magnetic) space group (G) of a solid; we use $\hat{g}$ to denote its representation in real space tensored with spin space. Its action on the position operator can always be decomposed as a point-group operation (an operation that preserves at least one point) and a translation [66]:

$$\hat{g}^{-1}\hat{r}_i\hat{g} = \breve{g}_{ij}\hat{r}_j + \delta_i, \quad \breve{g}^{-1} = \breve{g}^{t} \in \mathbb{R}, \tag{24}$$

which we shorten notationally as $\hat{g}^{-1}\hat{\boldsymbol{r}}\hat{g} = \breve{g}\hat{\boldsymbol{r}} + \boldsymbol{\delta}$. Here, we have introduced a real, orthogonal matrix $\breve{g}$ that represents the point-group component of g that acts in real space. For all symmetry elements in symmorphic space groups, a spatial origin may be chosen such that $\boldsymbol{\delta}$ is a Bravais-lattice vector [66]. To describe nonsymmorphic operations such as screw rotations and glide reflections, we allow $\boldsymbol{\delta}$ to be a rational fraction of a Bravais-lattice vector.

In addition to g that transforms space, we also consider g that reverses time. The time-reversal operation $g = T$ acts trivially on space ($\breve{g} = I, \boldsymbol{\delta} = 0$), and is represented by $\hat{T} = U_T K$, with U_T a unitary transformation and K the complex conjugation operation; $\hat{T}^2 = (-1)^F$, where $F = 0$ for integer-spin representations ($U_T = I$), and $F = 1$ for half-integer-spin representations ($U_T = -i\sigma_y$ in spinor space). It is useful to introduce a $\mathbb{Z}_2$ index that distinguishes between transformations which are purely spatial (and therefore have a unitary representation $\hat{g}$), and transformations which involve a time reversal, possibly composed with a spatial operation ($\hat{g}$ here is antiunitary):

$$\begin{aligned} g:\begin{pmatrix}\hat{\boldsymbol{r}}\\ t\end{pmatrix} &\to \begin{pmatrix}\breve{g} & 0\\ 0 & (-1)^{s(g)}\end{pmatrix}\begin{pmatrix}\hat{\boldsymbol{r}}\\ t\end{pmatrix} + \begin{pmatrix}\boldsymbol{\delta}\\ 0\end{pmatrix}; \\ s(g) &= \begin{cases}0, & \hat{g}\ \text{unitary},\\ 1, & \hat{g}\ \text{antiunitary}.\end{cases}\end{aligned} \tag{25}$$

As a useful example, we apply Eqs. (25) and (24) to derive

$$\hat{g}e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}\hat{g}^{-1} = e^{i(-1)^{s(g)}[\breve{g}\boldsymbol{k}]\cdot(\hat{\boldsymbol{r}}-\boldsymbol{\delta})}, \tag{26}$$

which implies that a Bloch function at wave vector $\boldsymbol{k}$, when operated upon by g, transforms in the representation

$$g\circ\boldsymbol{k} := (-1)^{s(g)}\breve{g}\boldsymbol{k}. \tag{27}$$

If g is a symmetry of the Hamiltonian ($[\hat{g},\hat{H}_0] = 0$), then

$$\begin{aligned}\hat{g}(\boldsymbol{k})\hat{H}_0(\boldsymbol{k})\hat{g}^{-1}(\boldsymbol{k}) &= \hat{H}_0(g\circ\boldsymbol{k}), \quad \text{with} \\ \hat{g}(\boldsymbol{k}) &:= e^{-i(g\circ\boldsymbol{k})\cdot\boldsymbol{\delta}}\hat{g}.\end{aligned} \tag{28}$$

This implies that if $|u_{m\boldsymbol{k}}\rangle$ is an eigenstate of $\hat{H}_0(\boldsymbol{k})$ with eigenvalue $\varepsilon_{m\boldsymbol{k}}$, then $\hat{g}(\boldsymbol{k})|u_{m\boldsymbol{k}}\rangle K^{s(g)}$ belongs to the eigenspace of $\hat{H}_0(\ g\circ\boldsymbol{k}\ )$ with the same energy $\varepsilon_{m\boldsymbol{k}}$; this is expressed as

$$\hat{g}(\boldsymbol{k})|u_{m\boldsymbol{k}}\rangle K^{s(g)} = |u_{n,g\circ\boldsymbol{k}}\rangle\breve{g}(\boldsymbol{k})_{nm}, \tag{29}$$

where $\breve{g}$, a unitary matrix that is block-diagonal with respect to the energy eigenspaces, expresses the ambiguity in our choice of basis vectors within each energy eigenspace. To clarify a possible source of confusion, $\langle\alpha|\hat{g}(\boldsymbol{k})|u_{m\boldsymbol{k}}\rangle K^{s(g)}$ is just a complex number—where $s(g) = 1$, there are two K operators in this expression: one explicit, and the other implicit in $\hat{g}$.

We refer to $\breve{g}(\boldsymbol{k})$ colloquially as the "sewing matrix," owing to its function in "sewing" together the cell-periodic functions by symmetry. Sewing matrices are the basic objects that encode symmetry constraints in the crystal-momentum representation, and they will play a prominent role in constraining the effective Hamiltonian. These matrices may be understood from a group-cohomological perspective [67]; the winding number of the sewing matrix over the Brillouin torus also plays a role in the topological classification of band insulators [68,69].

For our purpose of determining the symmetry constraints on the effective Hamiltonian, we will need to review a few properties of sewing matrices. Depending on the presence of spin-SU(2) symmetry, $\{\breve{g}(\boldsymbol{k})\}$ forms either an integer- or half-integer-spin representation of the space group [67,70]. A simple example might convey this point: let g be a glide operation ($\mathfrak{g}_{x,\vec{y}/2}$) that is composed of a reflection, that inverts $x \to -x$, and a translation by half a Bravais-lattice vector in $\vec{y}$ (denoted $\mathfrak{t}_{\vec{y}/2}$). In the space group, the multiplication rule for this element is $\mathfrak{g}^2_{x,\vec{y}/2} = \mathfrak{e}\,\mathfrak{t}_{\vec{y}}$, with $\mathfrak{e}$ a 2π rotation and $\mathfrak{t}_{\vec{y}}$ a full lattice translation; this is represented as [71]

$$\begin{aligned}&[\breve{\mathfrak{g}}_{x,\vec{y}/2}(-k_x,k_y,k_z)\breve{\mathfrak{g}}_{x,\vec{y}/2}(\boldsymbol{k})]_{mn} \\ &= \sum_{l}\langle u_{m\boldsymbol{k}}|e^{-ik_y/2}\hat{\mathfrak{g}}_{x,\vec{y}/2}\big|u_{l,(-k_x,k_y,k_z)}\big\rangle \\ &\quad\times\big\langle u_{l,(-k_x,k_y,k_z)}\big|e^{-ik_y/2}\hat{\mathfrak{g}}_{x,\vec{y}/2}|u_{n\boldsymbol{k}}\rangle \\ &= e^{-ik_y}\langle u_{m\boldsymbol{k}}|\hat{\mathfrak{g}}^2_{x,\vec{y}/2}|u_{n\boldsymbol{k}}\rangle = e^{-ik_y}(-1)^F\delta_{mn}.\end{aligned} \tag{30}$$

TABLE I. Examples of symmetries g of order N. a and $\boldsymbol{R}$ are defined through Eq. (31). m and p are quantities that are introduced later in Sec. VID 5: m is the number of cycles in the g orbit, and $p \sim p+N$ labels inequivalent extensions (by quasimomentum loop translations) of the point group generated by g. We have chosen the convention that $\mathfrak{o}$ is clockwise-oriented, and that $\mathfrak{c}_{nz}$ induces an anticlockwise rotation in $\boldsymbol{k}$ space; if both $\mathfrak{o}$ and $\mathfrak{c}_{nz}$ are anticlockwise-oriented, then the above values of p should be inverted in sign.

g	N	a	$\boldsymbol{R}$	m	p
$\mathfrak{i}$	2	0	$\boldsymbol{0}$	1	1
T	2	1	$\boldsymbol{0}$	1	1
$\mathfrak{c}_{nz}$	n	1	$\boldsymbol{0}$	1	1
$T\mathfrak{c}_{3z}$	6	1	$\boldsymbol{0}$	1	5
$T\mathfrak{c}_{4z}$	4	1	$\boldsymbol{0}$	1	3
$T\mathfrak{c}_{6z}$	6	1	$\boldsymbol{0}$	2	4
$\mathfrak{g}_{x,\vec{y}/2}$	2	1	$\vec{y}$	1	1

In the last equality, we employed that a rotation by 2π produces a representation-dependent, ± 1 factor, and also that $\mathfrak{t}_{\boldsymbol{R}}$ has a trivial action on cell-periodic functions.

More generally, for any nontrivial g which is not purely a translation, we may assign to g an order $N(g)$, which is the smallest integer in $\{2,3,4,6\}$ such that

$$g^{N(g)} = \mathfrak{e}^{a(g)}\mathfrak{t}_{\boldsymbol{R}(g)}, \quad a(g) \in \{0,1\}, \tag{31}$$

with $\mathfrak{t}_{\boldsymbol{R}}$ a translation by a Bravais-lattice vector $\boldsymbol{R}$ (possibly the zero vector) which depends on g. We have introduced a $\mathbb{Z}_2$ index $a(g)$ that equals 0 (resp. 1) if g^N is proportional to an odd (resp. even) multiple of a 2π rotation ($a=1$). In the case of $g=\mathfrak{g}^2_{x,\vec{y}/2}$ in Eq. (30), $N=2, a=1$, and $\boldsymbol{R}=\vec{y}$; other representative examples are summarized in Table I. If g reverses time, its order must be even:

$$s(g) = 1 \Rightarrow N(g) \in 2\mathbb{Z}. \tag{32}$$

This follows because g^N by assumption does not invert time [cf. Eq. (31)], and on the other hand it is the composition of T^N with a spatial transformation.

For any $\boldsymbol{k}_1$, we define

$$g \text{ orbit of } \boldsymbol{k}_1 := \{\boldsymbol{k}_i\}_{i=1}^N, \quad \text{with}$$
$$\boldsymbol{k}_{i+1} := g^i \circ \boldsymbol{k}_1 := \boldsymbol{k}_{i+N+1}, \tag{33}$$

which is not to be confused with Hamilton's semiclassical orbit; we are guaranteed that $\boldsymbol{k}_i = \boldsymbol{k}_{i+N}$ owing to Eq. (31). Equation (31) is represented with the sewing matrices as

$$\breve{g}_i := \breve{g}(\boldsymbol{k}_i), \quad \breve{g}_N K^s \ldots \breve{g}_2 K^s \breve{g}_1 K^s = (-1)^{Fa} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}. \tag{34}$$

When this equation is particularized to g which is unitarily represented, and to $\boldsymbol{k} = g\circ\boldsymbol{k}$, we obtain

$$\breve{g}(\boldsymbol{k})^N = (-1)^{Fa} e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}. \tag{35}$$

The N possible eigenvalues of $\breve{g}$ (at g-invariant wave vectors), corresponding to the N roots of $e^{i\pi Fa - i\boldsymbol{k}\cdot\boldsymbol{R}}$, label the different representations of g. More examples of sewing matrices are provided in the second column of Table II.

Finally, we consider how the sewing matrix transforms under basis transformations of the form in Eq. (22). From Eq. (29), we derive

$$\breve{g}(\boldsymbol{k}) \to V^\dagger(g\circ\boldsymbol{k})\breve{g}(\boldsymbol{k})K^{s(g)}V(\boldsymbol{k})K^{s(g)}. \tag{36}$$

For g-invariant wave vectors (defined through $\boldsymbol{k} = g\circ\boldsymbol{k}$ modulo a reciprocal vector), Eq. (36) particularizes to

$$\breve{g} \to \begin{cases} V^\dagger \breve{g} V, & \text{for unitary } g, \\ V^\dagger \breve{g} V^*, & \text{for antiunitary } g. \end{cases} \tag{37}$$

This distinction between unitary and antiunitary symmetries becomes relevant when we consider the symmetry constraints of the orbital moment in Sec. VIB.

For future reference, we employ the following notation for symmetry operations: T denotes time reversal, $\mathfrak{t}_{\vec{z}/2}$ a real-space translation by half a Bravais-lattice vector parallel to $\vec{z}$; $T\mathfrak{t}_{\vec{z}/2}$ is the composition of T and $\mathfrak{t}_{\vec{z}/2}$. $\mathfrak{i}$ denotes spatial inversion. $\mathfrak{r}_\alpha$ ($\mathfrak{g}_{\alpha,\vec{\beta}/2}$) is normal (glide) reflection that inverts the spatial coordinate α; the glide operation includes an additional translation by $\vec{\beta}/2$, which is half a Bravais-lattice vector in the β direction. $\mathfrak{c}_{nz}$ is an n-fold rotation about $\vec{z}$ ($n\in\{2,3,4,5\}$); $\mathfrak{s}_{nz,m}$ is a screw rotation that satisfies ${\mathfrak{s}_{nz,m}}^n = \mathfrak{t}_{m\boldsymbol{G}}$ with $\boldsymbol{G}$ the smallest reciprocal vector parallel to $\vec{z}$ and $m\in\{0,1,\ldots,n-1\}$. To describe half-integer-spin representations, we will employ the double-group formalism that identifies a 2π rotation with a group element ($\mathfrak{e}$) that is distinct from and squares to the identity.

D. Hamiltonian of a Bloch electron in a magnetic field

We study a Bloch electron immersed in a spatially homogeneous magnetic field $\boldsymbol{B}$, with corresponding vector potential

$$\boldsymbol{B} = \nabla\times\boldsymbol{A}, \quad \boldsymbol{a} = \frac{|e|}{c}\boldsymbol{A}. \tag{38}$$

The field-on Schrödinger Hamiltonian is defined as

$$\hat{H} = \frac{[\boldsymbol{p}+\boldsymbol{a}(\boldsymbol{r})]^2}{2m} + V(\boldsymbol{r}); \tag{39}$$

this is distinguished notationally from the zero-field Schrödinger Hamiltonian ($\hat{H}_0$) by having no subscript. Analogously, the field-on Pauli Hamiltonian is

$$\hat{H} = \frac{1}{2m}\left(\boldsymbol{p}+\boldsymbol{a}+\frac{\hbar}{4mc^2}\boldsymbol{\sigma}\times\nabla V\right)^2 + V(\boldsymbol{r}) + \frac{g_0}{2}\mu_B\boldsymbol{\sigma}\cdot\boldsymbol{B}, \tag{40}$$

with the free-electron g factor $g_0\approx 2$. The semiclassical equation of motion for a Bloch electron in a magnetic field is

$$\hbar\dot{\boldsymbol{k}}^\perp|_k = l^{-2}\epsilon^{\alpha\beta}\vec{\alpha}v^\beta(\boldsymbol{k}), \quad \alpha,\beta\in\{x,y\}, \tag{41}$$

which particularizes Eq. (1) to the case $\boldsymbol{B} = -B\vec{z}$. We refer to $\dot{\boldsymbol{k}}^\perp|_k$ as the *orbit velocity* at $\boldsymbol{k}$, which is distinguished from the field-independent band velocity $\boldsymbol{v}(\boldsymbol{k})$.

E. Field-free Bloch Hamiltonian in the Luttinger-Kohn representation

This subsection reviews a set of basis functions which are more convenient to employ near a conical band touching; this will be useful when we derive the effective Hamiltonian near a band degeneracy in Sec. VIII, and derive the Bohr-Sommerfeld quantization conditions in the presence of interband breakdown in Sec. IX.

TABLE II. The first column lists some commonly found symmetries in crystals, and we have employed the notation for symmetries that was introduced in Sec. III C; their corresponding sewing matrices form a representation of the space group, as described in the second column. The third and fourth columns describe general constraints on the orbital moment for the single- and multiband cases. In the last column, we describe the constraints which are specific to twofold spin-degenerate bands in a spin-orbit-coupled system, for which $\check{g}$ transforms in a half-integer-spin representation of the symmetry g; no restrictions of the symmetry representations have been made in the other columns; note that the multiband constraints in the fourth column also apply to spin-degenerate bands. A relation with the qualifier $|_{\boldsymbol{k}}$ applies to all wave vectors; $|_{\boldsymbol{k}=-\check{\mathfrak{c}}_{2z}\boldsymbol{k}}$ applies to any wave vector in the planes defined by $k_z = 0$ and π; finally, $|_{\boldsymbol{k}=-\boldsymbol{k}}$ applies to wave vectors which are invariant under inversion, modulo reciprocal-lattice translations. In the last two rows, we employ the notation $\boldsymbol{k}^{(i)} := \check{\mathfrak{c}}^i_{nz}\boldsymbol{k}$. In the last column, certain canonical choices for the sewing matrices are displayed; i.e., a basis may always be found where $\check{g}$ assumes the displayed forms, assuming that $\check{g}$ transforms in the half-integer-spin representation; by $M^z \propto \sigma_2$, we mean that it is proportional to the Pauli matrix σ_2 with a real, $\boldsymbol{k}$-dependent proportionality constant. As explained in the main text, the form of a symmetry constraint on M^z applies also to the Roth and Zeeman Hamiltonians (H_1^R and H_1^Z); in the latter case, we should fix $F = 1$ for half-integer-spin representations.

g	Space-group rule	Single-band $\boldsymbol{M}$	Multiband $\boldsymbol{M}$	Two-band constraints ($F = 1$)
T	$\check{g}\|_{-\boldsymbol{k}}\check{g}^*\|_{\boldsymbol{k}} = (-1)^F$, $\check{g}\|^T_{\boldsymbol{k}=-\boldsymbol{k}} = (-1)^F\check{g}\|_{\boldsymbol{k}=-\boldsymbol{k}}$	$\boldsymbol{M}\|_{-\boldsymbol{k}} = -\boldsymbol{M}\|_{\boldsymbol{k}}$	$\boldsymbol{M}\|_{-\boldsymbol{k}} = -\check{g}\boldsymbol{M}^*\check{g}^{-1}\|_{\boldsymbol{k}}$	For $\boldsymbol{k} = -\boldsymbol{k}$, $\check{g} = -i\sigma_2$, $\mathrm{Tr}[\boldsymbol{M}] = 0$
$T\mathfrak{i}$	$\check{g}\|_{\boldsymbol{k}}\check{g}^*\|_{\boldsymbol{k}} = (-1)^F$, $\check{g}\|^T_{\boldsymbol{k}} = (-1)^F\check{g}\|_{\boldsymbol{k}}$	$\boldsymbol{M}\|_{\boldsymbol{k}} = 0$	$\boldsymbol{M}\|_{\boldsymbol{k}} = -\check{g}\,\boldsymbol{M}^*\,\check{g}^{-1}\|_{\boldsymbol{k}}$	For all $\boldsymbol{k}$, $\check{g} = -i\sigma_2$, $\mathrm{Tr}[\boldsymbol{M}] = 0$
$T\mathfrak{c}_{2z}$	$\check{g}\|_{-\check{\mathfrak{c}}_{2z}\boldsymbol{k}}\check{g}^*\|_{\boldsymbol{k}} = I$, $\check{g}\|^T_{\boldsymbol{k}=-\check{\mathfrak{c}}_{2z}\boldsymbol{k}} = \check{g}\|_{\boldsymbol{k}=-\check{\mathfrak{c}}_{2z}\boldsymbol{k}}$	$\boldsymbol{M}\|_{-\check{\mathfrak{c}}_{2z}\boldsymbol{k}} = -\check{\mathfrak{c}}_{2z}\boldsymbol{M}\|_{\boldsymbol{k}}$, $M^z\|_{\boldsymbol{k}=-\check{\mathfrak{c}}_{2z}\boldsymbol{k}} = 0$	$\boldsymbol{M}\|_{-\check{\mathfrak{c}}_{2z}\boldsymbol{k}} = -\check{g}\,[\check{\mathfrak{c}}_{2z}\boldsymbol{M}^*]\,\check{g}^{-1}\|_{\boldsymbol{k}}$	For $k_z = 0$ or π, $\check{g} = I$, $M^z \propto \sigma_2$
$T\mathfrak{t}_{\vec{z}/2}$	$\check{g}\|_{-\boldsymbol{k}}\check{g}^*\|_{\boldsymbol{k}} = (-1)^F e^{-ik_z}$, $\check{g}\|^T_{\boldsymbol{k}=-\boldsymbol{k}} = (-1)^F e^{-ik_z}\check{g}\|_{\boldsymbol{k}=-\boldsymbol{k}}$	$\boldsymbol{M}\|_{-\boldsymbol{k}} = -\boldsymbol{M}\|_{\boldsymbol{k}}$	$\boldsymbol{M}\|_{-\boldsymbol{k}} = -\check{g}\boldsymbol{M}^*\check{g}^{-1}\|_{\boldsymbol{k}}$	(i) For $\{k_z = 0, \boldsymbol{k}^\perp = -\boldsymbol{k}^\perp\}$, $\check{g} = -i\sigma_2$, $\mathrm{Tr}[\boldsymbol{M}] = 0$ (ii) For $\{k_z = \pi, \boldsymbol{k}^\perp = -\boldsymbol{k}^\perp\}$, $\check{g} = I$, $\boldsymbol{M} \propto \sigma_2$
$\mathfrak{i}$	$\check{g}\|_{-\boldsymbol{k}}\check{g}\|_{\boldsymbol{k}} = I$	$\boldsymbol{M}\|_{-\boldsymbol{k}} = \boldsymbol{M}\|_{\boldsymbol{k}}$	$\boldsymbol{M}\|_{-\boldsymbol{k}} = \check{g}\boldsymbol{M}\check{g}^{-1}\|_{\boldsymbol{k}}$, $[\check{g}, \boldsymbol{M}]\|_{\boldsymbol{k}=-\boldsymbol{k}} = 0$	$\mathrm{Tr}[\boldsymbol{M}]\|_{-\boldsymbol{k}} = \mathrm{Tr}[\boldsymbol{M}]\|_{\boldsymbol{k}}$
$\mathfrak{r}_x$	$\check{g}\|_{\check{\mathfrak{r}}_x\boldsymbol{k}}\check{g}\|_{\boldsymbol{k}} = (-1)^F$	$\boldsymbol{M}\|_{\check{\mathfrak{r}}_x\boldsymbol{k}} = -[\check{\mathfrak{r}}_x\boldsymbol{M}]\|_{\boldsymbol{k}}$, $M^\alpha\|_{\check{\mathfrak{r}}_x\boldsymbol{k}=\boldsymbol{k}} = 0$, $\alpha\in\{y,z\}$	$\boldsymbol{M}\|_{\check{\mathfrak{r}}_x\boldsymbol{k}} = -\check{g}[\check{\mathfrak{r}}_x\boldsymbol{M}]\check{g}^{-1}\|_{\boldsymbol{k}}$, $[\check{g}, M^x]\|_{\check{\mathfrak{r}}_x\boldsymbol{k}=\boldsymbol{k}} = 0$	$\mathrm{Tr}[\boldsymbol{M}]\|_{\check{\mathfrak{r}}_x\boldsymbol{k}} = -\check{\mathfrak{r}}_x\mathrm{Tr}[\boldsymbol{M}]\|_{\boldsymbol{k}}$
$\mathfrak{g}_{x,\vec{y}/2}$	$\check{g}\|_{\check{\mathfrak{r}}_x\boldsymbol{k}}\check{g}\|_{\boldsymbol{k}} = (-1)^F e^{-ik_y}$	$\boldsymbol{M}\|_{\check{\mathfrak{r}}_x\boldsymbol{k}} = -[\check{\mathfrak{r}}_x\boldsymbol{M}]\|_{\boldsymbol{k}}$, $M^\alpha\|_{\check{\mathfrak{r}}_x\boldsymbol{k}=\boldsymbol{k}} = 0$, $\alpha\in\{y,z\}$	$\boldsymbol{M}\|_{\check{\mathfrak{r}}_x\boldsymbol{k}} = -\check{g}[\check{\mathfrak{r}}_x\boldsymbol{M}]\check{g}^{-1}\|_{\boldsymbol{k}}$, $[\check{g}, M^x]\|_{\check{\mathfrak{r}}_x\boldsymbol{k}=\boldsymbol{k}} = 0$	$\mathrm{Tr}[\boldsymbol{M}]\|_{\check{\mathfrak{r}}_x\boldsymbol{k}} = -\check{\mathfrak{r}}_x\mathrm{Tr}[\boldsymbol{M}]\|_{\boldsymbol{k}}$
$\mathfrak{c}_{nz}$	$\check{g}\|_{\boldsymbol{k}^{(n-1)}}\cdots\check{g}\|_{\boldsymbol{k}^{(1)}}\check{g}\|_{\boldsymbol{k}} = (-1)^F$	$\boldsymbol{M}\|_{\check{\mathfrak{c}}_{nz}\boldsymbol{k}} = [\check{\mathfrak{c}}_{nz}\boldsymbol{M}]\|_{\boldsymbol{k}}$	$\boldsymbol{M}\|_{\check{\mathfrak{c}}_{nz}\boldsymbol{k}} = \check{g}[\check{\mathfrak{c}}_{nz}\boldsymbol{M}]\check{g}^{-1}\|_{\boldsymbol{k}}$, $[\check{g}, M^z]\|_{\check{\mathfrak{c}}_{nz}\boldsymbol{k}=\boldsymbol{k}} = 0$	$\mathrm{Tr}[\boldsymbol{M}]\|_{\check{\mathfrak{c}}_{nz}\boldsymbol{k}} = \check{\mathfrak{c}}_{nz}\mathrm{Tr}[\boldsymbol{M}]\|_{\boldsymbol{k}}$
$\mathfrak{s}_{nz,m}$	$\check{g}\|_{\boldsymbol{k}^{(n-1)}}\cdots\check{g}\|_{\boldsymbol{k}^{(1)}}\check{g}\|_{\boldsymbol{k}} = (-1)^F e^{-imk_z}$	$\boldsymbol{M}\|_{\check{\mathfrak{c}}_{nz}\boldsymbol{k}} = [\check{\mathfrak{c}}_{nz}\boldsymbol{M}]\|_{\boldsymbol{k}}$	$\boldsymbol{M}\|_{\check{\mathfrak{c}}_{nz}\boldsymbol{k}} = \check{g}[\check{\mathfrak{c}}_{nz}\boldsymbol{M}]\check{g}^{-1}\|_{\boldsymbol{k}}$, $[\check{g}, M^z]\|_{\check{\mathfrak{c}}_{nz}\boldsymbol{k}=\boldsymbol{k}} = 0$	$\mathrm{Tr}[\boldsymbol{M}]\|_{\check{\mathfrak{c}}_{nz}\boldsymbol{k}} = \check{\mathfrak{c}}_{nz}\mathrm{Tr}[\boldsymbol{M}]\|_{\boldsymbol{k}}$

The Bloch functions are not an ideal basis for application near conical band touchings, owing to their discontinuity with respect to $\boldsymbol{k}$ at the touching point [61], which we set by convention to $\boldsymbol{0}$. Here, it is convenient to employ the Luttinger-Kohn functions $\{u_{nk_x0}(\boldsymbol{r})e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\}$, which are known to form a complete and orthonormal set of basis functions [72] and are analytically better behaved at $\boldsymbol{k} = \boldsymbol{0}$. To clarify the terminology we employ, Luttinger and Kohn (LK) considered in similar spirit the functions $\{u_{n\boldsymbol{0}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\}$ [72]; we take the liberty of referring to $\{\tilde{u}_{nk_x0}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\}$ as LK functions—the proof of completeness and orthonormality for $\{\tilde{u}_{nk_x0}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\}$ is nearly identical to that presented in Ref. [72].

The Bloch Hamiltonian in the LK representation has the form (with $\hbar = 1$)

$$\begin{aligned}&\langle u_{m,k_x,0}|\hat{H}_0(\boldsymbol{k})|u_{n,k_x,0}\rangle\\ &\quad= \tilde{H}(k_x,0)_{mn} + k_y\tilde{\Pi}^y_{mn}(k_x,0) + \frac{k_y^2}{2m}\delta_{mn},\end{aligned}\tag{42}$$

where we have applied the expansion Eq. (11) around the $k_y = 0$ line, which intersects the band-touching point at $\boldsymbol{k} = \boldsymbol{0}$. It is convenient to choose u_{nk_x0} to be eigenfunctions of the Bloch Hamiltonian $\hat{H}_0(k_x,0)$ [i.e., such that $\tilde{H}(k_x,0)$ is diagonal]; this choice for u_{nk_x0} will be emphasized notationally by adding a tilde: $u_{nk_x0} \to \tilde{u}_{nk_x0}$.

Let us restrict $\{\tilde{u}_{nk_x0}\}_{n=1}^D$ to the D-dimensional subspace projected by P, and further assume that this subspace of bands is D-fold degenerate at $\boldsymbol{k} = \boldsymbol{0}$; we further set the origin of energy such that $H_0(\boldsymbol{0}) = 0$. Applying the identity of Eq. (21) to $\boldsymbol{\nabla}^x_{\boldsymbol{k}}H_0|_{\boldsymbol{0}}$, the Hamiltonian to linear order in k_i simplifies to

$$\langle\tilde{u}_{m,k_x,0}|\hat{H}_0(\boldsymbol{k})|\tilde{u}_{n,k_x,0}\rangle = k_x\Pi^x_{mn}(\boldsymbol{0}) + k_y\Pi^y_{mn}(\boldsymbol{0}),\tag{43}$$

with Π^x and Π^y $D\times D$ diagonal matrices.

It would be useful to transform between the crystal-momentum and Luttinger-Kohn representations by the overlap matrix $\tilde{S}$ defined as

$$\begin{aligned}u_{n\boldsymbol{k}} &= \sum_{l=1}^{\infty}\tilde{u}_{l,k_x,0}\tilde{S}_{ln}(k_x,0,\boldsymbol{k})\\ &= \sum_{l=1}^{D}\tilde{u}_{l,k_x,0}S_{ln}(k_x,0,\boldsymbol{k}) + O(k_y/G_y),\quad\text{with}\\ \tilde{S}(k_x,0,\boldsymbol{k})_{mn} &= \langle\tilde{u}_{mk_x0}|u_{n\boldsymbol{k}}\rangle,\end{aligned}\tag{44}$$

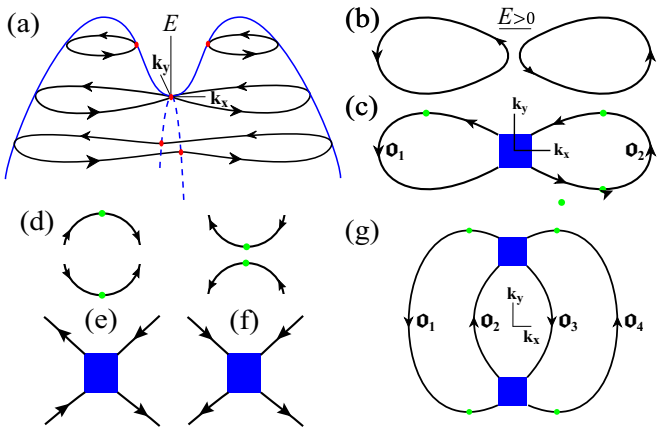

FIG. 1. (a) An example of a zero-field band dispersion with a saddle point. (b) A constant-energy band contour of (a). (c) The graph interpretation of (b); we refer to this as the "double-well" graph. (d) Turning points are degree-two vertices. (e) Intraband-breakdown vertex. (f) Interband-breakdown vertex. (g) The "butterfly" graph.

and G_y a reciprocal period. The projection of $\tilde{S}$ into the D-dimensional subspace is approximated, to an accuracy of $O(k_y/G_y)$, by a unitary matrix S defined by the eigenvalue equation:

$$\sum_{l=1}^{D}[k_x\Pi^x(\mathbf{0})+k_y\Pi^y(\mathbf{0})]_{ml}S_{ln}=\varepsilon_{n,\boldsymbol{k}}S_{mn}, \tag{45}$$

where $u_{n\boldsymbol{k}}$ diagonalizes $H_0(\boldsymbol{k})$ with energy eigenvalue $\varepsilon_{n,\boldsymbol{k}}$.

F. Graph-theoretic description of orbits

With the eventual goal of formulating quantization conditions (with and without breakdown) and their topological invariants, we will find it useful to formulate a graph-theoretic description of orbits. This section is written for easy reference of graph-theoretic terminology that we will eventually employ, and the reader may skip this on a first reading, and refer back to it when necessary.

Any zero-field band structure [as exemplified in Fig. 1(a)], when considered at fixed energy (and fixed k_z for 3D solids) [see Fig. 1(b)], may be represented as a directed graph [Fig. 1(c)]. A directed graph is composed of directed edges and vertices. A directed edge is a continuous line with an orientation; its beginning and end points are referred to as vertices. In our context, a directed edge is a section [black, arrowed trajectory in Fig. 1(c)] of a constant-energy band contour, and the vertices are either turning points (green dots), where the y component of the wave packet velocity vanishes, or breakdown regions (blue squares), where quantum tunneling between orbits is significant.

Each vertex is associated with a degree, which is the number of edges connected to the vertex: a turning vertex has degree two, and a breakdown vertex has degree four. Note that a breakdown region typically has dimension of order $1/l$, but its assumed smallness compared to the size of a typical orbit justifies our use of the term "breakdown vertex." From the orientation of the edges connected to the vertex, we might describe a turning vertex as one-in-one-out, and the breakdown vertex as two-in-two-out. It is useful to assign an orientation to each turning vertex, which might be clockwise or anticlockwise [Fig. 1(d)]—this determines the phase ($-i$ and $+i$, respectively) acquired by a wave packet as it turns, as we elaborate in Sec. V B. It is also useful to assign an orientation to distinguish two classes of breakdown vertices: in the case of the intraband-breakdown vertex [Fig. 1(e)], where tunneling occurs between orbits in the same band, the two incoming edges are parallel and lie on the same diagonal; for the interband-breakdown vertex [Fig. 1(f)], where tunneling occurs between different bands, the incoming edges lie on distinct diagonals. Intraband and interband breakdown is described respectively in Secs. VII and IX. For exemplification, Fig. 1(c) illustrates a "double-well" graph composed of six edges, four turning vertices, and a single intraband-breakdown vertex. Figure 1(g) illustrates a "butterfly" graph composed of eight edges, four turning vertices, and two intraband-breakdown vertices. The quantization conditions for these two graphs will be studied in Secs. VII C and VII D, respectively.

Two further comments regard the type of directed graphs that are relevant to the Bloch electron in a magnetic field. First, we are generally interested in directed multigraphs, which means that we allow for two vertices to be connected by more than one edge; e.g., the two breakdown vertices in the butterfly graph are connected by two edges lying in the middle of the graph. We also insist that our graphs are two-toroidal, by which we mean a graph may be drawn/embedded on a two-torus (here, the Brillouin torus) such that no edges cross.

The notion of connected components is intimately related to quantum tunneling. A connected component is a maximally connected subgraph—each vertex and edge belongs to exactly one connected component, and any two vertices in a connected component can be linked by a path. The appropriate description of orbits which are not linked by tunneling is a disconnected graph with multiple connected components; when the minimal separation in $\boldsymbol{k}$ space between two neighboring components is of order $1/l$, it becomes appropriate to connect the two components by an intraband- or interband-breakdown vertex. In the presence of tunneling, we may define a broken orbit in the following way: it is an oriented subgraph, composed only of directed edges and turning vertices, that forms a continuous path beginning and ending at a breakdown vertex. The beginning and ending vertex may be identical (as for the double well, which is composed of two broken orbits linked by a single breakdown vertex), or distinct (as for the butterfly graph, which is composed of four broken orbits linked by two vertices). The two-in-two-out rule for each breakdown vertex implies there are always two broken orbits which shoot out from the vertex, and another two broken orbits which terminate at the same vertex. The orientation of a broken orbit is determined from Hamilton's equation, and it generally comprises an N_s number of edges and (N_s-1) number of turning vertices, with $N_s \geqslant 1$; if an edge ν (resp. turning vertex p) belongs to a broken orbit $\mathfrak{o}_i$, we denote this by $\nu \in \mathfrak{o}_i$ (resp. $p \in \mathfrak{o}_i$).

For our formulation of a topological invariant in the quantization condition, it is useful to formulate a class of homotopically equivalent graphs. Two equivalent graphs may be continuously deformed into each other, given three rules for what is meant by "continuous": (i) one can neither break apart

a connected component, nor merge two connected components into one, (ii) a breakdown vertex is movable in $\boldsymbol{k}$ space, but unremovable from the graph, and (iii) the total number of turning vertices is not invariant, as explained in Sec. V B, however the net circulation of all turning vertices on a connected path is invariant.

Definition. A topological invariant in magnetic transport is a quantity that is invariant under continuous deformations of the zero-field Hamiltonian that preserve the homotopy class of the graph. A symmetry-protected topological invariant in magnetic transport is a quantity that is invariant under continuous deformations of the zero-field Hamiltonian that (i) preserves the homotopy class of the graph, as well as (ii) respects the symmetry of the zero-field Hamiltonian.

IV. REVIEW OF EFFECTIVE HAMILTONIAN IN THE ABSENCE OF INTERBAND BREAKDOWN

We are interested in semiclassical approximations to the exact Hamiltonians of a Bloch electron in a magnetic field, as shown in Eqs. (39) and (40); such approximations will be referred to as effective Hamiltonians. In this section, we particularize to cases where interband breakdown is negligible; the effective Hamiltonian that is valid near an interband degeneracy takes a different form that is described in Sec. IX.

In the presence of a field along $\vec{z}$, $\boldsymbol{k}^{\perp}=(k_x,k_y)$ is no longer a conserved quantity for the Bloch electron. In the lowest-order approximation, the effective Hamiltonian is obtained by the Peierls substitution [1]:

$$H_0(\boldsymbol{k}) \leftrightarrow H_0(\boldsymbol{K}), \tag{46}$$

which describes the unique, Weyl correspondence between a function of commuting variables (k_x,k_y), and a function of noncommuting variables:

$$\boldsymbol{K}=\boldsymbol{k}+\boldsymbol{a}(i\nabla_{\boldsymbol{k}}). \tag{47}$$

We refer to $\boldsymbol{K}$ as the kinetic quasimomentum operators, and their noncommutivity is manifest in

$$\boldsymbol{K}\times\boldsymbol{K}=-i\frac{e}{c}\boldsymbol{B}, \quad e<0. \tag{48}$$

Generally, a one-to-one correspondence exists between a classical "symbol" $[A(\boldsymbol{k})]$ and an operator $[A(\boldsymbol{K})]$, if $A(\boldsymbol{k})$ is a Fourier-invertible function of commuting variables, with the Fourier transform $\check{A}(\boldsymbol{r})$:

$$A(\boldsymbol{k})=\int d\boldsymbol{r}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\check{A}(\boldsymbol{r}). \tag{49}$$

Many of the functions we deal with, including the matrix $H_0(\boldsymbol{k})$, are periodic in reciprocal-lattice translations: $\boldsymbol{k}\to\boldsymbol{k}+\boldsymbol{G}$, in which case Eq. (49) particularizes to a Fourier-series expansion. The operator to which the symbol corresponds is

$$A(\boldsymbol{K}):=[A(\boldsymbol{k})]:=\int d\boldsymbol{r}e^{i\boldsymbol{K}\cdot\boldsymbol{r}}\check{A}(\boldsymbol{r}). \tag{50}$$

To make this definition rigorous, one assumes certain regularity conditions on A, and checks that the integral converges in some suitable sense [73]. The lowest-order effective Hamiltonian $H_0(\boldsymbol{K})$ was first derived in a tight-binding approximation [5]; its form may be argued from general principles of electromagnetic gauge invariance [74]. However, we cannot appeal to gauge invariance to predict the form of higher-order corrections, which may be organized in an asymptotic [52,75] expansion:

$$\mathcal{H}(\boldsymbol{K})=H_0(\boldsymbol{K})+H_1(\boldsymbol{K})+H_2(\boldsymbol{K})+\cdots, \tag{51}$$

where each term in the expansion corresponds to the symbol $H_j(\boldsymbol{k})=O(l^{-2j})$. By $O(l^{-2j})$, we mean that H_j is of the order $(a/l)^{2j}$, where a is a typical lattice period. H_j is obtained systematically [24] by expanding an eigenstate of $\hat{H}$, defined by

$$(\hat{H}-E)\Psi_E=0, \tag{52}$$

in a complete [50,76] basis of field-modified Bloch functions

$$\Psi_E(\boldsymbol{r})=\sum_{n\boldsymbol{k}}g_{n\boldsymbol{k}E}\phi_{n\boldsymbol{k}}(\boldsymbol{r}), \tag{53}$$

such that, in an asymptotic sense,

$$\sum_n[\mathcal{H}(\boldsymbol{K})_{mn}-E\delta_{mn}]g_{n\boldsymbol{k}E}=0. \tag{54}$$

Note that by $\sum_{\boldsymbol{k}}$ we really mean a continuous integral over the Brillouin torus. There are several different proposals for the best basis functions to formulate an effective Hamiltonian [14,23,24,48,50], but all these proposals agree [77] to lowest order in l^{-2} [5,51,78]:

$$\phi_{n\boldsymbol{k}}(\boldsymbol{r})=e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{n,\boldsymbol{k}+\boldsymbol{a}(\boldsymbol{r})}(\boldsymbol{r}). \tag{55}$$

This form of ϕ manifests the semiclassical intuition that for a slowly varying vector potential, the ordinary Bloch function is modified locally in space, but only through the wave vector dependence of the cell-periodic component: $u_{n\boldsymbol{k}}\to u_{n,\boldsymbol{k}+\boldsymbol{a}}$, which is no longer periodic in Bravais-lattice translations; we provide a further argument that motivates the form of ϕ in Appendix A 1. The Fourier transform of ϕ is a real-space function that may be obtained from applying the magnetic translation [79] to a Wannier function; in this real-space basis, the effective-Hamiltonian equation is a finite-difference equation for a wave function defined on a lattice [17], as famously exemplified by the Harper equation [80].

The value of Eq. (54) is that, in many cases of interest, the matrix elements between a single band and its complement (i.e., all other bands) have been removed perturbatively in the parameter l^{-2}; this decoupling of bands is asymptotic and fails if interband gaps become too small [52]. Assuming otherwise, we may truncate $\sum_n$ and solve for $[\mathcal{H}(\boldsymbol{K})_{nn}-E]g_{n\boldsymbol{k}E}=0$; in this sense we say $\mathcal{H}(\boldsymbol{K})_{nn}$ is a one-band effective Hamiltonian. The Weyl correspondence thus provides the link between the magnetic problem [given by $\mathcal{H}(\boldsymbol{K})_{nn}$] and band properties at zero field: $[H_0]_{nn}(\boldsymbol{k})$ describes the dispersion of a single band, which is renormalized [26,81,82] by the higher-order $\{H_j(\boldsymbol{k})\}$ as we eliminate degrees of freedom in the other bands. In other applications of Eq. (54), we may utilize a multiband effective Hamiltonian to describe a degenerate band subspace.

The effective-Hamiltonian theory has been rigorously justified for an energetically isolated nondegenerate band [50]; the justification for a finite family of (possibly magnetic) bands was achieved only recently [83]. While these impressive works go a long way in solving "one of the few unsolved problems of one-particle quantum mechanics" [84], they rely on the assumption of a strictly isolated band (or family of bands), i.e., that there exists a direct energy gap above and

below the band(s) in question. The complicated nature of bands in naturally occurring crystals often implies that indirect gaps are as common as direct gaps, except in the extreme tight-binding limit; this problem is especially severe in highly symmetrical crystals with many band touchings, of both the immovable [40] and movable kinds [85]. We further highlight a class of metallic systems where the nonexistence of a gap is guaranteed from topological principles—the surface states of certain topological insulators robustly interpolate between conduction and valence bands [54,55,86]; this phenomenon of spectral flow is familiar from the integer quantum Hall effect.

From physical grounds, one may expect that the effective Hamiltonian is valid for bands which are not energetically isolated over the Brillouin torus. That is, the existence of well-defined semiclassical orbits (at some energy E) at least validates $\mathcal{H}(\boldsymbol{K})$ in a local neighborhood of the orbit and at that energy E. Let us exemplify our perspective for graphene in a magnetic field—we will apply a single-band $\mathcal{H}$ at wave vectors sufficiently far (on the scale of l) from the Dirac point, even though it is impossible to symmetrically separate graphene's two bands (not counting spin). This impossibility is enforced by symmetry; i.e., the p_z bands of graphene form an elementary band representation [87–90] with two branches. Generally, the existence of semiclassical orbits leads to discrete magnetic energy levels (henceforth referred to as Landau levels) which may be determined by Bohr-Sommerfeld quantization rules [2–4,26,51,52]; this method has been verified numerically for simple models [91] and is predictive [32] of de Haas–van Alphen oscillations in metals [9,10]. We further substantiate our perspective for the single-band effective Hamiltonian in Sec. IV A, and for the multiband case in Sec. IV B.

A. Single-band effective Hamiltonian

The applicability of the single-band $\mathcal{H}_{nn}(\boldsymbol{K})$ generically depends on the wave vector and energy in question, and is contingent on the cell-periodic functions (in this one-band subspace) being smooth enough. By smooth enough, we mean that $\partial_k u = O(a)$ with a a typical lattice period; there may be isolated regions in $\boldsymbol{k}$ space where such smoothness cannot hold; e.g., where two bands touch at a conical degeneracy (a Dirac point), the Berry connection (with $\boldsymbol{\nabla}_{\boldsymbol{k}}$ in the azimuthal direction) diverges [61]. Even so, we may apply $\mathcal{H}_{nn}(\boldsymbol{K})$ at wave vectors sufficiently far (on the scale of l^{-1}) from the conical degeneracy. For notational convenience, we henceforth drop the subscript $\mathcal{H}(\boldsymbol{K})_{nn} \to \mathcal{H}(\boldsymbol{K})$. These are common scenarios in which we might consider a single-band effective Hamiltonian:

(i) Low-symmetry wave vectors where all bands are nondegenerate, e.g., a generic wave vector for a spin-orbit-coupled system without spacetime-inversion symmetry. By spacetime inversion, we mean a simultaneous inversion of both space and time (denoted Ti in later sections), which is known to result in spin-degenerate bands.

(ii) Spin-degenerate bands in the absence of spin-orbit and Zeeman couplings. Since S_z (spin component in the $\vec{z}$ direction) is conserved, electron dynamics in a field is effectively constrained within a single band.

Explicit expressions for the single-band effective Hamiltonian have been derived up to $H_2 = O(l^{-4})$ [24]; in this work, we derive the quantization conditions for the truncated $H_0 + H_1$, where

$$H_0(\boldsymbol{k}) = \varepsilon_{n\boldsymbol{k}} \tag{56}$$

is the energy-momentum dispersion of a band (labeled by an integer n), and H_1 may be split into gauge-dependent (H_1^B) and -independent (H_1^R, H_1^Z) terms as [24,92]

$$H_1(\boldsymbol{k}) = H_1^B(\boldsymbol{k}) + H_1^R(\boldsymbol{k}) + H_1^Z(\boldsymbol{k}), \tag{57}$$

$$H_1^B(\boldsymbol{k}) = l^{-2}\epsilon^{\alpha\beta}\mathfrak{X}^\beta v_n^\alpha, \tag{58}$$

$$H_1^R(\boldsymbol{k}) = \frac{1}{2l^2}\epsilon^{\alpha\beta}\sum_m \tilde{\mathfrak{X}}^\beta_{nm}(\tilde{\Pi}^\alpha - \tilde{v}^\alpha)_{mn}, \tag{59}$$

$$H_1^Z(\boldsymbol{k}) = -\frac{g_0\hbar^2}{4ml^2}\sigma^z. \tag{60}$$

Here, band indices m and n are not summed over unless explicitly stated; $\epsilon^{\alpha\beta}$ is the Levi-Cevita tensor with $\epsilon^{xy} = 1 = -\epsilon^{yx}$; $\boldsymbol{v}_n := \boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon_n$; and $\tilde{\boldsymbol{\Pi}}$, $\tilde{\mathfrak{X}}$, $\tilde{\boldsymbol{v}}$, $\boldsymbol{v}$, σ^z, and $\mathfrak{X}$ are $\boldsymbol{k}$-dependent matrices defined in Eqs. (13), (15), (16), and (20), respectively; in particular, $\mathfrak{X} = \tilde{\mathfrak{X}}_{nn}$ in this context. By gauge dependency, we refer to a phase ambiguity in the cell-periodic functions of band n [cf. Eq. (22) with $D = 1$], and hence also of the field-modified Bloch functions which form our basis [cf. Eq. (53)]; this results in H_1^B being not uniquely defined:

$$\begin{aligned} |u_{n\boldsymbol{k}}\rangle &\to |u_{n\boldsymbol{k}}\rangle e^{i\phi_n(\boldsymbol{k})} \\ &\Rightarrow H_1^B \to H_1^B + l^{-2}\epsilon^{\alpha\beta}\partial_\alpha\phi_n v_n^\beta. \end{aligned} \tag{61}$$

We will shortly demonstrate that the quantization condition is nevertheless gauge-invariant. If a symmetry (e.g., Ti) constrains the Berry curvature to vanish, a basis may be found such that the Berry connection $\mathfrak{X}_{nn}(\boldsymbol{k})$ (hence also H_1^B) vanishes at any $\boldsymbol{k}$; this basis may be continuously defined over the Brillouin torus unless there is a topological obstruction, which may originate from a Dirac point in 2D, or a line node in 3D [21].

On the other hand, H_1^R may be expressed in a manner that manifests its gauge invariance:

$$\sum_m \tilde{\mathfrak{X}}^\alpha_{nm}(\tilde{\Pi}^\beta - \tilde{v}^\beta)_{mn} = i\langle u_{n\boldsymbol{k}}|(\partial_\alpha Q)\hat{\Pi}^\beta|u_{n\boldsymbol{k}}\rangle, \tag{62}$$

where Q is the gauge-invariant, cell-periodic projections defined in Eq. (19), for $D = 1$. As we will demonstrate in Sec. V A, the WKB wave function of $\mathcal{H} = H_0 + H_1$ includes multiplicatively a geometric Berry phase factor [20] that originates from the gauge-dependent H_1^B (henceforth called the Berry term), and a nongeometric phase factor that originates from the gauge-independent H_1^R. We interpret H_1^R as a coupling ($-\boldsymbol{M}_n \cdot \boldsymbol{B}$) of the field to the band orbital moment, defined for band n as

$$\begin{aligned} M(\boldsymbol{k})_n^\alpha &= -\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{\bar{m}} \tilde{\mathfrak{X}}^\beta_{n\bar{m}}(\tilde{\Pi}^\gamma - \tilde{v}^\gamma)_{\bar{m}n} \\ &= i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{\bar{m}} \frac{\tilde{\Pi}^\beta_{n\bar{m}}\tilde{\Pi}^\gamma_{\bar{m}n}}{(\varepsilon_n - \varepsilon_{\bar{m}})} \\ &= -i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\langle\partial_\beta u_n|\hat{H}_0(\boldsymbol{k}) - \varepsilon_{n\boldsymbol{k}}|\partial_\gamma u_n\rangle. \end{aligned} \tag{63}$$

Here, we have used $\bar{m}$ to label bands which are orthogonally complementary to band n [cf. Eq. (19)]. These equivalent expressions for the orbital moment in Eqs. (62) and (63) are derived in Appendix A 2.

An expression identical to Eq. (63) appears in the correction to the energy of a wave packet in a Bloch band [22,33,34]. We, however, disagree with a claim in Ref. [22] that the orbital moment is absent in nonmagnetic Bloch bands (i.e., eigenstates of $\hat{H}_0$ without spontaneous time-reversal-symmetry breaking); we substantiate this point by a comprehensive symmetry analysis of H_1 in Sec. VI. The derivation of the semiclassical equations of motion, as corrected by H_1, was accomplished in Ref. [92]. H_1^R is sometimes referred to as the Rammal-Wilkinson term, and has been alternatively derived from a purely algebraic approach [93], as well as in a semiclassical treatment of the Harper-Hofstadter model [26,94]. Finally, we remark that terms analogous to H_1^R and H_1^B appear ubiquitously in the asymptotic theory of coupled-wave equations (i.e., multicomponent WKB theory) [27–29] as well as in space-adiabatic perturbation theory [30], which apply in a much wider variety of physical contexts than the present study.

B. Multiband effective Hamiltonian

Consider a multiband effective Hamiltonian that describes a D-fold degenerate band subspace projected by P [cf. Eq. (18)]. A common example of $D = 2$ arises in spin-orbit-coupled solids with spacetime-inversion symmetry—bands are spin-degenerate at generic wave vectors, and dynamics in a magnetic field is described by a two-band effective Hamiltonian $[\mathcal{H}(\boldsymbol{K})]$. $\mathcal{H}(\boldsymbol{K})$ loses its applicability near (on the scale of $1/l$) fourfold-degenerate band touchings which might occur in various contexts, e.g., (i) a 3D Dirac point, which is the critical point of a topological phase transition between trivial and topological insulators [41], or (ii) a symmetry-protected degeneracy that can be found in nonsymmorphic space groups [40,43].

For any D, the multiband generalization of Eq. (57) is [24]

$$H_1 = l^{-2}\epsilon_{\alpha\beta}\left[\frac{1}{2}\{\tilde{\Pi}^\alpha - \tilde{v}^\alpha, \tilde{\mathfrak{X}}^\beta\} + \mathfrak{X}^\beta v^\alpha\right] - \frac{g_0\hbar^2}{4ml^2}\sigma^z = H_1^R + H_1^B + H_1^Z, \tag{64}$$

where $\{a,b\} = ab + ba$, and we consider only matrix elements of H_1 within the P subspace. $H_1^B \propto \mathfrak{X}^\alpha v^\beta$ is just the product of two $D \times D$ matrices; in contrast, since $\boldsymbol{v}$ is the diagonal component of $\boldsymbol{\Pi}$ [cf. Eq. (13)], the first term in Eq. (64) involves *only* matrix summations between P and Q subspaces:

$$[(\tilde{\Pi}^\beta - \tilde{v}^\beta)\tilde{\mathfrak{X}}^\alpha]_{mn} = \sum_{\bar{l}}[\tilde{\Pi}^\beta - \tilde{v}^\beta]_{m\bar{l}}\tilde{\mathfrak{X}}^\alpha_{\bar{l}n}. \tag{65}$$

While H_1^Z has the advantage of looking more symmetric with respect to α and β, the following alternative expressions reveal a closer resemblance to the one-band H_1^Z in Eq. (57):

$$\frac{\epsilon_{\alpha\beta}}{2}\{\tilde{\Pi}^\beta - \tilde{v}^\beta, \tilde{\mathfrak{X}}^\alpha\} = \epsilon_{\alpha\beta}\tilde{\mathfrak{X}}^\alpha(\tilde{\Pi}^\beta - \tilde{v}^\beta) = -\epsilon_{\alpha\beta}(\tilde{\Pi}^\alpha - \tilde{v}^\alpha)\tilde{\mathfrak{X}}^\beta. \tag{66}$$

The multiband orbital moment, defined by $H_1^R = -\boldsymbol{M}\cdot\boldsymbol{B}$, therefore has a very similar form to Eq. (63):

$$M(\boldsymbol{k})^\alpha_{mn} = -\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{\bar{l}}\tilde{\mathfrak{X}}^\beta_{m\bar{l}}(\tilde{\Pi}^\gamma - \tilde{v}^\gamma)_{\bar{l}n} = i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{\bar{l}}\frac{\tilde{\Pi}^\beta_{m\bar{l}}\tilde{\Pi}^\gamma_{\bar{l}n}}{(\varepsilon_m - \varepsilon_{\bar{l}})}. \tag{67}$$

We stress that the multiband orbital moment nontrivially affects the Landau levels, which motivates a comprehensive symmetry analysis of H_1^R in Sec. VI B. Through H_1^B, the energy levels are also sensitive to the non-Abelian gauge structure in the subspace P, as we will demonstrate in the next section (Sec. V).

V. QUANTIZATION CONDITIONS FOR CLOSED ORBITS WITHOUT BREAKDOWN

As motivated in the last paragraphs of Sec. IV, we are interested in determining Landau levels from Bohr-Sommerfeld quantization rules. In this section we derive the rules for *closed* orbits, by which we mean orbits that do not extend beyond one unit cell in $\boldsymbol{k}$ space. These clearly do not exhaust all possible orbits [18], but they are sufficient to exemplify the results of this work; we will briefly remark on generalizations beyond closed orbits in Sec. V C 1. We further particularize to isolated orbits whose closest distance to any other orbit (if they exist) is much greater than $1/l$, with l the magnetic length. If this condition is violated, tunneling between orbits must be accounted for; generalized quantization conditions that incorporate tunneling are presented in Secs. VII and IX.

Let us first summarize our results, which we will derive in the subsequent subsections. For a closed orbit ($\mathfrak{o}$) corresponding to a nondegenerate band (labeled n) of the Pauli Hamiltonian [Eq. (4)], the Bohr-Sommerfeld quantization rule is

$$l^2 S[\mathfrak{o}] + \phi_M + \oint_{\mathfrak{o}}(\mathfrak{X} + \mathfrak{A})\cdot d\boldsymbol{k} + Z\oint_{\mathfrak{o}}\sigma^z\frac{dk}{v^\perp}\bigg|_{E=E_j,k_z} = 2\pi j + O(l^{-2}), \quad j \in \mathbb{Z}. \tag{68}$$

The left-hand side of Eq. (68) comprises five terms which we define in their order of appearance:

(i) The first term is a dynamical phase that is proportional to the $\boldsymbol{k}^\perp$-space area S bounded by $\mathfrak{o}_j$, with S being positive (resp. negative) for a clockwise-oriented (resp. anticlockwise) orbit.

(ii) The second term is a Maslov phase [95], e.g., $\phi_M = \pi$ for orbits which are deformable to a circle, and equals 0 for figure-of-eight orbits, as elaborated in Sec. V B. To leading order in the field, $l^2 S + \phi_M = 2\pi(j + 1/2)$ is a well-known result by Onsager and Lifshitz [2–4]. The Landau-level degeneracy ($\mathcal{N}$) may be obtained from the following semiclassical phase-space argument [96]: $\mathcal{N}$ equals the phase-space density of states $[(2\pi)^{-2}$ for two spatial dimensions, which we assume in this paragraph for simplicity], multiplied by the phase-space volume ($\delta\mathcal{V}$) in between two constant-energy hypersurfaces; these hypersurfaces correspond to nearest-neighbor Landau levels indexed by adjacent integers in the quantization rule,

hence

$$\mathcal{N} = \frac{\delta\mathcal{V}}{(2\pi)^2} = \frac{\bar{A}\delta S}{2\pi^2} = \frac{\bar{A}}{2\pi l^2}, \tag{69}$$

with $\bar{A}$ the real-space area of the 2D solid. This degeneracy simply reflects that a single-particle state undergoing localized cyclotron motion occupies an average area of $2\pi l^2$ in the semiclassical limit.

(iii) Beyond Onsager-Lifshitz, $\oint \mathfrak{X}$ is the single-band Berry phase acquired over a single cyclotron period [$\mathfrak{X} = \tilde{\mathfrak{X}}_{nn}$ as defined in Eq. (15)]. We might utilize Stoke's theorem to combine terms (i) and (iii) as

$$l^2 S[\mathfrak{o}] + \oint_{\mathfrak{o}} \mathfrak{X} \cdot d\boldsymbol{k} = l^2 \tilde{S},$$
$$\tilde{S} := \int |d^2\boldsymbol{k}|[1 - l^{-2}\mathcal{F}^z(\boldsymbol{k})], \tag{70}$$

with the Berry curvature defined by $\mathcal{F}^z = \epsilon_{\alpha\beta}\nabla^\alpha_{\boldsymbol{k}}\mathfrak{X}^\beta$. One may therefore interpret $(2\pi)^{-2}[1 - l^{-2}\mathcal{F}^z(\boldsymbol{k})]$ as the Berry-corrected, phase-space density of states for a 2D solid immersed in a spatially homogeneous field; i.e., a single-particle state occupies a volume in phase space that is modified by the coupling of the magnetic field to the Berry curvature. This correction to the phase-space density of states may alternatively be derived from a different route: through the semiclassical equations of motion [97–99], which ultimately also derives from the effective-Hamiltonian formalism [49,74].

(iv) The fourth term is the line integral of a one-form that encodes the orbital magnetic moment:

$$\mathfrak{A} \cdot d\boldsymbol{k} := \frac{[\tilde{\mathfrak{X}}^x(\tilde{\Pi}^y - \tilde{v}^y)]_{nn}}{2v_n^y} dk_x + (x \leftrightarrow y), \tag{71}$$

where $\tilde{\mathfrak{X}}^x(\tilde{\Pi}^y - \tilde{v}^y)$ is formally the product of two infinite-dimensional matrices expressed in Eq. (62). $\oint \mathfrak{A} \cdot d\boldsymbol{k}$ is not a geometric phase because it depends on the rate at which the orbit is traversed [recall that the orbit velocity is related to the band velocity through Eq. (41)]. Analogous expressions of $\oint \mathfrak{A} \cdot d\boldsymbol{k}$ have been called, in various contexts, the "no-name" phase [27,28,100], and sometimes the Ramal-Wilkinson phase. However, we will refer to it as the Roth phase to honor its first discoverer [24,25] in the context of Bloch electrons in a magnetic field.

(v) Finally, $Z \oint \sigma^z/v^\perp dk$ is the Zeeman energy of the nondegenerate band integrated over the orbit; the $\boldsymbol{k}$ dependence of the Zeeman energy originates from spin-orbit coupling. Note $Z := g_0\hbar/4m$, $dk = |d\boldsymbol{k}|$, $v^\perp := (v_x^2 + v_y^2)^{1/2}$, $\sigma^z(\boldsymbol{k}) := \sigma^z_{nn}(\boldsymbol{k}) = \langle u_{n\boldsymbol{k}}|\hat{\sigma}^z|u_{n\boldsymbol{k}}\rangle \in \mathbb{R}$.

All of (i)–(v) may be evaluated knowing the band structure at zero field; (i) and (ii)–(v) depend continuously on the energy of the orbit E; in three spatial dimensions, they depend additionally on the wave vector (k_z) parallel to the field. Equation (68) leads to discrete, macroscopically degenerate Landau levels labeled as E_j; more details about the spectrum, as well as consequences in dHvA oscillations, are described in Sec. V D.

In the absence of spin-orbit coupling, a spin-degenerate band results in spin-split Landau levels obtained from the two quantization conditions:

$$l^2 S[\mathfrak{o}] + \phi_M + \oint_{\mathfrak{o}} (\mathfrak{X} + \mathfrak{A}) \cdot d\boldsymbol{k} \pm \pi \frac{g_0}{2} \frac{m_c[\mathfrak{o}]}{m}\bigg|_{E=E_{\pm,j},k_z}$$
$$= 2\pi j + O(l^{-2}), \quad j \in \mathbb{Z}, \tag{72}$$

where $m_c := (\hbar^2/2\pi)\partial S/\partial E$ is the cyclotron mass for the orbit $\mathfrak{o}$. Despite the notational similarity of Eqs. (68) and (72), we remind the reader that the velocity matrix $\tilde{\boldsymbol{\Pi}}$ is defined differently when there is no spin-orbit coupling [cf. Eq. (10)]. In spite of the spin degeneracy of the bands at zero field, we analyze this case under the heading of "single band" (e.g., in Sec. V A 1) because the field-on Hamiltonian may be block-diagonalized with respect to the spin quantum number $S_z = \pm\hbar/2$; all "single-band" statements are then understood to apply to either of $S_z = \pm\hbar/2$, and all symmetries that we consider preserve S_z.

Equations (68) and (72) may be derived from the condition of continuity of the WKB wave function around the closed orbit [17,52]. This wave function is derived in Sec. V A, where we also demonstrate that the Berry and Roth phases are respectively generated from H_1^B and H_1^R. The additional phase of π on the right-hand side of Eq. (68) [and also of Eq. (72)] is a Maslov correction that we derive in Sec. V B; here, we argue that previous derivations [17,52] of the Maslov correction introduce an uncertainty of $O(l^{-2/3})$, which we reduce to $O(l^{-2})$ in an improved derivation. We combine these results in Sec. V C 1 to finally derive Eq. (68), and further discuss experimental signatures in quantum oscillations in Sec. V D.

For a closed orbit ($\mathfrak{o}$) corresponding to a D-fold-degenerate band subspace, the quantization condition is

$$l^2 S(E_{a,j},k_z) + \phi_M + \lambda_a(E_{a,j},k_z)$$
$$= 2\pi j + O(l^{-2/3}), \quad j \in \mathbb{Z},\ a \in \mathbb{Z}_D, \tag{73}$$

where $\{e^{i\lambda_a}\}_{a=1}^D$ is the spectrum of the unitary propagator

$$\mathcal{A}[\mathfrak{o}] = \overline{\exp}\left[i \oint_{\mathfrak{o}} \{(\mathfrak{A} + \mathfrak{X}) \cdot d\boldsymbol{k} + Z(\sigma^z/v^\perp)\, dk\} \right]. \tag{74}$$

Here, $\overline{\exp}$ denotes a path-ordered exponential, and we employ the same symbol $\mathfrak{X}$ for both the Abelian [as in Eq. (68)] and non-Abelian [as in Eq. (73)] Berry connection. The non-Abelian generalization of the Abelian Roth one-form [in Eq. (71)] is

$$(\mathfrak{A} \cdot d\boldsymbol{k})_{mn} = \frac{[\tilde{\mathfrak{X}}^x(\tilde{\Pi}^y - \tilde{v}^y)]_{mn}}{2v^y} dk_x + (x \leftrightarrow y), \tag{75}$$

with $m,n = 1,2,\ldots,D$. Due to the assumed degeneracy within P, the band velocity $\boldsymbol{v}_1 = \cdots = \boldsymbol{v}_D := \boldsymbol{v}$. Equation (73) leads to D sets of Landau levels (labeled by the a subscript on $\{E_{a,j}\}$). Landau levels within each set are locally periodic; i.e., the difference between two adjacent Landau levels ($E_{a,j+1} - E_{a,j}$) is approximately $2\pi/l^2(\partial S/\partial E)$ evaluated at $E_{a,j}$, as elaborated in Sec. V D.

The quantizations rule in Eqs. (73) and (74) may be compared with previous works. For Ti-symmetric, spin-orbit-coupled systems ($D = 2$), two-band quantization conditions have been derived [31] with an "equation-of-motion" method [25], which leads to formulating $\{e^{i\lambda_a}\}$ as eigenvalues of a

complex Ricatti equation [31]. Their method presupposes a special basis for the Bloch functions (i.e., a special gauge) in which the matrix exponent in Eq. (74) is traceless, as elaborated in Sec. C 4 b; note that $\{e^{i\lambda_a}\}$ are gauge-independent, so in principle their and our methods should converge to the same quantization rule for this symmetry class. In other formulations of the quantization rule for spin-degenerate bands, the Berry phase and/or orbital moment have either been neglected explicitly [19] or derived in a form that is difficult for comparison [101,102]. On the other hand, Eqs. (73) and (74) represent the quantization condition in its most general form, which would apply to any symmetry class, and to bands of any energy degeneracy (D). Since no special gauge was assumed in our expressions, they are useful for numerical computations where gauge fixing is often troublesome. One further contribution we make is a comprehensive, group-theoretic analysis of the propagator in Eq. (74), which determines in complete generality the symmetry constraints on the Landau levels (see Sec. VI D). In a complementary perspective, multiband wave packet theory has been derived in Ref. [35], and reviewed in Ref. [34] with notation that is closer to ours. They derived an equation of motion for a multicomponent wave packet that is also sensitive to the non-Abelian gauge structure; however, their dynamical equations are nontrivially coupled, and it is unclear to us whether a non-Abelian quantization rule can be derived in their approach.

Our results may plausibly be applied to charge-neutral, cold-atomic systems (e.g., optical lattices of bosonic cold atoms, degenerate Fermi gases) described by the field-on Schrödinger Hamiltonian [Eq. (39)] with an artificially induced gauge field. It is possible to mimic a magnetic field by (i) rotation of a Bose-Einstein condensate [103], (ii) coupling neutral fermionic atoms to slow light [104,105], and (iii) laser-assisted tunneling in an optical lattice [106]. In the limit of weak interactions, this leads to the quantization of energy levels which are analogous to Landau levels [107].

We remark on one caveat to the above discussion: energy quantization of closed orbits is never strictly correct in a solid. While the magnetic field tends to quantize electronic motion and form discrete levels, the crystalline potential tends to form bands. From the perspective of semiclassical orbits in $\boldsymbol{k}$ space, there generally exists a nonzero tunneling probability between closed orbits in distinct Brillouin zones. This leads to broadening of the Landau levels that cannot be accounted for with the above quantization rules. While this broadening is usually exponentially small in the field [14,108], it cannot be neglected in narrow energy ranges where the separation of orbits is of order $1/l$; this frequently occurs at saddle points at the Brillouin-zone edge [109,110].

A. WKB wave function of effective Hamiltonians

1. Single-band WKB wave function

We look for an eigenfunction of the single-band $\mathcal{H}(\boldsymbol{K}) = H_0 + H_1$ with the WKB ansatz

$$g_{\boldsymbol{k}} = e^{-i\psi_{\boldsymbol{k}}} \quad \text{with} \quad \psi = \psi_{-1} + \psi_0 + \psi_1 + \cdots, \tag{76}$$

where the subscript denotes the order in the WKB parameter l^{-2}. Any function that is asymptotically expandable as Eq. (76) will be called a WKB function. In the classically allowed regions, $\psi_{-1} \in \mathbb{R}$ is the integral of a classical action, while higher-order $\psi_j \in \mathbb{C}$. In the Landau gauge $\boldsymbol{A} = (By,0,0)$, the kinetic quasimomentum operators are

$$K_x = k_x + il^{-2}\partial_y, \quad K_y = k_y, \tag{77}$$

and we will look for wave functions over the circle parametrized by k_y, with k_x a good quantum number. We shall refer to this as the wave function in the (K_x,k_y) representation. In this representation, $\mathcal{H}$ may be solved as

$$[H_0(\boldsymbol{K}) + H_1(\boldsymbol{K}) - E]g^{\nu}_{\boldsymbol{k}E} = O(l^{-4}), \tag{78}$$

$$g^{\nu}_{\boldsymbol{k}E} = \frac{1}{\sqrt{|v^x_\nu|}} e^{ik_xk_yl^2} e^{-il^2\int[k^\nu_x - H^\nu_1(v^x_\nu)^{-1}]dk_y}. \tag{79}$$

Here, all quantities carrying a ν superscript or subscript depend on k_y and E; they also depend, in three spatial dimensions, on the wave vector k_z, but we shall henceforth omit this notationally. As a case in point, $k^\nu_x(k_y,E)$ should be distinguished from the continuous parameter k_x. k^ν_x describes an oriented *edge* (labeled by ν) of the zero-field band contour ($\mathfrak{o}$) at fixed energy E, with the orientation prescribed by Hamilton's equation; each k^ν_x corresponds to a single-valued solution of

$$H_0(k^\nu_x(k_y,E),k_y) = E. \tag{80}$$

The constant-energy contour of a single band may be divided into multiple edges; e.g., a closed contour has at least two edges. A more elaborate, graph-theoretic description of edges is provided in Sec. III F. For s_ν corresponding to a band index n, we further define

$$\begin{aligned} v^x_\nu(k_y,E) &:= v^x_n\big(k^\nu_x(k_y,E),k_y\big) \quad \text{and} \\ H^\nu_1(k_y,E) &:= H_1\big(k^\nu_x(k_y,E),k_y\big), \end{aligned} \tag{81}$$

as the band velocity and the first-order Hamiltonian [cf. Eq. (57)] evaluated on the edge ν. The single-band WKB wave function of H_0 was first derived by Zilberman [51]; Fischbeck later derived the corrections due to H_1 and H_2 [52], of which we have shown only the first-order correction in Eq. (79). We will therefore refer to Eq. (79) as the Zilberman-Fischbeck function; the same expression without the H_1 correction will be referred to as the Zilberman function. With sufficient hindsight, we may now identify the Roth, Berry, and Zeeman phases as being generated, respectively, by H^R_1, H^B_1, and H^Z_1:

$$\begin{aligned} l^2\int H_1\frac{dk_y}{v^x} &= l^2\int \big(H^R_1 + H^B_1 + H^Z_1\big)\frac{dk_y}{v^x} \\ &= \int (\mathfrak{A} + \mathfrak{X})\cdot d\boldsymbol{k} + \sigma^z\frac{Z\,dk}{v^\perp}; \end{aligned} \tag{82}$$

this expression is understood to be evaluated on a certain edge. To derive the last equality, we combine the definitions in Eq. (57) with the identity $0 = v^x\,dk_x + v^y\,dk_y$ (which is valid on a constant-energy contour); Hamilton's equation in Eq. (41) is also useful in identifying $-dk_y/v^x = dk/v^\perp$.

Let us derive the single-band, Zilberman-Fischbeck wave function. This serves a pedagogical purpose, but also warms us up for the slightly more complicated derivation of the multiband WKB wave function in Sec. V A 2, which in its most general form has not been seen.

Proof of Eq. (79). Applying the identity Eq. (B10) (derived in Appendix B 1) to Eq. (78), with the WKB ansatz $g = e^{-i\psi}$, we derive

$$E = H_0 + H_1 + l^{-2}\left[\frac{i}{2}\frac{\partial^2 H_0}{\partial k_x \partial k_y} + \psi_0'(k_y)\partial_x H_0\right] + \frac{i}{2}l^{-4}\psi_{-1}''\partial_x^2 H_0 + O(l^{-4}), \tag{83}$$

where H_j and its derivatives are evaluated at $(k_x + \psi_{-1}'/l^2, k_y)$, and ψ_j' is the first derivative of ψ_j with respect to k_y. A solution exists if ψ_{-1} can be found that satisfies the zeroth-order relation:

$$E = H_0(k_x + \psi_{-1}'/l^2, k_y). \tag{84}$$

For the purpose of deriving the quantization conditions, we will only need the WKB wave functions in the classically allowed regions, where $\psi_{-1}' \in \mathbb{R}$. Generally, there might be multiple single-valued and real solutions, which we label with ν as $k_x + \psi_{-1}^{\nu\,\prime}/l^2 := k_x^\nu(k_y, E)$ [compare Eq. (84) with Eq. (80)]. This implies

$$\psi_{-1}^\nu = -l^2 k_x k_y + l^2 \int k_x^\nu(k_y, E) dk_y \tag{85}$$

up to an irrelevant integration constant. Collecting the first-order terms in Eq. (83), and substituting the just-obtained expression for $\psi_{-1}^{\nu\,\prime\prime}$,

$$0 = l^2 H_1 + \frac{i}{2}\left(\frac{\partial^2 H_0}{\partial k_x \partial k_y} + \frac{\partial k_x^\nu}{\partial k_y}\frac{\partial^2 H_0}{\partial k_x^2}\right) + \psi_0^{\nu\prime}(k_y)\frac{\partial H_0}{\partial k_x}\bigg|_{\boldsymbol{k}\to(k_x^\nu, k_y)}. \tag{86}$$

Let us separate $\psi_0 = \psi_{0R} + i\psi_{0I}$ into real and imaginary parts. Setting the imaginary component of Eq. (86) to zero,

$$0 = \frac{1}{2}\frac{\partial v_\nu^x}{\partial k_y} + \psi_{0I}^{\nu\prime} v_\nu^x \Rightarrow e^{\psi_{0I}^\nu} \propto \frac{1}{\sqrt{|v_\nu^x|}}, \tag{87}$$

with v_ν^x defined in Eq. (81). Setting the real component of Eq. (86) to zero,

$$l^2 H_1 + \psi_{0R}^{\nu\,\prime}\frac{\partial H_0}{\partial k_x}\bigg|_{\boldsymbol{k}\to(k_x^\nu, k_y)} = 0 \Rightarrow \psi_{0R}^\nu = -l^2 \int H_1^\nu(k_y, E)\left(v_x^\nu\right)^{-1} dk_y, \tag{88}$$

with H_1^ν defined in Eq. (81).

2. Multiband WKB wave function

Let us define the multiband WKB wave function $\boldsymbol{f}$ as the eigenfunction of

$$[H_0(\boldsymbol{K}) + H_1(\boldsymbol{K}) - E]\boldsymbol{f}_{kE}^\nu = O(l^{-4}); \tag{89}$$

matrix summation is implicit in this expression, and $\boldsymbol{f}$ is a vector-valued function with as many components as the number (D) of bands in the degenerate subspace P. We would like to demonstrate that

$$\boldsymbol{f}_{kE}^\nu = \mathcal{A}_{kE}^\nu \boldsymbol{f}_{kE}^{0\nu} \tag{90}$$

with $\boldsymbol{f}^0$ the product of an as-yet-undetermined, k_y-independent vector $\boldsymbol{c}$ with the Zilberman function:

$$f_a^{0\nu} = c_a^\nu \frac{1}{\sqrt{|v_\nu^x|}} e^{ik_x k_y l^2} e^{-il^2 \int k_x^\nu dk_y}, \quad c_a^\nu \in \mathbb{C}, \tag{91}$$

with $a = 1, \ldots, D$, and $\mathcal{A}$ a unitary propagator defined as the path-ordered exponential

$$\mathcal{A}_{kE}^\nu = \overline{\exp}\left[il^2 \int H_1^\nu \left(v_\nu^x\right)^{-1} dk_y\right]. \tag{92}$$

Despite being a simple extension of the single-band wave function, we have not seen a multiband ansatz for the Roth effective Hamiltonian in the literature.

Proof. The assumed band degeneracy within P implies $[H_0(\boldsymbol{k})]_{mn} = \delta_{mn}[H_0(\boldsymbol{k})]$, and therefore

$$[H_0(\boldsymbol{K}) - E]\boldsymbol{f}^{0\nu} = O(l^{-4}), \tag{93}$$

with $\boldsymbol{f}^{0\nu}$ defined in Eq. (91), as a special case of Eq. (79) with $H_1 = 0$. We propose the ansatz $\boldsymbol{g} = \mathcal{A}\boldsymbol{f}^0$ with $\mathcal{A}_{\boldsymbol{k}}$ a $D \times D$ matrix that is differentiable with respect to k_y. Each matrix element $\mathcal{A}_{ab} \in \mathbb{C}$ is assumed to be of order one. The following identity is useful:

$$H_0(\boldsymbol{K})\mathcal{A}_{ab} = \sum_{\boldsymbol{R}} \check{H}_0(\boldsymbol{R}) e^{i\boldsymbol{k}\cdot\boldsymbol{R} - iR_x R_y l^2} \times \{\mathcal{A}_{ab} - l^{-2}R_x \partial_y \mathcal{A}_{ab} + O(l^{-4})\} e^{-l^{-2}R_x\partial_y}, \tag{94}$$

which is derivable from Eq. (B11). Letting Eq. (94), a matrix operator, act on the vector $\boldsymbol{f}^{0\nu}$, we obtain

$$\sum_c H_0(\boldsymbol{K})\mathcal{A}_{bc} f_c^{0\nu} = \sum_c \mathcal{A}_{bc} H_0(\boldsymbol{K}) f_c^{0\nu} + il^{-2}\sum_c \partial_y \mathcal{A}_{ac} v_\nu^x f_c^{0\nu} + O(l^{-4}), \tag{95}$$

with help from Eq. (B12). Here, we have introduced the band velocity v_ν^x on the edge labeled ν. By similar manipulations with the H_1 term, we derive

$$\sum_{bc} H_1(\boldsymbol{K})_{ab}\mathcal{A}_{bc} f_c^{0\nu} = \sum_{bc} \left(H_1^\nu\right)_{ab}\mathcal{A}_{bc}\boldsymbol{f}_c^{0\nu} + O(l^{-4}). \tag{96}$$

Inserting Eqs. (95) and (96) into Eq. (89), the zeroth-order terms cancel owing to Eq. (93); after factoring out a common multiplicative factor (the Zilberman function), what remains is

$$\sum_b il^{-2}\partial_y \mathcal{A}_{ab} v_\nu^x c_b^\nu + \sum_{bd}\left(H_1^\nu\right)_{ab}\mathcal{A}_{bd} c_d^\nu = O(l^{-4}).$$

We would like this equation to be true for arbitary $\boldsymbol{c}^\nu$; hence we are led to a simplified differential equation

$$\partial_y \mathcal{A} = il^2\left(v_\nu^x\right)^{-1} H_1^\nu \mathcal{A}, \tag{97}$$

which is solved by Eq. (92).

B. Maslov correction from turning points

For any closed orbit, as exemplified in Figs. 2(e)–2(h), there are at least two turning points for which, in their vicinity,

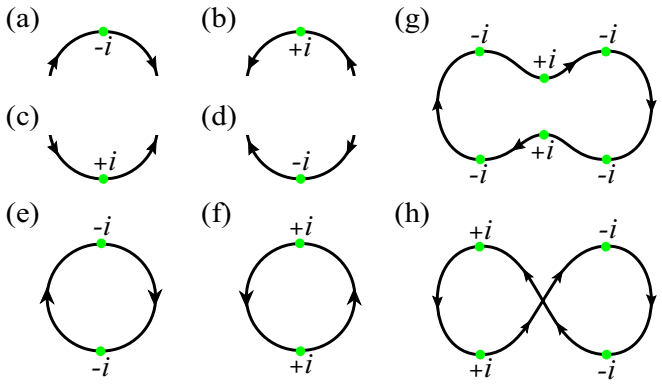

FIG. 2. Illustration of various turning points. Each turning point may be assigned a sense of circulation, which will be indicated by the sign of $\pm i$ next to each green dot. Each turning point may be divided into four classes illustrated by (a)–(d); these classes are distinguished by (i) the sign in $k_y \sim \pm k_x^2$ [i.e., whether the band contour forms an upright ($\smile$) or inverted ($\frown$) parabola], as well as (ii) an orientation determined by the direction of a semiclassical wave packet. If the band dispersion is expanded around each of the turning point as $\varepsilon_{\boldsymbol{k}} = u_y k_y + k_x^2/2m_x$, then in (a) $u_y, m_x > 0$, (b) $u_y, m_x < 0$, (c) $u_y > , m_x < 0$, and in (d) $u_y < 0, m_x > 0$. The correspondence between the sign of $\pm i$ and the four classes (a)–(d) of turning points is derived in Appendix B 2 a. (e)–(g) illustrate simple, closed orbits which are deformable to a circle, and (h) a nonsimple, closed orbit in the shape of a figure-of-eight.

the Zilberman-Fischbeck (ZF) wave function loses its validity due to strong quantum fluctuations. In the graph-theoretic language introduced in Sec. III F, a turning point is a *vertex* which is connected to two edges; alternatively stated, the beginning and end points of edges are vertices, and a turning point exemplifies a degree-two vertex.

It is well-known from the theory of caustics [95] that in passing around a turning point the WKB wave function *effectively* picks up a phase ϕ_r. For us, ϕ_r describes the phase difference between incoming and outgoing single-band ZF wave functions, which are valid sufficiently far from the turning point [111]; "incoming" and "outgoing" are interpretive characterizations of different edges of the ZF wave function—we may uniquely assign an orientation to each edge from Hamilton's equation [cf. Eq. (41)]. In analogy with a 1D Schrödinger particle reflecting off a wall, we might interpret the semiclassical wave packet for a Bloch electron as being reflected in the coordinate k_y—we therefore refer to ϕ_r as a reflection phase.

For a turning point in the orbit of a single band, we determine that $\phi_r = \pm\pi/2 + O(l^{-2})$, where the sign of $\pi/2$ is determined by the sense of circulation when passing the turning point: plus for anticlockwise, and minus for clockwise. We should clarify that this orientation is assigned locally to each turning point, and in a manner independent of the shape and orientation of the rest of the orbit. We may imagine minimally extending the parabolic contour at each turning point into a circle (e.g., $\curvearrowleft \to \circlearrowleft$, $\curvearrowright \to \circlearrowright$); we then assign the orientation by interpreting the circle as a clock face.

ϕ_r is derived by a divide-and-conquer approach—we approximately describe the turning region with an effective Hamiltonian that is linear in one momentum component and quadratic in the other. What distinguishes our approach from previous works [17,51,52] is that the effective Hamiltonian we adopt to describe the turning point is not just the Peierls-Onsager Hamiltonian, but includes the first-order correction by H_1. For simplicity, we consider a hard-wall boundary condition at the turning point; i.e., we ignore tunneling between closed orbits. We then match the asymptotic wave function of this small-momentum, effective Hamiltonian with the incoming and reflected WKB functions; the proof is completed in Appendix B 2.

The multiband analog of the calculation in Appendix B 2 is more involved. One may, however, avoid this calculation if one is willing to accept an uncertainty of $O(l^{-2/3})$ in the quantization condition; this viewpoint seems to be implicitly adopted in past works [17,51,52,101,102], though no attempt was made to quantify this uncertainty. To clarify, by exploiting the smallness of the turning regions relative to the rest of the semiclassical orbit, we might neglect the effect of H_1 in the turning regions, but account for it everywhere else on the orbit. In practice, this just means applying the zeroth-order Maslov correction (π for a closed orbit) to a quantization condition which already includes first-order corrections through the propagator of Eq. (92). Let us estimate the uncertainty in this approach. (i) If our asymptotic expansion of the quantization conditions is at all valid, we expect H_j to make an $O(l^{2-2j})$ contribution to the quantization condition; in particular, H_1 makes an $O(1)$ contribution. (ii) The length of the orbit lying within the turning region is of $O(l^{-2/3})$, as shown in Appendix B 2; generically, the length of the semiclassical orbit is on the order of the reciprocal period. The ratio of the turning length to that of the entire orbit is then of $O(l^{-2/3})$. Combining (i) and (ii), we expect that neglecting H_1 in the turning regions introduces an uncertainty of $O(l^{-2/3})$. For the multiband case, we therefore argue that the Maslov correction for a closed orbit is $\pi + O(l^{-2/3})$, for each of the D sets of sub-Landau levels [indexed by a in Eq. (73)].

C. Quantization conditions

1. Single-band quantization condition for closed orbits

Let us illustrate how to formulate the continuity condition for the circular closed orbit, which is composed of two edges (labeled by $\nu = \pm$) which touch at two turning points, as illustrated for two orientations in Figs. 2(e) and 2(f). We shall focus on the orbit that circulates as $\circlearrowleft$ in Fig. 2(e), which is expected of an electron pocket at the Fermi level. The quantization condition for an energy eigenstate at energy E and wave vector (k_x, k_z) is the continuity (with respect to k_y) of the wave function in the (K_x, k_y) representation.

The continuity condition on f may be formulated in a manner that emphasizes a semiclassical motion along the orbit; this motion is described by the time evolution of certain scalar amplitudes that we define for each edge:

$$a_{\nu,E}(t_\nu) := e^{-il^2 \int [k_x^\nu - H_1^\nu (v_\nu^x)^{-1}](dk_y/dt_\nu)dt_\nu}|_E. \tag{98}$$

This amplitude is simply the phase component of the Zilberman-Fischbeck wave function $g^\nu_{\boldsymbol{k}E}$ [cf. Eq. (79)], except without $e^{ik_x k_y l^2}$, which is trivially continuous over any closed orbit. The triviality of $e^{ik_x k_y l^2}$ would not be true of open orbits,

as we will substantiate in Sec. V C 3. We have parametrized each amplitude in Eq. (B30) by a timelike variable $t_\nu \in [0,1]$, which increases along the orbit in a direction consistent with Hamilton's equation. The end points ($t_\nu = 0$ and 1) correspond to distinct turning points that bound the edge ν. We may loosely interpret $\boldsymbol{k}(t_\nu)$ as the wave vector of the "moving wave packet" at time t_ν; we caution the reader that there are no turning points in conventional wave packet theory [7], but the language of a "moving wave packet" offers a convenient and visually appealing metaphor for the continuity and patching of wave functions in WKB theory.

With this caveat in mind, we will interpret

$$e^{i\theta_\nu(E,l^2)} = \frac{a_{\nu,E}(1)}{a_{\nu,E}(0)} := e^{-il^2\int_0^1[k_x^\nu - H_1^\nu(v_\nu^x)^{-1}](dk_y/dt_\nu)dt_\nu}\Big|_E \tag{99}$$

as the semiclassical phase acquired by a wave packet as it traverses the edge ν. As the wave packet approaches a turning point along the edge ν', it is reflected onto a distinct edge ν and picks up an additional phase of $\pm\pi/2$; the sign depends on the sense of circulation of the turning point, as we have illustrated in Figs. 2(a)–2(d). We implement this reflection phase in the boundary condition

$$a_\nu(0) = e^{i\phi_r^{\nu'}} a_{\nu'}(1), \quad e^{i\phi_r^{\nu'}} = \pm i. \tag{100}$$

For any closed orbit, the set of equations in Eq. (99) and Eq. (100) (for all ν,ν') may be combined into a single equation that is parametrized by E (and k_z in 3D solids). For our case study of the simplest closed orbit ↻, this single equation may be expressed in a manner that emphasizes its motional interpretation:

$$1 = (-i)e^{i\theta_-}(-i)e^{i\theta_+}|_{E,k_z,l^{-2}}. \tag{101}$$

Reading from right to left, a wave packet that begins at $t_+ = 0$ first accumulates the semiclassical phase $e^{i\theta_+}$ along edge $+$, is then reflected unto edge $-$, and accumulates $e^{i\theta_-}$ in traveling along this edge; a second reflection closes the loop, and the quantization rule states that the net phase (acquired by the wave packet around the loop) is an integer multiple of 2π. Equation (101) may equivalently be expressed as in Eq. (68), or as

$$-1 = \exp\left[il^2 S + i\oint_{\mathfrak{o}}(\mathfrak{A}+\mathfrak{X})\cdot d\boldsymbol{k} + iZ\oint_{\mathfrak{o}}(\sigma^z/v^\perp)dk\right]_{E_j,k_z}, \tag{102}$$

with S defined as the oriented area of the orbit:

$$S[\mathfrak{o}] := -\int_0^1 k_x^\nu(dk_y/dt_\nu)dt_\nu. \tag{103}$$

The other terms in the exponent are, collectively, the Roth-Berry-Zeeman phase originating from the H_1 term in $e^{i\theta_\nu}$ [cf. Eq. (82)]. The gauge ambiguity in the definition of H_1^B [recall Eq. (61)] is reflected in the above equations by the ambiguity in the Berry connection $\mathfrak{X}$. However, it is known [20] that the exponentiated loop integral of $i\mathfrak{X}$, as appears in Eq. (102), is gauge-invariant.

Being independent of k_x, Eq. (102) defines a set of discrete energy levels $\{E_j\}$ which are each macroscopically degenerate; we refer to them as Landau levels. The Zeeman term in Eq. (102) may be further simplified if spin-orbit coupling is absent; we might then replace $\sigma^z \to \pm 1$ and Eq. (102) reduces to Eq. (72) with the identification $Z\int dk/v^\perp = \pi(g_0/2)(m_c/m)$, with m_c the cyclotron mass.

The extension of Eq. (101) to the most general closed orbit, composed of $N_s \in 2\mathbb{Z}$ edges and an equal number of turning points, is

$$1 = \prod_{\nu=1}^{N_s} e^{i\phi_r^\nu} e^{i\theta_\nu}\Big|_{E,k_z,l^{-2}}, \tag{104}$$

where we identify the Maslov phase

$$\phi_M := \prod_{\nu=1}^{N_s}\phi_r^\nu \tag{105}$$

as the net reflection phase of all turning points. To derive the same quantization condition from more conventional means (i.e., continuity and patching of wave functions), we refer the interested reader to Appendix B 3. Equation (104) applies to any simple closed orbit, which we define as orbits that are deformable to a circle, e.g., Fig. 2(g). Equation (104) also applies to nonsimple closed orbits which are homotopically *in*equivalent to a circle, a case in point being the figure-of-eight illustrated in Fig. 2(h). A figure-of-eight pinches together an electron-like pocket with a hole-like pocket, and has recently been studied in the context of over-tilted Weyl/Dirac fermions [46,47]. The four turning points in a figure-of-eight have canceling circulations [as indicated by the sign of $\pm i$ in Fig. 2(h)], and therefore there is no Maslov correction, contrary to a claim in Ref. [46]; we shall extend our analysis of the over-tilted Weyl/Dirac fermion to include interband breakdown in Sec. IX.

The net reflection phase ($\prod_{\nu=1}^{N_s} e^{i\phi_r^\nu}$) of a closed orbit is invariant under continuous deformations of the orbit trajectory; we will argue for this by locally deforming the band contour near a turning point, while maintaining a closed orbit. If we invert the parabolic contour associated with a single turning point, we necessarily introduce a Mexican-hat wiggle with two additional turning points; e.g., Fig. 2(g) is a deformed version of Fig. 2(e). Since the net circulation (a notion we make precise in Sec. V B) of the final three points is always equal to that of the original point, there is no net change in the reflection phase. Having argued that the combined reflection phase is topologically invariant for a closed orbit, we may therefore evaluate this quantity for the simplest, homotopically equivalent representative. For the simple closed orbits, this is a circle [Figs. 2(e) and 2(f)], which has a reflection phase of π; this accounts for the π Maslov correction to the quantization condition. One implication of this argument: while Eq. (102) has been derived for a circular orbit, its final expression is generally valid for any simple closed orbit.

2. *Multiband quantization condition for closed orbits*

In the multiband case with $P(\boldsymbol{k})$ having rank D, we may analogously define a vector-valued amplitude for each edge

(labeled by ν) as

$$\boldsymbol{a}_{\nu E}(t_\nu) := e^{-il^2\int_0^{t_\nu} k_x^\nu (dk_y/dt_\nu')dt_\nu'}\,\mathcal{A}^\nu_{\boldsymbol{k}(t_\nu)E}\,\boldsymbol{a}_{\nu E}(0)\big|_E. \quad (106)$$

As defined in Eq. (92), $\mathcal{A}$ is a $D\times D$ unitary matrix acting on an as-yet-unspecified, constant vector $\boldsymbol{a}(0)$. $t_\nu\in[0,1]$ and $\boldsymbol{k}(t_\nu)$ have the same meaning as for the single-band case, as described below Eq. (98). We implement the boundary condition

$$\boldsymbol{a}_{\nu E}(0) = e^{i\phi_r^{\nu'}}\boldsymbol{a}_{\nu' E}(1), \quad e^{i\phi_r^{\nu'}} = \pm i, \quad (107)$$

for every two edges (ν and ν') that touch at a turning point.

For a closed orbit ($\mathfrak{o}$) comprising N_s edges (and an equal number of turning points), Eqs. (106) and (107) may be combined into a system of linear equations with D variables. The quantization condition is then equivalent to solving this system of equations; a solution exists upon satisfaction of the following determinantal equation:

$$\det\left[\left(\prod_{\nu=1}^{N_s} e^{i\phi_r^\nu}\right)e^{il^2S}\,\mathcal{A}[\mathfrak{o}]-I\right]\Big|_{E=E_{a,j},k_z}=0, \quad (108)$$

with $a\in\mathbb{Z}_D$, $j\in\mathbb{Z}$. $\mathcal{A}[\mathfrak{o}]$ is the propagator of Eq. (74) defined over the full orbit. Its solution corresponds to D sets of equidistant, macroscopically degenerate Landau levels. For a simple closed orbit, $\prod_{\nu=1}^{N_s} e^{i\phi_r^\nu}=-1$, and we are led directly to the multiband quantization conditions in Eqs. (73) and (74). In Sec. VI D 1, we show that Eq. (108) is invariant under the $U(D)$ gauge transformations [cf. Eq. (22)]; the transformation of the propagator under symmetry is further investigated in Sec. VI D 2, where we prove certain symmetry constraints for the Landau levels.

As an example, consider a spin system where $T\mathfrak{i}$ symmetry imposes a twofold degeneracy ($D=2$) in the zero-field band dispersion. In the presence of a field, the same symmetry imposes $\lambda_1=-\lambda_2 \bmod 2\pi$, as elaborated in Sec. VI D 2. If spin-orbit coupling is negligible, $|\lambda_1-\lambda_2|$ just equals the free-electron Zeeman splitting $[\pi(g_0/2)(m_c/m)$ from Eq. (72)].

3. Beyond closed orbits

We briefly comment on the quantization conditions for open orbits with negligible breakdown. One example would be a noncontractible orbit which extends across the Brillouin torus in a single direction; in the extended-zone scheme, these orbits traverse across different Brillouin zones. Since the phase factor $e^{ik_xk_yl^2}$ is not single-valued when $\boldsymbol{k}$ is advanced by a reciprocal vector, this phase cannot be neglected when one imposes continuity on the wave function in the (K_x,k_y) representation. When this phase is accounted for, it introduces a k_x dependence to the quantization condition, and consequently a loss of the (exponentially accurate) macroscopic degeneracy that characterizes closed orbits. We refer the reader to [18] for a more extensive discussion of open orbits.

D. Landau levels and de Haas–van Alphen oscillations for closed orbits

1. Single-band case

When the single-band quantization condition [Eq. (68)] is viewed at a fixed field (and a fixed wave vector k_z for a 3D solid), the energy difference between adjacent Landau levels is locally periodic as

$$E_{j+1}-E_j = \frac{2\pi}{l^2\partial S/\partial E}\bigg|_{E=E_j,k_z}+O(l^{-4}). \quad (109)$$

This follows from the assumption that the area of the orbit (S) as well as the Roth-Berry-Zeeman (RBZ) phase, collectively defined as

$$\lambda(E,k_z) := \oint_{\mathfrak{o}}(\mathfrak{A}+\mathfrak{X})\cdot d\boldsymbol{k} + Z\oint_{\mathfrak{o}}(\sigma^z/v^\perp)dk, \quad (110)$$

are smooth functions of energy on the scale of $E_{j+1}-E_j=O(l^{-2})$. Equivalently stated, we assume $\partial S/\partial E$ and $\partial\lambda/\partial E$ are $O(l^0)$ quantities. The quantity $(\hbar^2/2\pi)\partial S/\partial E$ that determines Landau-level differences has been referred to as the cyclotron mass; it coincides, in the free-electron limit ($V=0$), with the free-electron mass. Supposing E_n^0 are zeroth-order solutions of the quantization condition, the H_1 correction to E_n^0 is

$$\delta E_n = -l^{-2}\frac{\lambda}{\partial S/\partial E}\bigg|_{E_n^0}+O(l^{-4}). \quad (111)$$

If we view Eq. (68) at fixed energy (e.g., the Fermi energy E_F) and variable field, then the quantization condition is satisfied for a discrete set of fields (indexed by integer j), with corresponding magnetic lengths satisfying

$$l_{j+1}^2-l_j^2 = \frac{2\pi}{S(E_F,k_z)}+O(l^{-2}),$$
$$l_j^2 = \frac{2\pi j-\phi_M-\lambda}{S}\bigg|_{E=E_F,k_z}+O(l^{-2}). \quad (112)$$

The first equation forms the basis of quantum oscillatory phenomena of the de Haas–van Alphen (dHvA) type—they reflect how quasiperiodic Landau levels successively become equal to the Fermi energy as the reciprocal magnetic field is changed. Such l_j^2 is henceforth referred to as a dHvA level, and the set of all dHvA levels is referred to as the dHvA spectrum. The period in l_j^2 is not affected by the RBZ phase (λ), in accordance with the conventional theory of metals [6]. On the other hand, one may look to the phase offset of the dHvA oscillation to extract λ. In 3D metals, the curvature of the Fermi surface results in an additional Lifshitz-Kosevich correction [3,4] to this phase offset, which in sum equals

$$\gamma := l_j^2S+\phi_{\rm LK} \bmod 2\pi \equiv -\lambda-\phi_M+\phi_{\rm LK}, \quad (113)$$

with $\phi_{\rm LK}=\pm\pi/4$ depending on whether the orbit is maximal or minimal. Restated from the perspective of measurement, γ is the phase offset of the oscillations of the magnetization of a solid [8]. In the experimental literature, γ is often viewed graphically as the intercept of an extrapolated line connecting the discrete values of l^2 where the magnetization is peaked; we shall therefore refer to the quantity defined in Eq. (113) as the γ intercept.

A comprehensive symmetry analysis of the RBZ phase is performed in Sec. VI D; here we illustrate two highlights:

(i) Orbits which are mapped to themselves, up to a reversal in orientation, are said to be self-constrained. Graphene provides a paradigmatic example, for which an orbit encircling the Dirac point is invariant under $T\mathfrak{c}_{2z}$ symmetry; this leads to the vanishing of the Roth moment at each wave vector, as well

as the quantization of the Berry phase to π, which cancels the Maslov correction in the γ intercept.

(ii) Just as relevant are orbits which are mapped to distinct orbits by a symmetry; two related orbits are said to be mutually constrained. In a toy model of spinless graphene, two orbits which encircle different valley centers are mutually constrained by T symmetry; since each orbit does not encircle a T-invariant wave vector, it is not self-constrained by T symmetry. If the spatial symmetry ($\mathfrak{c}_{2z}$) is further broken (plausibly by epitaxial growth on certain substrates [112,113]), each valley-centered orbit develops an orbital moment; i.e., when integrated over the orbit, this moment results in a nontrivial Roth phase. Owing to T symmetry, the sum of Roth-Berry phases in two mutually constrained orbits cancels modulo 2π, but individually each phase should be measurable from the γ intercept of its corresponding orbit. Alternatively stated, two distinct but mutually constrained harmonics should appear in the magnetization oscillations. To make this toy model of graphene more realistic, one must incorporate the Zeeman effect, as described in Ref. [32].

The equations in this section directly apply to spin-orbit-coupled systems with nondegenerate bands; for Zeeman-coupled systems with negligible spin-orbit coupling, the above equations would apply to either of the two spin species, with σ_z replaced by ± 1; for intrinsically spinless systems described by a Schrödinger Hamiltonian, the above equations apply without the Zeeman term (i.e., set $Z = 0$).

2. Multiband case

Considering Eq. (73) at fixed field and k_z, we obtain, for each eigenvalue $e^{i\lambda_a}$ of the $D \times D$ propagator [Eq. (74)], a set of discrete energy levels $\{E_{a,j}\}_{j\in\mathbb{Z}}$. For fixed a, the energy difference between adjacent Landau levels is to leading order

$$E_{a,j+1} - E_{a,j} \approx \left.\frac{2\pi}{l^2 \partial S/\partial E}\right|_{E=E_{a,j},k_z}. \tag{114}$$

We have assumed here that $\partial S/\partial E$ and $\partial \lambda_a/\partial E$ are both $O(l^0)$. When Eq. (73) is viewed at fixed energy (E_F) and varying field, the quantization condition is satisfied for a discrete set of fields corresponding to

$$\begin{aligned} l^2_{a,j} &\approx \left.\frac{2\pi j - \phi_M - \lambda_a}{S}\right|_{E=E_F,k_z}, \\ l^2_{a,j+1} - l^2_{a,j} &\approx \frac{2\pi}{S(E_F,k_z)}. \end{aligned} \tag{115}$$

All equations in this section are accurate to $O(l^{-8/3})$, due to the $O(l^{-2/3})$ uncertainty in the multiband Maslov correction (derived in Sec. V B). The dHvA spectrum therefore divides into D sets of levels indexed by $a = 1,\ldots,D$; each set corresponds to a harmonic in the magnetization oscillations, with corresponding intercept $\gamma_a := l^2_{a,j} S + \phi_{\mathrm{LK}} \bmod 2\pi$.

Particularizing to spin-orbit-coupled bands with Ti symmetry ($D = 2$), Eq. (115) implies the existence of two harmonics in the magnetic oscillations. Absent any other symmetries, these harmonics are generally distinct, with $\lambda_1 = -\lambda_2 \bmod 2\pi$, as proven in the paragraph surrounding Eq. (146) in Sec. VI D. A more comprehensive symmetry analysis of λ_a is performed in the next section (Sec. VI).

VI. SYMMETRY IN THE FIRST-ORDER EFFECTIVE HAMILTONIAN THEORY

This section describes the effects of symmetry in the first-order effective theory; our analysis covers all possible symmetries that occur in crystals, i.e., in any space group or magnetic space group. We first identify in Sec. VI A the symmetries that are relevant to semiclassical orbits; we then describe how symmetry constrains the orbital magnetic moment and the Zeeman coupling (Sec. VI B), the first-order effective Hamiltonian (H_1) (Sec. VI C), and the propagator that is generated by H_1 (Sec. VI D) over an orbit. The eigenphases of this propagator enter the quantization conditions, from which one may determine the symmetry constraints on the Landau levels and dHvA oscillations. Our symmetry analysis is simplified by the classification of closed orbits into ten (and only ten) symmetry classes. These ten symmetry classes were first introduced and exemplified in Ref. [32]; in Sec. VI D, we provide a more detailed derivation which focuses on the possible types of symmetry representations. In addition, Sec. VI B may be used to analyze $\boldsymbol{k}$-resolved measurements of the orbital magnetic moment, e.g., through circular dichroism in photoemission [53].

A. Symmetries of semiclassical orbits

We encourage the reader to scan through Sec. III C, where we reviewed how symmetry constrains Bloch functions at zero field. We assume the reader is familiar with certain notations for symmetry transformations that were introduced therein. As a reminder, g denotes a symmetry in the (magnetic) space group (G) of a solid, and its representations in various contexts (cf. Sec. III C) are denoted by $\hat{g},\check{g},\check{g}$.

We would like to particularize to symmetries which are relevant to Bloch electrons in a field. Assuming that the field is oriented along $\vec{z}$, all semiclassical orbits are contained in quasimomentum planes orthogonal to $\vec{z}$, and we are interested in symmetries which relate one such orbit to another (or possibly an orbit to itself, up to a reversal in orientation). For 3D solids, g's action in $\boldsymbol{k}$ space may be block-diagonalized as

$$\begin{aligned} g : \boldsymbol{k} \to\ g \circ \boldsymbol{k} &:= (-1)^{s(g)} \check{g} \boldsymbol{k} \\ &= (-1)^{s(g)} (\check{g}^{\perp} \boldsymbol{k}^{\perp}, (-1)^{t(g)} k_z), \\ & t(g) \in \{0,1\}, \end{aligned} \tag{116}$$

where $t(g) = 0$ (resp. 1) for symmetries whose point-group operation preserves (resp. inverts) the coordinate parallel to the field. We distinguish between $\boldsymbol{k}$ which parametrizes the 3D Brillouin torus, and $\boldsymbol{k}^{\perp} = (k_x,k_y)$ which parametrizes a two-torus ($BT_{\perp}$) perpendicular to the field. We shall sometimes refer to $BT_{\perp}$ as a plane; symmetry operations that act only in $BT_{\perp}$ are described as planar.

In Eq. (116), we have also introduced $\check{g}^{\perp}$ as a real, orthogonal, 2×2 matrix; it represents the point-group operation that is restricted to $BT_{\perp}$. The determinant of $\check{g}^{\perp}$ defines a $\mathbb{Z}_2$ variable u as

$$\begin{aligned} (-1)^{u(g)} &:= \det[\check{g}^{\perp}], \quad u(g) \in \{0,1\}, \\ \det[\check{g}] &= (-1)^{t(g)+u(g)}. \end{aligned} \tag{117}$$

Let us demonstrate that $u(g) = 0$ (resp. 1) if g preserves (resp. inverts) the orientation of the semiclassical orbit; to

clarify, the orientation of an orbit is its sense of circulation, whether clockwise or anticlockwise, that is determined from Hamilton's equation [cf. Eq. (1)].

The symmetry constraint on the band velocities at $\boldsymbol{k}$ and $g\circ\boldsymbol{k}$ [recall Eq. (116)]

$$\boldsymbol{v}(\boldsymbol{k})=(-)^{s(g)}[\check{g}^T\boldsymbol{v}(g\circ\boldsymbol{k})],\quad \check{g}^T_{\alpha\beta}=\check{g}_{\beta\alpha}, \tag{118}$$

implies, through Eq. (41), an analogous relation between the orbit velocities [defined in Eq. (41)]:

$$\det[\check{g}^\perp]\dot{\boldsymbol{k}}^\perp|_{g\circ\boldsymbol{k}}=(-)^{s(g)}[\check{g}^\perp\dot{\boldsymbol{k}}^\perp|_{\boldsymbol{k}}]. \tag{119}$$

To interpret this equation, consider the map $\mathcal{Y}_g:\mathbb{R}^2\to\mathbb{R}^2$ between two planar wave vectors related by symmetry g:

$$\mathcal{Y}_g(\boldsymbol{k}^\perp)=(-1)^s\check{g}^\perp\boldsymbol{k}^\perp=(g\circ\boldsymbol{k})^\perp. \tag{120}$$

Equation (119) states that $\mathcal{Y}_g(\dot{\boldsymbol{k}}^\perp|_k)$ is equal in magnitude to the orbit velocity at $g\circ\boldsymbol{k}$, with a minus-sign difference iff $\det[\check{g}^\perp]=(-1)^{u(g)}=-1$. ■

All symmetries, whose point-group operation block-diagonalizes as in Eq. (116), will henceforth be referred to as *symmetries of the orbit configuration*. In deriving how these symmetries constrain the effective Hamiltonian, the following decomposition, valid for any symmetry of the orbit configuration, would be useful:

$$g=T^{s(g)}\,\mathfrak{t}_\delta\,\mathfrak{r}_z^{t(g)}\,\mathfrak{r}_x^{u(g)}\,\mathfrak{c}_{n(g),z}^{v(g)}\,\mathfrak{e}^{w(g)}, \tag{121}$$

where $\mathfrak{t}_\delta,T,\mathfrak{r},\mathfrak{c},\mathfrak{e}$ are symmetry operations defined in Sec. VI A; $s,t,u\in\{0,1\}$ have been previously defined [Eqs. (25), (116), and (117)]; $n\in\{2,3,4,6\}$ labels the possible discrete rotations; and we introduce here $w\in\{0,1\}$ and $v\in\{0,1,\ldots,n-1\}$. Equation (121) is really valid for a double group; for ordinary groups the same decomposition holds without the factor of $\mathfrak{e}^w$.

Proof of decomposition Eq. (121). If g inverts time, we decompose $g=Tg'$ such that g' is a purely spatial operation; otherwise, $g=g'$. Our shorthand for this is $g=T^sg'$. We further decompose g' into translational and point-preserving spatial transformations: $g'=\mathfrak{t}_\delta g''$. Applying Eq. (116), we find that g'' decomposes as $g''=\mathfrak{r}_z^t g^\perp$, such that the $g^\perp$ acts trivially on the coordinate orthogonal to the plane, i.e., $g^\perp:\boldsymbol{k}\to(\check{g}^\perp\boldsymbol{k}^\perp,k_z)$.

To complete the proof, we would need to show that any planar, spatial transformation may be expressed as $g^\perp=\mathfrak{r}_x^u\mathfrak{c}_{nz}^m\mathfrak{e}^w$. Any point group is built up of discrete rotations ($\mathfrak{c}$) and reflections ($\mathfrak{r}$), which are the fundamental covering operations [114]; 2D point groups are built up from planar rotations ($\mathfrak{c}_{nz}$) and reflection-invariant lines contained in the plane (in short: planar reflections); the latter are exemplified by $\mathfrak{r}_x$ and $\mathfrak{r}_y$. It is useful to distinguish between planar-proper ($\det\check{g}^\perp=+1$) and planar-improper ($\det\check{g}^\perp=-1$) transformations; all planar reflections (resp. rotations) are planar-improper (resp. proper). Some properties of successive transformations will be needed [114]: (i) the product of two planar rotations is another planar rotation, (ii) the product of two planar reflections is a planar rotation, and (iii) the product of a planar rotation with a planar reflection is another planar reflection. It follows from (i)–(iii) that any planar-proper transformation is proportional to $\mathfrak{c}_{nz}^m$ for some integers m and $n\in\{2,3,4,6\}$; the proportionality factor must act trivially in space, so it may be the identity operation or a 2π rotation. Therefore, if $g^\perp$ is planar-proper, it is expressible as $\mathfrak{c}_{nz}^m\mathfrak{e}^w$. Otherwise if $g^\perp$ is planar-improper, it can always be expressed as the product of (a) an arbitrarily chosen reflection (e.g., $\mathfrak{r}_x$), with (b) a proper transformation ($\mathfrak{c}_{nz}^m\mathfrak{e}^w$) that depends on our choice in (a). In summary, we may say $g^\perp=\mathfrak{r}_x^u\mathfrak{c}_{nz}^m\mathfrak{e}^w$ with $(-1)^u=\det[\check{g}^\perp]$ [cf. Eq. (117)]. This completes the proof. ■

Combining Eq. (116) with Eq. (121), g maps $\boldsymbol{k}^\perp\in BT_\perp$ to

$$g\circ\boldsymbol{k}^\perp=(-1)^s\check{g}^\perp\boldsymbol{k}^\perp,\quad\text{with}\quad \check{g}^\perp=\check{\mathfrak{r}}_x^u\,\check{\mathfrak{c}}_{nz}^v. \tag{122}$$

We say that $\boldsymbol{k}^\perp$ is g-invariant if $g\circ\boldsymbol{k}^\perp=\boldsymbol{k}^\perp$ up to a planar reciprocal vector. If g acts as a planar reflection, then its order must be even:

$$u(g)=1\Rightarrow N(g)\in 2\mathbb{Z}. \tag{123}$$

This follows because any odd power of a planar reflection is still a planar reflection, while by assumption $u(g^N)=0$ [cf. Eq. (31)].

It will be useful to classify symmetries according to the topology of the g-invariant points. Type-I symmetries are defined to leave every, generic $\boldsymbol{k}^\perp$ invariant; hence

$$(-1)^s\check{g}^\perp=I_2\Rightarrow\det\check{g}^\perp=1\Rightarrow u(g)=0, \tag{124}$$

with I_2 the 2×2 identity matrix. We may further distinguish type-I symmetries by s: either $u=s=v=0$, or $u=0,s=1$, and $\check{\mathfrak{c}}_{nz}^v=\check{\mathfrak{c}}_{2z}$.

Type-II symmetries are symmetries for which the generic $\boldsymbol{k}^\perp$ is not invariant; g-invariant $\boldsymbol{k}^\perp$ are isolated points if $u=0$; otherwise ($u=1$), they form isolated lines. To prove the last claim, if type-II g is planar-proper ($u=0$), then g maps $\boldsymbol{k}^\perp$ to $\check{\mathfrak{c}}_{2z}^s\check{\mathfrak{c}}_{nz}^v\boldsymbol{k}^\perp$, where we have identified $(-1)^s$ with $\check{\mathfrak{c}}_{2z}^s$. Being the product of two planar rotations, $\check{\mathfrak{c}}_{2z}^s\check{\mathfrak{c}}_{nz}^v$ must be again a planar rotation, whatever the values of s,n,v. Moreover, $\check{\mathfrak{c}}_{2z}^s\check{\mathfrak{c}}_{nz}^v$ cannot be the trivial rotation (identity transformation), due to our assumption that it is type II. The only rotationally invariant $\boldsymbol{k}^\perp$ are isolated points. Now if type-II g is planar-improper, g acts on $\boldsymbol{k}^\perp$ as the product of a planar reflection ($\check{\mathfrak{r}}_x$) and a planar rotation ($\check{\mathfrak{c}}_{2z}^s\check{\mathfrak{c}}_{nz}^v$); any such product is a planar reflection, possibly with the reflection-invariant line rotated.

Our classification of type-I and -II symmetries is extended to a classification of symmetric orbits in Sec. VI D.

B. Symmetry constraints of the orbital magnetic moment and the Zeeman coupling

We would like to identify the symmetry classes which allow for a $\boldsymbol{k}$-dependent orbital moment/Zeeman coupling. If the average of the orbital moment/Zeeman coupling over an orbit is nonzero, then the Landau levels are nontrivially affected, as explained in Sec. VI D 2. Our symmetry analysis is further motivated by recent experiments that are able to probe the orbital moment (at each wave vector) through circular dichroism in momentum-resolved photoemission [53]. We would like to determine whether dichroism is allowed by symmetry, and where dichroism may be found in the Brillouin torus.

Consider the orbital moment for a subspace of bands projected by $P(\boldsymbol{k})$ in Eq. (18), where, again, D the dimension of the subspace at each wave vector. For $D=1$, we refer to the single-band $\boldsymbol{M}_n$ defined in Eq. (63) for band n, but henceforth we will drop the n subscript; for $D>1$, we refer to the multiband $\boldsymbol{M}$ defined in Eq. (67), which is a $D\times D$

Hermitian matrix. To derive the symmetry constraints on the orbital moment, it is convenient to begin with its expression with the velocity matrix $\mathbf{\Pi}$ [second line of Eq. (67)], since the current operator $\hat{\mathbf{\Pi}} = -i[\hat{\boldsymbol{r}}, \hat{H}_0]/\hbar$ transforms simply under a symmetry (g). From the action of g on the position operator [cf. Eq. (24)], and the symmetry constraint on the cell-periodic functions [cf. Eq. (29)], we obtain

$$\Pi^{\alpha}_{mn}\big|_{g\circ\boldsymbol{k}} = (-1)^{s(g)}\breve{g}_{\alpha\beta}K^{s(g)}[\breve{g}^{*}\,\Pi^{\beta}\,\breve{g}^{T}]_{mn}K^{s(g)}|_{\boldsymbol{k}}. \quad (125)$$

Inserting this into Eq. (67) and after a little gymnastics (detailed in Appendix C 2),

$$\boldsymbol{M}|_{g\circ\boldsymbol{k}} = (-1)^{s(g)}\det[\breve{g}]\,\breve{g}\,K^{s(g)}\,(\breve{g}\boldsymbol{M})\,K^{s(g)}\,\breve{g}^{-1}|_{\boldsymbol{k}}. \quad (126)$$

While this expression is valid for any number of bands, it simplifies in the single-band case owing to (i) $\boldsymbol{M}$, being a Hermitian 1×1 matrix, is a real number, and (ii) $\breve{g}$, being a unitary 1×1 matrix, is a commuting phase factor that cancels with its Hermitian adjoint $\breve{g}^{-1}$. Therefore, the single-band orbital moment satisfies

$$\boldsymbol{M}|_{g\circ\boldsymbol{k}} = (-1)^{s(g)}\det[\breve{g}][\breve{g}\boldsymbol{M}]|_{\boldsymbol{k}}. \quad (127)$$

To interpret this equation, recall that each g corresponds to a certain action in spacetime, which may be decomposed as a point-preserving transformation and a translation [cf. Eq. (25)]; $(-1)^{s(g)}\det[\breve{g}]$ is the determinant of the matrix corresponding to the point-preserving transformation. Therefore, $\boldsymbol{M}$ transforms like the spatial components of a (3+1)-dimensional pseudovector, in addition to the transformation of its argument $\boldsymbol{k}$. Equation (127) is the full generalization (to any symmetry) of well-known constraints on the single-band moment with T and/or $\mathfrak{i}$ symmetry [25,33,74]. We exemplify Eqs. (126) and (127) with some naturally occurring, but certainly not exhaustive, symmetries in Table II; there, we employ certain notation for symmetries that have been introduced in Sec. III C.

The single-band orbital moment, in a certain direction α, vanishes at a specific $\boldsymbol{k}$, if there exists a symmetry that inverts $M^{\alpha}\to -M^{\alpha}$, and simultaneously maps said $\boldsymbol{k}$ to itself. This vanishing may occur at isolated points, e.g., $\boldsymbol{M}=0$ at inversion-invariant wave vectors (where $\boldsymbol{k}=-\boldsymbol{k}$) in systems with only time-reversal symmetry. In $T\mathfrak{i}$-symmetric systems, $\boldsymbol{M}=0$ in the entire torus. For further exemplification, $M^{z}=0$ for high-symmetry planes in systems with $T\mathfrak{c}_{2z}$ symmetry, and assuming no other symmetry we should not expect M^{x} or M^{y} to likewise vanish. For these examples we list the symmetry constraints on the single-band moment in the third column.

In the multiband case, symmetry cannot enforce that M^{α} vanishes as a matrix, but the analogous constraint is that its trace vanishes, as shown in the second-to-last column. Let us particularize the following discussion to spin-orbit-coupled systems, where bands are twofold spin-degenerate, and transform in a half-integer-spin representation ($F=1$). Any of T, $T\mathfrak{i}$, or $T\mathfrak{c}_{2z}$ symmetries constrains the trace of M^{α} to vanish for at least one α; i.e., M^{α} depends on (at most) three real parameters. The traceless condition for $T\mathfrak{i}$ symmetry has previously been observed by [115]. For $T\mathfrak{c}_{2z}$, the constraint on M^{z} is comparatively stronger, leading to M^{z} only depending on one real parameter. This distinction originates from $(T\mathfrak{c}_{2z})^{2}$ being a 4π rotation, and $(T\mathfrak{i})^{2}=T^{2}$ being a 2π rotation, as we proceed to explain.

(i) *Antiunitary representations that square to minus one.* Since $(T\mathfrak{i})^{2}=-I$ for a half-integer-spin representation, the corresponding sewing matrices satisfy $\breve{g}(\boldsymbol{k})K\breve{g}(\boldsymbol{k})K = \breve{g}(\boldsymbol{k})\breve{g}^{*}(\boldsymbol{k}) = -I$ which, in combination with the unitary of $\breve{g}$, implies that $\breve{g}$ is skew-symmetric. From physical grounds, we expect that $T\mathfrak{i}$ inverts the spin and should map any state to an orthogonal state. Mathematically, we understand this from the impossibility of finding a basis where the sewing matrix is diagonal; i.e., the effect of a basis transformation is to conjugate the sewing matrix by a complex, orthogonal matrix [cf. Eq. (37)], but no skew-symmetric matrix can be diagonalized by an orthogonal transformation. In the case of twofold-degenerate bands, we employ that any 2×2, unitary, skew-symmetric matrix is proportional to the Pauli matrix σ_2; the proportionality factor is an irrelevant phase factor that depends on the basis choice [cf. Eq. (22)]. The constraint on the orbital moment, $\boldsymbol{M}=-\sigma_2\boldsymbol{M}^{*}\sigma_2$ (from the fourth column), then implies $\boldsymbol{M}$ is traceless over the entire torus. A similar story unfolds for T symmetry at the inversion-invariant wave vectors.

(ii) *Antiunitary representations that square to one.* $(T\mathfrak{c}_{2z})^{2}=I$ implies that the corresponding $\breve{g}$ (in high-symmetry planes) is symmetric, and may be diagonalized by an orthogonal transformation. By phase redefinitions of the cell-periodic functions, it is always possible to find a basis where $\breve{g}=I$ (the 2×2 identity matrix); in this basis, $M^{z}=-M^{z*}$ (from the fourth column) implies that M^{z} is proportional to the Pauli matrix σ_2 with a real proportionality constant. One may verify that in whatever basis is chosen, M^{z} only depends only on one real parameter.

For any antiunitary representation that squares to a phase factor, the associativity of symmetry representations guarantees that this phase factor is either one [henceforth called type (ii)] or minus one [type (i)]. (Indeed, if $\breve{g}^{2}=e^{i\phi},\breve{g}^{3}=e^{i\phi}\breve{g}=\breve{g}e^{i\phi}\Rightarrow e^{i\phi}\in\mathbb{R}$.) In symmorphic space groups, all order-two symmetries that invert time may be classified, by their corresponding sewing matrices, into types (i) and (ii). This statement must be refined in nonsymmorphic, magnetic space groups where the multiplication rules for sewing matrices are wave-vector-dependent. A case in point is a spin-orbit-coupled system with $T\mathfrak{t}_{\vec{z}/2}$ symmetry, which arises in layered, antiferromagnetic compounds where the layers are stacked in the z direction; the ferromagnetic alignment in each layer alternates between every adjacent layer (separated by half a lattice vector: $\vec{z}/2$). Since $(T\mathfrak{t}_{\vec{z}/2})^{2}$ is the composition of a 2π rotation and a full lattice translation, the sewing matrix satisfies $\breve{g}|_{-\boldsymbol{k}}\breve{g}^{*}|_{\boldsymbol{k}}=-e^{-ik_z}$, which acts like a symmetry of type (i) where $k_z=0$, and of type (ii) where $k_z=\pi$; this leads to wave-vector-dependent constraints on the moment, as detailed in the last column.

The last class of symmetries are completely spatial and unitarily represented. Under basis transformations of the cell-periodic functions, the sewing matrices at high-symmetry points or lines may always be diagonalized by a unitary transformation [cf. Eq. (37)]; this is superfluously true for the single-band sewing matrix. The possible eigenvalues are discrete in phase, and they are determined by space-group rules in the second column. Alternatively stated, if a completely spatial g belongs to the group of the wave vector $\boldsymbol{k}$ [114], the eigenvalues of the sewing matrix label the representations of the bands at $\boldsymbol{k}$. Depending on the symmetry, different

components of the multiband orbital moment may be simultaneously diagonalized at the high-symmetry wave vectors: $\boldsymbol{M}$ for spatial inversion $\mathfrak{i}$, M^x for reflection $\mathfrak{r}_x$ and glide $\mathfrak{g}_{x,\vec{y}/2}$, M^z for rotations $\mathfrak{c}_{nz}$ and screws $\mathfrak{s}_{nz,m}$.

Let us add one final remark regarding the utility of Table II. Since the Roth Hamiltonian is defined by $H_1^R := -M^z B^z$, constraints that act on M^z apply directly to H_1^R. A case in point: since T symmetry imposes $M^z(-\boldsymbol{k}) = -M^z(\boldsymbol{k})$ in the single-band case (second column), it follows immediately that $H_1^R(-\boldsymbol{k}) = -H_1^R(\boldsymbol{k})$. For half-integer-spin representations ($F = 1$), the spin matrix σ^z transforms in the same way as M^z, as derived in Appendix C 3. Consequently, symmetry constraints on the Zeeman coupling $H_1^Z \propto \sigma^z B^z$ may also be deduced from the table, if particularized to $F = 1$.

C. Symmetry of the first-order effective Hamiltonian

As derived in Appendix C 3, the first-order effective Hamiltonian transforms under a symmetry (g) of the orbit configuration as

$$H_1|_{g\circ\boldsymbol{k}} = (-1)^{s(g)+u(g)}\breve{g}K^{s(g)}H_1K^{s(g)}\breve{g}^{-1} + i(-1)^{u(g)}l^{-2}\epsilon_{\alpha\beta}\,\breve{g}\nabla_{\boldsymbol{k}}^{\beta}\breve{g}^{-1}v^{\alpha}|_{\boldsymbol{k}} + l^{-2}\epsilon_{\alpha\beta}\delta^{\beta}v^{\alpha}|_{g\circ\boldsymbol{k}}. \tag{128}$$

This expression may be applied to (a) spin-orbit-coupled systems, (b) spinful systems with negligible spin-orbit coupling, and (c) plausibly to charge-neutral systems with effective magnetic fields. However, the meaning of H_1 is slightly different in each context:

(a) In spin-orbit-coupled systems, the Bloch functions form a half-integer-spin representations of the (magnetic) space group, and $H_1 = H_1^R + H_1^B + H_1^Z$ is contributed by Roth, Berry, and Zeeman [cf. Eq. (57)]. In systems with spacetime-inversion ($T\mathfrak{i}$) symmetry, bands are spin-degenerate and H_1 is a matrix with minimal dimension of two.

(b) H_0 is spin-SU(2)-symmetric in spinful systems with negligible spin-orbit coupling. It is therefore possible to work in a single-band basis that diagonalizes the Zeeman term; i.e., if the field points along $\vec{z}$, we work in the eigenbasis of $\hat{\sigma}^z$ where basis vectors are distinguished by the spin eigenvalue $s \in \pm 1$. In this basis, $H_1(\boldsymbol{k})$ is a diagonal 2×2 matrix:

$$H_1(\boldsymbol{k}) = \begin{pmatrix} H_1^+ & 0 \\ 0 & H_1^- \end{pmatrix},$$

$$H_1^{\pm}(\boldsymbol{k}) = H_1^R + H_1^B \mp \frac{g_0\hbar^2}{4ml^2}. \tag{129}$$

The symmetry analysis, when restricted to the $s = +1$ eigenspace (we could also have picked the -1 subspace; it matters not), is considerably simplified. The spin-restricted set of Bloch functions $\psi_{\boldsymbol{k}}(\boldsymbol{r},s = +)$ forms an integer-spin representation of the (magnetic) space group; the scalar Hamiltonians H_1^R and H_1^B are defined, just as in Eqs. (58) and (59), but with respect to Bloch functions in one spin eigenspace. Since we are ignoring the full spinor structure of the Bloch functions $\psi_{\boldsymbol{k}}(\boldsymbol{r},s \in \pm)$, we might colloquially refer to $\psi_{\boldsymbol{k}}(\boldsymbol{r},+)$ as "single-spin" Bloch functions, and $H_1^R + H_1^B$ as the "spin-independent" first-order-effective Hamiltonian. A symmetry operation in the (magnetic) space group that preserves the eigenvalue of $\hat{\sigma}^z$ is described as a single-spin symmetry. For example, while time reversal flips spin and is represented by $\hat{T} = -i\hat{\sigma}_y K$ satisfying $\hat{T}^2 = -I$, we may define a single-spin time-reversal operator T' that preserves $S_{\vec{z}}$ by composing T with a π-spin rotation about the $\vec{y}$ axis: $\hat{T}' = K$ squares to identity. While T constrains the full spin-dependent effective Hamiltonian [H_1 in Eq. (129)], it is T' that constrains the spin-independent Hamiltonian $H_1^R + H_1^B$. Rather than carry around two symbols (T' and T) for time reversal, it is simpler to talk about a single time reversal (T) which is represented on integer or half-integer spins, which we distinguish by $F \in \{0,1\}$: $\hat{T}^2 = (-1)^F$.

(c) Charge-neutral, cold-atomic systems are characterizable by the Berry phase and the Roth orbital moment: $H_1 = H_1^R + H_1^B$; we shall leave out of this discussion the Zeeman effect. Bloch functions of bosonic atoms (in optical lattices) form an integer-spin representations of the (magnetic) space group.

In all cases (a)–(c), the form of Eq. (128) may be motivated from the following two arguments:

(i) A field-free Bloch Hamiltonian having the same symmetry (g) transforms in nearly the same way to the *first* term on the right-hand side of Eq. (128) [recall Eq. (28)]; the only difference is an additional factor of $(-1)^{s+u}$ (which may be trivial) in Eq. (128). To understand this phase factor, let us consider an alternative definition of a symmetry of the orbit configuration (g) which is consistent with the original definition in Sec. VI A: it is an element of the space group (or magnetic space group) which induces a coordinate transformation where the magnetic field $\boldsymbol{B} \to (-1)^{s+u}\boldsymbol{B}$. In fact, the field-on Hamiltonian [Eqs. (39) and (40)] is invariant under g, if g acts not only on the electronic degrees of freedom in the solid through Eq. (25), but also on the magnetic field [116]. Since Eq. (128) describes a symmetry relation between electronic wave functions of the solid at a fixed field, we expect a compensating factor of $(-1)^{s+u}$.

(ii) The second term originates from the transformation of the Berry term, which we recall from Eq. (64) as being proportional to $\epsilon_{\alpha\beta}\mathfrak{X}^{\beta}v^{\alpha}$. Applying Eq. (29) to the definition of the non-Abelian Berry connection in Eq. (15),

$$\mathfrak{X}^{\alpha}|_{g\circ\boldsymbol{k}} = \breve{g}_{\alpha\beta}\left(\breve{g}K^{s(g)}\mathfrak{X}^{\beta}K^{s(g)}\breve{g}^{-1} + i(-1)^{s(g)}\,\breve{g}\nabla_{\boldsymbol{k}}^{\beta}\breve{g}^{-1}\right)\Big|_{\boldsymbol{k}} + \delta^{\alpha}. \tag{130}$$

The derivation of the above equation is aided by two identities [Eqs. (C20) and (C23)] proven in Appendix C 1. The first (resp. second) term in Eq. (130) contributes to the first (resp. second) term in Eq. (128).

Supposing the second and third terms in Eq. (128) were absent, we say that H_1 transforms covariantly under g. Given that the sewing matrix for a symmetry g depends on the basis chosen for the cell-periodic functions [cf. Eqs. (36) and (37)], one may ask if a basis exists where H_1 transforms covariantly under g for all $\boldsymbol{k}$ in the Brillouin torus. If the answer is yes, such a basis may be exploited to derive symmetry constraints on the Landau levels with relative ease. In short, the answer is no for a large class of band subspaces; we devote Appendix C 4 to a self-contained elaboration of "no," which originates from an obstruction in topologically nontrivial band subspaces. Some well-known obstructions forbid the construction of exponentially localized Wannier functions [117] (i.e., global

sections of the vector bundle), or symmetry-invariant Wannier functions [118]. In Appendix C 4, we will describe a novel type of obstruction—to symmetry covariance of H_1. The reader who is more interested in quantization conditions may transit immediately to the next section (Sec. VI D).

D. Symmetry of the first-order effective propagator

In the last section we dealt primarily with the symmetry and gauge transformations of the first-order effective Hamiltonian H_1, and argued that H_1 generically does not transform non-covariantly. A related observation is that the eigenspectrum of H_1 has no gauge-invariant meaning. This reflects how the effective-Hamiltonian description of a Bloch electron in a magnetic field is fundamentally a nondynamical gauge theory; in gauge theories, a known source of gauge-invariant observables comes from the spectrum of Wilson-loop operators [65]. In our context, we identify the analogous operator as the propagator $\mathcal{A}$, which we defined in Eq. (74) as the unitary generated by H_1 over the cyclotron period.

We will show that, unlike H_1, $\mathcal{A}$ behaves nicely under gauge and symmetry transformations. Precisely, we will show that $\mathcal{A}$ transforms covariantly under the $U(D)$ gauge transformation of the type Eq. (22) (in Sec. VI D 1), and covariantly under symmetry transformations of the type Eq. (128) (in Sec. VI D 2). One motivation for investigating these transformation behaviors is that $\mathcal{A}$ encodes the subleading corrections to the Bohr-Sommerfeld quantization conditions, as derived in Sec. V C 1. As we will show, the gauge covariance of $\mathcal{A}$ implies the gauge invariance of the Landau levels determined from the quantization conditions; the symmetry covariance of $\mathcal{A}$ implies certain symmetry constraints for the Landau levels that we will prove below.

We will use the same symbols ($\mathcal{A}$ and H_1) in a variety of contexts which are not necessarily mutually exclusive: (i) nondegenerate subspaces ($D = 1$), in which case $\mathcal{A}$ is a unimodular phase factor, (ii) degenerate subspaces ($D > 1$), in which case $\mathcal{A}$ is a $D \times D$ unitary, (iii) spin-orbit-coupled systems, (iv) spinful systems with negligible spin-orbit coupling, and (v) charge-neutral particles coupled to effective magnetic fields. We remind the reader that H_1 has a slightly different meaning in each of (iii)–(v), as detailed below Eq. (128). Unless D or the symmetry representation is explicitly specified in an equation, the reader may safely assume that the equation applies to all of (i)–(v).

We highlight one potentially confusing case where it is useful to have two related notions of $\mathcal{A}$: this is the case of spin-degenerate ($D = 2$) bands in solids with negligible spin-orbit coupling. We define the spin-dependent $\mathcal{A}_{F=1}$ as the unitary generated by the spin-dependent H_1 [the 2×2 matrix in Eq. (129)], and the spin-independent $\mathcal{A}_{F=0}$ as the unitary generated by the spin-independent, scalar $H_1^R + H_1^B$ [defined in Eq. (129)]. Both notions are related as

$$\mathcal{A}_{F=1} = \begin{pmatrix} \mathcal{A}_{F=0} e^{i\pi \frac{g_0}{2} \frac{m_c[\mathfrak{o}]}{m}} & 0 \\ 0 & \mathcal{A}_{F=0} e^{-i\pi \frac{g_0}{2} \frac{m_c[\mathfrak{o}]}{m}} \end{pmatrix},$$
$$\det \mathcal{A}_{F=1} = \mathcal{A}_{F=0}^2, \tag{131}$$

where the only spin-dependent component of $\mathcal{A}_{F=1}$ originates from the Zeeman coupling [g_0, m_c, m are defined in Eq. (72)]. The right equation in Eq. (131) expresses how the determinant of $\mathcal{A}_{F=1}$ is fully determined by the spin-independent $H_1^R + H_1^B$, and is not affected by the Zeeman splitting. Supposing g is a symmetry of the orbit configuration, it is represented differently when it acts on $\mathcal{A}_{F=1}$ vs $\mathcal{A}_{F=0}$; e.g., time reversal is represented as $\hat{T}^2 = (-1)^F$, as explained below Eq. (129).

1. Gauge covariance of the first-order propagator

One motivation to prove that the first-order propagator transforms covariantly: it follows that the spectrum obtained from the multiband quantization condition in Eq. (108) is gauge invariant; we remind the reader that the gauge invariance of the single-band quantization condition has been proven with less effort in Sec. V C 1.

To substantiate the multiband claim, we remind the reader that a matrix transforms covariantly if it is conjugated by the unitary V which reshuffles the bands within P [cf. Eq. (22)], as exemplified by the Roth and spin matrices [defined in Eq. (75), Eq. (17), and Eq. (20)]

$$\mathfrak{A} \to V^{-1}\mathfrak{A}V, \quad \boldsymbol{\sigma} \to V^{-1}\boldsymbol{\sigma}V. \tag{132}$$

In the single-band case, V is a commuting phase factor that cancels with V^{-1}; hence all covariant objects are also invariant. In contrast, the Berry connection transforms non-covariantly, as shown in Eq. (23). Nevertheless, we show that the propagator around a loop transforms as

$$\mathcal{A}[\mathfrak{o}] \to V(\boldsymbol{k}(0))^{-1}\mathcal{A}[\mathfrak{o}]V(\boldsymbol{k}(0)), \tag{133}$$

with $\boldsymbol{k}(0)$ the base point of the loop $\mathfrak{o}$. It follows that the quantization condition in Eq. (108) is gauge-invariant:

$$\begin{aligned} 0 &= \det\left[e^{il^2S}\mathcal{A}[\mathfrak{o}] + I\right] \\ &\to \det\left[e^{il^2S}V(\boldsymbol{k}(0))^{-1}\mathcal{A}[\mathfrak{o}]V(\boldsymbol{k}(0)) + I\right] \\ &= \det[V(\boldsymbol{k}(0))^{-1}]\det\left[e^{il^2S}\mathcal{A}[\mathfrak{o}] + I\right]\det[V(\boldsymbol{k}(0))^{-1}]. \end{aligned}$$

To prove Eq. (133), it is convenient to consider the propagator

$$\begin{aligned} &\mathcal{A}[\boldsymbol{k} \leftarrow \boldsymbol{k} - d\boldsymbol{k}] \\ &\quad\approx \exp[i(\mathfrak{A} + \mathfrak{X})\cdot d\boldsymbol{k} + i\sigma^z(Z/v^\perp)dk] + O(dk^2) \end{aligned} \tag{134}$$

over an infinitesimal path along the orbit, ending at $\boldsymbol{k}(t)$ and beginning at $\boldsymbol{k}(t - \delta t) = \boldsymbol{k}(t) - d\boldsymbol{k}(t)$; in short, we call such objects infinitesimal propagators. Applying Eqs. (132) and (23), the infinitesimal propagator transforms as

$$\begin{aligned} &\mathcal{A}[\boldsymbol{k} \leftarrow \boldsymbol{k} - d\boldsymbol{k}] \\ &\quad\to e^{iV^{-1}(\mathfrak{A}+\mathfrak{X})V\cdot d\boldsymbol{k} + iV^{-1}\sigma^z(Z/v^\perp)Vdk - V^{-1}\nabla_{\boldsymbol{k}}V\cdot d\boldsymbol{k}} \\ &\quad= V^{-1}(\boldsymbol{k})e^{i(\mathfrak{A}+\mathfrak{X})\cdot d\boldsymbol{k} + i\sigma^z(Z/v^\perp)dk}V(\boldsymbol{k})e^{-V^{-1}\nabla_{\boldsymbol{k}}V\cdot d\boldsymbol{k}} \\ &\quad= V(\boldsymbol{k})^{-1}\mathcal{A}[\boldsymbol{k} \leftarrow \boldsymbol{k} - d\boldsymbol{k}]V(\boldsymbol{k} - d\boldsymbol{k}), \end{aligned} \tag{135}$$

to linear order in dk. Consider a path-ordered multiplication of these infinitesimal propagators around a closed orbit beginning and ending at $\boldsymbol{k}(0)$; every V matrix that is not evaluated at $\boldsymbol{k}(0)$ is multiplied with its inverse. What remains of this path-ordered product, after taking the limit $\delta t \to 0$, is the right-hand side of Eq. (133).

2. Symmetry covariance of the first-order propagator

For a system having a symmetry (g) of the orbit configuration, we consider the infinitesimal propagator centered at wave vector $g\circ\boldsymbol{k}$ on an orbit, which is related through Eq. (128) to the infinitesimal propagator centered at $\boldsymbol{k}$:

$$e^{-iH_1\delta t/\hbar}|_{g\circ\boldsymbol{k}}=\left(\breve{g}K^s e^{-i(-1)^u H_1\delta t/\hbar}K^s\breve{g}^{-1}\right)\left(e^{(-1)^u l^{-2}\epsilon_{\alpha\beta}\breve{g}\nabla_{\boldsymbol{k}}^{\beta}\breve{g}^{-1}v^{\alpha}\delta t/\hbar}\right)\big|_{\boldsymbol{k}}e^{-il^{-2}\epsilon_{\alpha\beta}\delta^{\beta}v^{\alpha}\delta t/\hbar}\big|_{g\circ\boldsymbol{k}}+O(\delta t^2). \tag{136}$$

Hamilton's equation of motion [Eq. (1) particularized to $\boldsymbol{B}=-B\vec{z}$] informs us that $\delta t/\hbar l^2=-\delta k_y(\boldsymbol{k})/v^x(\boldsymbol{k})=\delta k_x(\boldsymbol{k})/v^y(\boldsymbol{k})$ [let us also define $\delta\boldsymbol{k}^{\perp}(\boldsymbol{k})=(\delta k_x,\delta k_y)$]; hence the above equation simplifies to

$$\begin{aligned}e^{-iH_1\delta t/\hbar}|_{g\circ\boldsymbol{k}}&=\left(\breve{g}K^s e^{-i(-1)^u H_1\delta t/\hbar}K^s\breve{g}^{-1}\right)\left(e^{-(-1)^u\breve{g}\nabla_{\boldsymbol{k}}\breve{g}^{-1}\cdot\delta\boldsymbol{k}^{\perp}}\right)\big|_{\boldsymbol{k}}e^{i\boldsymbol{\delta}\cdot\delta\boldsymbol{k}^{\perp}}\big|_{g\circ\boldsymbol{k}}+O(\delta t^2)\\&=\left(\breve{g}K^s e^{-i(-1)^u H_1\delta t/\hbar}K^s\right)\big|_{\boldsymbol{k}}\breve{g}^{-1}\big|_{\boldsymbol{k}-(-1)^u\delta\boldsymbol{k}^{\perp}(\boldsymbol{k})}e^{i\boldsymbol{\delta}\cdot\delta\boldsymbol{k}^{\perp}}\big|_{g\circ\boldsymbol{k}}+O(\delta t^2).\end{aligned} \tag{137}$$

If $u(g)=1$, the infinitesimal-time propagator centered at $\boldsymbol{k}$ (the right-hand side of the above equation) is reversed in orientation with respect to the semiclassical orbit, and vice versa. Equation (137) forms the basis to derive the symmetry constraints for any configuration of orbits.

Let us translate the symmetry constraint on infinitesimal propagators into a constraint for finite-time propagators, which are analogous to Wilson lines. Suppose $\boldsymbol{k}_i$ and $\boldsymbol{k}_f$ are boundary points of a curved line segment $(\mathcal{S})$ contained within an orbit; $\mathcal{S}$ is equipped with an orientation such that $\boldsymbol{k}_i$ (resp. $\boldsymbol{k}_f$) is the initial point (resp. final point), which we will denote as $\mathcal{S}:\boldsymbol{k}_f\leftarrow\boldsymbol{k}_i$. The propagator over $\mathcal{S}$ is defined by

$$\mathcal{A}[\mathcal{S}:\boldsymbol{k}_f\leftarrow\boldsymbol{k}_i]:=\overline{\exp}\left\{-i\int_{t(\boldsymbol{k}_i)}^{t(\boldsymbol{k}_f)}H_1(\boldsymbol{k}(t))\frac{dt}{\hbar}\right\}, \tag{138}$$

with $\overline{\exp}$ a path-ordered exponential, and $\boldsymbol{k}(t)$ and its inverse $t(\boldsymbol{k})$ determined by Hamilton's equation. We may define $-\mathcal{S}$ as the same line segment as $\mathcal{S}$ but with the opposite orientation; the corresponding propagator satisfies

$$\mathcal{A}[-\mathcal{S}:\boldsymbol{k}_i\leftarrow\boldsymbol{k}_f]=\mathcal{A}[\mathcal{S}:\boldsymbol{k}_f\leftarrow\boldsymbol{k}_i]^{-1}. \tag{139}$$

Let us define the symmetry-mapped segment $(g\circ\mathcal{S})$ as being bounded by initial point $g\circ\boldsymbol{k}_i$ and final point $g\circ\boldsymbol{k}_f$ [with $g\circ\boldsymbol{k}:=(-1)^s\breve{g}\boldsymbol{k}$]. The two corresponding segment propagators are related as

$$\begin{aligned}&\mathcal{A}[g\circ\mathcal{S}:g\circ\boldsymbol{k}_f\leftarrow g\circ\boldsymbol{k}_i]\\&=\breve{g}(\boldsymbol{k}_f)\,K^s\,\mathcal{A}[\mathcal{S}:\boldsymbol{k}_f\leftarrow\boldsymbol{k}_i]\,K^s\,\breve{g}^{-1}(\boldsymbol{k}_i)\,e^{i\boldsymbol{\delta}\cdot\int_{g\circ\mathcal{S}}d\boldsymbol{k}^{\perp}},\end{aligned} \tag{140}$$

which may be derived by a path-ordered multiplication of the infinitesimal propagators in Eq. (137), and taking the limit $\delta t\to 0$; in this process, every sewing matrix [originating from the right-hand side of Eq. (137)] is multiplied with its inverse, except for the sewing matrices at the boundary points.

Let us generalize Eq. (140) to a relation between propagators over closed orbits. If $\mathfrak{o}$ is a closed orbit with base point $\boldsymbol{k}_1$, then

$$\mathfrak{o}:\boldsymbol{k}_1\leftarrow\boldsymbol{k}_1,\quad \mathcal{A}[g\circ\mathfrak{o}]=\breve{g}(\boldsymbol{k}_1)\,K^s\,\mathcal{A}[\mathfrak{o}]\,K^s\,\breve{g}^{-1}(\boldsymbol{k}_1). \tag{141}$$

Note that $\int_{g\circ\mathfrak{o}}d\boldsymbol{k}^{\perp}=\boldsymbol{0}$ for a closed orbit; hence the $\boldsymbol{\delta}$-dependent phase factor on the right-hand side of Eq. (140) is trivial. When particularized to the case that $g\circ\mathfrak{o}$ and $\mathfrak{o}$ are identical orbits, up to a reversal in orientation that depends on $u(g)$,

$$g\circ\mathfrak{o}=(-1)^u\mathfrak{o},\quad \mathcal{A}[\mathfrak{o}]^{(-1)^u}=\breve{g}(\boldsymbol{k}_1)\,K^s\,\mathcal{A}[\mathfrak{o}]\,K^s\,\breve{g}^{-1}(\boldsymbol{k}_1), \tag{142}$$

where $\mathcal{A}^{(-1)^u}$ equals $\mathcal{A}$ (resp. $\mathcal{A}^{-1}$) if $u=0$ (resp. $u=1$). A topologically distinct possibility is that $g\circ\mathfrak{o}$ and $\mathfrak{o}$ are disconnected orbits. Let us define $\mathfrak{o}_1$ and $\mathfrak{o}_2$ as two disconnected orbits, whose orientations are determined by Hamilton's equation; $\mathfrak{o}_1$ is equipped with a base point $\boldsymbol{k}_1$, and $\mathfrak{o}_2$ with a base point $\boldsymbol{k}_2:=g\circ\boldsymbol{k}_1$. Then a simple generalization of Eq. (142) provides us with

$$\begin{aligned}g\circ\mathfrak{o}_1&=(-1)^u\mathfrak{o}_2,\\\mathcal{A}[\mathfrak{o}_2]^{(-1)^u}&=\breve{g}(\boldsymbol{k}_1)\,K^s\,\mathcal{A}[\mathfrak{o}_1]\,K^s\,\breve{g}^{-1}(\boldsymbol{k}_1).\end{aligned} \tag{143}$$

3. Ten classes of closed, elementary orbits

In a Brillouin two-torus $(BT_{\perp})$, any closed orbit configuration possessing a symmetry g may be divided into a set of elementary orbits $(\{E_i\})$. An elementary orbit E_i is defined to be the smallest possible closed orbit configuration that is closed under g; i.e., it cannot further be divided into smaller configurations which are closed under g. To clarify two distinct notions, "closed orbits" do not wrap around $BT_{\perp}$; if E_i is "closed under g," we mean that for every $\boldsymbol{k}^{\perp}\in E_i$, $g\circ\boldsymbol{k}^{\perp}\in E_i$ as well. We remind the reader that g maps $\boldsymbol{k}^{\perp}$ to $g\circ\boldsymbol{k}^{\perp}:=(-1)^{s(g)}\breve{g}^{\perp}\boldsymbol{k}^{\perp}$, with $\breve{g}^{\perp}$ the point-group component of g, as restricted to $BT_{\perp}$. Generally, E_i is composed of one or more closed orbits.

If there are multiple symmetries $(g,g',\ldots)$ in the group of the orbit configuration, the same orbit configuration may be divided into two (or more) distinct sets of elementary orbits ($\{E_i\}$ and $\{E_i'\}$), which are closed under g and g', respectively; each elementary orbit is therefore defined by its closure under a *single* symmetry, and we emphasize this by the paired notation (g,E_i). The motivation for this g-centric organization is that distinct symmetries impose distinct constraints on the propagators, which we classify into ten (and only ten) classes. In other words, any pair (g,E_i) falls into one of ten classes, and the propagator(s) over closed orbits $(\in E_i)$ satisfy one of ten classes of constraints, which we summarize in Table III. These ten classes were first introduced in Ref. [32] and exemplified by many existing materials; here we present a more mathematically oriented discussion that emphasizes the group-theoretic aspects of the ten classes. In addition, (i) a more thorough discussion is also provided for

TABLE III. The first column distinguishes between three topologically distinct mappings of $g : \boldsymbol{k} \to g \circ \boldsymbol{k}$, as summarized in Eqs. (144) and (145). The second and third columns subdivide the three mapping classes according to two $\mathbb{Z}_2$ indices defined in Eqs. (25) and (117); this gives ten classes in total. The fourth column describes the constraints on the propagator: for classes I and II-A (top six rows), $\mathcal{A}$ is the propagator for a single, elementary orbit $\mathfrak{o}$; for class II-B (last four rows), $\{\mathcal{A}_j\}_{j=1}^2$ is shorthand for $\mathcal{A}[\mathfrak{o}_j]$, with $\mathfrak{o}_{j+1} = g \circ \mathfrak{o}_j$ being symmetry-related, closed orbits. $\breve{g}$, $\{\breve{g}_i\}_{i=1}^N$, and $\tilde{g}$ are representations of the point group generated by g [cf. Eqs. (33)–(35)], as summarized in the sixth column. The fifth column describes the constraint imposed by g on the spectrum of $\mathcal{A}$, which we denote by $\sigma(\mathcal{A})$. We indicate the lack of a symmetry constraint with a $-$. $\sigma(\mathcal{A}) = \sigma(\mathcal{A})^*$ means the spectrum is invariant under complex conjugation; it follows immediately that $\det \mathcal{A} = \pm 1$. In some cases, the sign of this determinant is fully determined by specifying the band degeneracy (D) and the symmetry representation of the Bloch functions (whether integer- or half-integer-spin). Class-I symmetries have order two, and $\sigma(\mathcal{A}) = \sqcup_{i\in\pm}\sigma_i$ indicates that $\mathcal{A}$ is block-diagonal with respect to the two representations of the order-two symmetry. The last column lists some representative examples of the ten symmetry classes; the symbolic notation of various symmetries has been summarized in Sec. III C.

	$u(g)$	$s(g)$	Constraint on $\mathcal{A}$	Spectrum of $\mathcal{A}$	Representation of g	Ex. of g
(I)$\forall\, \boldsymbol{k}^\perp$,	0	0	$\mathcal{A} = \breve{g}\mathcal{A}\breve{g}^{-1}$	$\sigma(\mathcal{A}) = \sigma_+ \sqcup \sigma_-$	$\breve{g}^2 = e^{i\pi Fa - i\boldsymbol{k}_1\cdot\boldsymbol{R}}$	$\mathfrak{r}_z, \mathfrak{g}_{z,\vec{x}/2}$
$\boldsymbol{k}^\perp = g\circ\boldsymbol{k}^\perp$	0	1	$\mathcal{A} = \breve{g}\mathcal{A}^*\breve{g}^{-1}$	$\sigma(\mathcal{A}) = \sigma(\mathcal{A})^*$	$(\breve{g}K)^2 = e^{i\pi Fa - i\boldsymbol{k}_1\cdot\boldsymbol{R}}$	$T\mathfrak{i}, T\mathfrak{c}_{2z}$
(II-A)	0	0	$\mathcal{A} = \tilde{g}\mathcal{A}\tilde{g}^{-1}$	$-$	$\tilde{g}^N = \mathcal{A}^p e^{i\pi Fa}$	$\mathfrak{i}, \mathfrak{c}_{nz}$
$\boldsymbol{k}^\perp \in \mathfrak{o}$,	0	1	$\mathcal{A} = \tilde{g}\mathcal{A}^*\tilde{g}^{-1}$	$\sigma(\mathcal{A}) = \sigma(\mathcal{A})^*$	$(\tilde{g}K)^N = \mathcal{A}^p e^{i\pi Fa}$	$T, T\mathfrak{c}_{6z}$
$\lvert\mathfrak{o}\rvert = \lvert g\circ\mathfrak{o}\rvert$	1	0	$\mathcal{A} = \breve{g}\mathcal{A}^{-1}\breve{g}^{-1}$	$\sigma(\mathcal{A}) = \sigma(\mathcal{A})^*$	$\breve{g}^N = e^{i\pi Fa - i\boldsymbol{k}_1\cdot\boldsymbol{R}}$	$\mathfrak{r}_x, \mathfrak{r}_y$
	1	1	$\mathcal{A} = \breve{g}\mathcal{A}^t\breve{g}^{-1}$	$-$	$(\breve{g}K)^N = e^{i\pi Fa - i\boldsymbol{k}_1\cdot\boldsymbol{R}}$	$T\mathfrak{r}_x, T\mathfrak{r}_y$
(II-B)	0	0	$\mathcal{A}_2 = \breve{g}_1\mathcal{A}_1\breve{g}_1^{-1}$	$\sigma(\mathcal{A}_2) = \sigma(\mathcal{A}_1)$	$\breve{g}_N \ldots \breve{g}_1 = e^{i\pi Fa - i\boldsymbol{k}_1\cdot\boldsymbol{R}}$	$\mathfrak{c}_{nz}$
$\boldsymbol{k}^\perp \in \mathfrak{o}$,	0	1	$\mathcal{A}_2 = \breve{g}_1\mathcal{A}_1^*\breve{g}_1^{-1}$	$\sigma(\mathcal{A}_2) = \sigma(\mathcal{A}_1)^*$	$\breve{g}_N K \ldots \breve{g}_1 K = e^{i\pi Fa - i\boldsymbol{k}_1\cdot\boldsymbol{R}}$	T
$\lvert\mathfrak{o}\rvert \neq \lvert g\circ\mathfrak{o}\rvert$	1	0	$\mathcal{A}_2 = \breve{g}_1\mathcal{A}_1^{-1}\breve{g}_1^{-1}$	$\sigma(\mathcal{A}_2) = \sigma(\mathcal{A}_1)^*$	$\breve{g}_N \ldots \breve{g}_1 = e^{i\pi Fa - i\boldsymbol{k}_1\cdot\boldsymbol{R}}$	$\mathfrak{r}_x, \mathfrak{r}_y$
	1	1	$\mathcal{A}_2 = \breve{g}_1\mathcal{A}_1^t\breve{g}_1^{-1}$	$\sigma(\mathcal{A}_2) = \sigma(\mathcal{A}_1)$	$\breve{g}_N K \ldots \breve{g}_1 K = e^{i\pi Fa - i\boldsymbol{k}_1\cdot\boldsymbol{R}}$	$T\mathfrak{r}_x, T\mathfrak{r}_y$

g'-symmetric orbit configurations that lie within mirror/glide-invariant planes, where g' is an additional symmetry distinct from said mirror/glide. (ii) We also demonstrate how, with additional input about the symmetry representations of Bloch functions on the orbit, one may derive additional constraints on the spectrum of $\mathcal{A}$ that go beyond Table III.

The ten classes are partially distinguished by two $\mathbb{Z}_2$ indices which we have previously defined: $u(g)$ and $s(g)$. To remind the reader, $s(g) = 1$ if g contains a time reversal, and 0 otherwise [cf. Eq. (25)]; $u(g) = 0$ if the determinant of $\breve{g}^\perp$ (the point-group component of g, restricted to the xy plane) equals 1, and $u(g) = 1$ if $\det \breve{g}^\perp = -1$ [cf. Eq. (117)]; as explained in Sec. VI A, $u(g) = 0$ (resp. 1) if g preserves (resp. inverts) the orientation of the semiclassical orbit.

We shall subdivide the ten classes according to three topologically distinct mappings of $g : \boldsymbol{k}^\perp \to g \circ \boldsymbol{k}^\perp$ as

$$\text{(I)}\ \forall\, \boldsymbol{k}^\perp, \quad \boldsymbol{k}^\perp = g \circ \boldsymbol{k}^\perp, \tag{144}$$

$$\text{(II) generically,} \quad \boldsymbol{k}^\perp \neq g \circ \boldsymbol{k}^\perp,$$
$$\boldsymbol{k}^\perp \in \mathfrak{o}, \quad \begin{cases} \text{(II-A)} & \lvert\mathfrak{o}\rvert = \lvert g\circ\mathfrak{o}\rvert, \\ \text{(II-B)} & \lvert\mathfrak{o}\rvert \neq \lvert g\circ\mathfrak{o}\rvert. \end{cases} \tag{145}$$

In mappings of class I, all wave vectors in $BT_\perp$ are individually invariant under the symmetry, which implies that $u(g) = 0$, as proven in Eq. (124). There are therefore two classes of class-I elementary orbits which we distinguish by $s(g) \in \mathbb{Z}_2$. For mappings of class II, generic wave vectors are not invariant under g, but there exist closed submanifolds (isolated points/lines) of $BT_\perp$ which are invariant; we have shown in Sec. VI A that points occur iff $u = 0$, and lines iff $u = 1$. Suppose $\boldsymbol{k}^\perp$ is a point in a closed orbit $\mathfrak{o} \in E_i$; since E_i is closed under g, $g \circ \boldsymbol{k}^\perp \in g \circ \mathfrak{o} \in E_i$. We further distinguish between mappings where $g \circ \mathfrak{o}$ is identical to $\mathfrak{o}$ up to orientation (class II-A), or they are disconnected orbits (class II-B). We employ the notation that $\mathfrak{o}$ and $-\mathfrak{o}$ have opposing orientations, and $\lvert\mathfrak{o}\rvert$ as having no orientation. The defining characteristics of II-A and II-B may then be expressed as in Eq. (145). In class I and II-A, E_i is composed of a single orbit $\mathfrak{o}$, and we may say that $\mathfrak{o}$ is self-constrained by g; in II-B, E_i is composed of at least two closed orbits, which we say are mutually constrained by g. For class-II mappings, there are no constraints on s or u [as there were for class I in Eq. (124)]; hence there are four classes for each of II-A and II-B. This gives ten classes of elementary orbits in total, whose defining characteristics are summarized in the first three columns of Table III; representative examples of each class are given in the last column.

The rest of the table summarizes how the space-group symmetry g constrains the propagators $\mathcal{A}$; the operators (denoted by $\breve{g}$ in eight rows, and by $\tilde{g}$ in two) that constrain $\mathcal{A}$ form a representation of the space-group symmetry g, as shown in column six. $\breve{g}$ form either a linear or projective [119] representation of the point group (P_g) generated by g [120], while $\tilde{g}$ forms necessarily a projective representation. To clarify this comment, P_g is isomorphic to $\mathbb{Z}_N$ if g has order N; generally, P_g is a subgroup of the full point group of the space group. It is well known that symmorphic (resp. nonsymmorphic) space groups are split (resp. unsplit) extensions of point groups by discrete spatial translations [121]. Unsplit extensions may contain nonsymmorphic elements of order N—the corresponding multiplication rule is represented by $\breve{g}^N \propto e^{-i\boldsymbol{k}_1\cdot\boldsymbol{R}}$. Double space groups are known to correspond to a further extension by a 2π spin rotation; the multiplication rule for an order-N symmetry is represented by $\breve{g}^N \propto e^{iF\pi a}$. These two observations explain the form of the multiplication rules in all ten rows except for the third and fourth, where

respectively $\tilde{g}^N$ and $(\tilde{g}K)^N$ are proportional to $\mathcal{A}^p$, with $p \neq 0$ (mod N) and depending on (g, E_i). These two rules represent an unsplit extension of the point group by quasimomentum translations around the orbit E_i, which in the present context is a single loop; these translations are represented by the propagator $\mathcal{A}$, which generates a normal subgroup (isomorphic to $\mathbb{Z}$) of the extension. Extensions by quasimomentum loop translations are one key result of this section, and occur for all self-constrained orbits having no g-invariant points; this sharply delineates class-II-A orbits with $u = 0$ from the remaining eight classes, which are all linearly represented with respect to $\mathcal{A}$. To recapitulate, $\check{g}$ (or $\tilde{g}$), $e^{i\boldsymbol{k}_\perp\cdot\boldsymbol{R}}$, $e^{i\pi}$, and $\mathcal{A}$ generate a group; (i) the multiplication rules of this group (columns four and six), when combined with (ii) the spectral constraint on $\mathcal{A}$ (column five), uniquely distinguishes each of the ten classes. In other words, given (i) and (ii), one may uniquely determine the corresponding mapping type (I, II-A, II-B), u and s. We derive the table and discuss its implications in the following subsections, which are divided according to class-I mappings (Sec. VI D 4), class II-A (Sec. VI D 5), and class II-B (Sec. VI D 6). For some (and only some) classes, the above spectral constraints are further strengthened when given additional data about the band degeneracy D and the spin representation (whether integer or half-integer).

One last remark regards the application of Table III beyond the semiclassical theory of magnetotransport. All constraints in Secs. VI D 3–VI D 8 which are tabulated or expressed in labeled equations remain valid if we substitute $\mathcal{A} \to \mathcal{W}$, with $\mathcal{W}$ the purely geometric component of $\mathcal{A}$. That is, if we set the Roth and Zeeman Hamiltonians to zero, $\mathcal{A}$ reduces to $\mathcal{W}$—a path-ordered exponential of the Berry connection [20], which is non-Abelian for $D > 1$. Though generically the spectra of $\mathcal{A}$ and $\mathcal{W}$ are distinct, they satisfy the same type of constraints; e.g., if $\sigma(\mathcal{A}) = \sigma(\mathcal{A})^*$ from Table III, so would $\sigma(\mathcal{W}) = \sigma(\mathcal{W})^*$. $\mathcal{W}$ is the matrix representation of holonomy around the orbit $\mathfrak{o}$ [122] and has been called the Wilson loop of the Berry gauge field [65]. The commonality between $\mathcal{W}$ and $\mathcal{A}$ originates from their identical transformation behavior under symmetry [cf. Eqs. (140)–(143)]. The Wilson loop is a basic geometric characterization of bands that is intimately related to the topology of wave functions over the Brillouin torus [122].

4. Class-I elementary orbits

Let g be a symmetry such that every wave vector ($\boldsymbol{k}^\perp$) in a Brillouin two-torus $BT_\perp$ is g-invariant. Common examples include $g = T\mathfrak{i}, T\mathfrak{c}_{2z}$, and $\mathfrak{r}_z$; for 3D $T\mathfrak{i}$-symmetric solids, any two-torus embedded in the 3D Brillouin zone is invariant under $T\mathfrak{i}$, while for 3D solids with either $T\mathfrak{c}_{2z}$ or $\mathfrak{r}_z$ symmetry, we would particularize to the high-symmetry planes ($k_z = 0$ and π).

Since every $\boldsymbol{k}^\perp \in BT_\perp$ is g-invariant, if $\boldsymbol{k}^\perp \in \mathfrak{o}$ (a single closed orbit), then $\mathfrak{o} = g \circ \mathfrak{o}$, which further implies $\mathfrak{o}$ is itself an elementary orbit [of class I, as classified in Eqs. (144) and (145)]. If g is the only symmetry of $\mathfrak{o}$, there is no constraint on the shape of $\mathfrak{o}$. We have also proven that $u(g) = 0$ in Eq. (124), i.e., that class-I symmetries are orientation preserving. We further subdivide class-I orbits according to whether g includes a time reversal or not [$s(g) = 1$ or 0 respectively]; s distinguishes between two classes of constraints on the propagator $\mathcal{A}[\mathfrak{o}]$ over the oriented $\mathfrak{o}$. In contexts where we are discussing a single orbit, we employ $\mathcal{A}$ as a shorthand for $\mathcal{A}[\mathfrak{o}]$.

Class-I elementary orbits with $s(g) = 0$. This occurs when g is purely a spatial transformation; we ignore g that is purely a spatial translation, because it trivially constrains the propagator. To leave every wave vector in $BT_\perp$ invariant, $BT_\perp$ must be a mirror ($g = \mathfrak{r}_z$) or a glide (e.g., $g = \mathfrak{g}_{z,\vec{x}/2}$) plane. In either case, g is an order-two spatial symmetry [the order of a symmetry is defined in Eqs. (31)–(35)], which implies that g has two distinct representations. It is useful to block-diagonalize the Hilbert space ($\mathcal{L}$) with respect to the two representations of g; we shall denote this decomposition as $\mathcal{L} = \mathcal{L}_+ \oplus \mathcal{L}_-$. The corresponding block-diagonalization of $\mathcal{A}$ is denoted as $\mathcal{A} = \mathcal{A}_+ \oplus \mathcal{A}_-$ in the first row of Table III.

Suppose there exists a distinct symmetry g' in the group of the orbit configuration, whose operation preserves the decomposition $\mathcal{L}_+ \oplus \mathcal{L}_-$. That is, if a Bloch function $\psi \in \mathcal{L}_+$, then the symmetry-mapped Bloch function $g' \circ \psi$ belongs also in $\mathcal{L}_+$. To analyze how g' further constrains $\mathcal{A}_\pm$, we divide the orbit configuration into elementary orbits $\{(g', E_i')\}$; each of $\{(g', E_i')\}$ falls into one of the remaining nine classes. We may then apply any of the results in the bottom nine rows of Table III, with the understanding that $\mathcal{A}$ (as denoted in the table) is the propagator restricted to $\mathcal{L}_\pm$.

Let us particularize to g' that permutes the two representations of g, i.e., $g' \circ \mathcal{L}_\pm = \mathcal{L}_\mp$. Then if $\mathcal{A}_+$ is the propagator for a closed orbit $\mathfrak{o}$, it is symmetry-related to $\mathcal{A}_-$ which is the propagator for $g' \circ \mathfrak{o}$; in general $\mathfrak{o} \neq g' \circ \mathfrak{o}$. $\mathcal{A}_+[\mathfrak{o}]$ and $\mathcal{A}_-[g' \circ \mathfrak{o}]$ are mutually constrained in four possible ways, depending on the $\mathbb{Z}_2$ indices $u(g')$ and $s(g')$ which characterize g' (not g); these constraints are summarized in Table IV below, which applies regardless of whether $\mathfrak{o} = g' \circ \mathfrak{o}$ or not. Table IV summarizes one new result of this work.

The four classes of (g', E_i') in Table IV are essentially identical to the four classes of class-II-B elementary orbits (bottom four rows of Table III), if one relabels $\mathcal{A}_{1,2} \leftrightarrow \mathcal{A}_\pm$. The basic commonality is the existence of two distinct but symmetry-related vector bundles, each of which is defined over a 1D base

TABLE IV. Table of constraints for solids with (i) a class-I, unitarily represented, order-two symmetry g, and (ii) an additional symmetry g' that permutes the two representations of g. The second and third columns classify the constraints according to two $\mathbb{Z}_2$ indices [defined in Eqs. (25) and (117)] that characterize g' (not g). Fourth column describes the constraints on propagators $\mathcal{A}_\pm$ which are defined with respect to states in $\mathcal{L}_\pm$. In the entire table, $\mathcal{A}_+$ is shorthand for $\mathcal{A}_+[\mathfrak{o}]$, and $\mathcal{A}_-$ for $\mathcal{A}_-[g \circ \mathfrak{o}]$. The sixth column lists some representative examples of g', for the specific case of a half-integer-spin representation of $g = \mathfrak{r}_z$. For the nonsymmorphic examples of g' [$\mathfrak{s}_{2z,\vec{z}/2}$ and $T\mathfrak{g}_{x,\vec{z}/2}$], g' permutes the half-integer-spin representation of $\mathfrak{r}_z$ in the $k_z = \pi$ plane; for the remaining two symmorphic examples, this permutation occurs in both $k_z = 0$ and π planes.

	u'	s'	Constraint on $\mathcal{A}_\pm$	Spectrum of $\mathcal{A}_\pm$	g'
	0	0	$\mathcal{A}_+ = \check{g}'\mathcal{A}_-(\check{g}')^{-1}$	$\sigma(\mathcal{A}_+) = \sigma(\mathcal{A}_-)$	$\mathfrak{s}_{2z,\vec{z}/2}$
$g'\circ\mathcal{L}_\pm$	0	1	$\mathcal{A}_+ = \check{g}'\mathcal{A}_-^*(\check{g}')^{-1}$	$\sigma(\mathcal{A}_+) = \sigma(\mathcal{A}_-)^*$	$T, T\mathfrak{i}$
$= \mathcal{L}_\mp$	1	0	$\mathcal{A}_+ = \check{g}'\mathcal{A}_-^{-1}(\check{g}')^{-1}$	$\sigma(\mathcal{A}_+) = \sigma(\mathcal{A}_-)^*$	$\mathfrak{r}_x, \mathfrak{r}_y$
	1	1	$\mathcal{A}_+ = \check{g}'\mathcal{A}_-^t(\check{g}')^{-1}$	$\sigma(\mathcal{A}_+) = \sigma(\mathcal{A}_-)$	$T\mathfrak{g}_{x,\vec{z}/2}$

space (embedded in $BT_\perp$). In the case of (g', E_i'), the two vector bundles are distinct because the fibers transform in different representations of the order-two symmetry g; in the case of class-II-B elementary orbits, the two vector bundles are distinct because their base spaces ($\mathfrak{o}$ and $g\circ\mathfrak{o}$) are distinct. Given this broader perspective, the derivation of the four classes of constraints listed in Table IV are essentially identical to those for class II-B, which may be found in Sec. VI D 6 below.

Class-I elementary orbits with $s(g)=1$. If g includes a time reversal, as exemplified by $g=T\mathfrak{i}$ and $T\mathfrak{c}_{2z}$, we apply Eq. (142) to derive

$$\mathcal{A}=\breve{g}\mathcal{A}^*\breve{g}^{-1}\Rightarrow\sigma(\mathcal{A})=\sigma(\mathcal{A})^*\Rightarrow\det[\mathcal{A}]=\pm1. \quad (146)$$

The middle line states that the spectrum of $\mathcal{A}$ is invariant under complex conjugation. That $\det[\mathcal{A}]=-1$ might seem surprising for a contractible orbit, especially when one recalls that the U(1) Berry curvature $\mathcal{F}^z(\boldsymbol{k})=\epsilon_{\alpha\beta}\nabla^\alpha_{\boldsymbol{k}}\mathrm{Tr}[\mathfrak{X}^\beta(\boldsymbol{k})]$ vanishes almost everywhere—in the torus for the $T\mathfrak{i}$-symmetric case, and in the high-symmetry planes for the $T\mathfrak{c}_{2z}$-symmetric case. For $D=1$, the resolution is that the orbit must enclose a singularity in the curvature: the orbit is linked with an odd number of line nodes in the $T\mathfrak{i}$-symmetric case (a known result by Mikitik [21]), and encircles an odd number of Dirac points in the $T\mathfrak{c}_{2z}$-symmetric case; the latter is exemplified by graphene, as we have substantiated in Sec. VI D 1. To complete the argument that $\det[\mathcal{A}]=-1$ in these cases, the conical dispersion around a Dirac point/line node guarantees that the velocity $(\nabla_{\boldsymbol{k}}\varepsilon)$ is finite at the singular point; hence the nongeometric one-forms (Roth and Zeeman) negligibly contribute to $\mathcal{A}$ in the limit where the area of the loop (that encircles the singular point) vanishes.

For spin-orbit-coupled solids with bands which are spin-degenerate ($D=2$) owing to $T\mathfrak{i}$ symmetry, we may rule out $\det[\mathcal{A}]=-1$ because all time-reversal-symmetric orbits can be continuously contracted to a point; the argument for this is presented in Sec. VI D 7. The implications of this determinantal constraint for the quantization conditions and magnetic oscillations have been discussed, around Eqs. (73) and (115), respectively. We remark that $\det[\mathcal{A}]=+1$ may be alternatively derived if H_1 is traceless, as we have discussed in Sec. C 4 b.

5. Class-II-A elementary orbits

A class-II-A elementary orbit is a single closed orbit (denoted $\mathfrak{o}$), which is closed under g (i.e., $g\circ\mathfrak{o}=\mathfrak{o}$). Just as for class-I orbits, we define $\mathcal{A}[\mathfrak{o}]$ as the propagator over the oriented orbit $\mathfrak{o}$. At times we may suppress the argument of $\mathcal{A}$ notationally; in these cases $\mathcal{A}$ should be understood as $\mathcal{A}[\mathfrak{o}]$.

Class-II-A elementary orbits with $u(g)=1$. If $u(g)=1$, we have shown in Sec. VI A that g acts on $\boldsymbol{k}^\perp$ as a planar reflection, and therefore g-invariant $\boldsymbol{k}^\perp$ form isolated lines. Since $\mathfrak{o}$ is closed as an orbit, it must intersect a g-invariant line at minimally two points. For simple, closed orbits (which are equivalent to circle), there are only two intersections, which we denote by $\boldsymbol{k}_a$ and $\boldsymbol{k}_b$. There might be more intersections for nonsimple closed orbits (e.g., a figure-of-eight), but we shall identify the two intersection points that are farthest apart (on the g-invariant line) as $\boldsymbol{k}_a$ and $\boldsymbol{k}_b$. It is analytically convenient in derivations to let the base point of $\mathcal{A}$ lie on one of these invariant wave vectors (say, $\boldsymbol{k}_a$); we remark that the spectrum of $\mathcal{A}$ is independent of the position of the base point [123]. Particularizing Eq. (142) to the present context,

$$\mathcal{A}=\breve{g}K^s\,\mathcal{A}^{-1}\,K^s\,\breve{g}^{-1}\Rightarrow\sigma(\mathcal{A})=K^{1+s}\sigma(\mathcal{A})K^{1+s}, \quad (147)$$

with the sewing matrix $\breve{g}$ evaluated at $\boldsymbol{k}_a$. To clarify the above notation, $K^{1+s}\sigma(\mathcal{A})K^{1+s}=\sigma(\mathcal{A})^*$ iff $1+s$ is odd, and therefore $\sigma(\mathcal{A})$ is not constrained if $s(g)=0$.

To obtain another useful constraint, we might split the propagator into the product $\mathcal{A}=\mathcal{A}(\boldsymbol{k}_b)\mathcal{A}(\boldsymbol{k}_a)$, where $\mathcal{A}(\boldsymbol{k}_a)$ propagates through half the orbit beginning from $\boldsymbol{k}_a$ and ending at $\boldsymbol{k}_b$, and $\mathcal{A}(\boldsymbol{k}_b)$ completes the orbit. The constraint between $\mathcal{A}(\boldsymbol{k}_a)$ and $\mathcal{A}(\boldsymbol{k}_b)$ in Eq. (140) implies that

$$\begin{aligned}\mathcal{A}&=\mathcal{A}(\boldsymbol{k}_b)\mathcal{A}(\boldsymbol{k}_a)\\&=e^{i\boldsymbol{\delta}\cdot(g\circ\boldsymbol{k}_a-g\circ\boldsymbol{k}_b)}\breve{g}(\boldsymbol{k}_a)\,K^s\,\mathcal{A}(\boldsymbol{k}_a)^{-1}\,K^s\,\breve{g}^{-1}(\boldsymbol{k}_b)\,\mathcal{A}(\boldsymbol{k}_a).\end{aligned} \quad (148)$$

This is an additional constraint that has not been included in Table III. The spectra of unitaries with such a constraint have been studied by one of us in Refs. [123–125]; a common theme in these works is that, for certain symmetries $\{g\}$, the spectrum of $\mathcal{A}$ (or a subset thereof) may be robustly fixed to special values; the existence of such robust eigenvalues depends on the symmetry representations at the g-invariant wave vectors.

To provide a simple illustration, we consider a simple closed orbit that is invariant under the mirror symmetry $g=\mathfrak{r}_x$ ($u=1, s=0, \boldsymbol{\delta}=\boldsymbol{0}$). Since $\mathfrak{r}_x$ is an order-two symmetry, it has two distinct types of representations which we shall refer to as even and odd. For a nondegenerate band ($D=1$), Eq. (148) simplifies to $\mathcal{A}=\breve{\mathfrak{r}}_x(\boldsymbol{k}_a)\breve{\mathfrak{r}}_x^{-1}(\boldsymbol{k}_b)$, which equals $+1$ if the representations at $\boldsymbol{k}_a$ and $\boldsymbol{k}_b$ are identical, and -1 if the two representations are distinct. $\mathcal{A}=+1$ is exemplified by a band that is nondegenerate at all $\boldsymbol{k}^\perp$ bounded by $\mathfrak{o}$; due to continuity of the mirror representation along the g-invariant line, the representations at $\boldsymbol{k}_a$ and $\boldsymbol{k}_b$ must be identical. We may derive $\mathcal{A}=+1$ from an alternative argument: the nondegeneracy at all $\boldsymbol{k}^\perp$ implies that $\mathfrak{o}$ is continuously contractible to a point. $\mathcal{A}=-1$ occurs iff there is an odd number of band touchings along the segment of the mirror line contained within $\mathfrak{o}$; at each band touching (a Dirac point), the mirror representation flips discontinuously, and an odd number of flips implies that the representations at $\boldsymbol{k}_a$ and $\boldsymbol{k}_b$ are distinct. This is exemplified by the surface state of the SnTe class [55] of topological crystalline insulators. Dirac cones protected by glide or screw symmetry are also characterized by $\mathcal{A}=-1$ [59].

Class-II-A elementary orbits with $u(g)=0$. If $u(g)=0$, we have shown in Sec. VI A that g acts on generic $\boldsymbol{k}^\perp$ as a discrete rotation, while g-invariant (nongeneric) $\boldsymbol{k}^\perp$ are isolated points. Given that $g\circ\mathfrak{o}=\mathfrak{o}$, and that $\mathfrak{o}$ is closed as an orbit, $\mathfrak{o}$ must encircle a g-invariant point; however, $\mathfrak{o}$ itself contains no g-invariant points. In other words, g maps every wave vector on $\mathfrak{o}$ to a distinct wave vector on the same orbit. A commonly encountered example is $g=T$ or $\mathfrak{i}$, which maps $\boldsymbol{k}_1\to-\boldsymbol{k}_1$; for orbits that encircle an inversion-invariant point, $\{\boldsymbol{k}_1,-\boldsymbol{k}_1\}$ are distinct points lying on the same orbit.

Before stating the main result of this section, it would be useful to review and expand on the definition of order-N symmetries (g) and their corresponding g orbits. For any g which is not purely a translation, we may assign to g an

order $N(g) \in \{2,3,4,6\}$, a $\mathbb{Z}_2$ index $a(g)$, and a Bravais-lattice vector $\boldsymbol{R}(g)$, such that Eq. (31) is satisfied. A case in point is $g = \mathfrak{i}$, where $\mathfrak{i}^2 = I$ implies $N = 2, a = 0, \boldsymbol{R} = \boldsymbol{0}$, while $\mathfrak{c}_{nz}^n = \mathfrak{e}$ implies $N = n, a = 1, \boldsymbol{R} = \boldsymbol{0}$. Further examples are provided in Table I. Let $\boldsymbol{k}_1$ be an arbitrarily chosen base point in $\mathfrak{o}$, and define the g orbit of $\boldsymbol{k}_1$ as in Eq. (33); in particular, the g orbit of any $\boldsymbol{k} \in \mathfrak{o}$ also lies within $\mathfrak{o}$. For $g = \mathfrak{c}_{nz}$, there are N distinct points in the g orbit, which is a single cycle of length N. More generally, the g orbit may contain $m(g)$ cycle(s) of length $L(g) = N/m \in \mathbb{N}$; L is the smallest integer such that $g^L \circ \boldsymbol{k}^\perp = \boldsymbol{k}^\perp$ for all $\boldsymbol{k}^\perp$; $u = 1 \Rightarrow L \in 2\mathbb{Z}$ owing to $u(g^L) = 0$. m is a positive natural number that divides N, but is not equal to N; the latter inequality follows from the assumption that g is class II ($m = N$ would imply that generic wave vectors are invariant under g). For example $g = T\mathfrak{c}_{6z}$ has order $N = 6$, and its g-orbit is composed of $m = 2$ cycles of length $L = 3$; further examples are provided in Table I. It will be useful to define $\breve{g}_i$ as the sewing matrix that relates the Bloch functions at $\boldsymbol{k}_i$ to those at $\boldsymbol{k}_{i+1}$: in more detail, $\breve{g}_i := \breve{g}_{i+L} := \breve{g}(\boldsymbol{k}_i)$, as defined in Eq. (29). It follows from Eq. (29) and (31) that the sewing matrices form a representation of the space group, as shown in Eq. (34).

The main result of this section is that for every class-II-A symmetry (g) with $u(g) = 0$, there exists an equivalence class of operators $[\tilde{g}K^s]$ that constrains the propagator as

$$0 = [\,\tilde{g}K^s, \mathcal{A}\,]. \tag{149}$$

$\tilde{g}$ is a unitary defined with the equivalence

$$\tilde{g}^{-1} = \tilde{g}^\dagger, \quad \tilde{g}K^s \sim \mathcal{A}\tilde{g}K^s. \tag{150}$$

The motivation for this equivalence: if $\tilde{g}K^s$ were to be found that commutes with $\mathcal{A}$, it follows trivially that $\mathcal{A}\,\tilde{g}K^s$ would also commute with $\mathcal{A}$. $\tilde{g}$ and $\mathcal{A}$ are mutually constrained as

$$(\tilde{g}K^s)^N = \mathcal{A}^p\,(-1)^{Fa}, \quad p(g) \sim p + N, \tag{151}$$

where s, a, p, N, and $\boldsymbol{R}$ are g-dependent. Equations (149) and (151) may be viewed as multiplication rules in a group generated by $\mathcal{A}$, $\tilde{g}K^s$, and $(-1)^F$.

Observe that the group relation for $\tilde{g}K^s$ in Eq. (151) differs from the point-group relation for g only by a multiplicative factor of $\mathcal{A}^p$; we say that Eq. (151) represents an extension of the point group by the loop propagators $\mathcal{A}$. The exponent $p(g)$ is an integer defined with an equivalence $p \sim p + N$, which reflects $\tilde{g}K^s \sim \mathcal{A}\tilde{g}K^s$; the values of p for our list of representative symmetries are provided in Table I. Moreover, we prove in Sec. VI D 8 that

$$[p(g)] = [\nu\, m] \in \{[1],[2],\ldots,[N-1]\}, \quad \nu(g) \in \{1,-1\}, \tag{152}$$

where m (as defined above) is the number of cycles in the g orbit, and $\nu(g) = -1$ (resp. $+1$) if the g orbit has the same (resp. opposite) orientation as $\mathcal{A}$. Recall that m is a positive natural number that divides N but is less then N, and therefore p is not ~ 0. This implies that Eq. (151) represents an unsplit extension of the point group (generated by g) by the group of loop translations (generated by $\mathcal{A}$ and isomorphic to $\mathbb{Z}$). Equivalently stated, $\tilde{g}K^s \sim \mathcal{A}\tilde{g}K^s$ form an intrinsically projective representation [119] of a point group; inequivalent projective representations are classified by the second group cohomology [126], as we further develop in Sec. VI D 8. In addition to this general group-theoretic discussion, we provide a more detailed case study of the order-two symmetries T and $\mathfrak{i}$ in Sec. VI D 7.

The following constraint on the spectrum and determinant of $\mathcal{A}$ follows directly from Eq. (149):

$$\sigma(\mathcal{A}) = K^s\sigma(\mathcal{A})K^s \Rightarrow \quad \text{if } s = 1, \quad \det\mathcal{A} = \pm 1. \tag{153}$$

While the determinantal constraint (for $s = 1$) is a general result that applies independently of the band degeneracy and the symmetry representation, we may further restrict the determinant once these additional data are specified; we shall exemplify this claim with $g = T$. For bands which are nondegenerate along $\mathfrak{o}$ ($D = 1$), the determinantal constraint is merely a reality constraint on $\mathcal{A}$, a unimodular phase factor. The sign of $\mathcal{A} \in \mathbb{R}$ is determined by the symmetry representation as

$$g = T, \quad D = 1, \quad \mathcal{A} = (-1)^F. \tag{154}$$

$F = 0$ corresponds to integer-spin representations, which include single-spin bands in solids without spin-orbit coupling, and also charge-neutral bosonic systems. In the former case, $D = 1$ should be interpreted as the energy degeneracy of bands restricted to one spin subspace, and the reality constraint applies to the spin-independent propagator $\mathcal{A}_{F=0}$ defined in Eq. (131).

Next, let us consider spin-degenerate bands ($D = 2$) which transform in a half-integer-spin representation ($F = 1$) of time reversal (T). They may arise in (a) spin-orbit-coupled solids with $\mathfrak{i}$ symmetry (in addition to the assumed T symmetry), and (b) solids with negligible spin-orbit coupling. In these two cases, the constraint Eq. (153) particularizes to

$$\text{for spin-degenerate bands, } D = 2, F = 1, \det\mathcal{A} = 1. \tag{155}$$

The proof for case (b) follows: we have already shown in Eq. (131) how $\det\mathcal{A}$ is independent of the Zeeman effect, because the coupling to spin up exactly cancels the coupling to spin down. Consequently, $\det\mathcal{A}$ is completely determined by the Roth-Berry phase, which characterizes the zero-field Hamiltonian H_0. Due to the spin-SU(2) symmetry of H_0, $\det\mathcal{A}$ equals the square of the Roth-Berry phase factor of the scalar (i.e., nonspinor, spinless) wave function [cf. Eq. (131)]. To complete the proof, we utilize our general result in Eq. (153), which applies in particular to spinless, nondegenerate ($D = 1$) bands: the Roth-Berry phase factor is restricted to ± 1, owing to time-reversal symmetry.

The proof of Eq. (155) for case (a), as well as that of Eq. (154), is more involved and will be deferred to Sec. VI D 7.

6. Class-II-B elementary orbits

Let $\mathcal{A}_1 := \mathcal{A}[\mathfrak{o}_1]$ and $\mathcal{A}_2 := \mathcal{A}[\mathfrak{o}_2]$ denote the propagators for two disconnected closed orbits related by $g \circ \mathfrak{o}_1 = (-1)^u\mathfrak{o}_2$; the orientations of both $\mathfrak{o}_i$ are determined by Hamilton's equation. We denote the spectrum of $\mathcal{A}_i$ by $\sigma(\mathcal{A}_i) = \{\exp i\lambda_a^{(i)}\}_{a=1}^D$, where D is the band degeneracy (and may equal 1); $\{\lambda_a^{(1)}\}_{a=1}^D$ and $\{\lambda_a^{(2)}\}_{a=1}^D$ enter two independent quantization conditions having the same form as in Eq. (73). It follows from Eq. (143) that

$$\mathcal{A}_1 = \breve{g}K^s\mathcal{A}_2^{(-1)^u}K^s\breve{g}^{-1} \Rightarrow \sigma(\mathcal{A}_1) = K^{s+u}\sigma(\mathcal{A}_2)K^{s+u}. \tag{156}$$

For illustration, consider two disconnected orbits related by time-reversal symmetry, but neither orbit encircles a T-invariant point. We particularize to a spinless solid whose bands are nondegenerate ($D = 1$) along both of $\mathfrak{o}_i$. The above equations then simplify to the mutual constraint $\mathcal{A}_1 = \mathcal{A}_2^*$ or equivalently $\lambda_1 = -\lambda_2 \bmod 2\pi$. Since $\mathfrak{o}_i$ is not individually invariant under T, the average of the orbital moment over each orbit is generically nonzero; this leads to a nonzero Roth contribution to each of λ_i. The assumed absence of any stabilizing symmetry of $\mathfrak{o}_i$ implies that the Berry-phase contribution is not fixed to any special value. To recapitulate, there exists no constraints on individual values of λ_i; they satisfy only a mutual constraint. There are then two ladders of sub-Landau levels corresponding to two uncoupled orbits. In energetic units (locally defined) where the separation between adjacent levels (within one sub-Landau ladder) is one, the offset between the two ladders is $2\lambda_1/2\pi \bmod 1$ [cf. Eq. (111)]. This splitting should be observable as two mutually constrained harmonics in the dHvA oscillations, as exemplified by a toy model of distorted, spinless graphene in Sec. V D 1.

7. *Class-II-A orbits with time-reversal or spatial-inversion symmetry*

We provide a case study of class-II-A orbits with T ($u = 0, s = 1$) or $\mathfrak{i}$ ($u = 0, s = 0$) symmetry. We may study each symmetry independently, without assuming that the solid simultaneously has both symmetries. Both T and $\mathfrak{i}$ are order-two symmetries ($N = 2$), and their corresponding g orbits consist of a single cycle ($m = 1$).

First, we will provide an elementary derivation of Eqs. (149)–(152) particularized to $N = 2, m = 1, [p] = [1]$. In the following proof, equations with the symbol g apply to both $g = T$ and $g = \mathfrak{i}$; they are distinguished by $s(T) = 1$ and $s(\mathfrak{i}) = 0$. It is convenient to decompose the propagator as $\mathcal{A} = \mathcal{A}(-\boldsymbol{k}_1)\mathcal{A}(\boldsymbol{k}_1)$, where $\mathcal{A}(\boldsymbol{k}_1)$ propagates through half the loop beginning from $\boldsymbol{k}_1$ and ending at $-\boldsymbol{k}_1$, and $\mathcal{A}(-\boldsymbol{k}_1)$ completes the loop. Equation (140) constrains the half propagators as

$$\mathcal{A}(\pm\boldsymbol{k}_1) = \breve{g}(\pm\boldsymbol{k}_1)K^s\mathcal{A}(\mp\boldsymbol{k}_1)K^s\breve{g}^{-1}(\mp\boldsymbol{k}_1), \tag{157}$$

where $\breve{g}(\boldsymbol{k})K^s$ forms a representation of the space group:

$$\begin{aligned} &\text{for } g = T, \quad s(g) = 1, \quad \breve{g}(-\boldsymbol{k})\breve{g}(\boldsymbol{k})^* = (-1)^F, \\ &\text{for } g = \mathfrak{i}, \quad s(g) = 0, \quad \breve{g}(-\boldsymbol{k})\breve{g}(\boldsymbol{k}) = I. \end{aligned} \tag{158}$$

The above equations are the particularization of Eq. (34) for $N = 2$; they respectively represent $T^2 = \mathfrak{e}$ and $\mathfrak{i}^2 = I$; F distinguishes between integer-spin ($F = 0$) and half-integer-spin representations ($F = 1$). Owing to Eq. (157), the full propagator is constrained as

$$\begin{aligned} \mathcal{A} &= \mathcal{A}(-\boldsymbol{k}_1)\mathcal{A}(\boldsymbol{k}_1) \\ &= \breve{g}(-\boldsymbol{k}_1)K^s\mathcal{A}(\boldsymbol{k}_1)\mathcal{A}(-\boldsymbol{k}_1)K^s\breve{g}^{-1}(-\boldsymbol{k}_1) \\ &= \breve{g}(-\boldsymbol{k}_1)K^s\mathcal{A}^{-1}(-\boldsymbol{k}_1)\mathcal{A}(-\boldsymbol{k}_1)\mathcal{A}(\boldsymbol{k}_1)\mathcal{A}(-\boldsymbol{k}_1)K^s\breve{g}^{-1}(-\boldsymbol{k}_1) \\ &= \tilde{g}K^s\mathcal{A}K^s\tilde{g}^{-1}. \end{aligned} \tag{159}$$

We have introduced the unitary matrix $\tilde{g}$:

$$\tilde{g} := \breve{g}(-\boldsymbol{k}_1)K^s\mathcal{A}^{-1}(-\boldsymbol{k}_1)K^s, \quad \tilde{g}^{-1} = \tilde{g}^\dagger, \tag{160}$$

which satisfies

$$\begin{aligned} (\tilde{g}K^s)^2 &= \mathcal{A}^{-1}\,\breve{g}(-\boldsymbol{k}_1)K^s\breve{g}(\boldsymbol{k}_1)K^s \\ &= \begin{cases} (-1)^F\mathcal{A}^{-1}, & g = T, \\ \mathcal{A}^{-1}, & g = \mathfrak{i}. \end{cases} \end{aligned} \tag{161}$$

■

For $g = \mathfrak{i}$, Eq. (159) implies that $\tilde{g}$ and $\mathcal{A}$ are simultaneously diagonalizable, while Eq. (161) implies their eigenvalues are mutually correlated. A similar story occurs for $g = \mathfrak{c}_{nz}$: an operator $\tilde{\mathfrak{c}}_{nz}$ can be found that commutes with $\mathcal{A}$ and satisfies the extended group relation $\tilde{\mathfrak{c}}_{nz}^n = \mathcal{A}\,\mathfrak{e}$. The mutual constraints between $\mathcal{A}$ and $\tilde{g}$ do not constrain the spectrum of $\mathcal{A}$ for a *single* orbit; however, they may result in robust crossings in the spectra of a continuous family of rotationally invariant orbits, which cover a 2D Fermi surface embedded in a 3D Brillouin torus. Incidentally, such crossings are already known to exist in the spectra of a continuous family of Wilson loops ($\mathcal{W}$) that cover a Fermi surface; as mentioned earlier, $\mathcal{A}$ and $\mathcal{W}$ are similarly constrained; i.e., the above equations are valid with $\mathcal{A}$ replaced by $\mathcal{W}$. The presence of an odd number of crossings diagnoses the presence of a 3D Dirac point (protected by rotational symmetry [42]) enclosed by the Fermi surface [127,128].

Let us particularize to $g = T$, for which Eq. (159) implies $\det\mathcal{A} = \pm 1$. As noted in Eqs. (154) and (155) of Sec. VI D 5, the determinant is completely determined by the following additional data: band degeneracy (D) at generic $\boldsymbol{k}^\perp$, the symmetry representation (whether integer- or half-integer-spin, as specified by F). In the subsequent paragraphs, we derive Eq. (154), as well as Eq. (155) for solids with both T and $\mathfrak{i}$ symmetries.

$D = 1, F = 0$. By assumption, the $\boldsymbol{k}$-space loop $\mathfrak{o}_0$ encloses a time-reversal-invariant point, which we denote by $\breve{\boldsymbol{k}}$. We first offer a simplified argument for $\mathcal{A} = +1$ given two assumptions, which we will subsequently relax: our first assumption is that (a) the group of $\breve{\boldsymbol{k}}$ [denoted $G(\breve{\boldsymbol{k}})$] is only generated by T; hence all irreducible representations (irreps, in short) are one-dimensional. It follows that the group of a generic wave vector enclosed by $\mathfrak{o}$ is trivial. We may therefore conclude that the minimal, symmetry-enforced degeneracy at any wave vector within $\mathfrak{o}$ is unity. Our second assumption is that, (b) at any $\boldsymbol{k}$ within $\mathfrak{o}$, there are no accidental degeneracies between two one-dimensional irreps; we use "accidental" to generally describe degeneracies that are not enforced by symmetry, but require some fine-tuning of the Hamiltonian parameters. (a) and (b) imply that the band degeneracy is constant for all $\boldsymbol{k}$ within $\mathfrak{o}_0$, and consequently there exists a family of time-reversal-symmetric loops ($\mathfrak{o}_s$, parametrized by $s \in [0,1]$) that interpolates between $\mathfrak{o}_0$ and a zero-area loop $\mathfrak{o}_1$ which encircles $\breve{\boldsymbol{k}}$; these loops are just the constant-energy contours of the assumed-nondegenerate band dispersion. Correspondingly, there exists also a family of time-reversal-symmetric propagators $\mathcal{A}[\mathfrak{o}_s]$ which continuously interpolates between $\mathcal{A}[\mathfrak{o}_0]$ to $\mathcal{A}[\mathfrak{o}_1]$; in short, we say that $\mathcal{A}$ is contractible T-symmetrically. Since T is preserved throughout the interpolation, the sign of $\mathcal{A}[\mathfrak{o}_s]$ is independent of s, from which follows that $\mathcal{A}[\mathfrak{o}_0] = \mathcal{A}[\mathfrak{o}_1]$. To complete the argument, we would demonstrate that $\mathcal{A}[\mathfrak{o}_1] = +1$. Since the loop is of zero area and encloses no singularity in the Berry curvature, the Berry phase contribution to $\mathcal{A}[\mathfrak{o}_1]$ vanishes.

Such an argument cannot be applied to the nongeometric contributions (orbital moment and Zeeman coupling), owing to their inverse proportionality to the band velocity—which vanishes at the T-invariant point. Instead, by utilizing that time reversal inverts the angular momentum of states at $\pm\boldsymbol{k}^{\perp}$, we derive that the orbit-average of the nongeometric one-forms vanish. This completes the demonstration.

This result persists were we to relax our assumptions (a) and (b), as we proceed to explain. Let us consider the case where the band, which is presumed to be nondegenerate along $\mathfrak{o}_0$, is continuously connected to a band-touching point at $\check{\boldsymbol{k}}$ enclosed by $\mathfrak{o}_0$. This touching point may be of three types: (i) it may correspond to a higher-dimensional irrep of $G(\check{\boldsymbol{k}})$, that includes one or more point-group symmetries. (ii) The degeneracy might be an accidental touching between two one-dimensional irreps of T symmetry. It is also possible that (iii) the touching is an accidental degeneracy between multiple irreps, one or more of which has dimension greater than one due to a point-group symmetry. Due to the presence of this band touching at $\check{\boldsymbol{k}}$, we might question the T-symmetric contractibility of $\mathfrak{o}_0$. However, the reality constraint in Eq. (154) relies only on T symmetry; hence any T-symmetric perturbation to H_0 cannot change the sign of $\mathcal{A}$. We may choose our T-symmetric perturbation to remove any accidental or point-group-symmetry-enforced degeneracy at $\check{\boldsymbol{k}}$; in the latter case we would choose a perturbation that lowers the symmetry of $\check{\boldsymbol{k}}$. Analogously, we may also remove any degeneracy at generic wave vectors within $\mathfrak{o}_0$. To clarify our argument, our perturbation to H_0 may be arbitrarily small in magnitude, and the energetic splitting of the degeneracy also arbitrarily small—but strictly nonzero. We might define $\mathfrak{o}_0'$ as the band contour of the perturbed H_0, which lies at the same energy as $\mathfrak{o}_0$; $\mathfrak{o}_0' \to \mathfrak{o}_0$ as the strength of the perturbation vanishes. The smallness of the perturbation guarantees that the topology of $\mathfrak{o}_0$ does not change discontinuously, i.e., in the sense of a Lifshitz transition; the reality condition guarantees that under such continuous deformations, $\mathcal{A}[\mathfrak{o}_0] = \mathcal{A}[\mathfrak{o}_0']$. In this manner, we are once again able to construct the continuous, T-symmetric interpolation from $\mathcal{A}[\mathfrak{o}_0] \to \mathcal{A}[\mathfrak{o}_0'] \to +1$; in the second $\to$, the family of T-symmetric loops $\{\mathfrak{o}_s' | s \in [0,1]\}$ are just the constant-energy band contours of the perturbed H_0, with $\mathfrak{o}_1'$ the zero-area loop enclosing $\check{\boldsymbol{k}}$.

$D = 1, F = 1$. Let us consider Eq. (154) for half-integer-spin representations ($F = 1$); we restrict ourselves to solids without spatial inversion (i) symmetry; only then are bands nondegenerate at generic wave vectors. Since $\mathfrak{o}_0$ encloses a Kramers-degenerate wave vector ($\check{\boldsymbol{k}}$), $\mathcal{A}[\mathfrak{o}_0]$ is not contractible T-symmetrically. The linearly dispersing band touching at $\check{\boldsymbol{k}}$ contributes a Berry phase of π [20]; if linearly dispersing touchings occur elsewhere within $\mathfrak{o}_0$, they come always in time-reversed pairs; hence the net Berry phase for all touchings remains π. Furthermore, the Roth and Zeeman phases individually vanish, since T symmetry imposes $H_1^R(\boldsymbol{k}) = -H_1^R(-\boldsymbol{k})$ and $H_1^Z(\boldsymbol{k}) = -H_1^Z(-\boldsymbol{k})$ (cf. second column of Table II). We therefore conclude that $\mathcal{A} = -I$. Our assumption of a twofold, Kramers degeneracy at $\check{\boldsymbol{k}}$ may be challenged: the degeneracy may be further enhanced by point-group symmetry [40] and/or by fine-tuning of parameters in H_0. However, by T-symmetric perturbations which preserve the sign of $\mathcal{A}$, we may always remove all point-group symmetries and accidental touchings, and recover the minimal scenario of a single Dirac touching at $\check{\boldsymbol{k}}$.

$D = 2, F = 1$. In the case with spin-orbit coupling, we may argue for the stronger determinantal constraint of Eq. (155) in two different ways. The first is based on the observation that the only nontrivial symmetry of a generic wave vector is the combined space-time inversion $T\mathfrak{i}$; its half-integer-spin irrep is two-dimensional. The group of $\check{\boldsymbol{k}}$ (an inversion-invariant wave vector) is generated by T and $\mathfrak{i}$ individually; this group has only two inequivalent half-integer-spin irreps (corresponding to even and odd parities under $\mathfrak{i}$) which are both two-dimensional. Consequently, bands are twofold degenerate at every $\boldsymbol{k}$ lying in $\mathfrak{o}_0$, absent accidental touchings and any other point-group symmetry (beyond $\mathfrak{i}$) that may enhance the twofold degeneracy. As we have argued analogously above, these absences may be guaranteed by T- and $\mathfrak{i}$-symmetric perturbations that preserve the sign of $\det[\mathcal{A}]$. The constancy of band degeneracy at all $\boldsymbol{k}$ within $\mathfrak{o}_0$ implies that $\mathcal{A}[\mathfrak{o}_0]$ is T-symmetrically contractible to the 2×2 identity matrix, and therefore $\det[\mathcal{A}] = +1$.

In an alternative argument, we may exploit the existence of a continuous T-symmetric interpolation of the spin-degenerate subspace to a limit with vanishing spin-orbit coupling $\det[\mathcal{A}]$, being fixed to either of ±1, is invariant throughout this interpolation. Since we have independently proven $\det[\mathcal{A}] = +1$ in the case with vanishing spin-orbit coupling [cf. the discussion below Eq. (155)], we obtain a result consistent with the previous paragraph.

This unit determinant also applies to loops $\mathfrak{o}''$ that neither wrap around the Brillouin torus nor enclose an inversion-invariant point ($\check{\boldsymbol{k}}$). Absent other superfluous point-group symmetries, the group of any wave vector in $\mathfrak{o}''$ is generated by $T\mathfrak{i}$ and has a single inequivalent half-integer-spin irrep, which is two-dimensional; we may then apply the perturb-then-contract argument to obtain the desired result.

8. *Group-theoretic analysis of class-II-A orbits with $u(g) = 0$*

One goal of this section is to derive Eqs. (149)–(152). Before this, we shall elaborate on their implications on the group-theoretic structure of class-II-A orbits ($u = 0$). We have claimed that $\tilde{g}K^s \sim \mathcal{A}\,\tilde{g}K^s$ reflects an intrinsic ambiguity in how we represent symmetries of the propagator $\mathcal{A}$. The reader may be familiar with an analogous U(1)-phase ambiguity in the representation of symmetries of quantum Hamiltonians [119], which motivates the extension of groups by U(1) phase factors. In different contexts, these groups are known as ray or double groups, and have applications in magnetic translations [79] and in describing half-integer-spin systems [119]. Analogously, $\mathcal{A}^p$ in Eq. (151) originates from an extension of the point group by quasimomentum loop translations.

To elaborate on this extension, let us define $G_{\mathfrak{o}}$ as the subgroup of the (magnetic) space group (G) that stabilizes a contractible orbit $\mathfrak{o}$. $G_{\mathfrak{o}}$ is itself a (magnetic) space group, and its quotient with respect to its translational subgroup is a point group defined as $P_{\mathfrak{o}}$. Let $\mathcal{A} \in G_{\mathcal{A}} := \{\mathcal{A}^z | z \in \mathbb{Z}\}$ represent a single translation around $\mathfrak{o}$. The action of $P_{\mathfrak{o}}$ on $G_{\mathcal{A}}$ is defined through

$$g \in P_{\mathfrak{o}}, \quad \tilde{g}K^s\,\mathcal{A}\,K^s\tilde{g}^{-1} = \mathcal{A}^{(-1)^u}, \tag{162}$$

where g is a representative element in $P_{\mathfrak{o}}$. $\tilde{g}K^s$ is defined to be an operator that maps the propagator to itself, up to a reversal in orientation that is determined by $u(g)$. If $\tilde{g}K^s$ is found that satisfies Eq. (162), it follows trivially that $\mathcal{A}\tilde{g}K^s$ only satisfies Eq. (162). Therefore, $\tilde{g}K^s$ is only defined up to an equivalence $\tilde{g}K^s \sim \mathcal{A}\tilde{g}K^s$, and we say that the equivalence class $[\tilde{g}K^s]$ forms a (possibly projective) representation of $g \in P_{\mathfrak{o}}$. Alternatively stated, $\tilde{g}K^s$ and $\mathcal{A}$ are elements of a group which is an extension of $P_{\mathfrak{o}}$ by $G_{\mathcal{A}}$; the possible extensions are classified by the second group cohomology [121,126]: $H^2(G_{\mathfrak{o}},G_{\mathcal{A}})$. Extensions of the point group by *noncontractible* translations in $\boldsymbol{k}$ space were first studied by one of us in Ref. [70], to classify the symmetries of noncontractible Wilson loops that wrap around the Brillouin torus. The present program further demonstrates that extensions by *contractible* $\boldsymbol{k}$-space translations are needed in the group-theoretic description of closed orbits.

Proof of Eqs. (149)–(152). Let us define $S_i \equiv S_{i+L}$ as the minimal-length, oriented line segment (contained within $\mathfrak{o}$) that begins at $\boldsymbol{k}_i$ and ends at $\boldsymbol{k}_{i+1}$; recall that $\boldsymbol{k}_i$ are points on the g orbit of $\boldsymbol{k}_1$, and $\boldsymbol{k}_{i+1} = g \circ \boldsymbol{k}_i$, as defined in Eq. (34). For order-two symmetries such as T or $\mathfrak{i}$, there are two equal-length segments connecting $\boldsymbol{k}_1$ and $\boldsymbol{k}_2 := -\boldsymbol{k}_1$; in this case, either choice of segment is valid, and will not affect $[p(g)]$ in Eq. (152). We further define $\mathcal{A}_i \equiv \mathcal{A}_{i+L}$ as the propagator along S_i; in more detail, $\mathcal{A}_i := \mathcal{A}[S_i : \boldsymbol{k}_{i+1} \leftarrow \boldsymbol{k}_i]$ with segment propagator $\mathcal{A}[S]$ defined in Eq. (138). Let us introduce an index $\nu(g)$, which equals -1 (resp. $+1$) if $S_i(g)$ has the same (resp. reversed) orientation as $\mathfrak{o}$. Depending on ν, $\mathcal{A}$ is composed of a concatenation of $\{\mathcal{A}_i\}$ as

$$\mathcal{A}[\mathfrak{o}] = (\,\mathcal{A}_L \ldots \mathcal{A}_2\mathcal{A}_1\,)^{-\nu}. \tag{163}$$

Note that we have arbitrarily chosen the base point of $\mathfrak{o}$ as $\boldsymbol{k}_1$, but this choice does not affect the eigenvalues of $\mathcal{A}[\mathfrak{o}]$, which enter the quantization conditions (cf. Sec. V).

A particularization of Eq. (140) implies that

$$e^{i\boldsymbol{\delta}\cdot(g\circ\boldsymbol{k}_{j+1}-g\circ\boldsymbol{k}_j)}\breve{g}_{j+1}K^s\mathcal{A}_jK^s\breve{g}_j^{-1} = \mathcal{A}_{j+1} \tag{164}$$

$$\Leftrightarrow e^{i\boldsymbol{\delta}\cdot(g\circ\boldsymbol{k}_j-g\circ\boldsymbol{k}_{j+1})}\breve{g}_jK^s\mathcal{A}_j^{-1}K^s\breve{g}_{j+1}^{-1} = \mathcal{A}_{j+1}^{-1} \tag{165}$$

$$\Leftrightarrow e^{i\boldsymbol{\delta}\cdot(g\circ\boldsymbol{k}_j)}\breve{g}_jK^s\mathcal{A}_j^{-1} = e^{i\boldsymbol{\delta}\cdot(g\circ\boldsymbol{k}_{j+1})}\mathcal{A}_{j+1}^{-1}\breve{g}_{j+1}K^s. \tag{166}$$

Inserting Eq. (164) into Eq. (163) for $\nu = -1$,

$$\begin{aligned}\nu(g) = -1,\quad \mathcal{A}[\mathfrak{o}] &= \breve{g}_0K^s\mathcal{A}_{L-1}\ldots\mathcal{A}_1\mathcal{A}_LK^s\breve{g}_0^{-1}\\ &= \breve{g}_0K^s\mathcal{A}_0^{-1}\mathcal{A}_L\mathcal{A}_{L-1}\ldots\mathcal{A}_1\mathcal{A}_0K^s\breve{g}_0^{-1}\\ &= \tilde{g}K^s\mathcal{A}[\mathfrak{o}]K^s\tilde{g}^{-1},\end{aligned} \tag{167}$$

where $\tilde{g}$ in the last line is defined as

$$[\tilde{g}K^s] = \left[e^{i\boldsymbol{k}_1\cdot\boldsymbol{\delta}}\breve{g}_0K^s\mathcal{A}_0^{-1}\right],\quad \tilde{g}K^s \sim \mathcal{A}[\mathfrak{o}]\,\tilde{g}K^s. \tag{168}$$

Inserting Eq. (165) into Eq. (163) for $\nu = +1$,

$$\begin{aligned}\nu(g) = 1,\quad \mathcal{A}[\mathfrak{o}] &= \breve{g}_0K^s\mathcal{A}_0^{-1}\mathcal{A}_1^{-1}\ldots\mathcal{A}_{L-1}^{-1}K^s\breve{g}_0^{-1}\\ &= \left(\breve{g}_0K^s\mathcal{A}_0^{-1}\right)\mathcal{A}_1^{-1}\ldots\mathcal{A}_{L-1}^{-1}\mathcal{A}_L^{-1}\left(\mathcal{A}_0K^s\breve{g}_0^{-1}\right)\\ &= \tilde{g}K^s\mathcal{A}[\mathfrak{o}]K^s\tilde{g}^{-1},\end{aligned} \tag{169}$$

utilizing the same definition of $\tilde{g}$ in Eq. (168). In either case for ν, $\tilde{g}$ satisfies

$$\begin{aligned}(\tilde{g}K^s)^N &= \left(e^{i(g\circ\boldsymbol{k}_0)\cdot\boldsymbol{\delta}}\breve{g}_0K^s\mathcal{A}_0^{-1}\right)^N\\ &= \mathcal{A}_1^{-1}\left(e^{i(g\circ\boldsymbol{k}_1)\cdot\boldsymbol{\delta}}\breve{g}_1K^s\mathcal{A}_1^{-1}\right)^{N-1}e^{i(g\circ\boldsymbol{k}_1)\cdot\boldsymbol{\delta}}\breve{g}_1K^s\\ &= \mathcal{A}_1^{-1}\mathcal{A}_2^{-1}\left(e^{i(g\circ\boldsymbol{k}_2)\cdot\boldsymbol{\delta}}\breve{g}_2K^s\mathcal{A}_2^{-1}\right)^{N-2}\\ &\quad\times e^{i(g\circ\boldsymbol{k}_2)\cdot\boldsymbol{\delta}}\breve{g}_2K^se^{i(g\circ\boldsymbol{k}_1)\cdot\boldsymbol{\delta}}\breve{g}_1K^s\\ &= \left(\mathcal{A}_1^{-1}\mathcal{A}_2^{-1}\ldots\mathcal{A}_N^{-1}\right)\left(e^{i(g\circ\boldsymbol{k}_N)\cdot\boldsymbol{\delta}}\breve{g}_NK^s\ldots\right.\\ &\quad\left.e^{i(g\circ\boldsymbol{k}_2)\cdot\boldsymbol{\delta}}\breve{g}_2K^se^{i(g\circ\boldsymbol{k}_1)\cdot\boldsymbol{\delta}}\breve{g}_1K^s\right)\\ &= \mathcal{A}[\mathfrak{o}]^{\nu m}\,(-1)^{Fa(g)}.\end{aligned} \tag{170}$$

The second to fifth equalities are derived by an N number of iterative applications of Eq. (166); in the last line, we have employed Eq. (163), and the fact that the g orbit $\{\boldsymbol{k}_i\}_{i=1}^N$ contains m cycle(s) of length $L = N/m$, with $\boldsymbol{k}_L = \boldsymbol{k}_0$.

VII. INTRABAND BREAKDOWN

Intraband breakdown occurs in the vicinity of saddle points, which are the nuclei of Lifshitz transitions, i.e., changes in the topology of constant-energy band contours as a function of energy. In the neighborhood of a saddle point, the band contours approach each other as two arms of a hyperbola illustrated in Fig. 3. It is convenient to orient ourselves by parametrizing the zero-field, band dispersion as

$$\varepsilon_{\boldsymbol{k}} = \frac{k_x^2}{2m_1} - \frac{k_y^2}{2m_2}, \tag{171}$$

with $\boldsymbol{k} := (k_x,k_y)$ chosen so that both $m_j > 0$. In 3D solids, ε additionally depends on k_z as

$$\varepsilon_{k_x,k_y,k_z} = \frac{k_x^2}{2m_1} - \frac{k_y^2}{2m_2} + f(k_z). \tag{172}$$

Since k_z remains a conserved quantity in the presence of a magnetic field along $\vec{z}$, we may as well define $\varepsilon_{k_x,k_y} := \varepsilon_{k_x,k_y,k_z} - f(k_z)$ and work directly with Eq. (171).

For a fixed energy $\varepsilon_{\boldsymbol{k}} = E$, it is convenient to introduce the hyperbolic parameters

$$\begin{aligned}\frac{k_x^2}{a^2} - \frac{k_y^2}{b^2} &= \mathrm{sgn}[E],\\ a(E) &= \sqrt{2m_1|E|},\\ b(E) &= \sqrt{2m_2|E|},\end{aligned} \tag{173}$$

such that the hyperbolic asymptotes are diagonal lines parametrized by $k_y = \pm(b/a)k_x$. Figures 3(b) and 3(c) illustrate a discontinuous change in the band contours at $E = 0$.

A quantity of geometric significance is the area $(4ab)$ of the rectangle inscribed between the two hyperbolic arms (see Fig. 3); it is natural that the dimensionless parameter

$$\mu = \mathrm{sgn}[E]\tfrac{1}{2}abl^2 = \sqrt{m_1m_2}El^2 \tag{174}$$

determines the probability of tunneling between orbits; the exact form of μ will be motivated by the connection formula in Eq. (182). When $|\mu| = O(1)$, the minimal separation between two contours becomes of order l^{-1}, and tunneling between

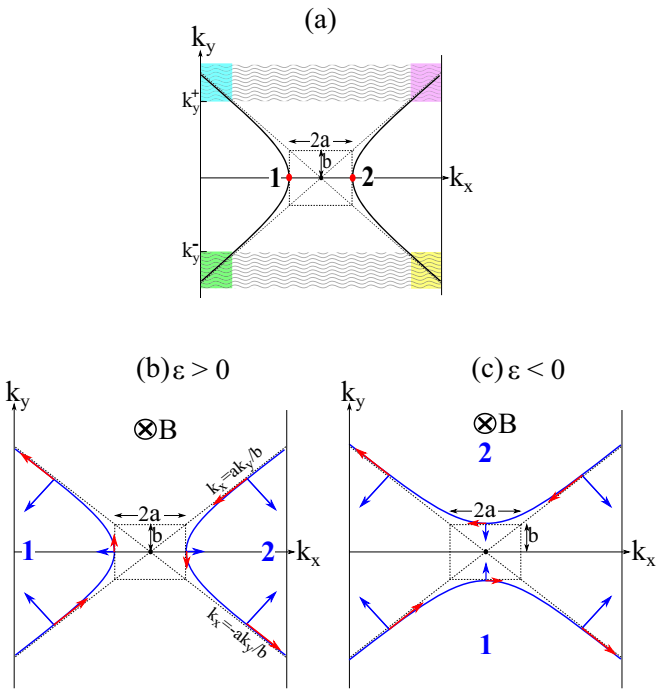

FIG. 3. (a) illustrates a region in $\boldsymbol{k}$ space where quantum fluctuations are strong; it shall be referred to as the breakdown region. The solid black lines are constant-energy band contours near a saddle point. The breakdown region overlaps with the semiclassical region (indicated by gray wavy lines). In (b) and (c), we representatively indicate the band and orbit velocities: respectively, $\nabla_{\boldsymbol{k}}\varepsilon$ is indicated by blue arrows, and $\dot{\boldsymbol{k}}$ (for a semiclassical wave packet with negative charge in a magnetic field $\boldsymbol{B} = -B\hat{z}$, $B > 0$) is indicated by red arrows. (b) shows the velocities at positive ε, and (c) for negative ε.

orbits—intraband breakdown—must be accounted for. One indication that the semiclassical equations of motion might fail is that the cyclotron mass $(\hbar^2/2\pi)\partial S/\partial E$ of the orbit diverges logarithmically as $E \to 0$ [129]; a related symptom is that both components of the band velocity $\nabla_{\boldsymbol{k}}\varepsilon$ vanish at the saddle point, as illustrated in Figs. 3(b) and 3(c). Both symptoms suggest that a hypothetical, Hamilton-obeying wave packet never reaches the saddle point in finite time.

A. Connection formula with intraband breakdown

The method to determine energy levels is similar in spirit to the divide-and-conquer approach of Sec. V B. The vicinity of the saddle point is a region of quantum fluctuations where the Zilberman-Fischbeck (ZF) function [Eq. (79)] loses its validity—as may be inferred from the diverging prefactor of $1/|v^x|^{1/2}$. What is needed is an approximate solution of the wave function in this breakdown region, with which to connect the two incoming ZF functions approaching along the $k_x = ak_y/b$ diagonal (see Fig. 3), with two outgoing ZF functions along the $k_x = -ak_y/b$ diagonal.

The main goal of this section is to derive this connection formula. The first step is to derive an effective Hamiltonian that is valid in the breakdown region; we must then derive the eigenfunctions (of this effective Hamiltonian) to the same order of accuracy (in inverse powers of l) as the Zilberman-Fischbeck function. For this purpose, one must go beyond the Peierls substitution of Eq. (171), which produces only the lowest-order, Peierls-Onsager Hamiltonian in the asymptotic expansion of Eq. (51). The Peierls-Onsager Hamiltonian forms the basis of previous treatments [11,130,131] of this problem, as briefly reviewed in Appendix D 1 a.

Let us elaborate on how this connection is done. As illustrated in Fig. 3(a), there exists an interval in $k_y \in [k_y^+, k_y^-]$, centered at the saddle point, where a semiclassical description breaks down; we shall refer to $[k_y^+, k_y^-]$ as the breakdown interval. It is convenient to define four directed edges which meet in the breakdown interval, which we label by the directions of their semiclassical motion along the hyperbolic asymptotes: $\nu \in \{\nwarrow, \swarrow\}$ above the breakdown interval, and $\nu \in \{\nearrow, \searrow\}$ below. By "directed edge," we are utilizing graph-theoretic language that is reviewed in Sec. III F. Each edge is parametrized by two single-valued functions $k_x^\nu(k_y, E)$ and $k_y^\nu(k_x, E)$; it is convenient to define for each edge the coordinate of closest approach $[\boldsymbol{k}_0^\nu(E) := (k_{x0}^\nu(E), k_{y0}^\nu(E))]$ to the saddle point, as indicated by red dots in Fig. 3(a).

Above the breakdown interval, the general analysis of Sec. V A 1 informs us that the wave function in the (K_x, k_y) representation [Eq. (77)] is a linear combination of at least two ZF functions (corresponding to the two edges above a saddle point):

$$f_{\boldsymbol{k}E}^+ = c_{\nwarrow E}\tilde{g}_{\boldsymbol{k}E}^{\nwarrow} + c_{\swarrow E}\tilde{g}_{\boldsymbol{k}E}^{\swarrow} + \cdots . \tag{175}$$

$c_{\nu E}$ are edge-dependent constants which are to be determined. As denoted vaguely by $\cdots$, there might in general be more edges in the above sum which correspond to constant-energy band contours far away from the saddle point (and therefore not illustrated in Fig. 3), but they will not be important in the matching procedure. The ZF functions in Eq. (175) are defined (up to a k_y-independent phase) as eigenfunctions of the effective Hamiltonian in the semiclassical region (denoted sm):

$$\text{For } \boldsymbol{k} \in sm, \quad [H_0(\boldsymbol{K}) + H_1(\boldsymbol{K}) - E]\tilde{g}_{\boldsymbol{k}E}^\nu = O(l^{-4}). \tag{176}$$

Precisely, we define

$$\tilde{g}_{\boldsymbol{k}E}^\nu := \frac{e^{ik_xk_yl^2}}{\sqrt{|v_x^\nu|}} e^{-il^2\int_{k_{y0}^\nu(E)}^{k_y}\left(k_x^\nu - \tilde{H}_1^\nu/v_\nu^x\right)dz}\Big|_{E\to\tilde{E}}, \tag{177}$$

$$\tilde{E} := E - H_1(\boldsymbol{0}), \quad \tilde{H}_1(\boldsymbol{k}) := H_1(\boldsymbol{k}) - H_1(\boldsymbol{0}); \tag{178}$$

one may verify that $\tilde{g}$ indeed satisfies the eigenvalue equation Eq. (176). Indeed, beginning from Eq. (176), one may redefine the origin of the energy as in Eq. (178), and utilize the known WKB solution from Eqs. (78) and (79). The reader may wonder what is the point of the redefinition of energetic origin, i.e., why not directly use the simpler expression

$$g_{\boldsymbol{k}E}^\nu := \frac{e^{ik_xk_yl^2}}{\sqrt{|v_x^\nu|}} \exp\left\{-il^2\int_{k_{y0}^\nu(E)}^{k_y}\left(k_x^\nu - \frac{H_1^\nu}{v_\nu^x}\right)dz\right\}\Big|_E, \tag{179}$$

which is also a solution of Eq. (176) in the semiclassical region. Indeed, $g_{\boldsymbol{k}E}^\nu - \tilde{g}_{\boldsymbol{k}E}^\nu = O(l^{-2})$ in the semiclassical region, as proven in Appendix D 1 c. However, at the coordinate of the saddle point, the phase of g [which includes a term proportional to $H_1(\boldsymbol{0})\log|E|$] diverges logarithmically as $|E| \to 0$, while the phase of $\tilde{g}$ is continuous across $E = 0$. For this reason, we

will find that $\tilde{g}$ is a better WKB function to formulate quantization conditions that are valid in the vicinity of a saddle point.

Below the breakdown interval, we analogously have

$$f^{-}_{\boldsymbol{k}E} = c_{\nearrow E}\tilde{g}^{\nearrow}_{\boldsymbol{k}E} + c_{\searrow E}\tilde{g}^{\searrow}_{\boldsymbol{k}E} + \cdots . \tag{180}$$

Assuming the non-WKB wave function in the breakdown region has been solved for, we may utilize this wave function as a bridge to coherently relate $\{c_{\nwarrow E}, c_{\swarrow E}\}$ (defined above the breakdown region) to $\{c_{\nearrow E}, c_{\searrow E}\}$ (defined below). For the purpose of deriving quantization conditions in Sec. VII B, we find it intuitive to express this relation as a scattering-matrix equation connecting incoming to outgoing sections:

$$\begin{pmatrix} c_{\nwarrow E} \\ c_{\searrow E} \end{pmatrix} = \mathbb{S}(E,k_z)\begin{pmatrix} c_{\nearrow E} \\ c_{\swarrow E} \end{pmatrix}. \tag{181}$$

The scattering matrix in the Peierls-Onsager approximation is known to be [111,130,131]

$$\begin{aligned}
\mathbb{S}^{(0)}(E,l^2) &= \begin{pmatrix} \mathcal{T} & \mathcal{R} \\ \mathcal{R} & \mathcal{T} \end{pmatrix}\Bigg|_{El^2}, \\
\mathcal{T}(\mu) &= e^{i\phi(\mu)}\frac{e^{\pi\mu/2}}{\sqrt{2\cosh(\pi\mu)}}, \\
\mathcal{R}(\mu) &= -i\,e^{i\phi(\mu)}\frac{e^{-\pi\mu/2}}{\sqrt{2\cosh(\pi\mu)}}, \\
\phi(\mu) &= \arg[\Gamma(1/2 - i\mu)] + \mu\log|\mu| - \mu,
\end{aligned} \tag{182}$$

with μ defined in Eq. (174) and Γ the Gamma function. Alternatively stated, $\mathbb{S}^{(0)}$ is the connection formula for Zilberman functions without higher-order corrections [i.e., Eq. (177) with $H_1(\boldsymbol{k}) = 0$].

The derivation of Eq. (182) is reviewed in Appendix D 1 a, where we elaborate on a useful analogy: magnetic tunneling of a Bloch electron near a saddle point is mathematically equivalent to a Schrödinger particle tunneling across an inverted parabolic barrier—a problem first studied by Kemble [132]. In particular, it is well known [132] that the tunneling probability at the barrier maximum is half of unity, which is reflected in Eq. (182) by $|\mathcal{T}|^2 = |\mathcal{R}|^2 = 1/2$ for $\mu = 0$ [see Fig. 4(a)]; we shall refer to this as the Kemble limit. We refer to ϕ as the intraband scattering phase and plot it in Fig. 4(b); ϕ has the following properties: (a) it is an odd function of μ that vanishes at zero and the limits $\pm\infty$, and (b) its first-order derivative diverges logarithmically as $\mu \to 0$. In all cases we have studied, property (b) does not lead to any irregularity in the Landau levels, owing to the cancellations of logarithmic divergences in $(\partial\phi/\partial E)$ and the cyclotron mass $(\propto \partial S/\partial E)$; we will exemplify this cancellation in Sec. VII C. For quick reference in the future,

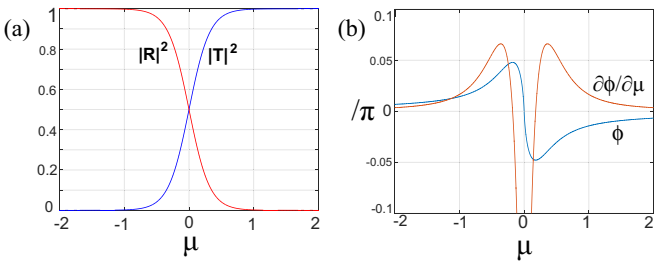

FIG. 4. Plots vs μ of (a) $|\mathcal{R}|^2$ (red) and $|\mathcal{T}|^2$ (blue), (b) the scattering phase ϕ (blue) and its derivative with respect to μ (red).

$$\mu \to +\infty, \quad \mathcal{T} \to 1, \quad \mathcal{R} \to 0, \quad \phi \to 0; \tag{183}$$

$$\mu \to 0, \quad \mathcal{T} \to 1/\sqrt{2}, \quad \mathcal{R} \to -i/\sqrt{2}, \quad \phi \to 0; \tag{184}$$

$$\mu \to -\infty, \quad \mathcal{T} \to 0, \quad \mathcal{R} \to -i, \quad \phi \to 0. \tag{185}$$

The transition from $\{|\mathcal{T}| = 1, |\mathcal{R}| = 0\}$ to $\{|\mathcal{T}| = 0, |\mathcal{R}| = 1\}$ reflects a Lifshitz transition of the band contours. While the change in the band contour is discontinuous across $E = 0$, the scattering parameters are continuous in energy; there is, as noted, a first-order non-analyticity in ϕ. The $\mu \to -\infty$ limit corresponds to the absence of tunneling (in the $\vec{y}$ direction) between the two semiclassical orbits drawn in Fig. 3(c). In this limit, $\mathcal{R} = -i$ is the phase acquired by a wave packet as it approaches the saddle point and is reflected with unit probability; the point of closest approach to the saddle point may therefore be identified as a turning point, just as discussed in Sec. V B. Indeed, we have demonstrated in Sec. V B that a wave packet rounding a turning point with a clockwise orientation picks up a phase of $-i$, which we consistently identify with $\mathcal{R} = -i$ here.

Let us argue generally that the scattering matrix, for any form of breakdown, should transform covariantly under gauge transformations within P. Viewed broadly, the scattering matrix describes the phase-coherent amplitudes for Feynman trajectories through the breakdown region. There is in general a phase ambiguity in how we define the incoming and outgoing scattering states, whose wave functions have the Zilberman-Fischbeck form in Eq. (177); in particular, the phase difference between any two states connected by a tunneling trajectory has no gauge-invariant meaning. To appreciate this point, consider a phase redefinition of the cell-periodic function projected by P: $|u_{\boldsymbol{k}}\rangle \to |u_{\boldsymbol{k}}\rangle e^{i\phi(\boldsymbol{k})}$. The resultant non-covariant transformation of the Berry connection ($\mathfrak{X} \to \mathfrak{X} - \boldsymbol{\nabla}_{\boldsymbol{k}}\phi$) occurring in $\tilde{H}_1$ [cf. Eq. (177)] results in the scattering-state wave function transforming as

$$g^{\nu\pm}_{\boldsymbol{k}E} \to g^{\nu\pm}_{\boldsymbol{k}E} e^{-i\phi^{\nu}(k_y)+i\phi(\boldsymbol{k}^{\nu}_0)}, \tag{186}$$

where in the last expression $\phi^{\nu}(k_y)$ equals $\phi(\boldsymbol{k})$ evaluated on the section ν, and at the coordinate k_y. If hypothetically the scattering matrix were insensitive to phase redefinitions of the scattering states, as is $\mathbb{S}^{(0)}$ [cf. Eq. (182)], one would conclude that the quantization condition depended on the phase difference $\phi(\boldsymbol{k}^{\nu}_0) - \phi(\boldsymbol{k}^{\mu}_0)$, which is generally nonzero for a tunneling trajectory connecting the edges ν and μ (note $\boldsymbol{k}^{\nu}_0 \neq \boldsymbol{k}^{\mu}_0$). The coefficients $c_{\nu E}$, defined in Eqs. (175)–(180), should transform with a *canceling* phase factor

$$\begin{aligned}
&c_{\nu E} \to c_{\nu E} e^{-i\phi(\boldsymbol{k}^{\nu}_0)} \\
&\Rightarrow \mathbb{S} \to \begin{pmatrix} e^{-i\phi(\boldsymbol{k}^{\nwarrow}_0)} & 0 \\ 0 & e^{-i\phi(\boldsymbol{k}^{\searrow}_0)} \end{pmatrix}\mathbb{S}\begin{pmatrix} e^{-i\phi(\boldsymbol{k}^{\nearrow}_0)} & 0 \\ 0 & e^{-i\phi(\boldsymbol{k}^{\swarrow}_0)} \end{pmatrix}.
\end{aligned}$$

Equivalently stated, the scattering matrix must transform gauge-covariantly. We see from this argument that the necessity of gauge covariance follows from the existence of tunneling trajectories, which is a characteristic feature of both intraband and interband breakdown—but not of turning points. We believe that our argument should broadly apply to any quantum tunneling phenomenon within a subspace of states (bands, in our context) that is nontrivially embedded in a larger space of states; this point has been overlooked in conventional treatments [111] of tunneling with scattering matrices.

Let us show that the next-order corrections to $\mathbb{S}^{(0)}$ restore the essential gauge covariance. For this purpose, it is sufficient to consider the correction by the Berry term H_1^B alone:

$$\mathbb{S}(E,l^2) \overset{H_1=H_1^B}{=} \begin{pmatrix} \mathcal{T}(\mu)\, e^{i\int_{-b}^{b}\mathfrak{X}^y(0,k_y)dk_y} & \mathcal{R}(\mu)\, e^{-i\int_{-a}^{a}\mathfrak{X}^x(k_x,0)dk_x} \\ \mathcal{R}(\mu)\, e^{i\int_{-a}^{a}\mathfrak{X}^x(k_x,0)dk_x} & \mathcal{T}(\mu)\, e^{-i\int_{-b}^{b}\mathfrak{X}^y(0,k_y)dk_y} \end{pmatrix}, \tag{187}$$

neglecting terms of order $O(l^{-2},(b/G)^2,(a/G)^2)$. $a(E)$ and $b(E)$ are the hyperbolic parameters defined in Eq. (173), and G is a typical reciprocal period. For positive E, $\int_{-b(E)}^{b(E)}\mathfrak{X}^y(0,k_y)dk_y$ is the integral of the Berry connection along the shortest-length tunneling trajectory that connects $\boldsymbol{k}_0^{\nearrow}(E)$ to $\boldsymbol{k}_0^{\searrow}(E)$ through the classically forbidden region [e.g., the vertical dashed line in Fig. 5(b)]. That this tunneling trajectory is of the shortest length should not be taken too seriously; a slightly deformed trajectory within the breakdown region gives a correction [of $O(l^{-2})$] that is beyond the accuracy of our calculation (detailed in Appendix D 1). Under a phase redefinition $|u_{\boldsymbol{k}}\rangle \to |u_{\boldsymbol{k}}\rangle e^{i\phi(\boldsymbol{k})}$, the open-line Berry integral transforms as

$$\int_{-b}^{b}\mathfrak{X}^y dk_y \to \int_{-b}^{b}\mathfrak{X}^y dk_y - \phi(\boldsymbol{k}_0^{\searrow}) + \phi(\boldsymbol{k}_0^{\nearrow}), \tag{188}$$

which implies that the scattering matrix transforms covariantly as

$$\mathbb{S} \to \begin{pmatrix} e^{-i\phi(\boldsymbol{k}_0^{\nwarrow})} & 0 \\ 0 & e^{-i\phi(\boldsymbol{k}_0^{\searrow})} \end{pmatrix} \mathbb{S} \begin{pmatrix} e^{-i\phi(\boldsymbol{k}_0^{\nearrow})} & 0 \\ 0 & e^{-i\phi(\boldsymbol{k}_0^{\swarrow})} \end{pmatrix}. \tag{189}$$

Indeed, Eq. (187) is minimally corrected from $\mathbb{S}^{(0)}$ to ensure gauge covariance; our calculation shows that the minimally corrected matrix completely accounts for corrections by H_1^B. In solids where the Roth (H_1^R) and Zeeman (H_1^Z) terms vanish by symmetry (e.g., $\mathfrak{c}_{2z}T$ symmetry; cf. Sec. VI B), there are no further leading-order corrections to the scattering matrix.

Unlike the Berry correction to the scattering matrix, the Roth and Zeeman corrections cannot be argued for from gauge covariance; a calculation is necessary, which we detail in Appendix D 1. When all three corrections are accounted for, we find that the scattering matrix takes the form

$$\mathbb{S}(E,l^2) = \begin{pmatrix} \mathcal{T}(\tilde{\mu})e^{i\delta_y(\tilde{E})} & \mathcal{R}(\tilde{\mu})e^{-i\delta_x(\tilde{E})} \\ \mathcal{R}(\tilde{\mu})e^{i\delta_x(\tilde{E})} & \mathcal{T}(\tilde{\mu})e^{-i\delta_y(\tilde{E})} \end{pmatrix} + O(l^{-2},(b/G)^2,(a/G)^2), \tag{190}$$

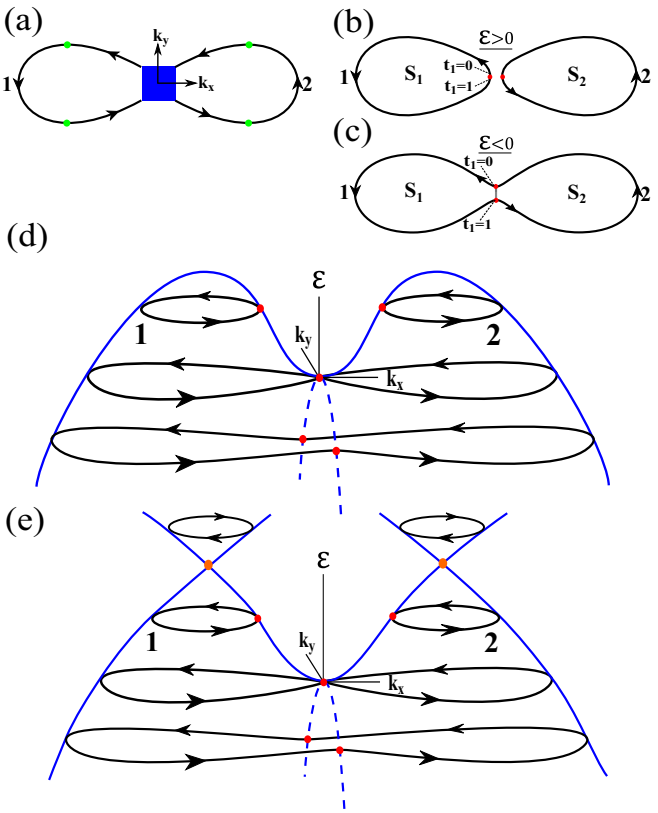

FIG. 5. (a) The double-well graph consists of two broken orbits (labeled $i = 1,2$) linked by a single breakdown vertex. Each broken orbit comprises three sections and two turning points. (b) and (c) illustrate the band contours at positive and negative energies, respectively. (d) and (e) are possible realizations of the double-well graph: (d) depicts a band dispersion with two nearby peaks, and (e) illustrates two Dirac points (indicated by orange dots) in close proximity.

$$\tilde{E} := E - H_1(\boldsymbol{0}), \quad \tilde{\mu} := \sqrt{m_1 m_2}\tilde{E}l^2, \tag{191}$$

$$\delta_y(E) := 2m_1\, H_{1x}\, b(E)\, l^2, \quad \delta_x(E) := 2m_2\, H_{1y}\, a(E)\, l^2, \tag{192}$$

where $H_1(\boldsymbol{0})$, H_{1x}, and H_{1y} are defined as coefficients in the low-momentum expansion of $H_1 = H_1^B + H_1^R + H_1^Z$ about the saddle point:

$$\begin{aligned} H_1(\boldsymbol{k}) &:= H_1(\boldsymbol{0}) + H_{1x}k_x + H_{1y}k_y + \cdots, \\ H_1^B(\boldsymbol{k}) &= l^{-2}(\mathfrak{X}^y v^x - \mathfrak{X}^x v^y) \\ &= l^{-2}\left(\mathfrak{X}^y\frac{k_x}{m_1} + \mathfrak{X}^x\frac{k_y}{m_2}\right) := H_{1x}^B k_x + H_{1y}^B k_y, \end{aligned} \tag{193}$$

$$H_1(\boldsymbol{0}) = H_1^R(\boldsymbol{0}) + H_1^Z(\boldsymbol{0}). \tag{194}$$

Note that $H_1^B(\boldsymbol{0}) = 0$ because the saddle point is an extremum in the band dispersion, and the second line follows from particularizing the definition of H_B^1 [cf. Eq. (58)] to the saddle point.

In Eq. (192), we have defined δ_x and δ_y as phase corrections to the scattering matrix. Their respective proportionality to $a(E)$ and $b(E)$ identifies them as phases acquired in the

tunneling trajectories parallel to $\vec{x}$ and $\vec{y}$. This tunneling phase includes the open-line Berry phase from our previous result in Eq. (187); e.g., we may identify

$$\delta_y = \int_{-b}^{b} \mathfrak{X}^y(0,k_y)dk_y + 2m_1\left(H_{1x}^R + H_{1x}^Z\right) b\, l^2 + O\left(\frac{b^2}{G^2}\right), \tag{195}$$

with aid from Eq. (193). Under a phase redefinition $|u_{\boldsymbol{k}}\rangle \to |u_{\boldsymbol{k}}\rangle e^{i\phi(\boldsymbol{k})}$, δ_y transforms just like Eq. (188) owing to the gauge invariance of H_1^R and H_1^Z, and therefore $\mathbb{S}(E,l^2)$ in Eq. (190) transforms covariantly, just as in Eq. (189). We might further motivate the form of Eq. (195) by rewriting it completely in terms of H_1, v^x, and b:

$$\delta_y \approx \int_{-b}^{b} \left\{ \frac{\overline{H_1(\boldsymbol{k}) - H_1(\boldsymbol{0})}}{v^x(\boldsymbol{k})} \right\}_{k_y} dk_y, \tag{196}$$

where $\{\bar{\cdot}\}_{k_y}$ denotes the k_x average of the quantity $\cdot$ over a fixed-k_y cross section of the classically forbidden region. With some creative license, one might interpret Eq. (196) as the Roth-Berry-Zeeman phase averaged over all possible tunneling trajectories parallel to $\vec{y}$.

As expressed in (195), δ_y may be separated into gauge-dependent (δ_y^B with B for Berry) and gauge-invariant (δ_y^{RZ} for Roth and Zeeman) terms. δ_y^{RZ} may be dropped if we are willing to accept an $O(1)$ accuracy for the scattering phase. Indeed, we make the following estimate for the size of δ_y^{RZ}: since the tunneling trajectory has length $2b$ with b the hyperbolic parameter [cf. Eq. (173)], δ_y^{RZ} is of order $O(b/G)$ with G the reciprocal period. We might further bound $b \leqslant O(1/l)$, which is the width of the breakdown region. An analogous argument allows us to approximate

$$\begin{aligned} \delta_x &= \delta_x^B + O(1/l), \quad \delta_x^B := \int_{-a}^{a} \mathfrak{X}^x(k_x,0)dk_x, \\ \delta_y &= \delta_y^B + O(1/l), \quad \delta_y^B := \int_{-b}^{b} \mathfrak{X}^y(0,k_y)dk_y. \end{aligned} \tag{197}$$

Since δ_i^B is gauge-dependent, there is no sense in which we might similarly conclude it is small.

B. Quantization condition for closed orbits with intraband breakdown

We summarize a few salient points from the previous subsection (Sec. VII A): in the presence of intraband breakdown, we divide the Brillouin torus into overlapping subregions. A breakdown region is a strip centered at a saddle point in the energy-momentum dispersion, as illustrated in Fig. 3(a); wave functions therein are eigenfunctions of an approximate effective Hamiltonian in the (K_x,k_y) representation. In the semiclassical subregions, the Zilberman-Fischbeck wave functions [$\tilde{g}_{\boldsymbol{k}E}$ in Eq. (177)] are asymptotically valid in the limit of weak fields. Both types of wave functions are matched where the breakdown and semiclassical subregions overlap; matching conditions are known as connection formulas, and may be expressed with the scattering matrix in Eq. (190).

The condition for an energy eigenstate at energy E and wave vector k_x is the continuity (with respect to k_y) of the wave function in the (K_x,k_y) representation. This continuity condition has a simple graphical interpretation, which we will now develop. We view a closed-orbit configuration (which is presumably close to at least one saddle point) as a graph, which is composed of breakdown vertices and broken orbits. A breakdown vertex is region of dimension $1/l$ and centered at the coordinate of a saddle point, as illustrated in a blue patch in Fig. 5. A broken orbit is an orbit over a smooth trajectory that begins at a breakdown vertex and ends at a (possibly distinct) breakdown vertex (a precise definition is provided in Sec. III F). The continuity condition is conveniently expressed as a system of linear equations whose variables are scalar amplitudes ($\{A_{iE}\}$, defined in the next paragraph) which are associated to broken orbits (denoted $\{\mathfrak{o}_i\}$). We will find it useful to parametrize each broken orbit ($\mathfrak{o}_i$) by a timelike variable $t_i \in [0,1]$, which increases along the orbit in a direction consistent with Hamilton's equation. $t_i = 0$ corresponds to the point of closest approach to the saddle point of origin, and $t_i = 1$ to the point of closest approach to the destined saddle point, as illustrated for the graph in Figs. 5(b) and 5(d). We caution the reader that (i) these points of closest approach are zero-field band characterizations of each breakdown vertex, which is equipped with more internal structure than a point, and (ii) t_i should be distinguished from t_ν, which we introduced in Eq. (98) to parametrize an edge ν.

To each point on the broken orbit $\mathfrak{o}_i$ we assign a scalar amplitude $A_{i,E}(t_i)$; while in principle we may specify its full functional dependence on $t_i \in [0,1]$, it is simplest in practice to just specify the ratio of the amplitudes at the end points:

$$e^{i\Theta_i(E,l^2)} := \frac{A_{i,E}(1)}{A_{i,E}(0)} := \prod_{p\in\mathfrak{o}_i} e^{i\phi_r^p} \prod_{\nu\in\mathfrak{o}_i} e^{i\tilde{\theta}_\nu}\Big|_{E,l^2}. \tag{198}$$

In the moving-wave packet description (introduced in Sec. V C 1), $A_{iE}(t_i)$ may be interpreted as the time-evolving amplitude for a wave packet moving within $\mathfrak{o}_i$. Θ_i is then the net phase acquired by a wave packet in traversing the full length of $\mathfrak{o}_i$. Θ_i includes the sum of semiclassical phases $\tilde{\theta}_\nu$ acquired along each edge $\nu \in \mathfrak{o}_i$:

$$e^{i\tilde{\theta}_\nu(E,l^2)} := e^{-il^2\int_{k_{yi}^\nu}^{k_{yf}^\nu}[k_x^\nu - \tilde{H}_1^\nu (v_\nu^x)^{-1}]dk_y}\Big|_{\tilde{E},l^2}, \tag{199}$$

with $\tilde{E}$ and $\tilde{H}$ defined in Eq. (178), and k_{yi}^ν (resp. k_{yf}^ν) defined as the k_y coordinate of the wave packet as it enters (resp. leaves) the oriented edge ν; precisely, if the edge ν is bounded by two turning vertices, k_{yi}^ν and k_{yf}^ν are coordinates of these two vertices; if the edge ν enters a breakdown vertex, k_{yf}^ν is the coordinate of closest approach to the saddle point. We further add to Θ_i a reflection phase ϕ_r^p for each turning vertex $p \in \mathfrak{o}_i$. As discussed in Sec. V B, $\phi_r^p = \pm\pi/2$, with the sign depending on the sense of circulation of each turning vertex.

The connection formula at each saddle point (labeled s) may be expressed as a scattering matrix (which in general depends on s) that maps two incoming amplitudes to two outgoing amplitudes:

$$\begin{pmatrix} A_{i_{\nwarrow},E}(0) \\ A_{i_{\searrow},E}(0) \end{pmatrix} = \mathbb{S}_s(E,l^2) \begin{pmatrix} A_{i_{\nearrow},E}(1) \\ A_{i_{\swarrow},E}(1) \end{pmatrix}. \tag{200}$$

The expression for $\mathbb{S}$ may be found in Eqs. (187) and (190). $i_\nearrow$ labels the broken orbit that is approaching the saddle point from the $\nearrow$ direction, i.e., in the direction of increasing k_x and

k_y; take care that $\{i_\nwarrow, i_\searrow, i_\nearrow, i_\swarrow\}$ do not necessarily correspond to four distinct broken orbits.

Combining Eqs. (198) and (200) for all broken orbits in the graph, we obtain a system of linear equations with the variables $\{A_{i,E}(0)\}$, which is then solved by standard algebraic methods. A solution exists upon satisfaction of a determinantal equation that is parametrized by energy E (and wave vector k_z in 3D solids); this is the generalized Bohr-Sommerfeld quantization condition. For comparison, Eq. (108) shows an analogous determinantal equation for a simple, closed orbit without breakdown. Let us follow this algorithm to determine the quantization conditions for two case studies.

C. Case study: The double-well graph, applied to conventional and topological metals

The simplest graph with a single breakdown vertex describes a Lifshitz transition where two orbits merge into one, as illustrated in Figs. 5(a)–5(c). Scattering from a saddle point is analogous to a Schrödinger particle scattering from an inverted parabolic potential. The semiclassical motion of wave packets on either side of the saddle point is reminiscent of a Schrödinger particle in a double well; hence we shall refer to Fig. 5(a) as the double-well graph.

We offer two topologically distinct realizations of the double-well graph illustrated in Fig. 5(a): Fig. 5(d) illustrates a conventional metal whose band dispersion has two nearby maxima—this has also been referred to as "necking" in Ref. [11]; Fig. 5(e) illustrates two Dirac/Weyl points in close proximity, which materializes in topological metals near a metal-insulator transition. The double-well graph is a good description of both conventional and topological metals for an interval of energy centered at their respective saddle points; however, their difference in Berry phase leaves a signature in the Landau levels which we will investigate. The quantization condition for the double-well graph was first derived by Azbel in the Peierls-Osager approximation [11]; here, we derive also the subleading corrections to the quantization condition that encode the Berry phase, the orbital moment, and the Zeeman effect. A particular expression of this corrected condition was presented previously in Ref. [36] assuming certain crystalline point-group symmetries; here, we shall assume no such symmetries and derive the most general form of the quantization condition.

1. *Quantization condition for the asymmetric double well*

The two broken orbits in the double-well graph are denoted by $\mathfrak{o}_i$, with $i = 1,2$ indicated in Fig. 5(a). Corresponding to these orbits are two scalar amplitudes (A_{1E}, A_{2E}), which are related by the scattering matrix as

$$\begin{pmatrix} A_{1E}(0) \\ A_{2E}(0) \end{pmatrix} = \mathbb{S}(E,l^2) \begin{pmatrix} A_{1E}(1) \\ A_{2E}(1) \end{pmatrix} \Rightarrow \det \left[\mathbb{S} \begin{pmatrix} e^{i\Theta_1} & 0 \\ 0 & e^{i\Theta_2} \end{pmatrix} - I \right] \Big|_{E,l^2} = 0, \quad (201)$$

with $\Theta_j(E)$ defined in Eq. (198). The above equation may be interpreted thus: a wave packet that traverses the full length of $\mathfrak{o}_i$ accumulates a phase Θ_i; as it passes through the breakdown region, the incoming wave packet splits into two outgoing wave packets with amplitudes determined by the scattering matrix. The determinantal equation in Eq. (201) expresses the condition that these amplitudes are everywhere single-valued.

Employing the expression for the scattering matrix [Eq. (190)] and the identity $\mathcal{T}^2 - \mathcal{R}^2 = e^{i2\phi}$, the determinantal equation may be expressed trigonometrically as

$$\cos \left[\frac{\Omega_1 + \Omega_2}{2} \Big|_{E,l^2} + \phi(\tilde{\mu}) \right] = |\mathcal{T}(\tilde{\mu})| \cos \left[\frac{\Omega_1 - \Omega_2}{2} \Big|_{E,l^2} \right]. \quad (202)$$

$\tilde{\mu}$ has been defined in Eq. (191), and $\Omega_j \in \mathbb{R}$ may be expressed, modulo 2π, as

$$\begin{aligned} &\Omega_j(E,l^2) \\ &:= \Theta_j(E,l^2) + (-1)^{j+1}\delta_y(\tilde{E},l^2) \\ &= \pi + \left\{ l^2 S_j + l^2 \int_0^1 \frac{\tilde{H}_1^{\nu(t_j)}}{v_{\nu(t_j)}^x} \frac{dk_y}{dt_j} dt_j + (-1)^{j+1}\delta_y^B \right\}_{\tilde{E},l^2}, \end{aligned} \quad (203)$$

with $\boldsymbol{k}(t_j,E)$ defined as the point on $\mathfrak{o}_j$ at timelike t_j and energy E, and $\nu(t_j,E)$ labels uniquely the edge that contains $\boldsymbol{k}(t_j,E)$. For $E > 0$, Ω_j is simply the phase (Θ_j) acquired by a wave packet as it traverses the full length of $\mathfrak{o}_j$ [cf. Eq. (198)]; $E < 0$, $\mathfrak{o}_j$ is not closed [see Fig. 5(c)], and Ω_j includes an additional Berry phase (δ_y^B) acquired in a tunneling trajectory that connects the two end points of $\mathfrak{o}_j$. In more detail, let us describe the four terms in Eq. (203) in their order of appearance:

(i) The π term originates from the two turning vertices on each broken orbit, as indicated by green dots in Fig. 5(a). Each turning vertex has an anticlockwise circulation and contributes a $+\pi/2$ reflection phase, as discussed in Sec. V B.

(ii) We have previously employed $S(E)$ to denote the oriented area of a simple, closed orbit in Sec. V; S is positive for clockwise-oriented orbits and vice versa. In the presence of intraband breakdown, $S_j(E)$ denotes analogously the oriented area of a closed Feynman trajectory (denoted $\bar{\mathfrak{o}}_{j,E}$), which is a "minimally modified closure" of the broken orbit $\mathfrak{o}_j$ at energy E. That is, we extend the broken orbit by the shortest possible path to form a closed loop. For $E > 0$, $\mathfrak{o}_j$ is already closed [see Fig. 5(b)]; for $E < 0$, we add an oriented vertical line [dashed line in Fig. 5(c)] of length $2b$ across the classically forbidden region. We shall refer to this added line as a tunneling trajectory.

(iii) The two additional terms that contribute to Ω_j [in Eq. (203)] represent the leading-order corrections to the Peierls-Onsager approximation. The first corrective term is a phase acquired over $\mathfrak{o}_j$, and is generated by the Roth-Berry-Zeeman correction to the Peierls-Onsager Hamiltonian [H_1 in Eq. (57)].

(iv) The second corrective term [$\pm\delta_y^B$ in Eq. (203)] is defined to vanish for $E > 0$, but for $E < 0$ it is the Berry phase acquired over the tunneling trajectory that connects the boundary points of $\mathfrak{o}_j$ [cf. Eq. (197)]. We may combine δ_y^B with the Berry contribution to $\int \tilde{H}_1 dt$ to obtain an integral of the Berry connection over the closed loop $\bar{\mathfrak{o}}_{j,E}$. Thus, a gauge transformation of the type Eq. (23) (with $D = 1$) may modify Ω_j by any integer multiple of 2π, but does not

affect the quantization condition in Eq. (202). This concretely exemplifies how the gauge-covariance of the scattering matrix (originating from δ_y^B) results in the gauge-invariance of the quantization condition.

Equations (202) and (203) are the main result of this section. This quantization condition provides an algebraic approach to determine the Landau levels for any tunneling strength, and without recourse [45–47] to large-scale numerical diagonalization. There are two limits $\mu \to \pm\infty$ where the Landau levels determined by Eq. (202) are locally periodic in the sense of Eq. (109). For $E > 0$ and in the limit of weak field, we combine Eqs. (183) and (202) to obtain $\sin(\Omega_j/2) = 0$, which are independent quantization conditions for two orbits with negligible tunneling, as illustrated in Fig. 5(b). Each condition may be cast more familiarly as

$$2\pi(n+1/2) \approx l^2 S_j(E_n) + \oint_{\bar{\mathfrak{o}}_j} (\mathfrak{A}+\mathfrak{X})\cdot d\boldsymbol{k} + Z(\sigma^z/v^\perp)|d\boldsymbol{k}||_{E_n}, \tag{204}$$

with $v^\perp := (v^x+v^y)^{1/2}$; this expression is the anticlockwise-oriented analog of the single-band quantization condition in Eq. (68) for simple closed orbits. The last three terms on the right-hand side are the Roth, Berry, and Zeeman contributions, as we have defined below Eq. (68). In deriving Eq. (204), we have employed a well-known expression for the cyclotron mass [7] ($\partial S/\partial E = -\oint |d\boldsymbol{k}|/v^\perp$) and the identity

$$\text{for } |E| > 0,$$
$$\left\{ S_j + H_1(\mathbf{0}) \int_0^1 \frac{dt_j}{v^x_{v(t_j)}} \frac{dk_y}{dt_j} \right\}_{E-H_1(\mathbf{0})} = \left\{ S_j - H_1(\mathbf{0}) \int_{\bar{\mathfrak{o}}_j} \frac{|d\boldsymbol{k}|}{v^\perp} \right\}_{E-H_1(\mathbf{0})} = S_j(E) + O(l^{-4}).$$

A different, locally periodic spectrum emerges in the weak-field limit for $E < 0$: combining Eqs. (185) and (202), we obtain a single quantization condition for the combined orbit illustrated in Fig. 5(c): $\cos(\Omega_1/2 + \Omega_2/2) = 0$. This condition is equivalent to Eq. (204) with the replacements $S_j \to S_1 + S_2$ and $\oint_{\bar{\mathfrak{o}}_j} \to \oint_{\bar{\mathfrak{o}}_1+\bar{\mathfrak{o}}_2}$.

For general μ and not assuming any symmetry, the spectrum of Eq. (202) is neither locally periodic nor completely random. Corresponding to the two distinct arguments of the cosine functions in Eq. (202), there are generally two, distinct harmonics that competitively produce a quasirandom [18] spectrum, i.e., a spectrum that is intermediate between that of an ordered and disordered system. Consequently, magnetic oscillatory patterns (e.g., of the de Haas–van Alphen type) are not completely smeared out, but retain a regularity that reflects the long-range correlations in a quasirandom spectrum [18]. We will refer to linearly independent arguments of trigonometric functions in the quantization condition as "trigonometric harmonics," to distinguish them from the related concept of dHvA harmonics in the magnetization.

While our quantization condition is valid for any tunneling strength, we may anyway gain some intuition about quasirandomness in a weak-tunneling parameter regime where one trigonometric harmonic is dominant over the other. The dominant harmonic determines a semiclassical Landau fan in the absence of tunneling; to clarify, a Landau fan describes discrete energy levels $\{E_j^0(B)\}_{j\in\mathbb{Z}}$ whose separation $(E_{j+1}-E_j)$ increases with the magnetic field; i.e., the levels fan out. To leading order in a tunneling parameter (specified below), the tunneling correction to the fan $\delta E_j(B)$ oscillates with the frequency corresponding to the weaker harmonic. Such a perturbative treatment of quasirandom spectra is developed generally in Sec. IX E. As an example, let us perturbatively treat the regime $\mu \ll 0$, where $(\Omega_1+\Omega_2)/2$ dominates over $(\Omega_1-\Omega_2)/2$. The dominant harmonic determines the semiclassical Landau fan through $\cos(\Omega_1/2+\Omega_2/2) = 0$; Landau levels are indexed by $j \in \mathbb{Z}$ as

$$\left.\frac{\Omega_1+\Omega_2}{2}\right|_{E_j^0} = \frac{\pi}{2} + j\pi. \tag{205}$$

To leading order in $|\mathcal{T}|$ and ϕ, the correction to the Landau fan is

$$\delta E_j(B) = \left.\frac{\phi + (-1)^j|\mathcal{T}|\cos[\,(\Omega_1-\Omega_2)/2\,]}{(-1/2)[\,\partial(\Omega_1+\Omega_2)/\partial E\,]}\right|_{E_j^0}, \tag{206}$$

where the factor $(-1)^j$ originated from our evaluation of $\sin[(\Omega_1+\Omega_2)/2]$ at E_j^0. The above equation is valid assuming $|\mathcal{T}|$ and ϕ are small and slowly varying on the scale of δE. Indeed, the typical scale of variation for $|\mathcal{T}(\mu)|$ and $\phi(\mu)$ is $\Delta\mu \sim 1$ [see Figs. 4(a) and 4(b)], which implies an energy scale $\Delta E \sim 1/\sqrt{m_1 m_2}l^2$ from the defining relation $\mu = \sqrt{m_1 m_2}El^2$. It follows that

$$\frac{\delta E_j}{\Delta E} \sim \left.\frac{\sqrt{m_1 m_2}}{\partial(S_1+S_2)/\partial E}\left[\phi + (-1)^j|\mathcal{T}|\cos\frac{\Omega_1-\Omega_2}{2}\right]\right|_{E_j^0},$$

which vanishes for small enough field or large enough $|E_j^0|$.

2. *Quantization condition for the symmetric double well*

Next, we discuss how certain (magnetic) point-group symmetries may simplify the quantization condition, and make contact with the simpler expressions found in Ref. [36].

(i) Consider a time-reversal-symmetric (T), spin-orbit-coupled solid with a twofold rotational axis ($\mathfrak{c}_{2z}$) parallel to the field, but lacking spatial inversion symmetry. The latter implies bands are nondegenerate at generic wave vectors. We shall assume the Weyl points and saddle points [Figs. 5(c) and 5(d)] lie on generic wave vectors in a plane (e.g., $k_z = 0$) that is invariant under both rotation and time reversal. Weyl points in a rotationally invariant plane are not uncommon, as exemplified by TaAs [133–135]. The combined symmetry $T\mathfrak{c}_{2z}$ ensures that $H_1^R = H_1^B = 0$ at any $\boldsymbol{k}$ in this plane (cf. Sec. VI B); hence Eq. (203) simplifies to

$$\Omega_j(E,l^2) = \pi + l^2 S_j(E) + \oint_{\bar{\mathfrak{o}}_j(E)} \mathfrak{X}\cdot d\boldsymbol{k}, \tag{207}$$

with the right-hand side evaluated at $E = \tilde{E}$ [recall that $\tilde{E}$ differs from E by $H_1^R(\mathbf{0}) + H_1^Z(\mathbf{0})$].

(ii) Suppose a mirror symmetry ($x \to -x$) relates the two maxima in Fig. 5(c) and the two Weyl points in Fig. 5(d); the saddle point lies on the mirror line where $H_1^R = H_1^Z = 0$ (cf. Sec. VI B); hence $E = \tilde{E}$ also. Note however that the Roth

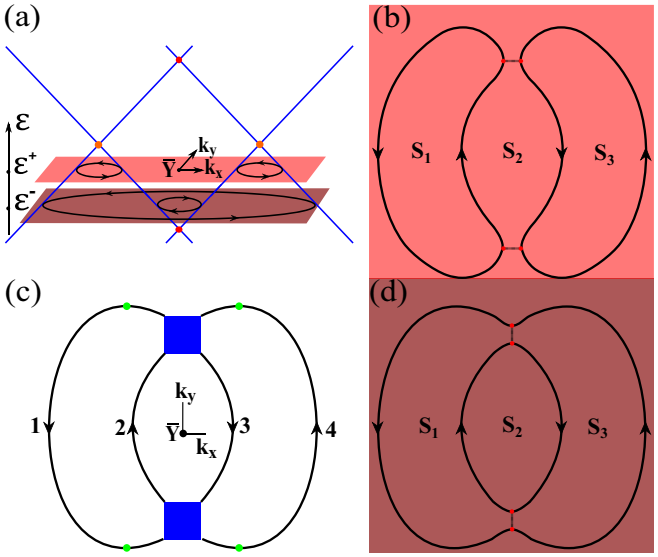

FIG. 6. Surface-state band contours in the SnTe class of topological crystalline insulators. The corresponding graph consists of four broken orbits (labeled $i = 1,2,3,4$) linked by two breakdown vertices. Orbits 1 and 4 each comprise three edges and two turning vertices; orbits 2 and 3 each comprise a single edge.

and Zeeman terms are not constrained to vanish at generic wave vectors away from the mirror line; thus

$$\Omega_j(l^2,E) = \pi + l^2 S_j(E) + \oint_{\bar{\mathfrak{o}}_j(E)} \mathfrak{X}\cdot d\boldsymbol{k} + \oint_{\mathfrak{o}_j(E)} \mathfrak{A}\cdot d\boldsymbol{k} + Z(\sigma^z/v^\perp)|d\boldsymbol{k}|. \quad (208)$$

The Landau levels and dHvA oscillations for both cases (i) and (ii) have been studied in Ref. [36].

In our next case study, we will apply the algorithm developed in Sec. VII B to derive the quantization condition for a relatively more complicated graph.

D. Case study of topological crystalline insulators: The butterfly graph

The butterfly graph illustrated in Fig. 6(c) is materialized on the 001 surface of the SnTe class of topological crystalline insulators [55,136,137], which has the same symmetry as rocksalt. The 001 surface is symmetric under the point group C_{4v}, which is generated by the fourfold rotation $\mathfrak{c}_{4z}$ and the reflection $\mathfrak{r}_x$. We focus on the vicinity of the $\mathfrak{c}_{2z}$-invariant wave vector $\bar{Y}$, which is an intersection of two orthogonal reflection-invariant ($\mathfrak{r}_x$ and $\mathfrak{r}_y$) lines. Along the $\mathfrak{r}_x$-invariant line, the dispersion of the surface states is plotted with blue lines in Fig. 6(a). The four surface bands intersect at four Dirac points, two of which (indicated by red dots) are robust due to Kramers degeneracy, and the other two (brown dots) are robust due to $\mathfrak{r}_x$ symmetry. We shall distinguish them by calling the former T-Dirac points, and the latter $\mathfrak{r}_x$-Dirac points. At energy ε^+ just below the $\mathfrak{r}_x$-Dirac points, the band contours form two nonconcentric circles (within the red plane); at energy ε^- just above the lower T-Dirac point, the band contours form two concentric circles (within the brown plane). At an intermediate, critical energy, there is necessarily a Lifshitz transition [55] facilitated by two saddle points, as illustrated in Figs. 6(b)–6(d).

Following our algorithm to determine the quantization condition, we first identify four broken orbits and label them as 1,2,3,4 in Fig. 6(c). Corresponding to these orbits are four scalar amplitudes, which are related by the scattering matrices as

$$\begin{pmatrix} A_{1E}(0) \\ A_{3E}(0) \end{pmatrix} = \mathbb{S}(E) \begin{pmatrix} A_{2E}(1) \\ A_{4E}(1) \end{pmatrix},$$
$$\begin{pmatrix} A_{2E}(0) \\ A_{4E}(0) \end{pmatrix} = \mathbb{S}(E) \begin{pmatrix} A_{1E}(1) \\ A_{3E}(1) \end{pmatrix}$$
$$\Rightarrow \det\left[\mathbb{S}\begin{pmatrix} e^{i\Theta_2} & 0 \\ 0 & e^{i\Theta_4} \end{pmatrix} \times \mathbb{S}\begin{pmatrix} e^{i\Theta_1} & 0 \\ 0 & e^{i\Theta_3} \end{pmatrix} - I\right]\Big|_E = 0. \quad (209)$$

Here, the scattering matrices corresponding to the two saddle points are identical owing to $\mathfrak{r}_y$ symmetry; we remind the reader that Θ_i is the semiclassical phase acquired by a wave packet in traversing the full length of $\mathfrak{o}_i$, as defined in Eq. (198). In spin-orbit-coupled systems with $\mathfrak{c}_{2z}T$ symmetry, both *single*-band Roth and Zeeman terms vanish (i.e., $H_1^Z = H_1^R = 0$); this follows from particularizing the general symmetry constraints in Eqs. (C35) and (C36). The Berry term is, however, nonnegligible due to the Dirac cones present in this band structure.

Let us then insert the Berry-corrected scattering matrix [Eq. (187)] into Eq. (209) and perform the necessarily algebraic manipulations, with aid from the identity $\mathcal{T}^2 - \mathcal{R}^2 = e^{i2\phi}$. The result may be stated intuitively in this manner: let us define for each of the three delineated regions in Fig. 6(c) a closed Feynman trajectory [a concept described below Eq. (203)], which we denote respectively as $\bar{\mathfrak{o}}_{1,E}, \bar{\mathfrak{o}}_{2,E}$, and $\bar{\mathfrak{o}}_{3,E}$. The semiclassical phase acquired from traversing each Feynman trajectory in a direction determined by Hamilton's equation is, respectively,

$$\Omega_1(E,l^2) = l^2 S_1(E),$$
$$\Omega_2(E,l^2) = l^2 S_2(E) + \pi, \quad (210)$$
$$\Omega_3(E,l^2) = l^2 S_3(E),$$

with negative S_1 and S_3 (due to the anticlockwise orientations of $\bar{\mathfrak{o}}_{1,E}$ and $\bar{\mathfrak{o}}_{3,E}$) and positive S_2. Each of $\{\bar{\mathfrak{o}}_{j,E}\}_{j=1}^3$ encircles a Dirac point (as illustrated in Fig. 6), and is therefore characterized by a Berry phase of π; once again, the robustness of π is due to $\mathfrak{c}_{2z}T$ symmetry. There are two turning points on each of $\bar{\mathfrak{o}}_{1,E}$ and $\bar{\mathfrak{o}}_{3,E}$, as indicated by green dots in Fig. 6(c); the resultant Maslov correction cancels the Berry-phase correction in Ω_1 and Ω_3. Finally, we should exploit that the areas of left and right boundaries ($\bar{\mathfrak{o}}_{1,E}$ and $\bar{\mathfrak{o}}_{3,E}$) are identical due to $\mathfrak{r}_x$ symmetry; hence $\Omega_1 = \Omega_3$. Putting all this together, the quantization condition may be expressed as a competition of two trigonometric harmonics:

$$0 = e^{-i2\phi} + e^{i(\Omega_1+\Omega_3+2\phi)} + |\mathcal{R}|^2[e^{i(\Omega_1-\Omega_2+\Omega_3)} + e^{i\Omega_2}] - |\mathcal{T}|^2[e^{i\Omega_1} + e^{i\Omega_3}]$$
$$\Rightarrow 0 = \cos(\Omega_1 + 2\phi) + |\mathcal{R}|^2\cos(\Omega_1 - \Omega_2) - |\mathcal{T}|^2.$$

In the three limits of μ described in Eqs. (183)–(185),

$$\begin{aligned} &\mu \to +\infty, \quad l^2 S_1 = 2n\pi; \\ &\mu = 0, \quad 1 = 2\cos[l^2 S_1] - \cos[l^2(S_1 + S_2)]; \\ &\mu \to -\infty, \quad l^2(2S_1 + S_2) = 2m\pi, \quad l^2 S_2 = 2n\pi; \end{aligned} \tag{211}$$

with $m,n \in \mathbb{Z}$. There are two semiclassical limits of the quantization condition: for $\mu \to +\infty$ (resp. $\mu \to -\infty$), we obtain independent quantization conditions for two nonconcentric (resp. concentric) simple orbits; in these cases, the Maslov and Berry corrections sum to zero modulo 2π. Except in these two semiclassical limits, the spectrum is quasirandom, and may be analyzed with the perturbative techniques developed in Appendix F 4.

VIII. EFFECTIVE HAMILTONIAN FOR GENERAL BAND TOUCHINGS

Band touchings have long provided endless entertainment in condensed-matter physics [138,139]. There are two senses in which bands may robustly touch at a point in $\boldsymbol{k}$ space. In one sense, the touching is movable, but alone it is unremovable. A 3D Weyl point exemplifies a linearly dispersing touching between two bands which is free to move in the Brillouin torus [37–39,41] but can never be removed unless it meets a Weyl point with an opposite chirality [122,140]. The freedom of one Weyl point to move but not to gap out may be understood from the following argument: in the absence of symmetry, a touching between two bands is described locally (in $\boldsymbol{k}$ space) by a two-dimensional Hamiltonian having no constraints. For a generic 2×2 Hermitian matrix, three real parameters must be tuned to impose a degeneracy. In general, we refer to the number of real Hamiltonian parameters needed to tune a degeneracy as the co-dimension (p) of the Hamiltonian [141]; the co-dimension depends on the symmetry class of the Hamiltonian, which in the present discussion is trivial. In 3D solids, the Brillouin torus affords us three parameters; hence perturbations of the $\boldsymbol{k}$-dependent Hamiltonian of a Weyl fermion merely moves the Weyl point but cannot gap it out.

Imposing a point-group symmetry (of both symmorphic and nonsymmorphic kinds), often in combination with time-reversal symmetry, may reduce the co-dimension. If such symmetry exists in the groups of all wave vectors in the 3D Brillouin torus, then line nodes are stable. More generally, the stable nodes form a $(d-p)$-dimensional submanifold of a d-dimensional manifold in $\boldsymbol{k}$ space; p is the symmetry-dependent co-dimension of the Hamiltonian, d is the dimension of manifold where this symmetry acts locally (i.e., maps $\boldsymbol{k} \to \boldsymbol{k}$). d may be less than the spatial dimension of the solid. For example, 3D Weyl points are stable in a 2D submanifold ($d = 2$) that is invariant under the composition of twofold rotation and time reversal (which enforces $p = 2$) [133]; other examples where $d = p = 1$ may be found in the literature [42,59,62].

In the other sense of robustness, a band touching may be both immovable and unremovable. It occurs at high-symmetry points or lines, and is attributed to a high-dimensional irreducible representation of the little group at such a point. In time-reversal-invariant, spin-orbit-coupled systems, the possible dimensions of these irreducible representations are 3,4,6,8 [40,43].

The physical phenomena that are attributed to all these band touchings form an immense literature; much of this literature focuses on their unusual magnetic response [36,37,46,142,143]. Any theoretical understanding of these magnetic phenomenon begins in the formulation of an effective Hamiltonian that is applicable to band degeneracies; however, this formulation is complicated by the discontinuity [61] of the band eigenfunction at a touching point. The standard lore is to operationally implement the Peierls substitution in a $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonian. To our knowledge, such a lowest-order effective Hamiltonian has only been justified for a two-band touching with a linear dispersion [12,17]; i.e., no justification exists for (a) twofold degeneracies with nonlinear dispersions (e.g., the multi-Weyl points in Ref. [62]), and (b) higher-fold band degeneracies [40].

Moreover, there has been no attempt to derive higher-order (in l^{-2}) corrections to the Peierls-substituted Hamiltonian. A Peierls-subtituted Hamiltonian for a low-energy band subspace (that touches) accurately determines Landau levels if this subspace is from all other bands by an energy gap that is large compared to the cyclotron energy. However, in naturally occurring solids, the band-touching subspace is typically embedded in a larger space of bands which disperse like spaghetti, and energy gaps between bands are typically small [40]. In some cases [44,144], the band-touching subspace is connected (in the sense of a graph [85,145]) to a larger-rank elementary band representation [87–90]. Simply stated, symmetry enforces that there are other bands close by.

This section addresses the above issues by presenting an effective Hamiltonian that is applicable to *any* type of band touching, including all cases mentioned above. The lowest-order effective Hamiltonian confirms the standard lore that the Peierls substitution works, if correctly done. Motivated by applications to spaghetti bands, we also derive the subleading corrections to the Peierls-type Hamiltonian, which encode the band-degenerate generalization of the orbital moment and the geometric phase.

A. Basis functions in the vicinity of a band degeneracy

In the rest of this section, we use $\boldsymbol{k} := (k_x,k_y)$ to denote a two-component wave vector, with the understanding that k_z (for 3D solids) is a conserved quantity for a field aligned in $\vec{z}$. The majority of band touchings occur at isolated wave vectors in the constant-k_z plane; these point degeneracies are 2D Dirac points. Even generically dispersing line nodes may be viewed as a point degeneracy, when we restrict the line node to a constant-k_z plane [146].

The effective Hamiltonian [cf. Eq. (51)] in a basis of field-modified Bloch functions [cf. Eq. (55)] is not applicable near a point degeneracy $\bar{\boldsymbol{k}}$. Indeed, Eq. (51) is an asymptotic expansion in the parameter l^{-2}, and each power of l^{-1} is accompanied with a derivative (with respect to $\boldsymbol{k}$) of either the Bloch Hamiltonian, the band dispersion, or the cell-periodic energy eigenfunction ($u_{n\boldsymbol{k}}$) [24]. The validity of this expansion thus relies on $l^{-1}\boldsymbol{\nabla}_{\boldsymbol{k}} u_{n\boldsymbol{k}}$ being of order a/l (with a a lattice period), lest there is no sense in which H_{j+1} is smaller than H_j. However, this would not be true in the vicinity of the Dirac

point, where the Berry connection (for a $\boldsymbol{k}$-space derivative in the azimuthal direction) diverges [61,63]. This directly invalidates the first-order Berry term [H_1^B in Eq. (58)] in the expansion.

The appropriate basis functions near a band-touching point are either field-modified Wannier functions (in a generalized sense [17,50]) or field-modified Luttinger-Kohn functions [12]. Luttinger-Kohn functions are well known from the effective-mass theory [72,147] and have been reviewed in Sec. III E. We will adopt the latter approach by Slutskin, which produces an effective Hamiltonian that acts on wave functions over quasimomentum space.

Previous derivations [12,17] of the effective Hamiltonian have only been carried out to lowest order in the field, and only for a conventional Dirac point with a conical dispersion. Here, no assumptions will be made about the degeneracy or the band dispersion. We will employ an ansatz for the wave function that is inspired by Slutskin [12]:

$$\Psi(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n,K_x,0}(\boldsymbol{r}) f_{n\boldsymbol{k}}, \tag{212}$$

with K_x [the kinetic quasimomentum operator defined in Eq. (77)] acting on $f_{n\boldsymbol{k}}$, which we refer to as the wave function in the $(K_x,0)$ representation; shortly we will derive an effective Hamiltonian for $f_{n\boldsymbol{k}}$. In Eq. (212) and henceforth, we suppress the spin index and assume $\bar{\boldsymbol{k}} = \boldsymbol{0}$ for notational simplicity. u_{nK_x0} is defined by replacing k_x in u_{nk_x0} by the kinetic quasimomentum K_x. Explicitly, employing the Wannier-function expansion of u_{nk_x0} in Eq. (7) [148], we replace k_x in the exponent of Eq. (7) by K_x:

$$u_{nK_x0}(\boldsymbol{r}) := \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} e^{-iK_x(x-R_x)} W_n(\boldsymbol{r}-\boldsymbol{R}). \tag{213}$$

The boundary conditions on $f_{n\boldsymbol{k}}$ are determined (as detailed in Appendix E 2) by the condition that the expansion in Eq. (212) is independent of the unit cell in $\boldsymbol{k}$ space, i.e.,

$$\alpha(\boldsymbol{k},\boldsymbol{r}) := \sum_n e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{nK_x0}(\boldsymbol{r}) f_{n\boldsymbol{k}} = \alpha(\boldsymbol{k}+\boldsymbol{G},\boldsymbol{r}) \tag{214}$$

for any reciprocal vector $\boldsymbol{G}$. For definiteness, we will choose $\sum_{\boldsymbol{k}}$ to be an integral over the first Brillouin zone.

The main motivation for our ansatz is that $\{u_{nk_x0}\}$ can be chosen to be smooth with respect to k_x (even at the band-touching point), and we might therefore anticipate that the resultant effective Hamiltonian is well-behaved analytically. An example of a smooth basis would be the energy eigenfunctions of $\hat{H}_0(k_x,0)$, with corresponding energy functions $\{\varepsilon_{nk_x0}\}$ that are smooth across $k_x = 0$; we refer to this as the "energy basis."

Under certain formal assumptions, our ansatz in Eq. (212) is equivalent to an expansion in Slutskin's basis functions [12] (cf. Appendix E 2). An analogy can also be made with Roth's basis [24] of field-modified Bloch functions [cf. Eqs. (53)–(55)]. Indeed, Roth's ansatz is equivalent to Eq. (212) with $u_{nK_x0}(\boldsymbol{r})$ replaced by $u_{n\boldsymbol{K}}(\boldsymbol{r})$, as we demonstrate in Appendix A 1.

We will demonstrate that our basis functions are complete and orthonormal with respect to functions in $\mathbb{R}^d$; neither of these properties were proven in the previous works [12,17], and instead a variational argument was used. The question of completeness is, can any function over $\mathbb{R}^d$ be written in the form of Eq. (212)? We may make the following argument for the positive claim: if l^{-2} is set to zero in Eq. (212), it reduces to an expansion over Luttinger-Kohn functions: $u_{nk_x0}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$, which are known to form a complete and orthonormal set of basis functions [72]. For sufficiently small fields, it is plausible that the completeness and orthonormality relations are preserved; the latter property should presently be understood as an operator relation

$$\int d\boldsymbol{r}\; u^\dagger_{mK_x0}(\boldsymbol{r}) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} e^{i\boldsymbol{k}'\cdot\boldsymbol{r}} u_{nK'_x0}(\boldsymbol{r}) = \delta(\boldsymbol{k}-\boldsymbol{k}')\delta_{mn}, \tag{215}$$

with $\boldsymbol{k}$ and $\boldsymbol{k}'$ restricted to the first Brillouin zone. Let us prove our claim.

Some well-known properties of Luttinger-Kohn functions will be useful, including the completeness and orthonormality of $\{u_{nk_x0}\}$ with respect to cell-periodic functions [reviewed in Eq. (6)]. These properties are simply generalized to the operator relations

$$\sum_n u_{n,K_x0}(\boldsymbol{\tau}) u^\dagger_{n,K_x0}(\boldsymbol{\tau}') = \delta(\boldsymbol{\tau}-\boldsymbol{\tau}'), \tag{216}$$

$$\int d\boldsymbol{\tau}\, u^\dagger_{m,K_x0}(\boldsymbol{\tau}) u_{l,K_x0}(\boldsymbol{\tau}) = \delta_{ml}. \tag{217}$$

We remind the reader that $\boldsymbol{\tau}$ is the cell-periodic position coordinate, and $\int d\boldsymbol{\tau}$ is the integral over a unit cell; we will often decompose $\boldsymbol{r} = \boldsymbol{\tau} + \boldsymbol{R}$, with $\boldsymbol{R}$ labeling a Bravais-lattice cell. The adjoint operation in Eq. (217) is defined as $u^\dagger_{nK_x0} := [u^*_{nk_x0}]$; i.e., we first complex-conjugate the symbol and then symmetrize it. We will employ that the Bloch functions [denoted $\{v_{n\boldsymbol{k}}(\boldsymbol{r})e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\}_{n\in\mathbb{Z}}$ with v cell-periodic] are complete with respect to functions of $\boldsymbol{r}\in\mathbb{R}^d$; i.e., any $\Psi(\boldsymbol{r})$ may be expressed as

$$\Psi(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} v_{n\boldsymbol{k}}(\boldsymbol{\tau}) g_{n\boldsymbol{k}} \tag{218}$$

for some function $g_{n\boldsymbol{k}}$. By expressing $v_{n\boldsymbol{k}}(\boldsymbol{\tau}) = \int d\boldsymbol{\tau}'\delta(\boldsymbol{\tau}-\boldsymbol{\tau}')v_{n\boldsymbol{k}}(\boldsymbol{\tau}')$ and inserting Eq. (216), we arrive at

$$\Psi(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{m\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{mK_x0}(\boldsymbol{\tau}) \sum_n \langle u_{mK_x0}|v_{n\boldsymbol{k}}\rangle g_{n\boldsymbol{k}}, \tag{219}$$

from which we identify the wave function in the $(K_x,0)$ representation as $f_{m\boldsymbol{k}} = \sum_n \langle u_{mK_x0}|v_{n\boldsymbol{k}}\rangle g_{n\boldsymbol{k}}$. This proves completeness. To prove the orthonormality, we exploit the translational symmetry of $u_{nK_x0}(\boldsymbol{r}) = u_{nK_x0}(\boldsymbol{r}+\boldsymbol{R})$ to express the left-hand side of Eq. (215) as

$$\int d\boldsymbol{\tau}\; u^\dagger_{mK_x0}(\boldsymbol{\tau}) \left\{ \sum_{\boldsymbol{R}} e^{i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{R}} \right\} e^{i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{\tau}} u_{nK'_x0}(\boldsymbol{\tau}). \tag{220}$$

The sum over $\boldsymbol{R}$ produces $\delta(\boldsymbol{k}-\boldsymbol{k}')$; from Eq. (217), we derive that the integral over $\boldsymbol{\tau}$ produces δ_{mn}. The orthonormality condition implies that, given any $\Psi(\boldsymbol{r})$, we may extract its wave function in the $(K_x,0)$ representation by

$$f_{n\boldsymbol{k}} = \frac{1}{\sqrt{N}} \int d\boldsymbol{r} e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} u^\dagger_{nK_x0}(\boldsymbol{r}) \Psi(\boldsymbol{r}); \tag{221}$$

here, we have assumed $\boldsymbol{k}$ lies in the first Brillouin zone.

B. Effective Hamiltonian in the vicinity of a band degeneracy

Our goal is to derive an effective Hamiltonian in the $(K_x,0)$ representation, i.e., acting on the wave function $f_{n\boldsymbol{k}}$ which we introduced in Eq. (212). Due to the periodicity of $\alpha(\boldsymbol{k},\boldsymbol{r})$ [cf. Eq. (214)], the position operator acts in a simple manner:

$$\hat{\boldsymbol{r}}\Psi(\boldsymbol{r}) = \frac{1}{\sqrt{N}}\sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}(i\boldsymbol{\nabla}_{\boldsymbol{k}})u_{nK_x0}f_{n\boldsymbol{k}}, \tag{222}$$

and therefore the mechanical momentum acts as

$$\{\hat{\boldsymbol{p}} + \boldsymbol{a}(\hat{\boldsymbol{r}})\}\Psi(\boldsymbol{r}) = \frac{1}{\sqrt{N}}\sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\{\hat{\boldsymbol{p}} + \boldsymbol{K}\}u_{nK_x0}f_{n\boldsymbol{k}} \tag{223}$$

with $\boldsymbol{K} = \boldsymbol{k} + \boldsymbol{a}(i\boldsymbol{\nabla}_{\boldsymbol{k}})$ the kinetic quasimomentum operator. It follows that the field-on Hamiltonian [Eqs. (39) and (40)] acts as

$$\hat{H}\Psi(\boldsymbol{r}) = \frac{1}{\sqrt{N}}\sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\hat{H}_0(\boldsymbol{K})u_{nK_x0}f_{n\boldsymbol{k}}. \tag{224}$$

Applying the operation $(1/N)\int d\boldsymbol{r}e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}u^{\dagger}_{mK_x0}$ to the time-independent Hamiltonian equation, $(\hat{H}-E)\Psi = 0$, we obtain an effective Hamiltonian equation

$$\sum_n \{\tilde{\mathcal{H}}_{mn} - E\delta_{mn}\}f_{n\boldsymbol{k}} = 0. \tag{225}$$

The $E\delta_{mn}$ term in Eq. (225) is simply obtained from the wave function extraction of Eq. (221); determining $\tilde{\mathcal{H}}$ requires a calculation that we detail in Appendix E 3; its complete form is

$$\tilde{\mathcal{H}} = \Bigg[\tilde{H}_0 + k_y\tilde{\Pi}^y + \frac{k_y^2}{2m} - \frac{1}{l^2}\left(\tilde{\mathfrak{X}}^x\tilde{\Pi}^y + \frac{k_y}{m}\tilde{\mathfrak{X}}^x\right) + \frac{1}{2ml^4}(\tilde{\mathfrak{X}}^x\tilde{\mathfrak{X}}^x - i\partial_{k_x}\tilde{\mathfrak{X}}^x)\Bigg]_{K_x,0}. \tag{226}$$

Here, $\tilde{H}_0$, $\tilde{\boldsymbol{\Pi}}$, and $\tilde{\mathfrak{X}}$ are matrices defined respectively in Eqs. (12), (13), and (15); the notation $[\tilde{H}_0]_{K_x,0}$ is shorthand for the operator $[\tilde{H}_0(k_x,0)]$. To simplify the presentation, we assume that $\hat{H}$ corresponds to the Schrödinger Hamiltonian minimally coupled to the electromagnetic field [cf. Eq. (39)]; the Pauli case [cf. Eq. (40)] is a simple generalization of the present equations.

While Eq. (226) is formally an infinite-dimensional matrix equation that is valid over the entire Brillouin torus, we are pragmatically interested in a few-band, effective Hamiltonian that corresponds to a low-energy subspace projected by P; in the $\boldsymbol{k}$ region of interest, it is assumed there are no band touchings between P and its orthogonal complement. To achieve an effective few-band Hamiltonian, we need to transform $\mathcal{H}$ as

$$S^{\dagger}\tilde{\mathcal{H}}S = \tilde{\mathcal{H}}', \tag{227}$$

such that $\tilde{\mathcal{H}}'$ is block-diagonal with respect to the decomposition $P\oplus Q$. From the wave function perspective, we are modifying our ansatz in Eq. (212) as

$$\Psi'(\boldsymbol{r}) = \frac{1}{\sqrt{N}}\sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{n,K_x,0}(\boldsymbol{r})S(\boldsymbol{K})f_{n\boldsymbol{k}}. \tag{228}$$

One aspect of the block diagonalization is well known: for any Luttinger-Kohn-type basis functions which are evaluated at $k_y = 0$, we expect that any few-band, effective Hamiltonian should be valid only for small k_y. Consequently, we would treat k_y/G_y (with G_y a reciprocal period) as a small parameter. Using standard Löwdin partitioning techniques which are well known in $\boldsymbol{k}\cdot\boldsymbol{p}$ theory [72,149,150], the block diagonalization may then be carried out perturbatively in k_y.

However, a nontrivial generalization of Löwdin partitioning techniques is required, since every term in Eq. (227) is a function of noncommuting variables ($\boldsymbol{K}$). The major difficulty lies in evaluating a product of matrix functions of $\boldsymbol{K}$. To overcome this, we borrow an insight from past constructions of effective Hamiltonians [14,23,24]; namely, we will organize $\tilde{\mathcal{H}}$ and S in an expansion in powers of l^{-2}, such that each term in the series is a symmetrized function of $\boldsymbol{K}$. Once this organization is performed, we may then exploit well-known rules for the calculus of symmetrized operators. Of particular utility is the following product rule [24]:

$$A(\boldsymbol{K})B(\boldsymbol{K}) = \left[e^{(i/2)l^{-2}\epsilon_{\alpha\beta}\nabla^{\alpha}_{\boldsymbol{k}}\nabla^{\beta}_{\boldsymbol{k}'}}A(\boldsymbol{k})B(\boldsymbol{k}')\big|_{\boldsymbol{k}=\boldsymbol{k}'}\right], \tag{229}$$

which we derive in Eq. (E8). Equation (229) is a particularization of a Moyal expansion, which is familiar from the correspondence between quantum and classical physics [151]. As it stands, our expression for $\mathcal{H}$ in Eq. (226) is not organized in the above sense, but this will be rectified in Sec. VIII C.

The upshot of the last two paragraphs is that both k_y and l^{-2} should be taken as independent, small parameters. To our knowledge, partitioning the Hilbert space simultaneously with these two parameters has never been done. In Sec. VIII D, we formulate an algorithm for this partitioning, which may in principle be carried out to any order in k_y and l^{-2}. When this algorithm is carried out to the lowest nontrivial order, we derive the following effective Hamiltonian:

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1^R + \mathcal{H}_1^B + O\big(k_y l^{-2}, k_y^2, l^{-4}\big), \tag{230}$$

$$\mathcal{H}_0 = H_0(K_x,0) + \tfrac{1}{2}[\{k_y,\Pi^y(k_x,0)\}], \tag{231}$$

$$\mathcal{H}_1^R = \frac{1}{2l^2}[\{\Upsilon^y,\tilde{\Pi}^x\} - \{\mathring{\mathfrak{X}}^x,\tilde{\Pi}^y\}]_{K_x,0}, \tag{232}$$

$$\mathcal{H}_1^B = -\frac{1}{2l^2}\{\mathfrak{X}^x,\Pi^y\}_{K_x,0}, \tag{233}$$

where $[\{A,B\}] := [AB + BA]$. We have retained our convention that the infinite-dimensional matrices $\{\tilde{H}_0,\tilde{\mathfrak{X}}^x,\tilde{\boldsymbol{\Pi}}\}$, when restricted to the D-dimensional vector space projected by P, are to be denoted by the same symbols without the tilde accent; cf. Eq. (20). $\mathcal{H}_1^R$ in Eq. (232) should be understood as the D-rank projection of infinite-dimensional matrices; two of those matrices, which are both off-block-diagonal with respect to $P\oplus Q$, are defined for the first time here: (a) $\mathring{\mathfrak{X}}^x$ is the off-block-diagonal component of

$$\tilde{\mathfrak{X}}^x = \mathring{\mathfrak{X}}^x + \dot{\mathfrak{X}}^x, \quad \mathring{\mathfrak{X}}^x_{mn} = \mathring{\mathfrak{X}}^x_{\bar{m}\bar{n}} = 0,$$
$$i\mathring{\mathfrak{X}}^x_{m\bar{n}}(k_x,0) = -\frac{\tilde{\Pi}^x_{m\bar{n}}(k_x,0)}{\varepsilon_{mk_x0} - \varepsilon_{\bar{n}k_x0}}, \tag{234}$$

while $\dot{\mathfrak{X}}^x$ is block-diagonal. (b) $\tilde{\Upsilon}^y$ is defined by its elements: for any m,n (labeling bands projected by P) and $\bar{m},\bar{n}$ (labeling

bands projected by Q),

$$\tilde{\Upsilon}^y_{mn} = \tilde{\Upsilon}^y_{\bar{m}\bar{n}} = 0,$$
$$i\tilde{\Upsilon}^y_{\bar{m}n}(k_x) = -\frac{\tilde{\Pi}^y_{\bar{m}n}(k_x,0)}{\varepsilon_{\bar{m}k_x0} - \varepsilon_{nk_x0}}, \tag{235}$$
$$i\tilde{\Upsilon}^y_{m\bar{n}}(k_x) = -\frac{\tilde{\Pi}^y_{m\bar{n}}(k_x,0)}{\varepsilon_{mk_x0} - \varepsilon_{\bar{n}k_x0}}.$$

These particular expressions for $\mathring{\tilde{\mathfrak{X}}}^x$ and $\tilde{\Upsilon}^y$ are valid in a certain basis for the cell-periodic functions—namely, where $\{u_{nk_x0}\}$ from our ansatz [cf. Eq. (228)] correspond to energy bands, and are also smooth in k_x; in this basis, $H_0(k_x,0)$ is a diagonal matrix, with diagonal elements equal to energy functions $\{\varepsilon_{nk_x0}\}_{n=1}^D$ which are also smooth in k_x. For any line (at fixed $k_y = 0$) that does not form a loop (around the Brillouin torus), such an "energy basis" can always be found.

Let us discuss the possible band structures for which Eq. (230) may be applied. While we have motivated the choice of our basis functions (in Sec. VIII A) by their utility in the vicinity of a point degeneracy, we should clarify that the derivation of the effective Hamiltonian [cf. Eq. (230)] makes no assumptions about the presence of a point degeneracy, and is therefore also applicable to nondegenerate bands. If there exists multiple touchings between bands in the subspace of P, Eq. (230) is applicable if there exists an orthogonal coordinate system where all touchings occur on the straight line of fixed $k_y = 0$. The range of k_x for which $\mathcal{H}$ is valid is only restricted by the existence of a smooth (in k_x) "energy basis"; in some cases, this smooth basis may be found over the entire circle of fixed $k_y = 0$. For applications to a single point degeneracy, the essential physics is often captured by an effective Hamiltonian that is linearized in k_x around said point, in which case Eq. (233) particularizes to

$$\mathcal{H} = H_0 + K_x\Pi^x + iK_x[\mathfrak{X}^x, H_0] + k_y\Pi^y + \frac{1}{2l^2}(\{\tilde{\Upsilon}^y, \tilde{\Pi}^x\} - \{\mathring{\tilde{\mathfrak{X}}}^x, \tilde{\Pi}^y\} - \{\mathfrak{X}^x, \Pi^y\}) + O\left(k_i l^{-2}, k_i^2, l^{-4}\right), \tag{236}$$

where all matrices above are evaluated at $\bar{\boldsymbol{k}} = \boldsymbol{0}$. The above equation is derived by utilizing the identity in Eq. (21). In particular, if $\bar{\boldsymbol{k}} = \boldsymbol{0}$ is a point of degeneracy for all D bands projected by $P(\boldsymbol{k})$, then $[\mathfrak{X}^x, H_0]$ vanishes and Eq. (236) further simplifies to

$$\mathcal{H} = H_0 + K_x\Pi^x + k_y\Pi^y + \frac{1}{2l^2}(\{\tilde{\Upsilon}^y, \tilde{\Pi}^x\} - \{\mathring{\tilde{\mathfrak{X}}}^x, \tilde{\Pi}^y\} - \{\mathfrak{X}^x, \Pi^y\}) + O\left(k_i l^{-2}, k_i^2, l^{-4}\right). \tag{237}$$

$\mathcal{H}_0$ in Eq. (231) [as well as the first three terms in Eq. (237)] shall be referred to as the Peierls-Onsager Hamiltonian in the $(K_x,0)$ representation; its form is closely analogous to the Peierls-Onsager Hamiltonian in the (K_x,k_y)-representation [cf. Eq. (56)]. Indeed, we may arrive at the first three terms in Eq. (237) by the Peierls substitution $\boldsymbol{k} \to \boldsymbol{K}$ of the Bloch Hamiltonian in the Luttinger-Kohn representation: $H_0(\boldsymbol{k}) = H_0(\boldsymbol{0}) + k_x\Pi^x(\boldsymbol{0}) + k_y\Pi^y(\boldsymbol{0}) + O(k_ik_j)$ [derived in Eq. (43)]. In the presence of a point degeneracy, this Peierls substitution is only valid for a Luttinger-Kohn basis that is smooth (in k_x) across the degeneracy. A case in point is the Peierls-Onsager Hamiltonian for the Dirac point in graphene: $\mathcal{H} = vK_x\tau_1 + vK_y\tau_2$, where τ_j are Pauli matrices that span a vector space corresponding to the two sublattices. Here, we may identify $\tau_1 = \pm 1$ as labeling the two Luttinger-Kohn functions $(u_{\pm,k_x,0})$, which depend smoothly on k_x across the Dirac point. Going beyond two-band touchings with conical dispersions, we emphasize that Eq. (237) proves the lowest-order validity of the Peierls-Onsager Hamiltonian for band touchings of any kind, including (a) those with nonlinear dispersions, e.g., the double-Weyl point in Ref. [62] disperses quadratically in two directions, as well as (b) higher-degeneracy touchings, e.g., the "spin-one Weyl" point described by $K_xL_x + K_yL_y + k_zL_z$ [20,40], where $\boldsymbol{L}$ are the generators of SO(3) in the spin-one representation.

Going beyond the leading-order Peierls substitution, we view $\mathcal{H}_1^R$ [in Eq. (232)] as the direct generalization of the Roth orbital moment, and the $\mathcal{H}_1^B$ [in Eq. (233)] as the direct generalization of the Berry term; their implications on the Landau levels will be investigated in a future work. In their original formulation [14,23,24], the Roth and Berry terms describe the first-order corrections to the Peierls-Onsager effective Hamiltonian for either (i) a single, nondegenerate band (as reviewed in Sec. IV A), or (ii) a subspace of degenerate bands (reviewed in Sec. IV B). Here, we are claiming that $\mathcal{H}_1^R$ and $\mathcal{H}_1^B$ are applicable to multiple bands, degenerate or nondegenerate, which disperse in any fashion—possibly touching at isolated wave vectors. The broadness of our claim suggests that if we particularize Eq. (230) to cases (i) or (ii), we should be able to recover an analog of the previously derived effective Hamiltonians; we demonstrate this in Appendix E 4.

C. Derivation of symmetrized effective Hamiltonian in the $(K_x,0)$ representation

As motivated in the paragraph containing Eq. (229), the goal of this subsection is to derive Eq. (225) with $\tilde{\mathcal{H}}$ expressed in a power series in l^{-2}, such that each term is symmetrized with respect to $\boldsymbol{K}$. $\tilde{\mathcal{H}}$ is defined implicitly through

$$\sum_n \tilde{\mathcal{H}}(\boldsymbol{K})_{mn} f_{n\boldsymbol{k}} = \frac{1}{\sqrt{N}} \int d\boldsymbol{r}\; u^\dagger_{mK_x0}(\boldsymbol{r}) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \hat{H}\Psi(\boldsymbol{r}), \tag{238}$$

with Ψ having the ansatz form in Eq. (212), and $\boldsymbol{k}$ assumed to lie in the integral domain of $\sum_{\boldsymbol{k}}$ in Eq. (212).

In the first step, we show that $\tilde{\mathcal{H}}$ has the more explicit form:

$$\tilde{\mathcal{H}}(\boldsymbol{K})_{mn} = \int d\boldsymbol{\tau} u^\dagger_{mK_x0}(\boldsymbol{\tau}) \hat{H}_0(\boldsymbol{K}) u_{nK_x0}(\boldsymbol{\tau}). \tag{239}$$

Beginning from the right-hand side of Eq. (238), we employ Eq. (224) and the translational symmetry of the operator $u_{nK_x0}(\boldsymbol{r}) = u_{nK_x0}(\boldsymbol{r} + \boldsymbol{R})$ to derive

$$\begin{aligned}
&\frac{1}{\sqrt{N}} \int d\boldsymbol{r}\; u^\dagger_{mK_x0}(\boldsymbol{r}) e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} \hat{H}\Psi(\boldsymbol{r}) \\
&= \frac{1}{N} \int d\boldsymbol{r} \sum_{nk'} u^\dagger_{mK_x0}(\boldsymbol{r}) e^{i(k'-k)\cdot\boldsymbol{r}} \hat{H}_0(\boldsymbol{K}') u_{nK'_x0}(\boldsymbol{r}) f_{nk'} \\
&= \frac{1}{N} \int d\boldsymbol{\tau} \sum_{nk'} u^\dagger_{mK_x0}(\boldsymbol{\tau}) e^{i(k'-k)\cdot\boldsymbol{\tau}} \left\{ \sum_{\boldsymbol{R}} e^{i(k'-k)\cdot\boldsymbol{R}} \right\} \\
&\quad \times \hat{H}_0(\boldsymbol{K}') u_{nK'_x0}(\boldsymbol{\tau}) f_{nk'} \\
&= \int d\boldsymbol{\tau} \sum_n u^\dagger_{mK_x0}(\boldsymbol{\tau}) \hat{H}_0(\boldsymbol{K}) u_{nK_x0}(\boldsymbol{\tau}) f_{n\boldsymbol{k}},
\end{aligned} \tag{240}$$

from which Eq. (239) directly follows. In the second equality, we have split the integral $\int d\boldsymbol{r}\, f(\boldsymbol{r})$ as $\sum_{\boldsymbol{R}} \int d\boldsymbol{\tau}\, f(\boldsymbol{\tau}+\boldsymbol{R})$; i.e., we integrate over the cell-periodic position coordinate $\boldsymbol{\tau}$ and sum over all unit cells labeled by the Bravais lattice vectors $\boldsymbol{R}$.

The right-hand side of Eq. (239) involves a triple product of symmetrized operators; we evaluate it utilizing the product rule of Eq. (229). We first consider the product of $\hat{H}_0(\boldsymbol{K})$ with any other symmetrized operator; owing to $\hat{H}_0(\boldsymbol{k})$ being quadratic in $\boldsymbol{k}$, the expansion of Eq. (229) continues at most to second order:

$$\hat{H}_0(\boldsymbol{K})B(\boldsymbol{K}) = [\hat{H}(\boldsymbol{k})B(\boldsymbol{k})] + \frac{i}{2l^2}\epsilon_{\alpha\beta}[\hat{\Pi}^{\alpha}(\boldsymbol{k})\partial_{\beta}B(\boldsymbol{k})] - \frac{1}{8ml^4}\epsilon_{\alpha\beta}\epsilon_{\alpha\nu}[\partial_{\beta}\partial_{\nu}B(\boldsymbol{k})] = \left[\hat{H}\left(k_{\alpha} + (i/2)l^{-2}\epsilon_{\alpha\beta}\nabla^{\beta}_{\boldsymbol{k}}\right)B(\boldsymbol{k})\right]. \quad (241)$$

To evaluate a triple product of the form $A(\boldsymbol{K})\hat{H}_0(\boldsymbol{K})B(\boldsymbol{K})$, we may first evaluate $\hat{H}_0 B$ using Eq. (241), then apply Eq. (E8) to $\{A\{\hat{H}_0 B\}\}$. In this manner, we derive the symbol of $\tilde{\mathcal{H}}(\boldsymbol{K})$ as

$$\begin{aligned}\tilde{\mathcal{H}}_{mn}(\boldsymbol{k}) &= e^{(i/2)l^{-2}\epsilon_{\alpha\beta}\nabla^{\alpha}_{\boldsymbol{k}}\nabla^{\beta}_{\boldsymbol{k}'}}\int d\boldsymbol{\tau}\, u^{*}_{mk_x0}(\boldsymbol{\tau})\hat{H}_0\left(k'_{\alpha} + (i/2)l^{-2}\epsilon_{\alpha\beta}\nabla^{\beta}_{\boldsymbol{k}'}\right)u_{nk'_x0}(\boldsymbol{\tau})\big|_{\boldsymbol{k}'=\boldsymbol{k}} \\ &= \langle u_{mk_x0}|\hat{H}_0(\boldsymbol{k})|u_{nk_x0}\rangle + \frac{i}{2l^2}\left\{\langle\nabla^{x}_{\boldsymbol{k}}u_{mk_x0}|\hat{\Pi}^{y}(\boldsymbol{k})|u_{nk_x0}\rangle - \langle u_{mk_x0}|\hat{\Pi}^{y}(\boldsymbol{k})|\nabla^{x}_{\boldsymbol{k}}u_{nk_x0}\rangle\right\} + \frac{1}{2ml^4}\langle\nabla^{x}_{\boldsymbol{k}}u_{mk_x0}|\nabla^{x}_{\boldsymbol{k}}u_{nk_x0}\rangle, \\ &= \langle u_{mk_x0}|\hat{H}_0(\boldsymbol{k})|u_{nk_x0}\rangle - \frac{1}{2l^2}\left\{\tilde{\mathfrak{X}}^{x}, \left(\tilde{\Pi}^{y} + \frac{k_y}{m}\right)\right\}_{mn;k_x0} + \frac{1}{2ml^4}(\tilde{\mathfrak{X}}^{x})^2_{mn;k_x0}. \end{aligned} \quad (242)$$

In the last equality, we employed $\hat{\boldsymbol{\Pi}}(\boldsymbol{k}) = \hat{\boldsymbol{\Pi}} + \boldsymbol{k}/m$, which follows from the definition of the velocity operator in Eqs. (9) and (10). We might also express

$$\langle u_{mk_x0}|\hat{H}_0(\boldsymbol{k})|u_{nk_x0}\rangle = (\tilde{H}_0 + k_y\tilde{\Pi}^{y})_{mn;k_x0} + \frac{k_y^2}{2m}\delta_{mn}, \quad (243)$$

using the identity in Eq. (11). The highest-order term in Eq. (242) is $O(l^{-4})$ due to the following two reasons: (i) $e^{(i/2)l^{-2}\epsilon_{\alpha\beta}\nabla^{\alpha}_{\boldsymbol{k}}\nabla^{\beta}_{\boldsymbol{k}'}}$ acts on a function which depends quadratically on k_y [through $\hat{H}_0(\boldsymbol{k})$], and (ii) the operator $\hat{H}_0(k'_{\alpha} + (i/2)l^{-2}\epsilon_{\alpha\beta}\nabla^{\beta}_{\boldsymbol{k}'})$ is at most of order l^{-4}.

As motivated towards the end of Sec. VIII B, we should consider k_y/G_y as a small parameter, independent of and in addition to l^{-2}. For any function of $\boldsymbol{k}$ and l^{-2}, we may indicate its order in k_y by a superscript:

$$G^{a}_{b}(\boldsymbol{k},l^{-2}) = O\left(k_y^{a}l^{-2b}\right); \quad (244)$$

we retain our convention of indicating the order in l^{-2} through the subscript; for matrices, we would have additional subscripts to indicate the row and column indices: $\{G^{a}_{b}\}_{mn}$. For symmetrized operators $G(\boldsymbol{K})$, we may also label them as $G^{a}_{b}(\boldsymbol{K})$ if their corresponding symbols satisfy Eq. (244). We are now ready to organize the effective Hamiltonian in Eq. (242) in a power series in the two small parameters, with the aid of Eq. (243):

$$\tilde{\mathcal{H}}(\boldsymbol{K}) = \tilde{\mathcal{H}}^{0}_{0} + \tilde{\mathcal{H}}^{1}_{0} + \tilde{\mathcal{H}}^{0}_{1} + \tilde{\mathcal{H}}^{1}_{1} + \tilde{\mathcal{H}}^{2}_{0} + \tilde{\mathcal{H}}^{0}_{2}, \quad (245)$$

$$\tilde{\mathcal{H}}^{0}_{0}(\boldsymbol{K}) = \tilde{H}_0(K_x,0), \quad (246)$$

$$\tilde{\mathcal{H}}^{1}_{0}(\boldsymbol{K}) = \tfrac{1}{2}[\{k_y,\tilde{\Pi}^{y}(k_x,0)\}], \quad (247)$$

$$\tilde{\mathcal{H}}^{0}_{1}(\boldsymbol{K}) = -\frac{1}{2l^2}\{\tilde{\mathfrak{X}}^{x}(K_x,0),\tilde{\Pi}^{y}(K_x,0)\}, \quad (248)$$

$$\tilde{\mathcal{H}}^{1}_{1}(\boldsymbol{K}) = -\frac{1}{2ml^2}[\{\tilde{\mathfrak{X}}^{x}(k_x,0),k_y\}], \quad (249)$$

$$\tilde{\mathcal{H}}^{2}_{0}(\boldsymbol{K}) = \frac{k_y^2}{2m}, \quad (250)$$

$$\tilde{\mathcal{H}}^{0}_{2}(\boldsymbol{K}) = \frac{1}{2ml^4}(\tilde{\mathfrak{X}}^{x})^2\Big|_{K_x,0}. \quad (251)$$

D. Block-diagonalization of effective Hamiltonian

Our goal is to find a transformation (S) that block-diagonalizes the effective Hamiltonian [$\tilde{\mathcal{H}}$ in Eq. (245)] with respect to the decomposition $P \oplus Q$; recall Eq. (227). We will carry out this transformation perturbatively in the two small parameters k_y and l^{-2}; our approach thus marries the traditional Löwdin partitioning in $\boldsymbol{k}\cdot\boldsymbol{p}$ Hamiltonians (which utilizes $\boldsymbol{k}$ as a small parameter) [72,149,150] with the lesser-known block-diagonalization procedures of effective Hamiltonians (which utilize l^{-2} as a small parameter) [14,23,24].

Let us expand S in a series organized in powers of k_y and l^{-2}, where each term in the series is a symmetrized function of $\boldsymbol{K}$:

$$S(\boldsymbol{K}) = I + {\sum_{i,j}}' S^{i}_{j}(\boldsymbol{K}), \quad S^{i}_{j}(\boldsymbol{K}) = \left[S^{i}_{j}(\boldsymbol{k})\right],$$

$$S^{i\,\dagger}_{j}(\boldsymbol{K}) = \left[S^{i\,\dagger}_{j}(\boldsymbol{k})\right], \quad S^{i}_{j}(\boldsymbol{k}) = O\left(k_y^{i}l^{-2j}\right). \quad (252)$$

By $\sum'_{i,j}$, we mean to sum over all nonnegative integers but exclude the single case of $i=j=0$. S is formally an infinite-dimensional matrix operator, as are $\tilde{\mathcal{H}}$ and $\tilde{\mathcal{H}}'$. We have chosen the lowest-order term in S to be the identity operation, since $\mathcal{H}^{0}_{0}$ is already block diagonal [cf. Eq. (246)], and requires no further modification.

From a wave function perspective, we are modifying our ansatz $\Psi \to \Psi'$ as in Eq. (228). Following essentially the same steps as outlined in Sec. VIII B, the modified, time-independent Hamiltonian equation $[(\hat{H}-E)\Psi' = 0]$ is equivalent to

$$\{S^{\dagger}\tilde{\mathcal{H}}S - ES^{\dagger}S\} * f_{\boldsymbol{k}} = 0, \quad (253)$$

where, again, $A * B$ denotes a matrix multiplying a vector with implicit index summation. To maintain the structure of an eigenvalue equation, we insist on the unitary condition $S^{\dagger}S = I$. In practice, this unitarity condition will be imposed

perturbatively. That is, from

$$S^\dagger(\boldsymbol{K})S(\boldsymbol{K}) = I + \sum_{i,j}{}' S_j^{i\,\dagger}(\boldsymbol{K}) + \sum_{a,b}{}' S_b^a(\boldsymbol{K}) + \sum_{i,j,a,b}{}' S_j^{i\,\dagger}(\boldsymbol{K}) S_b^a(\boldsymbol{K}),$$

we impose the conditions order by order, e.g.,

$$\begin{aligned} S_1^0 &= -S_1^{0\dagger}, \quad S_0^1 = -S_0^{1\dagger}, \\ 0 &= S_1^1 + S_1^{1\dagger} + S_1^{0\dagger} S_0^1 + S_0^{1\dagger} S_1^0, \\ 0 &= S_2^{0\dagger} + S_2^0 + S_1^{0\dagger} S_1^0, \\ 0 &= S_0^{2\dagger} + S_0^2 + S_0^{1\dagger} S_0^1, \ldots . \end{aligned} \tag{254}$$

We may also expand

$$\begin{aligned} S^\dagger \tilde{\mathcal{H}} S = \tilde{\mathcal{H}}_0^0 &+ \left(\tilde{\mathcal{H}}_0^1 + \left[\tilde{\mathcal{H}}_0^0, S_0^1\right]\right) + \left(\tilde{\mathcal{H}}_1^0 + \left[\tilde{\mathcal{H}}_0^0, S_1^0\right]\right) \\ &+ \left(\tilde{\mathcal{H}}_0^2 + \tilde{\mathcal{H}}_0^0 S_0^2 + S_0^{2\dagger}\tilde{\mathcal{H}}_0^0 + \left[\tilde{\mathcal{H}}_0^1, S_0^1\right] + S_0^{1\dagger}\tilde{\mathcal{H}}_0^0 S_0^1\right) \\ &+ \cdots . \end{aligned} \tag{255}$$

The commutators in the above expansion were derived by utilizing the anti-Hermiticity of S_0^1 and S_1^0 [cf. Eq. (254)]. The commutator of two symmetrized operators is not just the symmetrized commutator of their corresponding symbols, e.g.,

$$\left[\tilde{\mathcal{H}}_0^0(\boldsymbol{K}), S_j^i(\boldsymbol{K})\right] = \left[\left[\tilde{H}_0(k_x,0), S_j^i(\boldsymbol{k})\right]\right] + \sum_{n=1}^{i} C_{j+n}^{i-n}. \tag{256}$$

The additional terms $\{C_{j+n}^{i+n}\}_{n=1}^{i}$ on the right-hand side originate from a Moyal expansion, which we express in a more general form in Eq. (E9). The regularity of C_{j+n}^{i+n} implies that for every unit increase in the power in l^{-2}, there is a corresponding unit decrease in the power of k_y. After all, the Moyal expansion is an expansion in $l^{-2}\epsilon_{\alpha\beta}\nabla_{\boldsymbol{k}}^\alpha \nabla_{\boldsymbol{k}}^\beta$. Each of $\{C_{j+n}^{i-n}\}$ renormalizes $\{\tilde{\mathcal{H}}_{j+n}^{i-n'}\}$ in the block-diagonalization procedure.

In principle, we may block-diagonalize $\tilde{\mathcal{H}}$ to any order in k_y or l^{-2}; this will be demonstrated explicitly for $\tilde{\mathcal{H}}_j^i$ with $(i,j) = (0,1)$ and $(1,0)$. Two identities will be useful for this purpose, which are particularizations of Eq. (256):

$$\left[\tilde{\mathcal{H}}_0^0(\boldsymbol{K}), S_0^1(\boldsymbol{K})\right] = \left[\left[\tilde{H}_0(k_x,0), S_0^1(\boldsymbol{k})\right]\right] + \frac{i}{2l^2}\left[\left\{\tilde{\Pi}^x(k_x,0), \nabla_{\boldsymbol{k}}^y S_0^1\right\}\right], \tag{257}$$

$$\left[\tilde{\mathcal{H}}_0^0(\boldsymbol{K}), S_1^0(\boldsymbol{K})\right] = \left[\left[\tilde{H}_0(k_x,0), S_1^0(\boldsymbol{k})\right]\right]. \tag{258}$$

Equation (257) has one more term than (258) because S_0^1 is linear in k_y while S_1^0 is independent of k_y. We will exemplify how the last term in Eq. (257) renormalizes $\tilde{\mathcal{H}}_1^0$.

Employing the identities in Eqs. (257) and (258), the first two brackets in Eq. (255) may be expressed as

$$\begin{aligned} \tilde{\mathcal{H}}_0^1 + \left[\tilde{\mathcal{H}}_0^0, S_0^1\right]\big|_{\boldsymbol{K}} &= \left[\tilde{\mathcal{H}}_0^1(\boldsymbol{k}) + \left[\tilde{H}_0(k_x,0), S_0^1(\boldsymbol{k})\right] + \frac{i}{2l^2}\left\{\tilde{\Pi}^x(k_x,0), \nabla_{\boldsymbol{k}}^y S_0^1\right\}\right], \\ \tilde{\mathcal{H}}_1^0 + \left[\tilde{\mathcal{H}}_0^0, S_1^0\right]\big|_{\boldsymbol{K}} &= \left[\tilde{\mathcal{H}}_1^0(\boldsymbol{k}) + \left[\tilde{H}_0(k_x,0), S_1^0(\boldsymbol{k})\right]\right], \end{aligned} \tag{259}$$

leading to

$$\tilde{\mathcal{H}}_0^{\prime 1} = \left[\tilde{\mathcal{H}}_0^1(\boldsymbol{k}) + \left[\tilde{H}_0(k_x,0), S_0^1(\boldsymbol{k})\right]\right], \tag{260}$$

$$\tilde{\mathcal{H}}_1^{\prime 0} = \left[\tilde{\mathcal{H}}_1^0(\boldsymbol{k}) + \left[\tilde{H}_0(k_x,0), S_1^0(\boldsymbol{k})\right] + \frac{i}{2l^2}\left\{\tilde{\Pi}^x(k_x,0), \nabla_{\boldsymbol{k}}^y S_0^1\right\}\right]. \tag{261}$$

To block-diagonalize $\tilde{\mathcal{H}}_0^{\prime 1}$ with respect to $P \oplus Q$, we choose S_0^1 so that the off-block-diagonal elements of $[\tilde{H}_0, S_0^1]$ exactly cancel the off-block-diagonal elements of $\tilde{\mathcal{H}}_0^1$. A simple expression for $S_0^1(\boldsymbol{k})$ exists if we employ a basis for cell-periodic functions $\{u_{nk_x0}\}$ that (i) corresponds to energy bands, and (ii) retains our initial assumption of smoothness in k_x. The cancellation of off-block-diagonal elements in the energy basis leads to the following condition:

$$\begin{aligned} \left\{\tilde{\mathcal{H}}_0^1(\boldsymbol{k})\right\}_{\bar{m}n} &= -\left(\varepsilon_{\bar{m}k_x0} - \varepsilon_{nk_x0}\right)\left\{S_0^1(\boldsymbol{k})\right\}_{\bar{m}n} \\ \Rightarrow \left\{S_0^1(\boldsymbol{k})\right\}_{\bar{m}n} &= -k_y \frac{\tilde{\Pi}_{\bar{m}n}^y(k_x,0)}{\varepsilon_{\bar{m}k_x0} - \varepsilon_{nk_x0}}, \end{aligned} \tag{262}$$

with $\bar{m}$ labeling bands in Q and n the bands in P; clearly the above equality also holds with $\bar{m}$ and n interchanged. In the last equality, we extracted $\tilde{\mathcal{H}}_0^1$ from Eq. (247). The block-diagonal elements of S_0^1 may be any smooth function of $\boldsymbol{k}$ that is linear in k_y; in practice we will set all of them to zero, such that

$$S_0^1(\boldsymbol{k}) = -ik_y \tilde{\Upsilon}^y(k_x), \tag{263}$$

with $\tilde{\Upsilon}$ defined in Eq. (235).

Inserting this equation, as well as Eq. (248), into Eq. (261) we obtain

$$\begin{aligned} \tilde{\mathcal{H}}_1^{\prime 0} &= \left[\left[\tilde{H}_0(k_x,0), S_1^0(\boldsymbol{k})\right] - \frac{1}{2l^2}\{\tilde{\mathfrak{X}}^x, \tilde{\Pi}^y\}_{k_x,0} + \frac{1}{2l^2}\{\tilde{\Pi}^x, \tilde{\Upsilon}^y\}_{k_x,0}\right] \qquad (264) \\ &= \left[\left[\tilde{H}_0(k_x,0), S_1^0(\boldsymbol{k})\right] + \frac{1}{2l^2}(\{\tilde{\Upsilon}^y, \tilde{\Pi}^x\} - \{\mathring{\tilde{\mathfrak{X}}}^x, \tilde{\Pi}^y\})_{k_x,0} - \frac{1}{2l^2}\{\dot{\tilde{\mathfrak{X}}}^x, \tilde{\Pi}^y\}_{k_x,0}\right]. \qquad (265) \end{aligned}$$

In the last equality, we have separated $\tilde{\mathfrak{X}}^x$ into its block-diagonal ($\dot{\tilde{\mathfrak{X}}}^x$) and off-block-diagonal ($\mathring{\tilde{\mathfrak{X}}}^x$) components [cf. Eq. (234)]. In similar fashion to the case of $\mathcal{H}_0^{1\prime}$, we will cancel the off-block-diagonal elements of $\mathcal{H}_1^{0\prime}$ by a judicious choice of S_1^0:

$$\begin{aligned} \left\{S_1^0(\boldsymbol{k})\right\}_{\bar{m}n} &= \frac{(\{\tilde{\Upsilon}^y, \tilde{\Pi}^x\} - \{\mathring{\tilde{\mathfrak{X}}}^x, \tilde{\Pi}^y\} - \{\dot{\tilde{\mathfrak{X}}}^x, \tilde{\Pi}^y\})_{\bar{m}n;k_x0}}{-2l^2\left(\varepsilon_{\bar{m}k_x0} - \varepsilon_{nk_x0}\right)}, \\ \left\{S_1^0(\boldsymbol{k})\right\}_{mn} &= S_1^0(\boldsymbol{k})\}_{\bar{m}\bar{n}} = 0. \end{aligned} \tag{266}$$

This completes the block-diagonalization to order k_y and l^{-2}; what remains is to define finite-dimensional matrices, having dimension D equal to the rank of $P(\boldsymbol{k})$, to replace the infinite-dimensional matrices. We then finally obtain the effective Hamiltonian of Eq. (230).

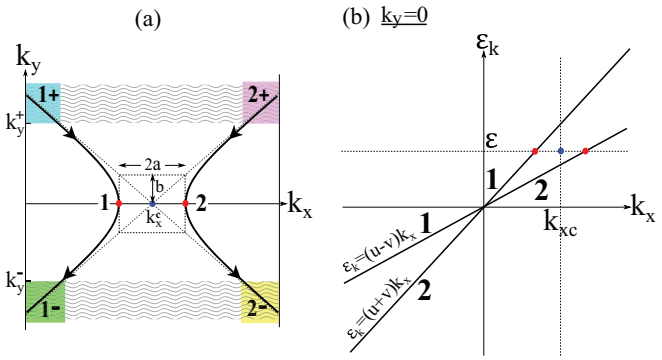

FIG. 7. (a) illustrates the interband breakdown region, which overlaps with the semiclassical region (indicated by gray wavy lines). Solid black lines illustrate the band contours at positive energy. The orbit velocity $\dot{\boldsymbol{k}}$ is indicated by arrows, and determined by Hamilton's equation. (b) Band dispersion at fixed $k_y = 0$ for the Hamiltonian in Eq. (267), with $u > v > 0$ and $w > 0$. The two bands are labeled by indexes 1 and 2.

IX. INTERBAND BREAKDOWN

Interband breakdown occurs where two constant-energy band contours—belonging to distinct bands—become anomalously close. As illustrated in Fig. 7(b), the two contours approach each other as two arms of a hyperbola, just as in the case for intraband breakdown. What distinguishes the two cases are the orientations [152] of traveling wave packets (as determined by Hamilton's equation) on both arms: opposite for interband [Fig. 7(b)], and identical for intraband [Figs. 3(b) and 3(c)]. Another distinguishing feature is that for interband breakdown, only one of two in-plane components of the band velocity ($\boldsymbol{\nabla}_{\boldsymbol{k}}\varepsilon$) becomes anomalously small, whereas this is true for both in-plane components in the intraband case.

To systematize the derivation of quantization conditions, it will be useful to formalize the above discussion in the language of graph theory (cf. Sec. III F). Each region of strong tunneling is identified with a degree-four, two-in-two-out vertex. For the intraband- (resp. interband-) breakdown vertex, the two incoming edges are (resp. not) diametrically opposite to each other. To simplify notation, any "breakdown vertex" in the rest of this section should be understood as an interband-breakdown vertex.

A. Symmetry analysis and Bloch Hamiltonian near a II-Dirac point

Let us identify the symmetry classes which stabilize band touchings of the kind that leads to interband breakdown. To begin, how are band touchings (of any kind) stabilized in a Brillouin two-torus ($BT_\perp$) parametrized by $\boldsymbol{k} = (k_x, k_y)$; we shall again assume the field is aligned in $\vec{z}$. Our present discussion is restricted to $BT_\perp$, but we will eventually comment on how $BT_\perp$ is embedded in a Brillouin three-torus. Applying the argument in the introduction of Sec. VIII, robust band touchings occur on points or curves, if the co-dimension of the Hamiltonian is two or one, respectively. We shall investigate the former case, and postpone the latter case to future studies. We shall also assume throughout this section that the point touching occurs between two nondegenerate bands; touchings between spin-degenerate bands are briefly discussed in Sec. X.

We focus on point degeneracies which lie at the tip of an energy-momentum cone; i.e., the band degeneracy splits at linear order in $\boldsymbol{k}$ (originating from the band touching), and the constant-energy contours intersect as an "X". From a general classification of Fermi surfaces near conical band touchings [56], this "X" must correspond to a type-II Dirac point [57–59]. A II-Dirac point is minimally modeled by the following Hamiltonian in the Luttinger-Kohn representation [cf. Eq. (43)]:

$$\begin{aligned} H_0(\boldsymbol{k}) &= (u + v\gamma_3)k_x + wk_y\gamma_1, \\ \Pi^x(\boldsymbol{0}) &= u + v\gamma_3, \\ \Pi^y(\boldsymbol{0}) &= w\gamma_1, \end{aligned} \tag{267}$$

with $\boldsymbol{k} = (k_x, k_y)$ originating from the point of degeneracy; γ_j are Pauli matrices spanning a (pseudo)spin-half basis. The linearized band dispersion is shaped as a "Dirac" cone which is rotationally invariant if $u = 0$. If u is continuously increased till $|u| > |v|$, the cone tilts over $E = 0$ (the energy of the Dirac point), and the zero-energy band contour changes discontinuously from a point to an "X". Precisely, for a finite-energy window near zero, the corresponding contours form a family of hyperbolic curves:

$$\begin{aligned} &\frac{(k_x - k_{xc})^2}{\bar{a}^2} - \frac{k_y^2}{\bar{b}^2} = 1, \quad \text{with} \quad k_{xc} := \frac{uE}{u^2 - v^2}, \\ &\bar{a} := \frac{vE}{u^2 - v^2}, \quad \bar{b} := \frac{vE}{w\sqrt{u^2 - v^2}}, \end{aligned} \tag{268}$$

with k_{xc} the center of the hyperbola; the "X" corresponds to the hyperbolic asymptotes: $k_y = \pm(\bar{b}/\bar{a})(k_x - k_{xc})$.

While Eq. (267) is not the most general form of a II-Dirac Hamiltonian [56], its simplicity manifests the physics we will describe. We may further motivate Eq. (267) as the most general Hamiltonian (up to unitary equivalence) satisfying the symmetry constraints:

$$[\hat{g}_1, H_0(\boldsymbol{k})] = 0, \quad \{\hat{g}_1, i\} = 0, \quad \hat{g}_1^2 = +I, \tag{269}$$

$$\begin{aligned} &\hat{g}_2 H_0(\boldsymbol{k})\hat{g}_2^{-1} = H_0(k_x, -k_y), \quad [\hat{g}_2, i] = 0, \\ &\mathrm{Tr}[\hat{g}_2] = 0, \quad \hat{g}_2^2 = \begin{cases} +I, & [\hat{g}_1, \hat{g}_2] = 0, \\ -I, & \{\hat{g}_1, \hat{g}_2\} = 0. \end{cases} \end{aligned} \tag{270}$$

g_1 is a spacetime transformation that maps $\boldsymbol{k} \to \boldsymbol{k}$ (within the plane), and has an antiunitary representation that squares to $+I$. For example, g_1 could be $T\mathfrak{i}$ in an integer-spin representation; alternatively, $g_1 = T\mathfrak{c}_{2z}$ in either half-integer or integer-spin representation, in which case $BT_\perp$ is identified with either of the high-symmetry planes: $k_z = 0$ or π. In all these cases, g_1 lowers the co-dimension of the Hamiltonian to two, and hence stabilizes Dirac points in $BT_\perp$. g_2 is an order-two, spatial symmetry which maps $\boldsymbol{k} \to (k_x, -k_y)$. A touching between two orthogonal representations of g_2 occurs at the Dirac point; the commutation relations between $\hat{g}_1$ and $\hat{g}_2$ in Eq. (270) imply that each representation of g_2 is invariant under g_1, so that the band touching may split away from $\boldsymbol{0}$. Examples of g_1 include the reflection $\mathfrak{r}_y$, or the glide $\mathfrak{g}_{y,\vec{x}/2}$. Incidentally, $T\mathfrak{i}$ and $\mathfrak{g}_{y,\vec{x}/2}$ are the symmetries of the monolayer MTe_2 (M = W, Mo) [153], which serves as a toy model for II-Dirac fermions [59]. For the pseudospin basis chosen

in Eq. (267), $\hat{g}_1 = K$ and $\hat{g}_2 = \sigma_3$. We clarify that $\boldsymbol{k} = \boldsymbol{0}$ is *not* an inversion-invariant wave vector, but a generic point on either g_2-invariant line ($k_y = 0$ or π). Lacking a symmetry (e.g., $\mathfrak{c}_{nz}$, T, $\mathfrak{r}_x$) that nontrivially transforms k_x in the sense of Eq. (28), the Hamiltonian term uk_x is legal and tilts the Dirac cone in the direction parallel to the g_2-invariant lines [154].

There are two topologically distinct ways to embed $BT_\perp$ (containing a II-Dirac point) in a 3D Brillouin torus: (i) If the degeneracy splits away from $BT_\perp$, it is a genuine 3D point degeneracy of the II-Weyl type; this may be modeled by adding $tk_z\gamma_2$ to the Hamiltonian in Eq. (267). (ii) If the degeneracy persists away from $BT_\perp$, the II-Dirac point should be identified as a point on a line degeneracy; this may be modeled by adding $tk_z\gamma_3$ to Eq. (267), such that the line degeneracy lies on the intersection of two planes: $k_y = 0$ and $tk_z + vk_x = 0$. Interband breakdown in solids with line nodes was first studied by Slutskin [12], but the conception of Weyl/Dirac points did not exist at his time (1967).

A quantity of geometric significance is the area $(4\bar{a}\bar{b})$ of the rectangle inscribed between the two hyperbolic arms (see Fig. 3). It is natural that the dimensionless parameter

$$\bar{\mu} = \frac{1}{2}\bar{a}\bar{b}l^2 = \frac{v^2E^2l^2}{2w(u^2 - v^2)^{3/2}} \geqslant 0 \tag{271}$$

determines the probability of tunneling between orbits: tunneling is negligible where $\bar{\mu} \gg 1$, and significant otherwise. The exact form of $\bar{\mu}$ will be motivated by the connection formula in Eqs. (283)–(285), which is the key result of this section.

B. Effective Hamiltonian for interband breakdown, and the Landau-Zener analogy

Following the divide-and-conquer strategy that we have employed for the turning point and the saddle point, we would likewise need to formulate an effective Hamiltonian that is valid at the interband-breakdown region, and solve for its wave function nonperturbatively. In the Landau electromagnetic gauge where k_x is a good quantum number, the breakdown region is an interval $k_y \in [k_y^+, k_y^-]$ centered at the II-Dirac point where quantum tunneling is significant, as illustrated in Fig. 7(a).

One complication for interband breakdown that did not occur in the previous two cases: a point degeneracy in the band dispersion invalidates the use of the field-modifed Bloch functions [cf. Eq. (55)] as basis functions. As motivated in Sec. VIII A, we will instead employ a set of field-modified [12] Luttinger-Kohn functions [72] which are analytic with respect to $\boldsymbol{k}$ at the degeneracy; these are basis functions in what we call the $(K_x,0)$ representation. At energies where interband breakdown is relevant, our ansatz for the wave function is

$$\Psi_{k_x}(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{k_y \in [k_y^+, k_y^-]} \sum_{n=1}^{2} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \tilde{u}_{n,K_x,0}(\boldsymbol{r}) \tilde{f}_{n\boldsymbol{k}} + \frac{1}{\sqrt{N}} \sum_{k_y \notin [k_y^+, k_y^-]} \sum_{n=1}^{2} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n,K_x,k_y}(\boldsymbol{r}) g_{n\boldsymbol{k}} + \cdots, \tag{272}$$

with $n = 1,2$ labeling bands in the band-touching subspace, as illustrated in Fig. 7. K_x, the kinetic quasimomentum operator defined in Eq. (77), acts on $\tilde{f}_{n\boldsymbol{k}}$ and $g_{n\boldsymbol{k}}$, which are wave functions in the $(K_x,0)$ and (K_x,k_y) representations, respectively. They are respectively valid in the breakdown and semiclassical intervals. An assumption (on the band parameters [12]) is made that the two domains of validity overlap. In this region of overlap [indicated by wavy lines in Fig. 7(a)], the wave functions in the two representations may be matched as

$$\tilde{f}_{m\boldsymbol{k}} = \sum_{n=1}^{2} \langle \tilde{u}_{m,K_x,0} | u_{n,K_x,k_y} \rangle g_{n\boldsymbol{k}} + O\left(\frac{k_y}{G_y}, l^{-2}\right), \tag{273}$$

where $n = 1,2$ are indices for the band-touching subspace. The proof of this relation is closely analogous to the proof of completeness in Eqs. (218) and (219); instead of employing Bloch functions which are complete with respect to functions of $r \in \mathbb{R}^d$, we would likewise use that field-modified Bloch functions are complete with respect to functions of $r \in \mathbb{R}^d$ [50,76]. As denoted vaguely by ... in Eq. (272), there might generally be more contributions to Ψ that are associated with edges far away from the II-Dirac point (and therefore not illustrated in Fig. 7); these contributions will not be important in the matching procedure described in Sec. IX C.

As derived in Sec. VIII B, the effective Hamiltonian in the $(K_x,0)$ representation is obtained from Eq. (237) as

$$\mathcal{H}_0(\boldsymbol{K}) = H_0 + K_x\Pi^x + k_y\Pi^y, \tag{274}$$

where H_0, Π_j are 2×2 matrices evaluated at the point of degeneracy ($\boldsymbol{k} = \boldsymbol{0}$). When Eq. (274) is particularized to our minimal model in Eq. (267),

$$\sum_{n=1}^{2} ([\mathcal{H}_0]_{mn} - E\delta_{mn})\tilde{f}_{n\boldsymbol{k}}, \quad \text{with}$$
$$\mathcal{H}_0(\boldsymbol{K}) = K_x(u + v\gamma_3) + wk_y\gamma_1. \tag{275}$$

To simplify the notation, we would further assume u,v,w are all positive. Our neglect of the first-order-in-l^{-2} corrections in Eq. (237) is only justified if, near $\boldsymbol{k} = \boldsymbol{0}$, a large energetic gap separates the two-band subspace (involved in the degeneracy) from every other band [155].

Transforming the wave function as

$$\tilde{f}_{n\boldsymbol{k}} = \alpha_{\boldsymbol{k}E} \sum_{\bar{m}=\bar{1}}^{\bar{2}} \bar{T}_{n\bar{m}} \bar{f}_{\bar{m}\boldsymbol{k}}, \quad \text{with}$$
$$\alpha_{\boldsymbol{k}E} = e^{i(k_x - k_{xc}(E))k_y l^2}, \tag{276}$$
$$\bar{T} = \frac{1}{\sqrt{2}} \begin{pmatrix} (u+v)^{-1/2} & (u+v)^{-1/2} \\ (u-v)^{-1/2} & -(u-v)^{-1/2} \end{pmatrix},$$

the effective eigenvalue equation describes the Landau-Zener dynamics of a two-level system:

$$0 = \left(\bar{a}\tau_1 + \frac{\bar{a}}{\bar{b}}k_y\tau_3 + \frac{i}{l^2}\frac{\partial}{\partial k_y}\right)\bar{f}_{\bar{k}}. \tag{277}$$

Here, τ_j are Pauli matrices, and Eq. (277) should be interpreted as a matrix differential equation acting on a two-component vector wave function $\bar{f}_{\bar{m}\boldsymbol{k}}$. A more general transformation to a Landau-Zener dynamical equation is described in

Appendix F 1, which would apply to a larger class of matrix Hamiltonians than assumed for our minimal model.

In the Landau-Zener analogy, k_y is interpreted as a time variable, and $\{k_x^n(k_y,E)\}_{n=1}^2$ as two "Landau-Zener energy" branches:

$$k_x^n(k_y,E) = k_{xc} + (-1)^n|\bar{a}|\sqrt{1+\frac{k_y^2}{\bar{b}^2}}, \quad k_x^n(0,0) = 0, \quad (278)$$

with k_{xc},$\bar{a}$, and $\bar{b}$ being E-dependent hyperbolic parameters defined in Eq. (268). "Energy" in the Landau-Zener analogy should not be confused with the actual energy (E) in the magnetic problem. We will refer to the zero-field energy bands labeled by $n = 1,2$ as the k_y-dependent "adiabatic basis" in the Landau-Zener analogy; as long as $E \neq 0$, there exists an adiabatic limit ($l^{-2} \to 0$) where the band is a conserved quantity. If $E = 0$, such an adiabatic limit does not exist and the probability of tunneling is unity; this has been described as a momentum-space analog of Klein tunneling [46]. At $E = 0$, it is more convenient to employ a diabatic basis which corresponds to the maximal-tunneling trajectories

$$k_x^{\bar{m}}(k_y) = (-1)^{\bar{m}}\left|\bar{a}/\bar{b}\right|k_y, \quad (279)$$

as illustrated in Fig. 8(c); the diabatic basis shall be labeled by $\bar{m} \in \{\bar{1},\bar{2}\}$, just as we have done for the two-component vector $\bar{f}_{\bar{m}k}$ in Eq. (277). We see that the matrix $\bar{T}_{n\bar{n}}$ introduced in Eq. (276) transforms between the adiabatic and diabatic bases.

C. Connection formula and quantization condition for interband breakdown

The Bohr-Sommerfeld quantization rule is the continuity condition on wave functions in both representations [$\tilde{f}_{nk}$ and g_{nk} in Eq. (272)], with the understanding that $\tilde{f}$ and g are related through Eq. (273) where the semiclassical and breakdown intervals overlap. In this section, we will derive a scattering-matrix formula that connects g across the breakdown interval; in effect, we may forget about $\tilde{f}$ and impose continuity on g, which satisfies certain wave-function-matching conditions.

Once $\tilde{f}$ is forgotten, we may proceed in close analogy to the graph-theoretic formulation of the quantization condition for intraband breakdown in Secs. III F–VII B. We assume the reader has some familiarity with these sections; we shall therefore avoid a lengthy exposition on similar-sounding generalities, in preference for a heuristic derivation of the quantization condition for a single II-Dirac point, as illustrated in Fig. 11 below.

A crucial ingredient to quantization conditions with interband breakdown is the connection formula for a single interband-breakdown vertex, which we will subsequently derive. We will actually derive two connection formulas: (a) the first formula, as summarized in Eqs. (283)–(285), is applicable for $E \neq 0$ and connects Zilberman-Fischbeck (ZF) functions in the adiabatic basis, and (b) for $E = 0$, the second formula [Eqs. (287) and (288)] connects ZF functions in the diabatic basis. These formulas extend a previous formula [12] to include the effect of the Berry phase.

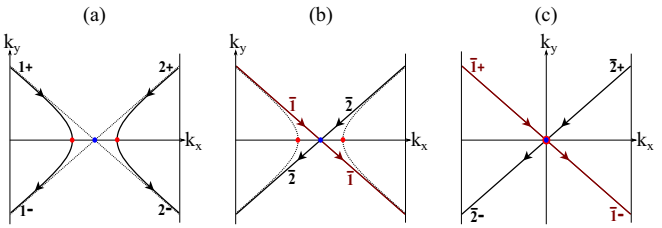

FIG. 8. (a) Labels for the four edges that meet at an interband-breakdown vertex. Black solid lines correspond to constant-energy band contours at zero field and nonzero energy ($E = 0$ being the energy of the II-Dirac point); we shall refer to energy bands for $E \neq 0$ as the adiabatic basis. (b) and (c) illustrate the diabatic basis for $E \neq 0$ and $E = 0$, respectively. For $E \neq 0$, the diabatic basis coincides with energy bands only for $|k_y| \gg |\bar{b}|$. For (and only for) $E = 0$, the center of the hyperbola [indicated by blue dot in (c)] coincides with the wave vector ($\boldsymbol{k} = 0$) of the II-Dirac point, and the diabatic basis coincides with the energy bands for all k_y.

1. Connection formula for $E \neq 0$

As illustrated in Fig. 8(a), the four edges which connect to the breakdown vertex are distinguished by the labels (1+,1−,2+,2−). We will eventually formulate the connection formula as a scattering matrix relating (1+,2+) to (1−,2−). Given an arbitrary graph with one or more breakdown vertices, it is important to correctly identify the four labels for each individual vertex. 1+ and 2+ label the two edges which are oriented toward the vertex, and 1− and 2− label the edges oriented away. We remind the reader that the orientation of each edge is the direction of a hypothetical wave packet, which is determined by Hamilton's equation with the convention $\boldsymbol{B} = -|B|\hat{z}$. Let us set down local coordinates centered on each vertex, such that + lies to the north and − to the south; we may then assign 1 to the west, and 2 to the east, as exemplified by the three graphs in Figs. 9(b)–9(d). It is sometimes convenient to define (as we have already done) a right-handed coordinate system where k_y increases in the direction from − to +, and

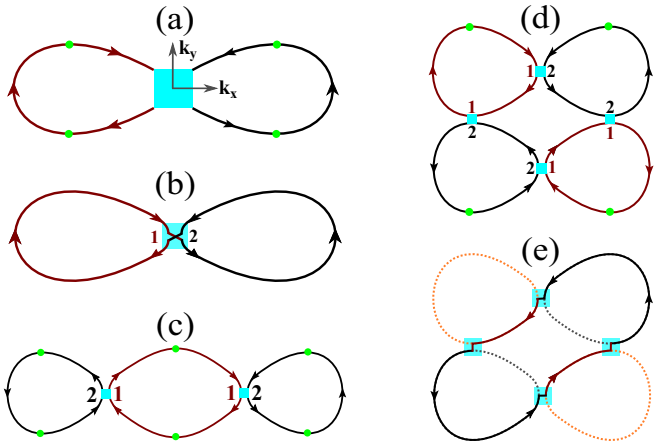

FIG. 9. (a) Graph for a single II-Dirac point. (b) Anticrossing figure-of-eight trajectory. (c) is a graph that typically occurs in band-inverted nonsymmorphic metals [59]. (b) and (e) illustrate closed Feynman trajectories. Brown and black pockets have opposite circulations.

k_x increases in the direction from 1 to 2, as exemplified in Fig. 9(a). Since western and eastern edges also correspond to distinct bands, we might also view 1,2 as band indices which are locally defined at each vertex.

For $E \neq 0$, the two edges belonging to band n form a smooth curve given by $k_x^n(k_y, E)$ in Eq. (278). The corresponding Zilberman-Fischbeck (ZF) wave function in the (K_x, k_y) representation is

$$w_{n\boldsymbol{k}} = \frac{1}{\sqrt{|v_n^x|}} e^{ik_x k_y l^2} e^{-il^2 \int_0^{k_y} [k_x^n - H_1^n (v_n^x)^{-1}] dz}, \tag{280}$$

following the general analysis of Sec. V A 1; H_1^n (the Roth-Berry-Zeeman Hamiltonian) and v_n^x (the band-diagonal velocity) are single-band quantities evaluated on the nth band contour. The H_1 term in Eq. (280) is further simplified owing to the assumed space-time symmetry g_1 in Eq. (269): (i) The orbital moment vanishes, as may be deduced from Table II. (ii) In spin-orbit-coupled systems, the Zeeman coupling also vanishes owing to g_1 (cf. Table II). (iii) In solids with negligible spin-orbit coupling, we work in the eigenbasis of the spin operator $\hat{\sigma}_z$, and the Zeeman splitting results in a constant term in H_1 which we will not write out explicitly. What remains of $\int H_1^n (v_n^x)^{-1}$ is the integral of the single-band Berry connection $\mathfrak{X}_n$ along the nth band contour, i.e.,

$$w_{n\boldsymbol{k}} = \frac{1}{\sqrt{|v_n^x|}} e^{ik_x k_y l^2} e^{-il^2 \int_0^{k_y} k_x^n dz} \, \mathcal{W}_{nk_y}, \tag{281}$$

with $\mathcal{W}$ defined as

$$\mathcal{W}_{nk_y} := \exp\left[i \int_{k_x^n(0),0}^{k_x^n(k_y),k_y} \mathfrak{X}_n \cdot d\boldsymbol{k}' \right] = \begin{cases} \mathcal{W}_{n+}, & k_y \gg +|\bar{b}|, \\ \mathcal{W}_{n-}, & k_y \ll -|\bar{b}|. \end{cases} \tag{282}$$

$\mathfrak{X}_n$ is only well defined everywhere along $k_x^n(k_y, E)$ for $E \neq 0$. For $E = 0$, the Berry connection (for a $\boldsymbol{k}$-space derivative in the azimuthal direction) diverges at the II-Dirac point; it is appropriate here to employ different ZF functions in a diabatic basis, which will be described in Sec. IX C 2 below. We will henceforth refer to $w_{n\boldsymbol{k}}$ defined in Eq. (280) as the adiabatic ZF functions, and restrict our attention to $E \neq 0$ in the remainder of this section.

Since the (K_x, k_y) representation is not valid for an interval of k_y in the breakdown region [illustrated by the white region in Fig. 7(b)], we introduce a label to distinguish (K_x, k_y) wave functions that are valid above (+) and below (−) the II-Dirac point:

$$\text{For } E \neq 0, \quad g_{1\boldsymbol{k}}^{\pm} = c_1^{\pm} w_{1\boldsymbol{k}} + \cdots, \quad g_{2\boldsymbol{k}}^{\pm} = c_2^{\pm} w_{2\boldsymbol{k}} + \cdots. \tag{283}$$

Here ... indicates contributions by edges far away from the II-Dirac point; they will not play a role in deriving the connection formula. The derivation proceeds in three steps: (i) In the breakdown interval, we solve for the eigenfunction of the effective Hamiltonian in the $(K_x, 0)$ representation [cf. Eq. (275)]. (ii) In the interval of overlap, we transform the eigenfunction of (i) to a wave function in the (K_x, k_y) representation through Eq. (273), and (iii) match the resultant wave function to the WKB wave functions $(g_{n\boldsymbol{k}}^{\pm})$ defined above. Step (i) is elaborated in Appendix F 1, and (ii) and (iii) in Appendix F 2. In this manner, we obtain a scattering-matrix equation relating incoming (at positive k_y) to outgoing (negative k_y) amplitudes:

$$\text{For } E \neq 0, \quad \begin{pmatrix} c_1^- \\ c_2^- \end{pmatrix} = \mathbb{S} \begin{pmatrix} c_1^+ \\ c_2^+ \end{pmatrix},$$
$$\mathbb{S}(E, l^2) = \begin{pmatrix} \sqrt{1-\rho^2} e^{i\omega} & -e^{i(\theta_1 - \theta_2)} \rho \\ e^{i(\theta_2 - \theta_1)} \rho & \sqrt{1-\rho^2} e^{-i\omega} \end{pmatrix}, \tag{284}$$
$$\rho(\bar{\mu}) = e^{-\pi \bar{\mu}},$$

with $\bar{\mu}$ the dimensionless tunneling parameter defined in Eq. (271). ω is the interband scattering phase plotted in Fig. 10, and defined by

$$\omega(\bar{\mu}) = \bar{\mu} - \bar{\mu} \ln \bar{\mu} + \arg[\Gamma(i\bar{\mu})] + \pi/4 \to \begin{cases} -\pi/4, & \bar{\mu} \to 0, \\ 0, & \bar{\mu} \to +\infty, \end{cases} \tag{285}$$

with Γ the Gamma function. In particular, $\omega = -\pi/4$ at the energy of the II-Dirac point, which may alternatively be derived by perturbation theory [156].

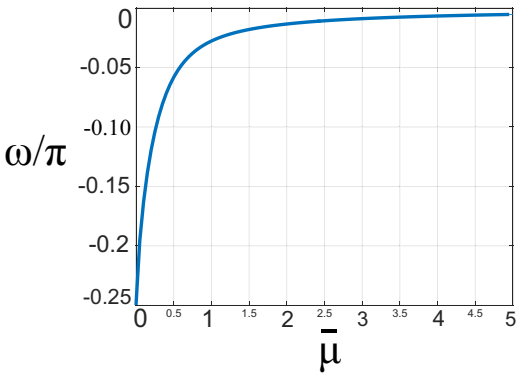

FIG. 10. Interband scattering phase ω vs $\bar{\mu}$.

The interband tunneling amplitude $(\mathbb{S}_{12})$ may be viewed [157] as the exponentiated action of a tunneling trajectory that encircles a Kohn branch point [158] in complex-k_y space; $|\mathbb{S}_{12}|^2 = e^{-2\pi\bar{\mu}}$ is the famous Landau-Zener tunneling probability. The unspecified phase $(\theta_1 - \theta_2)$ in $\mathbb{S}_{12}$ reflects an intrinsic phase ambiguity between two nondegenerate bands. This ambiguity was implicit in the expansion of Eq. (283), where we might have arbitrarily redefined $w_{n\boldsymbol{k}} \to w_{n\boldsymbol{k}} e^{i\theta_n}$ by a $\boldsymbol{k}$-independent but band-dependent phase. This arbitrary phase should not, however, affect the quantization condition for closed orbits, owing to the following argument: the quantization condition is a function of phases acquired by wave packets along closed Feynman trajectories. Let us consider those closed trajectories that involve interband tunneling. The two-in-two-out rule at an interband vertex guarantees that a wave packet must traverse an even number ($p \in 2\mathbb{Z}$) of pockets before forming a loop. For illustration, $p = 2$ for the figure-of-eight trajectory in Fig. 9(b), and $p = 4$ for the trajectory (indicated by solid lines) in Fig. 9(e). Each pocket corresponds to a single band, and a wave packet that tunnels between two pockets (labeled i_1 and i_2) picks up the interpocket phase difference $(\theta_{i_2} - \theta_{i_1})$. Since the wave packet must eventually

return to the pocket it originated, the sum of all interpocket phase differences acquired in a closed trajectory vanishes. This shall be made more explicit in our subsequent case study of Figs. 9(a) and 9(b) (Sec. IX D below).

2. Connection formula for $E = 0$

As motivated in the discussion below Eq. (282), we would like to define a different set of ZF functions (henceforth referred to as diabatic ZF functions) which are applicable at the energy of the II-Dirac point. By requiring that these functions are continuous along the maximal-tunneling trajectories $[k_x^{\bar{n}}(k_y)$, as defined in Eq. (279)], they assume the form

$$w_{\bar{n}\boldsymbol{k}} = \frac{1}{\sqrt{|v_{\bar{n}}^x|}} e^{ik_x k_y l^2} e^{-il^2 \int_0^{k_y} k_x^{\bar{n}} dz}\, \mathcal{W}_{\bar{n}k_y}, \tag{286}$$

with $\mathcal{W}_{\bar{n}}$ defined just as in Eq. (282), except the Berry connection is integrated over $k_x^{\bar{n}}(k_y)$. Following essentially the same argument as in Eq. (283), we define the coefficients $c_{\bar{n}}^{\pm}$ through

$$g_{\boldsymbol{k},E=0}^{\pm} = c_{\bar{1}}^{\pm} w_{\bar{1}\boldsymbol{k}} + c_{\bar{2}}^{\pm} w_{\bar{2}\boldsymbol{k}} + \cdots. \tag{287}$$

These coefficients correspond to the relabeled edges illustrated in Fig. 8(c), and are related simply as

$$\text{For } E = 0, \quad \begin{pmatrix} c_{\bar{1}}^- \\ c_{\bar{2}}^- \end{pmatrix} = \bar{\mathbb{S}} \begin{pmatrix} c_{\bar{1}}^+ \\ c_{\bar{2}}^+ \end{pmatrix}, \quad \bar{\mathbb{S}}(l^2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{288}$$

which states that Landau-Zener tunneling occurs with unit probability, independently of the strength of the field. This connection formula may be derived from solving Eq. (277), which decouples (for $E = 0$) to two scalar, first-order differential equations.

Let us compare this zero-energy connection formula $[\bar{\mathbb{S}}(l^2)]$ to the finite-energy formula $[\mathbb{S}(E,l^2)$ in Eq. (284)] in the limit $E \to 0^{\pm}$, with $0^{\pm}$ a vanishingly small positive/negative quantity. Ignoring the $(\theta_2 - \theta_1)$ phase (whose irrelevance was argued for in Sec. IX C 1), $\mathbb{S} \to -i\tau_2$, with τ_2 a Pauli matrix; this implies $c_{\bar{1}}^+ = c_{\bar{1}}^-$ and $c_{\bar{2}}^+ = -c_{\bar{2}}^-$, which differs from Eq. (288) by a minus sign. This apparent discontinuity in the connection formula does not imply that the quantization condition is also discontinuous at $E = 0$; we will see how this tension is resolved—by the Berry phase—in the following case study.

D. Case study: Single II-Dirac point

We study the simplest example of a single II-Dirac point, where the Fermi surface closes off as in Fig. 11(a); this may be modeled by adding a cubic term to the II-Dirac Hamiltonian [cf. Eq. (267)]

$$H_0(\boldsymbol{k}) = (u + v\gamma_3)k_x + wk_y\gamma_1 - t(1 - \gamma_3)k_x^3,$$
$$u,v,t,w > 0, \quad u > v. \tag{289}$$

This model has various realizations in the literature [46,47,159]. The corresponding graph in Fig. 11(b) comprises four edges, four turning vertices, and a single interband-breakdown vertex. The Landau levels of this model were first studied in Refs. [46,47] using a combination of semiclassical analysis (for single-band transport) and large-scale numerical

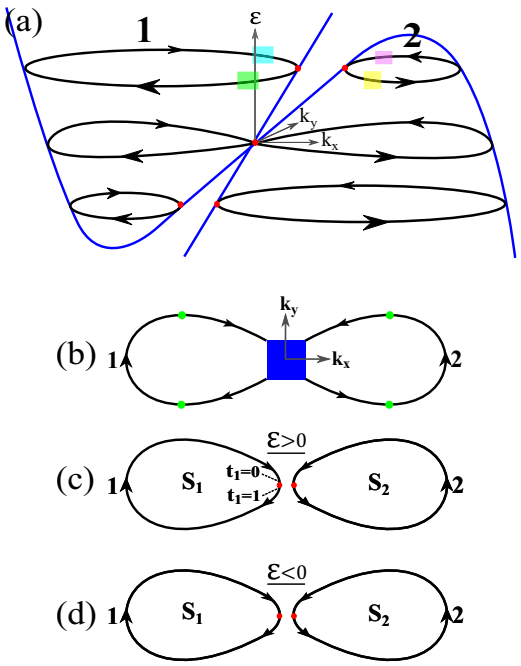

FIG. 11. (a) Energy-momentum dispersion of a single II-Dirac fermion. (b) shall be referred as the II-Dirac graph.

diagonalization; a quantization condition that determines this Landau level for any tunneling strength was first formulated by us in Ref. [36]. In this section, we derive the same quantization condition in greater detail; we hope to equip the interested reader with the technical know-how to construct quantization conditions for other, possibly more complicated, graphs. It should be clarified at the onset that we are constructing quantization conditions that apply to homotopy classes of graphs (a definition of homotopy equivalence is provided in Sec. III F), of which Eq. (289) merely describes one representative.

1. Quantization condition for the II-Dirac graph

The II-Dirac graph is similar to the double-well graph of Sec. VII C 1 in having two broken orbits linked by a single degree-four vertex (these graph-theoretic terms have been defined Sec. III F). The major differences between inter- and intraband breakdown lie in (i) the scattering amplitudes [contrast Eq. (284) with Eq. (190)], and in (ii) the orientations of the four edges adjacent to the breakdown vertex (see Fig. 1). This orientation demonstrably affects the signs of the semiclassical phases acquired along the broken orbits.

To each of our broken orbits (labeled $\{\mathfrak{o}_i\}_{i=1}^2$) we assign a scalar amplitude $A_{i,E}(t_i)$, with $t_i \in [0,1]$ a timelike variable that increases along the $\mathfrak{o}_i$ in a direction consistent with Hamilton's equation. $t_i = 0$ corresponds to the point of closest approach to the initial hyperbolic center $[\boldsymbol{k} = (k_{xc}(E),0)]$, and $t_i = 1$ to the point of closest approach to the final hyperbolic center, as illustrated in Fig. 11(c). In general, the initial and final hyperbolas may correspond to distinct II-Dirac points; in the present case study they are identical.

For $E \neq 0$, the Bloch functions can be made first-order differentiable with respect to $\boldsymbol{k} \in \mathfrak{o}_j$, and consequently the Berry connection is well defined; this is not true at $E = 0$,

where Bloch functions are discontinuous at the cusp of $\mathfrak{o}_j$. Let us then define Ω_j for $E \neq 0$ as the net phase acquired by a wave packet in traversing $\bar{\mathfrak{o}}_j$: this has the form

$$\Omega_j(E,l^2) = l^2 S_j(E) + \oint_{\bar{\mathfrak{o}}_j} \mathfrak{X} \cdot d\boldsymbol{k} + \pi, \tag{290}$$

where S_j is the oriented area of $\mathfrak{o}_j$; note S_1 and S_2 have opposite signs. π in the above equation corresponds to the Maslov correction for simple closed orbits, and the Berry phase contribution is fixed to π or 0, corresponding respectively to whether $\bar{\mathfrak{o}}_j$ encircles the II-Dirac point or not (Sec. VI D 2). The robust quantization of the Berry phase is a result of the symmetry g_1 [cf. Eq. (269)], which additionally ensures that the Roth and Zeeman contributions to Ω_j vanish for spin-orbit-coupled solids (cf. Sec. VI B). Since the orbit $\mathfrak{o}_j$ that encloses the II-Dirac point changes discontinuous across $E = 0$, Ω_j is also necessarily discontinuous:

$$\Omega_1(E,l^2) = \begin{cases} l^2 S_1(E), & E > 0, \\ l^2 S_1(E) + \pi, & E < 0, \end{cases}$$
$$\Omega_2(E,l^2) = \begin{cases} l^2 S_2(E) + \pi, & E > 0, \\ l^2 S_2(E), & E < 0. \end{cases} \tag{291}$$

Since loop integrals of the Berry connection are only uniquely defined modulo 2π, some phase convention has been chosen in the above expressions; such a choice will not matter to the quantization condition, which is only a function of $\exp[i\Omega_j]$ [as justified in Eqs. (292) and (293) below].

The following determinantal equation expresses the condition that the amplitudes $\{A_i\}$ are everywhere single-valued:

$$\begin{aligned} \begin{pmatrix} A_{1E}(0) \\ A_{2E}(0) \end{pmatrix} &= \mathbb{S}(E,l^2) \begin{pmatrix} A_{1E}(1) \\ A_{2E}(1) \end{pmatrix} \\ &= \mathbb{S}(E,l^2) \begin{pmatrix} e^{i\Omega_1} & 0 \\ 0 & e^{i\Omega_2} \end{pmatrix} \begin{pmatrix} A_{1E}(0) \\ A_{2E}(0) \end{pmatrix} \\ &\Rightarrow \det \left[\mathbb{S} \begin{pmatrix} e^{i\Omega_1} & 0 \\ 0 & e^{i\Omega_2} \end{pmatrix} - I \right] \Big|_{E,l^2} = 0. \end{aligned} \tag{292}$$

Employing the expression for the scattering matrix [Eq. (284)] at nonzero E, the determinantal equation may be expressed as

$$0 = 1 + e^{i(\Omega_1+\Omega_2)} - \sqrt{1-\rho^2}(e^{i(\Omega_1+\omega)} + e^{i(\Omega_2-\omega)}). \tag{293}$$

The three phases occurring above may be identified with the phases acquired by a wave packet in traversing three closed Feynman trajectories; e.g., $\Omega_1 + \Omega_2$ corresponds to the figure-of-eight trajectory illustrated in Fig. 9(b). As we have argued generally in Sec. IX C 1, the phase $(\theta_1 - \theta_2)$ in the tunneling matrix element should not affect the quantization condition since it expresses an arbitrary phase difference between the electron and hole pocket. We should see this directly from our case study: the figure-of-eight trajectory includes an electron-to-hole tunneling trajectory [occurring with amplitude $\mathbb{S}_{21} = \rho e^{i(\theta_2-\theta_1)}$], and also the reverse hole-to-electron tunneling trajectory [with amplitude $\mathbb{S}_{12} = -\rho e^{i(\theta_1-\theta_2)}$]. Equation (293) is equivalent to

$$\cos \frac{\Omega_1 + \Omega_2}{2}\Big|_{E,l^2} = \sqrt{1-\rho^2} \cos \left[\frac{\Omega_1 - \Omega_2}{2}\Big|_{E,l^2} + \omega(\bar{\mu}) \right], \tag{294}$$

which we have previously analyzed in Ref. [36].

Here, we focus on resolving the tension originating from a discontinuity of our connection formula at $E = 0$ (see Sec. IX C 2); the upshot is that a simultaneous discontinuity in the Berry phase ensures that the quantization condition remains continuous at $E = 0$. We remind the reader that Eq. (294) has been derived utilizing the connection formula for nonzero E. In the limit $\bar{\mu} \to 0^{\pm}$ (equivalently, $E \to 0^{\pm}$ at finite field), $\rho \to 1$, and Eq. (294) simplifies to

$$\frac{l^2(S_1 + S_2)}{2}\Big|_{E_n^0} = n\pi. \tag{295}$$

Despite our proximity to a band degeneracy, the form of Eq. (295) is reminiscent of an Onsager-Lifshitz-Roth quantization condition for *single-band* magnetotransport; the resultant Landau levels are also locally periodic. We may ascribe this emergent periodicity (in the Landau spectrum) to the periodic motion of a wave packet over the figure-of-eight illustrated in Fig. 9(b). Over one cyclotron period, the wave packet accumulates (a) a trivial Maslov phase from four turning points with vanishing net circulation [cf. Fig. 2(h)], (b) a net π-Berry phase of the two pockets (owing to a pseudospin argument in Fig. 12), and (c) a net π phase from two Landau-Zener tunnelings. The last phase is obtained by multiplying the two off-diagonal elements of the scattering matrix: $[\mathbb{S}(0^{\pm}) = -i\tau_2]$.

At strictly zero energy, exactly the same quantization condition [Eq. (295)] may be derived with the connection formula of Eq. (288). Here, the emergent periodicity (in the Landau specrum) is ascribed to periodic motion over a topologically distinct figure-of-eight (illustrated in Fig. 12). The two figures-of-eight differ in the vicinity of the II-Dirac point: bands cross at $E = 0$, but anticross at $E = 0^{\pm}$. For the crossing figure-of-eight, (a) the scattering matrix is trivially identity, (b) the Maslov phase vanishes [for the same reason described in Fig. 2(h)], and (c) the Berry phase is also trivial, owing to a pseudospin argument given in Fig. 12. One practical implication of this discussion is that the two limiting values of Eq. (294) as $E \to 0^{\pm}$ are equal, so one may as well extend the domain of Eq. (294) to include $E = 0$; this extended quantization condition is then continuous in E.

Another aspect of the $|E| \to 0$ limit is worth discussing: the second-order derivatives (with respect to E) of $l^2(S_1 - S_2)/2$ [occurring in the right argument of Eq. (294)] diverge logarithmically. This divergence is a generic property of hyperbolic curves at the point of intersection [129]; physically stated, it originates from the transition from crossing to anticrossing orbits at the II-Dirac point. This divergence does not lead to any irregularity in the quantization condition, due to a canceling logarithmic divergence of the scattering phase ω [which also occurs in the right argument of Eq. (294)]. It is remarkable that an analogous cancellation of divergences occurs for intraband breakdown. This is exemplified by the quantization condition for the double-well graph [cf. Eq. (202)], where the *first*-order derivative of $l^2(S_1 + S_2)/2$ diverges logarithmically, but is also

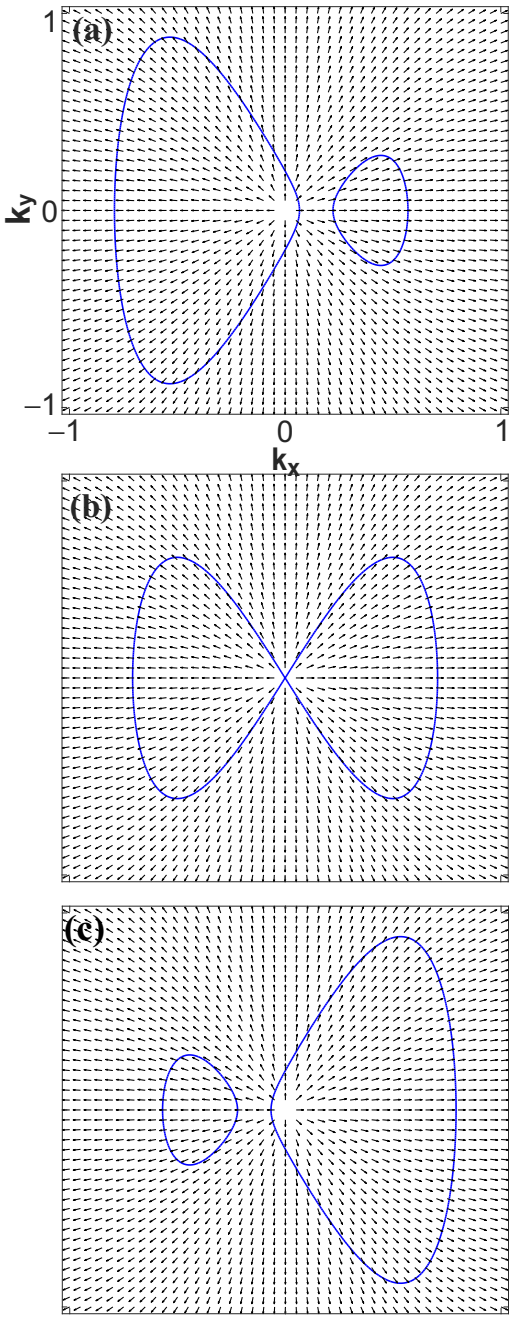

FIG. 12. Expressing $H_0(\boldsymbol{k}) = d_0(\boldsymbol{k}) + d_1(\boldsymbol{k})\gamma_1 + d_3(\boldsymbol{k})\gamma_3$ from Eq. (289), we plot the two-vector $[\boldsymbol{d} = (d_1, d_3)]$ as black arrows over $\boldsymbol{k}$ space, for $E > 0$, $E = 0$, and $E < 0$, respectively. The horizontal (resp. vertical) component of each arrow is proportional to d_3 (resp. d_1). The eigenfunctions of $H_0(\boldsymbol{k})$ are two pseudospinors which are parallel and antiparallel to $\boldsymbol{d}$. The parallel transport condition for a wave function is that its pseudospin remains parallel/antiparallel to $\boldsymbol{d}$ at all times. The Berry phase (ϕ_B) of an orbit may be deduced by evaluating the winding of the pseudospin over said orbit. (a) The pseudospin winds by 2π for the left orbit (hence $\phi_B = \pi$ owing to Berry's argument [20]), but does not wind for the right orbit ($\phi_B = 0$). (b) The pseudospin does not wind for the crossing figure-of-eight orbit, hence $\phi_B = 0$. Take care that as the wave function is parallel-transported across the II-Dirac point, $\boldsymbol{d}$ flips sign but the pseudospin does not.

canceled by the diverging scattering phase (ϕ). These two case studies suggest that quantum tunneling, of both interband and intraband types, tends to smoothen out non-analyticities in the classical action function.

E. Perturbative treatment of quasirandom spectrum

The typical spectrum of quantization conditions with tunneling is quasirandom, unless symmetry imposes commensuration of phases in the quantization condition [36]. The goal of this section is formulate a general perturbation theory to treat quasirandom spectra, and then apply it to our II-Dirac case study.

The general structure of the perturbation theory may be formulated in this manner. Let the quantization condition be expressed as

$$f(E, B; \tau(E, B)) = 0 \to E_n(B), \tag{296}$$

which is an implicit equation for the discrete solutions $E_n(B)$; τ is a tunneling parameter whose functional form depends on the type of breakdown (whether inter- or intraband) and the type of graph. We consider a semiclassical limit of the quantization condition where $\tau(E, B) \to \tau_0$ (a constant), such that

$$f(E, B; \tau_0) = 0 \to E_n^0(B) \tag{297}$$

determines a locally periodic spectrum that forms a Landau fan. Let $\tau = \tau_0 + \delta\tau$ and consider a perturbative expansion in $\delta\tau$. To linear order,

$$0 = f(E, B; \tau_0) + \delta\tau(E, B)\, f_1(E, B) + O(\delta\tau^2); \tag{298}$$

the first-order-corrected energy levels are defined by $E_n^1 = E_n^0 + \delta E_n^1 + O(\delta\tau^2)$ with the assumption $\delta E^1 = O(\delta\tau)$. Substituting E_n^1 into the above equation,

$$\begin{aligned} O(\delta\tau^2) &= f\left(E_n^1, B; \tau_0\right) + \delta\tau f_1\big|_{E_n^0, B} \\ &= \left\{ \delta E_n^1 \frac{\partial f}{\partial E}\bigg|_{\tau_0} + \delta\tau f_1 \right\}_{E_n^0, B}, \\ \Rightarrow \delta E_n^1 &= -\delta\tau \frac{f_1}{(\partial f/\partial E)|_{\tau_0}}\bigg|_{E_n^0, B}. \end{aligned} \tag{299}$$

This equation is valid assuming that $\delta\tau$ is small and slowly varying on the scale of δE_n^1; this should be checked for self-consistency. Equations (297)–(299) have been exemplified for intraband breakdown in Eqs. (205) and (206), and we shall now apply it to our case study of interband breakdown. One key equation [Eq. (300)] in the subsequent section has been presented in Ref. [36], but the reader may benefit from a more detailed discussion.

1. *Case study: Quasirandom spectrum of the II-Dirac graph*

Since no (magnetic) space-group symmetry relates an electron to a hole pocket, the two distinct arguments in the cosine functions of Eq. (294) competitively produce a quasirandom Landau spectrum. In the regime $\bar{\mu} \approx 0$, the dominant trigonometric harmonic $(\Omega_1 + \Omega_2)/2$ determines a semiclassical Landau fan indexed by $n \in \mathbb{Z}$ [cf. Eq. (295)].

The largeness of $|S_1|$ and $|S_2|$ relative to l^{-2} justifies the semiclassical approximation; there is, however, no need for $|S_1 + S_2|$ to be large. In particular, the zeroth ($n = 0$) Landau level is nondispersive and occurs at an energy (E_0^0) where electron and hole pockets are perfectly compensated: $(S_1 + S_2)|_{E_0^0} = 0$; this energy does not necessarily lie at the II-Dirac point [46].

To leading order in $\sqrt{1-\rho^2}$, the correction to the Landau fan is derived in Appendix F 4 a to be

$$\delta E_n^1 = 2(-1)^{n+1}\mathrm{sgn}[E]\frac{\sqrt{1-\rho^2}}{l^2(S_1+S_2)'} \times \sin\left[\omega + \frac{l^2(S_1-S_2)}{2}\right]\Bigg|_{E_n^0}, \tag{300}$$

with the shorthand $O' = \partial O/\partial E$. In particular, the correction to the zeroth Landau level is a sinusoid enveloped by a function $\propto B^{1/2}$:

$$\delta E_0^1 = -2\sqrt{\pi}\frac{v}{\sqrt{w}(u^2-v^2)^{3/4}}\frac{E_0^0}{l(S_1+S_2)'} \times \sin\left[\omega + \frac{l^2(S_1-S_2)}{2}\right]\Bigg|_{E_0^0} + O(\bar{\mu}^{3/2}l^{-2}). \tag{301}$$

For the nonzeroth Landau levels, we show in Appendix F 4 a that the envelop function grows as $B^{1/2}$ at weak field, but eventually crosses over to a $B^{3/2}$ dependence at a scale that depends on the band parameters.

For Eqs. (295)–(300) to be consistent, $\sqrt{1-\rho^2}$ should be small and slowly varying on the scale of δE_n^1. Indeed, the typical scale of variation for $\sqrt{1-\rho^2}$ is $\Delta\bar{\mu} \sim 1$, which implies an energy scale

$$\Delta E \sim \frac{\sqrt{w}(u^2-v^2)^{3/4}}{v}\frac{1}{l}, \tag{302}$$

from the definition of $\bar{\mu}$ in Eq. (271). It follows that

$$\frac{\delta E_n^1}{\Delta E} \sim \frac{v^2}{w(u^2-v^2)^{3/2}}\frac{E_n^0}{(S_1+S_2)'} \tag{303}$$

vanishes for small enough E_n^0. One additional remark is that the typical spacing of the Landau fan is small compared to the energy interval where breakdown is significant:

$$\frac{E_{n+1}^0 - E_n^0}{\Delta E} \sim \frac{2\pi}{l(S_1+S_2)'|_{E_n^0}}\frac{v}{\sqrt{w}(u^2-v^2)^{3/4}} = O(1/l). \tag{304}$$

X. DISCUSSION AND OUTLOOK

We have provided the recipe to cook up quantization rules for a large class of closed orbits: (i) In the absence of breakdown, our rules apply to band subspaces of arbitrary energy degeneracy. (ii) For band subspaces which are nondegenerate at generic wave vectors, we have accounted for intraband breakdown associated with saddle points and interband breakdown associated with conical touching points between two bands (II-Dirac points).

This certainly does not exhaust all types of band touchings: not all point touchings are conical; e.g., the band dispersion around a multi-Weyl point [62] is quadratic in $\boldsymbol{k}$. Not all band touchings occur between two bands; e.g., the spin-one Weyl point [40] is a touching of three bands. If bands are spin-degenerate at generic wave vectors, a touching point is minimally fourfold degenerate, e.g., an over-tilted 3D Dirac point [160,161]. Spin-degenerate orbits may also intersect at fourfold-degenerate saddle points. Moreover, band touchings occur not just at isolated points, but also along lines. The connection formulas in all the above cases are unknown, but we hope that this work lays the groundwork for their future derivation. One necessary ingredient would be an effective Hamiltonian that is valid at any type of band-touching point, as we have derived in Sec. VIII. The connection formula should be derivable by matching the eigenfunctions of this effective Hamiltonian to semiclassical WKB wave functions. For spin-degenerate bands, the matching should be performed for the multicomponent WKB wave function derived in Sec. V.

The quantization rules in this work apply only to closed orbits, and include the complete subleading-in-B correction. For an energy-nondegenerate band in the absence of breakdown, higher-order corrections to the quantization rule have been derived with various methods: beginning from the effective-Hamiltonian formalism, higher-order corrections may be obtained from an equation-of-motion method [25] as well as with WKB methods [52]; alternatively, these corrections may be derived from the zero-field, zero-temperature magnetic response functions [162]. However, a higher-order theory for energy-degenerate bands has not been developed. Finally, it would be interesting to generalize this work to open orbits, i.e., noncontractible orbits that extend across the Brillouin torus.

ACKNOWLEDGMENTS

The authors are grateful to Nicholas Read, Meng Cheng, Titus Neupert, Lee Chinghua, Xi Dai, T. O'Brien, Lukas Muechler, Wang Zhijun, Kadigrobov, Qian Niu, and Chen Fang for various discussions. We especially thank Judith Höller for mathematical advice, and Wang Chong for a critical reading of the manuscript. We acknowledge support by the Yale Postdoctoral Prize Fellowship and NSF DMR Grant No. 1603243.

APPENDIX A: APPENDIX TO SEC. IV: "REVIEW OF EFFECTIVE HAMILTONIAN IN THE ABSENCE OF INTERBAND BREAKDOWN"

1. Introduction to field-modified Bloch functions

We provide a pedagogical introduction to field-modified Bloch functions, and derive a few useful identities which will be used throughout the main text.

Let us first motivate the form of the field-modified Bloch functions in Eq. (55) by an argument [74] involving gauge invariance. Suppose at zero field the energy eigenfunctions are expressed in Bloch form: $\hat{H}_0\psi_{n\boldsymbol{k}'} = \varepsilon_{n\boldsymbol{k}'}\psi_{n\boldsymbol{k}'}$. A zero field is expressible as the curl of a constant vector potential $\boldsymbol{a}_0$, hence by gauge invariance,

$$0 = [\hat{H}[\boldsymbol{a}_0] - \varepsilon_{n\boldsymbol{k}'}]e^{i(\boldsymbol{k}'-\boldsymbol{a}_0)\cdot\boldsymbol{r}}u_{n\boldsymbol{k}'}. \tag{A1}$$

We see that $\boldsymbol{k} = \boldsymbol{k}' - \boldsymbol{a}_0$ is the quantity that determines the change in phase of the wave function under discrete translations, and $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{n\boldsymbol{k}+\boldsymbol{a}_0}$ has energy eigenvalue $\varepsilon_{n\boldsymbol{k}+\boldsymbol{a}_0}$. In a weak field, i.e., for $\boldsymbol{a}(\boldsymbol{r})$ that slowly varies in space, the appropriate basis functions to describe field-induced dynamics within the band n is just $e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{n\boldsymbol{k}+\boldsymbol{a}(\boldsymbol{r})}$ to leading order in the field [cf. Eq. (55)]; to our knowledge, these types of basis functions were first proposed by Zilberman [51].

Let us derive equivalent expression for the field-modified Bloch functions which is more amenable to algebraic manipulations:

$$\begin{aligned} u_{n\boldsymbol{K}^*}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} &= \int d\boldsymbol{r}'\breve{u}_{n\boldsymbol{r}'}e^{-i\boldsymbol{K}^*\cdot\boldsymbol{r}'}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \\ &= \int d\boldsymbol{r}'\breve{u}_{n\boldsymbol{r}'}e^{-i[\boldsymbol{k}+\boldsymbol{a}(-i\nabla_{\boldsymbol{k}})]\cdot\boldsymbol{r}'}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \\ &= \int d\boldsymbol{r}'\breve{u}_{n\boldsymbol{r}'}e^{-i[\boldsymbol{k}+\boldsymbol{a}(\boldsymbol{r})]\cdot\boldsymbol{r}'}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = u_{n,\boldsymbol{k}+\boldsymbol{a}(\boldsymbol{r})}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \end{aligned} \tag{A2}$$

Here, $\breve{u}_{n\boldsymbol{r}'}$ is the Fourier transform of $u_{n\boldsymbol{k}}$, $\boldsymbol{K}$ are the kinetic quasimomentum operators defined in Eq. (47), and the second-to-last equality is valid in the symmetric gauge, where $[\boldsymbol{k}\cdot\boldsymbol{r}', \boldsymbol{a}(-i\nabla_{\boldsymbol{k}})\cdot\boldsymbol{r}'] = 0$. An arbitrary state may be expanded in field-modified Bloch functions as in Eq. (53), which is equivalently expressed as

$$\Psi(\boldsymbol{r}) = \sum_{n\boldsymbol{k}}(u_{n\boldsymbol{K}^*}e^{i\boldsymbol{k}\cdot\boldsymbol{r}})g_{n\boldsymbol{k}} = \sum_{n\boldsymbol{k}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}(u_{n\boldsymbol{K}}g_{n\boldsymbol{k}}), \tag{A3}$$

where $\sum_{\boldsymbol{k}}$ is really a continuous integral. After the above "integration by parts,' the basis functions effectively become operators acting on the wave function $g_{n\boldsymbol{k}}$.

This "integration by parts" formula [Eq. (A3)] was proven in Ref. [24]. Here, we offer a more explicit proof for pedagogy.

Proof. For a constant magnetic field, the vector potential can be written in the linear gauge as $\boldsymbol{a}(\boldsymbol{r}) = \boldsymbol{b}^j r_j$, or equivalently $a_i(\boldsymbol{r}) = b_i^j r_j$. A useful identity in this context is then

$$e^{(1/2)[\pm i\boldsymbol{v}\cdot\boldsymbol{b}^j\nabla_{k_j},\boldsymbol{v}\cdot\boldsymbol{k}]} = e^{\pm(i/2)v_i b_i^j v_j}. \tag{A4}$$

By the Baker-Campbell-Hausdorff lemma,

$$\begin{aligned} e^{-i(\boldsymbol{r}-\boldsymbol{R})\cdot[\boldsymbol{k}+\boldsymbol{a}(\pm i\nabla_{\boldsymbol{k}})]} &= e^{-i(\boldsymbol{r}-\boldsymbol{R})\cdot\boldsymbol{k}}e^{-i(\boldsymbol{r}-\boldsymbol{R})\cdot\boldsymbol{a}(\pm i\nabla_{\boldsymbol{k}})} \\ &\quad\times e^{\mp(i/2)(r_i-R_i)b_i^j(r_j-R_j)}. \end{aligned} \tag{A5}$$

Sandwiching Eq. (A5) in two different ways (also with opposite signs in the argument of $\boldsymbol{a}$), we obtain an identity

$$\begin{aligned} &e^{i\boldsymbol{k}\cdot\boldsymbol{r}}e^{-i(\boldsymbol{r}-\boldsymbol{R})\cdot[\boldsymbol{k}+\boldsymbol{a}(i\nabla_{\boldsymbol{k}})]}e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \\ &= e^{-(i/2)(r_i+R_i)b_i^j(r_j-R_j)} \\ &= e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}e^{-i(\boldsymbol{r}-\boldsymbol{R})\cdot[\boldsymbol{k}+\boldsymbol{a}(-i\nabla_{\boldsymbol{k}})]}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \end{aligned} \tag{A6}$$

which will be used in the following. We apply the Fourier expansions

$$\begin{aligned} g_{n\boldsymbol{k}} &= \sum_{\boldsymbol{R}}\breve{g}_{n\boldsymbol{R}}e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}, \quad \text{and} \\ u_{n\boldsymbol{K}} &= \sum_{\boldsymbol{R}}W_n(\boldsymbol{r}-\boldsymbol{R})e^{-i\boldsymbol{K}\cdot(\boldsymbol{r}-\boldsymbol{R})}, \end{aligned} \tag{A7}$$

to express

$$\begin{aligned} \Psi &= \sum_{n\boldsymbol{k}}g_{n\boldsymbol{k}}(u_{n\boldsymbol{K}^*}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}) \\ &= \sum_{n\boldsymbol{k},\boldsymbol{R},\boldsymbol{R}'}\breve{g}_{n\boldsymbol{R}'}W_n(\boldsymbol{r}-\boldsymbol{R})e^{-i\boldsymbol{k}\cdot\boldsymbol{R}'}e^{-i\boldsymbol{K}^*\cdot(\boldsymbol{r}-\boldsymbol{R})}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \\ &= \sum_{n\boldsymbol{k},\boldsymbol{R},\boldsymbol{R}'}\breve{g}_{n\boldsymbol{R}'}W_n(\boldsymbol{r}-\boldsymbol{R})e^{-i\boldsymbol{k}\cdot\boldsymbol{R}'}e^{-i[\boldsymbol{k}+\boldsymbol{a}(-i\nabla_{\boldsymbol{k}})]\cdot(\boldsymbol{r}-\boldsymbol{R})}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \\ &= \sum_{n\boldsymbol{k},\boldsymbol{R},\boldsymbol{R}'}\breve{g}_{n\boldsymbol{R}'}W_n(\boldsymbol{r}-\boldsymbol{R})e^{-i\boldsymbol{k}\cdot\boldsymbol{R}'}e^{-i[\boldsymbol{k}+\boldsymbol{a}(\boldsymbol{r})]\cdot(\boldsymbol{r}-\boldsymbol{R})}e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \\ &\propto \sum_{\boldsymbol{k}}e^{i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')}. \end{aligned} \tag{A8}$$

The delta function allows us to express the above equation as

$$\begin{aligned} \Psi = N\sum_{n\boldsymbol{k},\boldsymbol{R}}\breve{g}_{n\boldsymbol{R}'}W_n(\boldsymbol{r}-\boldsymbol{R})e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \\ \times e^{-i[\boldsymbol{k}+\boldsymbol{a}(-i\nabla_{\boldsymbol{k}})]\cdot(\boldsymbol{r}-\boldsymbol{R})}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \end{aligned} \tag{A9}$$

We would like to show that this equals

$$\begin{aligned} \sum_{n\boldsymbol{k}}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{n\boldsymbol{K}}g_{n\boldsymbol{k}} = N\sum_{n\boldsymbol{k},\boldsymbol{R}}\breve{g}_{n\boldsymbol{R}}W_n(\boldsymbol{r}-\boldsymbol{R})e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \\ \times e^{-i[\boldsymbol{k}+\boldsymbol{a}(i\nabla_{\boldsymbol{k}})]\cdot(\boldsymbol{r}-\boldsymbol{R})}e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}. \end{aligned} \tag{A10}$$

In deriving the above equality, we have reduced the double summation over $\boldsymbol{R}$ to a single summation, by similar manipulations. Comparing the last two equations, and applying the identity Eq. (A6), we thus derive the desired relation.

2. Equivalent expressions for the orbital magnetic moment

a. Single-band orbital moment

The gauge-independent orbital moment, in the spatial direction $\vec{\alpha}$ ($\alpha = x, y, z$), for a band labeled n, is defined as [cf. Eq. (62)]

$$M(\boldsymbol{k})_n^\alpha = -\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}[\mathfrak{X}^\beta(\Pi^\gamma - v^\gamma)]_{nn}. \tag{A11}$$

Applying the identity Eq. (14),

$$[\mathfrak{X}^\alpha(\Pi^\beta - v^\beta)]_{nn} = \sum_{l;\varepsilon_l\neq\varepsilon_n}\mathfrak{X}_{nl}^\alpha\Pi_{ln}^\beta = \sum_{l;\varepsilon_l\neq\varepsilon_n}\frac{\Pi_{nl}^\alpha\Pi_{ln}^\beta}{i(\varepsilon_n-\varepsilon_l)}, \tag{A12}$$

we derive an equivalent expression

$$\begin{aligned} M(\boldsymbol{k})_n^\alpha &= i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{l;\varepsilon_l\neq\varepsilon_n}\frac{\Pi_{nl}^\beta\Pi_{ln}^\gamma}{\varepsilon_n-\varepsilon_l} \\ &= i\frac{|e|}{2m^2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{l;\varepsilon_l\neq\varepsilon_n}\frac{p_{nl}^\beta p_{ln}^\gamma}{\varepsilon_n-\varepsilon_l}; \end{aligned} \tag{A13}$$

in the last equality, $\boldsymbol{p}_{mn}(\boldsymbol{k})$ is the canonical momentum matrix: $\langle u_{m\boldsymbol{k}}|\hat{p}|u_{n\boldsymbol{k}}\rangle$. Equation (A13) coincides with the correction $(-\boldsymbol{M}\cdot\boldsymbol{B})$ to the energy of a wave packet in Ref. [22]. We offer yet another equivalent expression which is identical in form (but carrying a different name) to that found in the

WKB treatment of coupled-wave [27,28] and coupled-channel equations [100]:

$$\begin{aligned} M(\boldsymbol{k})_n^\alpha &= -i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\langle\partial_\beta u_n|\hat{H}_0(\boldsymbol{k})-\varepsilon_{n\boldsymbol{k}}|\partial_\gamma u_n\rangle \\ &= -i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_m[\langle\partial_\beta u_n|\hat{H}_0(\boldsymbol{k})|u_m\rangle\langle u_m|\partial_\gamma u_n\rangle \\ &\quad -\varepsilon_n\langle\partial_\beta u_n|u_m\rangle\langle u_m|\partial_\gamma u_n\rangle] \\ &= -i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{m;m\neq n}(\varepsilon_m-\varepsilon_n)\langle\partial_\beta u_n|u_m\rangle\langle u_m|\partial_\gamma u_n\rangle \\ &= -i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{m;m\neq n}(\varepsilon_m-\varepsilon_n)\mathfrak{X}^\beta_{nm}\mathfrak{X}^\gamma_{mn} \\ &= i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\sum_{m;m\neq n}\frac{\Pi^\beta_{nm}\Pi^\gamma_{mn}}{\varepsilon_n-\varepsilon_m}. \end{aligned} \tag{A14}$$

Let us compare these expressions to the gauge-dependent moment corresponding to the Berry term in the effective Hamiltonian; i.e., we express $H_1^B=-\tilde{\boldsymbol{M}}\cdot\boldsymbol{B}$ with

$$\begin{aligned} \tilde{M}(\boldsymbol{k})_n^\alpha &= -\frac{|e|}{\hbar c}\epsilon^{\alpha\beta\gamma}[\mathfrak{X}^\beta\boldsymbol{v}^\gamma]_{nn} = -\frac{|e|}{\hbar c}\epsilon^{\alpha\beta\gamma}\mathfrak{X}^\beta_{nn}\partial_\gamma\varepsilon_n \\ &= -\frac{|e|}{\hbar c}\epsilon^{\alpha\beta\gamma}\left[\partial_\gamma\left(\mathfrak{X}^\beta_{nn}\varepsilon_n\right)-i\langle\partial_\gamma u_n|\partial_\beta u_n\rangle\varepsilon_n\right] \\ &= -\frac{|e|}{\hbar c}\epsilon^{\alpha\beta\gamma}\left[-\partial_\beta\left(\mathfrak{X}^\gamma_{nn}\varepsilon_n\right)+i\langle\partial_\beta u_n|\partial_\gamma u_n\rangle\varepsilon_n\right]. \end{aligned} \tag{A15}$$

The total derivative (i.e., the first term in brackets in the above equation) cannot be ignored: it makes this quantity independent of the zero of energy. The sum of the two moments is then

$$\begin{aligned} [M+\tilde{M}](\boldsymbol{k})_n^\alpha &= -i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\langle\partial_\beta u_n|\hat{H}_0(\boldsymbol{k})+\varepsilon_n|\partial_\gamma u_n\rangle \\ &\quad +\frac{|e|}{\hbar c}\epsilon^{\alpha\beta\gamma}\partial_\beta\left(\mathfrak{X}^\gamma_{nn}\varepsilon_n\right). \end{aligned} \tag{A16}$$

For insulators with vanishing Chern number (C_1) in the Brillouin two-torus ($T_\perp$) perpendicular to the field, a first-order-differentiable basis for $u_{n\boldsymbol{k}}$ may be found over $T_\perp$. This implies that $\mathfrak{X}_{nn}$ is continuous over $T_\perp$ (ε_n clearly also satisfies this property), and therefore integrating the total moment over $T_\perp$,

$$\begin{aligned} &\int_{T_\perp}\frac{d^2k}{(2\pi)^2}[M+\tilde{M}](\boldsymbol{k})_n^\alpha \\ &\overset{C_1=0}{=} i\frac{|e|}{2\hbar c}\epsilon^{\alpha\beta\gamma}\int_{T_\perp}\frac{d^2k}{(2\pi)^2}\langle\partial_\beta u_n|\hat{H}_0(\boldsymbol{k})+\varepsilon_n|\partial_\gamma u_n\rangle. \end{aligned} \tag{A17}$$

The right-hand side seems at first sight to depend on the zero of energy, but note that the effect of such a shift is proportional to C_1, which vanishes by assumption. Equation (A17) is identical to the zero-temperature expression obtained for the orbital magnetization using various methods: (a) a Wannier representation for bands was used in Refs. [163,164], (b) quantum-mechanical perturbation theory in Ref. [165], and (c) a Green's function approach in Ref. [166].

b. Orbital magnetic moment for any number of bands

Let us derive equivalent expressions for the single- and multiband orbital magnetic moments, which manifest how they transform under basis changes of the form Eq. (22). The basis transformations we consider preserve both P and Q [recall Eqs. (18) and (19)]; i.e., the unitary V in Eq. (22) is block-diagonal with respect to the decomposition into P and Q. From the simple identities,

$$\begin{aligned} \partial_\alpha P &= \sum_n |u_n\rangle\langle\partial_\alpha u_n| + |\partial_\alpha u_n\rangle\langle u_n|, \\ (\partial_\alpha P)Q &= \sum_{n,\bar{m}}|u_n\rangle\langle\partial_\alpha u_n|u_{\bar{m}}\rangle\langle u_{\bar{m}}|, \end{aligned} \tag{A18}$$

we derive

$$\begin{aligned} [(\Pi^\beta-v^\beta)\mathfrak{X}^\alpha]_{mn} &= i\sum_{\bar{l}}\Pi^\beta_{m\bar{l}}\langle u_{\bar{l}}|\partial_\alpha u_n\rangle \\ &= i[P\hat{\Pi}^\beta Q\partial_\alpha P]_{mn}, \end{aligned} \tag{A19}$$

$$\begin{aligned} [\mathfrak{X}^\alpha(\Pi^\beta-v^\beta)]_{mn} &= i\sum_{\bar{l}}\langle u_m|\partial_\alpha u_{\bar{l}}\rangle\Pi^\beta_{\bar{l}n} \\ &= i[P(\partial_\alpha Q)\hat{\Pi}^\beta P]_{mn}. \end{aligned} \tag{A20}$$

For the single-band orbital moment for band n, the last equality reduces to $i\langle u_n|(\partial_\alpha Q)\hat{\Pi}^\beta|u_n\rangle$, which leads directly to Eq. (62).

APPENDIX B: APPENDIX TO SEC. V: "QUANTIZATION CONDITIONS FOR CLOSED ORBITS WITHOUT BREAKDOWN"

1. Identities for Weyl-symmetrized operators

The following identities may be generalized to nonperiodic functions of $\boldsymbol{k}$ by replacing the Fourier sum with a Fourier integral.

Let $\check{A}_j(\boldsymbol{k})$ be the Fourier transform of $A_j(\boldsymbol{k})=O(l^{-2j})$, and applying the definition of a Weyl-symmetrized operator [cf. Eq. (50)],

$$\begin{aligned} &\sum_{\boldsymbol{R}}\check{A}_j(\boldsymbol{R})e^{i\boldsymbol{K}\cdot\boldsymbol{R}}e^{-i\psi(k_y)} \\ &\quad=\sum_{\boldsymbol{R}}\check{A}_j(\boldsymbol{R})e^{ik_xR_x}e^{-(R_x/l^2)\partial_y+ik_yR_y}e^{-i\psi(k_y)}. \end{aligned} \tag{B1}$$

Applying the Baker-Campbell-Hausdorff identity for a central commutator,

$$\begin{aligned} e^{A+B} &= e^Ae^Be^{(1/2)[B,A]}\Rightarrow e^{-(R_x/l^2)\partial_y+ik_yR_y} \\ &= e^{ik_yR_y}e^{-(R_x/l^2)\partial_y}e^{-iR_xR_y/2l^2}, \end{aligned} \tag{B2}$$

and an identity valid for any function $f(k_y)$,

$$e^{-(R_x/l^2)\partial_y}f(k_y)=f(k_y-R_x/l^2)e^{-(R_x/l^2)\partial_y}, \tag{B3}$$

we derive that Eq. (B1) equals

$$\sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i\boldsymbol{K}\cdot\boldsymbol{R}} e^{-i\psi(k_y)} = \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{ik_xR_x + ik_yR_y - iR_xR_y/2l^2} e^{-i\psi(k_y - R_x/l^2)} e^{-(R_x/l^2)\partial_y} \tag{B4}$$

$$= \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{ik_xR_x + ik_yR_y - iR_xR_y/2l^2} e^{-i\psi(k_y) + i\{\psi'_{-1}(k_y) + \psi'_0(k_y)\}R_x/l^2 - (i/2)\psi''_{-1}(k_y)R_x^2/l^4 + O(l^{-4})} e^{-(R_x/l^2)\partial_y} \tag{B5}$$

$$= e^{-i\psi(k_y)} \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i(k_x + \psi'_{-1}/l^2)R_x + ik_yR_y - iR_xR_y/2l^2} e^{i\psi'_0R_x/l^2 - (i/2)\psi''_{-1}R_x^2/l^4 + O(l^{-4})} e^{-(R_x/l^2)\partial_y} \tag{B6}$$

$$= e^{-i\psi(k_y)} \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i(k_x + \psi'_{-1}/l^2)R_x + ik_yR_y - iR_xR_y/2l^2} \{1 + i\psi'_0R_x/l^2 - (i/2)\psi''_{-1}R_x^2/l^4 + O(l^{-4})\} e^{-(R_x/l^2)\partial_y}. \tag{B7}$$

If $\psi = O(1)$, the above equation particularizes to

$$\sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i\boldsymbol{K}\cdot\boldsymbol{R}} e^{-i\psi(k_y)} = e^{-i\psi(k_y)} \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i\boldsymbol{k}\cdot\boldsymbol{R} - iR_xR_y/2l^2} \{1 + i\psi'_0R_x/l^2 + O(l^{-4})\} e^{-(R_x/l^2)\partial_y}. \tag{B8}$$

Letting the operator Eq. (B7) act on the identity function,

$$\sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i\boldsymbol{K}\cdot\boldsymbol{R}} e^{-i\psi(k_y)} 1 \tag{B9}$$

$$= e^{-i\psi(k_y)} \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i(k_x + \psi'_{-1}/l^2)R_x + ik_yR_y} \{1 + i(iR_x)(iR_y)/2l^2 + \psi'_0(iR_x)/l^2 + (i/2)\psi''_{-1}(iR_x)^2/l^4 + O(l^{-4})\}$$

$$= e^{-i\psi(k_y)} \left\{ A_j(\boldsymbol{k}) + \frac{i}{2l^2}\frac{\partial^2 A}{\partial k_x \partial k_y} + \frac{\psi'_0}{l^2}\frac{\partial A}{\partial k_x} + \frac{i\psi''_{-1}}{2l^4}\frac{\partial^2 A}{\partial^2 k_x} + O(l^{-4-2j}) \right\}_{\boldsymbol{k} \to \boldsymbol{k} + \vec{x}\psi'_{-1}/l^2}. \tag{B10}$$

By similar manipulations, we may derive an identity that is closely analogous to Eq. (B7):

$$f(k_y) = O(1), \quad \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i\boldsymbol{K}\cdot\boldsymbol{R}} f(k_y) = \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i\boldsymbol{k}\cdot\boldsymbol{R} - iR_xR_y/2l^2} \left\{ f(k_y) - \frac{R_x}{l^2} f'(k_y) + O(l^{-4}) \right\} e^{-(R_x/l^2)\partial_y}. \tag{B11}$$

If we let the operator in Eq. (B11) act on $e^{-i\psi}$,

$$\sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i\boldsymbol{K}\cdot\boldsymbol{R}} f(k_y) e^{-i\psi(k_y)} = \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) e^{i\boldsymbol{k}\cdot\boldsymbol{R} - iR_xR_y/2l^2} \left\{ f(k_y) - \frac{R_x}{l^2} f'(k_y) + O(l^{-4}) \right\} e^{-i\psi(k_y - R_x/l^2)}$$

$$= f(k_y) A_j(\boldsymbol{K}) e^{-i\psi(k_y)} + il^{-2} f'(k_y) \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) \frac{\partial e^{i\boldsymbol{k}\cdot\boldsymbol{R}}}{\partial k_x} e^{-iR_xR_y/2l^2} e^{-i\psi(k_y - R_x/l^2)} + O(l^{-4-2j})$$

$$= f(k_y) A_j(\boldsymbol{K}) e^{-i\psi(k_y)} + il^{-2} f'(k_y) e^{-i\psi(k_y)} \sum_{\boldsymbol{R}} \check{A}_j(\boldsymbol{R}) \frac{\partial}{\partial k_x} e^{i(k_x + \psi'_{-1}/l^2)R_x + ik_yR_y} + O(l^{-4-2j})$$

$$= f(k_y) A_j(\boldsymbol{K}) e^{-i\psi(k_y)} + il^{-2} f'(k_y) e^{-i\psi(k_y)} \frac{\partial A}{\partial k_x}\bigg|_{\boldsymbol{k} \to \boldsymbol{k} + \vec{x}\psi'_{-1}/l^2} + O(l^{-4-2j}). \tag{B12}$$

2. Appendix to Sec. V B: "Maslov correction from turning points"

Here we derive the Maslov correction to the single-band quantization conditions from a WKB approach. After reviewing the solution of the Peierls-Onsager Hamiltonian at the turning point in Appendix B 2 a, we derive the first-order-corrected effective Hamiltonian and its solution in Appendix B 2 b. By wave function matching with the Zilberman-Fischbeck functions, we may determine the "reflection phase" (ϕ_r) at each turning point; the sum of all reflection phases is the desired Maslov correction. We pay careful attention to assigning a sense of circulation to each turning point in Appendix B 2 a; this determines the sign of each ϕ_r, which is important to keep track of when we perform the sum $\sum \phi_r$. Finally, in Appendix B 2 c, we estimate the size of the turning region where quantum fluctuations render the Zilberman-Fischbeck wave functions invalid.

a. Review of solution to the Peierls-Onsager Hamiltonian at the turning point

Let us review the Peierls-Onsager solution at the turning point, which was first derived by Zilberman [51]. We assume that the reader has some familiarity with the WKB theory of turning points, and shall keep the review brief. We will go one small step beyond [51] by defining a sense of circulation for

each turning point, which determines the sign of the relative phase between incoming and outgoing WKB solutions.

We assume that the field-free Hamiltonian may be approximated by

$$H_0(\boldsymbol{k}) = E + u_y k_y + \frac{k_x^2}{2m_x}, \tag{B13}$$

with momentum coordinates originating from the turning point at energy E. The constant-energy band contour in the vicinity of the turning point may be split into two sections that touch at the same point; we use $\nu = +\,(-)$ to denote the section to the right (left) of the point:

$$k_x^{\pm}(k_y, E) = \pm\sqrt{-2m_x u_y k_y}. \tag{B14}$$

The sign of $m_x u_y$ determines whether the classical region lies at positive or negative k_y, as illustrated in Figs. 2(a)–2(d).

$H_0(\boldsymbol{k})$ is in Weyl correspondence with the Peierls-Onsager Hamiltonian $H_0(\boldsymbol{K}) := [H_0(\boldsymbol{k})]$; we shall assume the Landau gauge $K_x = k_x + il^{-2}(\partial/\partial k_y)$ and $K_y = k_y$. $H_0(\boldsymbol{K})$ becomes independent of k_x after the the basis transformation $e^{ik_x k_y l^2}$:

$$e^{-ik_x k_y l^2} H_0(\boldsymbol{K}) e^{ik_x k_y l^2} = E + u_y k_y - \frac{1}{2m_x l^4}\frac{\partial^2}{\partial k_y^2}. \tag{B15}$$

We shall separately tackle the two cases corresponding to different signs of $m_x u_y$.

(i) $m_x u_y > 0$; band contour is an inverted parabola $\frown$, i.e., $k_y \sim -k_x^2$.

Equation (B15) is an Airy differential equation with the dimensionless variable $z = (2m_x u_y l^4)^{1/3} k_y$. In the limit $z \ll 0$ (i.e., within the classical region, and sufficiently far from the turning point), and assuming a hard-wall boundary condition, the Airy function has the asymptotic form [167]

$$\lim_{z \ll 0} \mathrm{Ai}(z) = \frac{1}{|z|^{1/4}}\left(e^{i(2/3)|z|^{3/2}+i\pi/4} - e^{-i(2/3)|z|^{3/2}-i\pi/4}\right), \tag{B16}$$

which is then matched with the Zilberman functions [Eq. (79) without the H_1 correction]; some assumption must be made on the band parameters and the field for this matching region to exist [51]. The prefactor $|z|^{-1/4}$ is proportional to $|v_\nu^x|^{-1/2}$ for both $\nu = \pm$. The phase factor in the Zilberman function is

$$e^{-il^2\int_0^{k_y} k_x^{\nu}(t,E)dt}, \tag{B17}$$

with k_y negative in the classical region; we remind the reader that this sign is determined by the sign of $m_x u_y$. From Eq. (B14), $k_x^- \leqslant 0$ and $k_x^+ \geqslant 0$, so we identify

$$\lim_{z \ll 0} \mathrm{Ai}(z) \propto \frac{1}{|v^x(k_y)|^{1/2}}\left(c_+^{\frown}\, e^{-il^2\int_0^{k_y} k_x^+} + c_-^{\frown}\, e^{-il^2\int_0^{k_y} k_x^-}\right),$$
$$c_+^{\frown} := e^{i\pi/4}, \quad c_-^{\frown} := e^{i3\pi/4}. \tag{B18}$$

From Hamilton's equation [Eq. (41)], $\hbar\dot{k}_x = l^{-2}u^y$; $(u_y > 0, m_x > 0)$ thus corresponds to a wave packet circulating in the clockwise sense: $\curvearrowright$ [illustrated in Fig. 2(a)], and $(u_y < 0, m_x < 0)$ to $\curvearrowleft$ [Fig. 2(b)]. For the locally clockwise (resp. locally anticlockwise) trajectory, the relative phase factor between outgoing and incoming WKB wave is then $c_+^{\frown}/c_-^{\frown} = -i$ (resp. $c_-^{\frown}/c_+^{\frown} = +i$); this may be interpreted as the phase acquired by a wave packet as it is reflected (in k_y) from the turning point.

(ii) $m_x u_y < 0$; band contour is an upright parabola $\smile$, i.e., $k_y \sim +k_x^2$.

Equation (B15) is an Airy differential equation with the dimensionless variable $z = -(2|m_x u_y|l^4)^{1/3} k_y$, which differs from the previous case in the sign of z/k_y. The Airy solution in the classical region $(k_y \gg 0, z \ll 0)$ has the same asymptotic form as in Eq. (B16). However, now that k_y is positive in the classical region (with k_x^- and k_x^+ retaining their original signs), we switch the identification of $\nu = \pm$ Zilberman functions in the Airy function:

$$\lim_{z \ll 0} \mathrm{Ai}(z)$$
$$= \frac{1}{|z|^{1/4}}\left(e^{i(2/3)|z|^{3/2}+i\pi/4} - e^{-i(2/3)|z|^{3/2}-i\pi/4}\right)$$
$$\propto \frac{1}{|v^x(k_y)|^{1/4}}\left(c_-^{\smile} e^{-il^2\int_0^{k_y} k_x^-} + c_+^{\smile} e^{-il^2\int_0^{k_y} k_x^+}\right),$$
$$c_-^{\smile} := e^{i\pi/4}, \quad c_+^{\smile} := e^{i3\pi/4}. \tag{B19}$$

A wave packet that circulates the turning point in the locally clockwise sense $(u_y < 0, m_x > 0)$ thus picks up a phase factor $c_-^{\smile}/c_+^{\smile} = -i$ [illustrated in Fig. 2(d)]; the locally anticlockwise wave packet $(u_y > 0, m_x < 0)$ picks up $+i$ [Fig. 2(c)].

b. First-order-corrected wave function at the turning point

To account for H_1 in the above matching procedure, we first need to derive a first-order-corrected effective Hamiltonian $(\mathcal{H} = H_0 + H_1)$ in the turning region. Let us expand H_1 around the turning point as

$$H_1(\boldsymbol{k}) = H_1(\boldsymbol{0}) + H_{1x}k_x + H_{1y}k_y + H_{1xx}k_x^2 + \cdots. \tag{B20}$$

We argue that only the terms which are written explicitly above are relevant to $\mathcal{H}$ in the limit of small field. Indeed, the neglected terms (δH_1) are bounded by their values at the boundary of the turning region: $\delta H_1(\Delta\boldsymbol{k}) = O(l^{-4})$, with our estimates of $\Delta\boldsymbol{k}$ in the above paragraph. One may verify that the explicit terms in Eq. (B20), when evaluated on the boundary, are greater than $O(l^{-4})$. When these explicit terms are added to H_0, the result is an effective Hamiltonian that is identical in form to Eq. (B13):

$$[\mathcal{H}(\boldsymbol{q})] = \mathcal{H}(\boldsymbol{Q}) = E + \tilde{u}_y Q_y + \frac{Q_x^2}{2\tilde{m}_x} + O(l^{-4}), \tag{B21}$$

but is shifted in velocity $\tilde{u}_y = u_y + H_{1y}$, mass $\tilde{m}_x = m_x - 2m_x^2 H_{1xx}$, and the momentum variables

$$q_x = k_x + mH_{1x}, \quad q_y = k_y + H_1(\boldsymbol{0})/u_y \quad \leftrightarrow$$
$$Q_x = K_x + mH_{1x}, \quad Q_y = Q_y + H_1(\boldsymbol{0})/u_y. \tag{B22}$$

We assume $m_x u_y > 0$ in this derivation, which is simply generalized for the other sign. $\mathcal{H}$ may be solved with the same techniques; the Airy eigenfunction may be expressed as a sum of Zilberman functions:

$$f_{\boldsymbol{k}E} = e^{iq_x k_y l^2} \sum_{\nu=\pm} c_\nu^{\frown} \frac{1}{\sqrt{|v_\nu^x|}} e^{-il^2\int_0^{q_y} q_x^\nu(z,E)dz}, \tag{B23}$$

with $c_+^\frown/c_-^\frown = -i$, and q_x^ν describes a section of $\mathcal{H}$ at energy E:

$$0 = \mathcal{H}\left(q_x^\nu(q_y,E),q_y\right) - E \Rightarrow q_x^\pm = \pm\sqrt{-2\tilde{m}_x\tilde{u}_y q_y}. \quad \text{(B24)}$$

This function is related to the zero-field band contour $k_x^\pm = \pm(-2m_x u_y k_y)^{-1/2}$ by

$$q_x^\nu(z,E) - k_x^\nu(z,E) = -\frac{H_{1y}k_y + H_{1xx}\left(k_x^\nu\right)^2}{v_\nu^x} + O(l^{-4}). \quad \text{(B25)}$$

Inserting this, as well as the left-hand side of Eq. (B22), into Eq. (B23), we express f in terms of the original $\boldsymbol{k}$ coordinates and the zero-field band contour:

$$\begin{aligned} f_{\boldsymbol{k}E} = e^{ik_xk_yl^2}e^{im_xH_{1x}k_yl^2}\sum_{\nu=\pm}c_\nu^\frown\frac{1}{\sqrt{\left|v_\nu^x\right|}} \\ \times \exp\Bigg[-il^2\int_0^{k_y}k_x^\nu dz + il^2\int_0^{k_y} \\ \times \left[H_{1y}z + H_{1xx}\left(k_x^\nu\right)^2\right]\frac{dz}{v_\nu^x} \\ - il^2H_1(\mathbf{0})k_x^\nu(k_y,E)/u_y + O(l^{-2})\Bigg]. \end{aligned} \quad \text{(B26)}$$

This complicated expression may be simplified with the identification

$$\begin{aligned} \int_0^{k_y}\frac{H_1^\nu}{v_\nu^x}dz = m_xk_yH_{1x} + \int_0^{k_y}\frac{H_{1y}z + H_{1xx}\left(k_x^\nu\right)^2}{v_\nu^x}dz \\ - k_x^\nu(k_y,E)H_1(\mathbf{0})u_y^{-1} + O(l^{-4}); \end{aligned} \quad \text{(B27)}$$

here, our estimation of $O(l^{-4})$ was made by evaluating the neglected terms at the boundary of the turning region. Therefore, we arrive at

$$f_{\boldsymbol{k}E} = \sum_{\nu=\pm}c_\nu^\frown\frac{1}{\sqrt{\left|v_\nu^x\right|}}e^{-il^2\int_0^{k_y}dz(k_x^\nu - k_x - H_1^\nu/v_\nu^x) + O(l^{-2})}, \quad \text{(B28)}$$

which implies that the incoming and reflected Zilberman-Fischbeck functions are related by the reflection phase factor $e^{i\phi_r} = c_+^\frown/c_-^\frown = -i + O(l^{-2})$. For an analogous result in the coupled-channel equations in nuclear physics, we refer the reader to [100].

c. Estimation of size of the turning region

It is useful to estimate the size of the region in $\boldsymbol{k}$ space $(\Delta k_x \Delta k_y)$, in the vicinity of the turning point, where the Zilberman-Fischbeck wave functions are invalid; equivalently, this is where the asymptotic limits of the Airy functions would not apply; we have called this the turning region. From $z = O(1)$, we obtain $\Delta k_y = O(l^{-4/3})$. The two sections $s_\pm$ of the band contour that meet at the turning point are described by $k_x^\pm = \pm(-2m_xu_yk_y)^{-1/2}$. Combining this with our estimate of Δk_y, we obtain $\Delta k_x = O(l^{-2/3})$; note that $\Delta k_x \Delta k_y = O(l^{-2})$. We may further estimate the length of the semiclassical orbit that lies within the turning region as

$$\begin{aligned} 2\int_0^{\Delta k_x}\sqrt{1 + (dk_y/dk_x)^2}dk_x \\ = 2\int_0^{\Delta k_x}\sqrt{1 + \frac{k_x^2}{(u_ym_x)^2}}dk_x = O(l^{-2/3}). \end{aligned} \quad \text{(B29)}$$

3. Quantization condition for the simplest closed orbit, from conventional means

We review the conventional determination [51] of the quantization conditions without breakdown, through the simplest case study of the closed orbit $\mathfrak{o}$ in Fig. 2(e); it is composed of two edges (labeled $\nu = \pm$) that touch at two turning points. Let us define the wave function in the (K_x,k_y) representation as $f_{\boldsymbol{k}E}$; the quantization condition is the condition of continuity of $f_{\boldsymbol{k}}$ with respect to k_y.

For the interval of k_y within the classical region and sufficiently far from the two turning points, f is the sum of two Zilberman-Fischbeck (ZF) functions which correspond to the two edges: $f_{\boldsymbol{k},E} = \sum_{\nu=\pm}c_\nu g_{\boldsymbol{k}E}^\nu$, with g defined in Eq. (79). To impose continuity, it is convenient to introduce the gauge-transformed wave function

$$\tilde{f}_{k_y,E} := e^{-ik_xk_yl^2}f_{\boldsymbol{k}E} = \sum_{\nu=\pm}c_\nu\left|v_\nu^x(k_y)\right|^{-1/2}a_\nu(k_y),$$

where a_ν are scalar amplitudes which we define for each edge ν as

$$a_\nu(k_y) := e^{-il^2\int[k_x^\nu - H_1^\nu(v_\nu^x)^{-1}]dk_y}. \quad \text{(B30)}$$

As mentioned in Sec. V C 1, the phase $k_xk_yl^2$ is trivially continuous over a closed orbit.

$\tilde{f}_{k_yE}$ may be analytically continued into the turning region, such that its domain extends up to but excludes the turning point; here, the velocity prefactor diverges. The function that facilitates this continuation is the leading asymptotic term of the modified Airy wave function at the turning point, which we derive in Eq. (B28).

By analytic continuation to the top turning point (at wave vector k_{y1}), we arrive at the following expression for

$$\begin{aligned} \tilde{f}_{k_yE} = c_-^\frown\frac{1}{|v_-^x(k_y)|^{1/2}}\frac{a_-(k_y)}{a_-(k_{y1})} + c_+^\frown\frac{1}{|v_+^x(k_y)|^{1/2}}\frac{a_+(k_y)}{a_+(k_{y1})}, \\ \frac{c_+^\frown}{c_-^\frown} = -i. \end{aligned} \quad \text{(B31)}$$

By analytic continuation to the bottom turning point (at wave vector k_{y2}), we obtain a different expression,

$$\begin{aligned} \tilde{f}_{k_yE} = c_-^\smile\frac{1}{|v_-^x(k_y)|^{1/2}}\frac{a_-(k_y)}{a_-(k_{y2})} + c_+^\smile\frac{1}{|v_+^x(k_y)|^{1/2}}\frac{a_+(k_y)}{a_+(k_{y2})}, \\ \frac{c_+^\smile}{c_-^\smile} = +i. \end{aligned} \quad \text{(B32)}$$

The continuity condition is then equivalent to the identity of Eqs. (B31) and (B32). Equating the right-hand side of these two equations and eliminating c_ν, we derive

$$-1 = \frac{a_+(k_{y2})}{a_+(k_{y1})}\frac{a_-(k_{y1})}{a_-(k_{y2})}. \quad \text{(B33)}$$

By reparametrizing a by the timelike parameters $t_\pm$ [cf. Eqs. (98) and (99)], the above condition may be identified with Eq. (101).

APPENDIX C: APPENDIX TO SEC. VI: "SYMMETRY IN THE FIRST-ORDER EFFECTIVE HAMILTONIAN THEORY"

1. Symmetry in Bloch Hamiltonians

The aim of this section is to expand on the review of symmetries in Sec. III C and further derive some identities which will be useful in deriving symmetry constraints on the effective Hamiltonian. These identities all involve cell-periodic functions and their symmetry constraints [cf. Eqs. (C19), (C20), and (C23)].

To begin, let us recall some notation from Sec. III. A cell-periodic function may be expanded as

$$\langle\alpha|u_{n\boldsymbol{k}}\rangle = u_{n\boldsymbol{k}}(\alpha), \quad |u_{n\boldsymbol{k}}\rangle = \sum_\alpha u_{n\boldsymbol{k}}(\alpha)|\alpha\rangle,$$
$$\langle u_{n\boldsymbol{k}}|\alpha\rangle = u_{n\boldsymbol{k}}(\alpha)^*, \quad \langle u_{n\boldsymbol{k}}| = \sum_\alpha u_{n\boldsymbol{k}}(\alpha)^*\langle\alpha|, \tag{C1}$$

where α is a shorthand for $(\boldsymbol{\tau},s)$, with s a spin index and $\boldsymbol{\tau}$ the cell-periodic position coordinate that is defined with the equivalence $\boldsymbol{\tau}\sim\boldsymbol{\tau}+\boldsymbol{R}$ ($\boldsymbol{R}$ being a Bravais-lattice vector). $\sum_\alpha$ should be interpreted as an integration of $\boldsymbol{\tau}$ over the unit cell, in addition to a sum over the spin index σ. The overlap of bra with ket is defined as

$$\langle u|v\rangle = \sum_\alpha u^*(\alpha)v(\alpha), \quad \langle\alpha|\beta\rangle = \delta_{\alpha\beta}, \tag{C2}$$

where $\delta_{\alpha\beta}$ is a shorthand for the product of a Dirac delta function in real space and a Kronecker delta function in spin space. We remark that the final results of this section, and the way they are derived, are essentially unchanged if we interpret α as a discrete label for a basis of Löwdin orbitals [168,169] in tight-binding methods.

Let a symmetry operation g act on the cell-periodic variable as

$$\hat{g}|\alpha\rangle = |\beta\rangle[U_g]_{\beta\alpha}K^{s(g)}, \quad U_g^{-1} = U_g^\dagger, \tag{C3}$$

with $s(g)$ defined in Eq. (25), repeated indices are summed, and K is the complex-conjugation operation that leaves the basis vector invariant:

$$KzK = z^*, \quad K|\alpha\rangle K = |\alpha\rangle, \quad K^2 = I. \tag{C4}$$

To clarify, Eq. (C3) is shorthand for

$$\hat{g}|\alpha\rangle = \begin{cases} |\beta\rangle[U_g]_{\beta\alpha}, & g \text{ unitary}, \\ |\beta\rangle[U_g]_{\beta\alpha}K, & g \text{ antiunitary}. \end{cases} \tag{C5}$$

For example, consider $g=M_x$ as a reflection that maps $x\to -x$, in which case

$$\hat{M}_x|\tau_x,\tau_y,\tau_z,s\rangle = -i|-\tau_x,\tau_y,\tau_z,-s\rangle. \tag{C6}$$

Here, s labels the eigenvalue of spin component S_z; we have used that M_x is a product of a spatial inversion with a twofold rotation about $\vec{x}$: $\hat{M}_x = \mathrm{i}C_{2x} = \mathrm{i}e^{-iJ_x\pi} = \mathrm{i}e^{-iL_x\pi}(-i\sigma_x)$. If g is the spatial translation by $\boldsymbol{R}$, then U_g is the identity operation, due to the just-mentioned equivalence $\boldsymbol{\tau}\sim\boldsymbol{\tau}+\boldsymbol{R}$. The triviality of spatial translations implies that $\{U_gK^{s(g)}|\ g\in G\}$ forms a representation of the point group of the crystal, i.e., the quotient of the full space group G (or magnetic space group) over the subgroup of discrete real-space translations.

Bear in mind that $\hat{g}$ acts on complex numbers as

$$\hat{g}z = K^{s(g)}zK^{s(g)}\hat{g}. \tag{C7}$$

We further define $\hat{g}^*$ by

$$\hat{g}^*|\alpha\rangle = |\beta\rangle[U_g]^*_{\beta\alpha}K^{s(g)}, \tag{C8}$$

such that

$$K\hat{g}K = \hat{g}^*. \tag{C9}$$

The inverse operation is

$$\hat{g}^{-1}|\alpha\rangle = K^{s(g)}|\beta\rangle[U_g]^\dagger_{\beta\alpha}, \tag{C10}$$

from which one may verify $\hat{g}\hat{g}^{-1} = \hat{g}^{-1}\hat{g} = I$. From Eq. (24),

$$\hat{g}e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}\hat{g}^{-1} = e^{[(-1)^{s(g)}i]\boldsymbol{k}\cdot[\breve{g}^{-1}(\hat{\boldsymbol{r}}-\boldsymbol{\delta})]} = e^{i[g\circ\boldsymbol{k}]\cdot(\hat{\boldsymbol{r}}-\boldsymbol{\delta})}. \tag{C11}$$

Consequently, a Bloch function at wave vector $\boldsymbol{k}$, when operated upon by g, transforms with a possibly distinct wave vector

$$\boldsymbol{k}' := g\circ\boldsymbol{k}, \quad \frac{\partial k'_\alpha}{\partial k_\beta} = (-1)^{s(g)}\breve{g}_{\alpha\beta}, \tag{C12}$$

as may be ascertained from

$$\hat{g}e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}|u_{n\boldsymbol{k}}\rangle = e^{i\boldsymbol{k}'\cdot\hat{\boldsymbol{r}}}\hat{g}(\boldsymbol{k})|u_{n\boldsymbol{k}}\rangle. \tag{C13}$$

Here, we have combined $\hat{g}$ and the nonsymmorphic phase factor in Eq. (C11) as

$$\hat{g}(\boldsymbol{k}) := e^{-i(g\circ\boldsymbol{k})\cdot\boldsymbol{\delta}}\hat{g}. \tag{C14}$$

Combining Eq. (C1) with Eq. (C3),

$$\langle\alpha|\hat{g}|u\rangle = \langle\alpha|\sum_\beta K^{s(g)}u(\beta)K^{s(g)}|\delta\rangle[U_g]_{\delta\beta}K^{s(g)}$$
$$= \sum_\beta K^{s(g)}u(\beta)K^{s(g)}[U_g]_{\alpha\beta}K^{s(g)}. \tag{C15}$$

If g is a symmetry of the Hamiltonian, then, applying Eq. (C11),

$$\hat{g}(\boldsymbol{k})\hat{H}_0(\boldsymbol{k})\hat{g}(\boldsymbol{k})^{-1}$$
$$= \hat{g}e^{-i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}\hat{g}^{-1}\hat{H}_0\hat{g}e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}\hat{g}^{-1}$$
$$= e^{-i[g\circ\boldsymbol{k}]\cdot(\hat{\boldsymbol{r}}-\boldsymbol{\delta})}\hat{H}_0e^{i[g\circ\boldsymbol{k}]\cdot(\hat{\boldsymbol{r}}-\boldsymbol{\delta})} = \hat{H}_0(\ g\circ\boldsymbol{k}\ ). \tag{C16}$$

This implies that if $|u_{m\boldsymbol{k}}\rangle$ is an eigenstate of $\hat{H}_0(\boldsymbol{k})$ with eigenvalue $\varepsilon_{m\boldsymbol{k}}$, then $\hat{g}(\boldsymbol{k})|u_{m\boldsymbol{k}}\rangle K^{s(g)}$ belongs to the eigenspace of $\hat{H}_0(\ g\circ\boldsymbol{k}\ )$ with the same energy $\varepsilon_{m\boldsymbol{k}}$; the ambiguity in how we pick basis vectors within each energy eigenspace is expressed as

$$\hat{g}(\boldsymbol{k})|u_{m\boldsymbol{k}}\rangle K^{s(g)} = |u_{n,g\circ\boldsymbol{k}}\rangle\breve{g}(\boldsymbol{k})_{nm}, \tag{C17}$$

where $\breve{g}$ is a "sewing matrix" that is block-diagonal with respect to the energy eigenspaces, such that each distinct block

corresponds to a distinct energy. Equation (C17) is a shorthand for

$$e^{-i(g\circ\boldsymbol{k})\cdot\boldsymbol{\delta}}\sum_{\beta}K^{s(g)}u_{m\boldsymbol{k}}(\beta)K^{s(g)}[U_g]_{\alpha\beta}=u_{n,g\circ\boldsymbol{k}}(\alpha)\breve{g}(\boldsymbol{k})_{nm}. \tag{C18}$$

Equation (C17) implies

$$\hat{g}(\boldsymbol{k})|u_{m\boldsymbol{k}}\rangle K^{s(g)}\breve{g}^{-1}(\boldsymbol{k})_{mn}=|u_{n,g\circ\boldsymbol{k}}\rangle, \tag{C19}$$

from which one obtains

$$\begin{aligned}\left|\nabla^{\alpha}_{\boldsymbol{k}}u_{n,\boldsymbol{k}}\right\rangle\big|_{\boldsymbol{k}\to g\circ\boldsymbol{k}}&=\frac{\partial k_\beta}{\partial k'_\alpha}\nabla^{\beta}_{\boldsymbol{k}}[\hat{g}(\boldsymbol{k})|u_{m\boldsymbol{k}}\rangle K^{s}\breve{g}^{-1}(\boldsymbol{k})_{mn}]\\&=(-1)^{s}\breve{g}_{\alpha\beta}\left(\left|\nabla^{\beta}_{\boldsymbol{k}}u_{m\boldsymbol{k}}\right\rangle K^{s}\breve{g}^{-1}(\boldsymbol{k})_{mn}+|u_{m\boldsymbol{k}}\rangle K^{s}\nabla^{\beta}_{\boldsymbol{k}}\breve{g}^{-1}(\boldsymbol{k})_{mn}\right)-i\delta^{\alpha}\hat{g}(\boldsymbol{k})|u_{m\boldsymbol{k}}\rangle K^{s}\breve{g}^{-1}(\boldsymbol{k})_{mn}.\end{aligned} \tag{C20}$$

In the last equality we substituted $(\partial k_\beta/\partial k'_\alpha)$ with Eq. (C12). Taking the complex conjugate of Eq. (C18),

$$e^{i(g\circ\boldsymbol{k})\cdot\boldsymbol{\delta}}\sum_{\beta}K^{s(g)}u_{m\boldsymbol{k}}(\beta)^{*}K^{s(g)}[U_g]^{\dagger}_{\beta\alpha}=\breve{g}(\boldsymbol{k})^{\dagger}_{mn}u_{n,g\circ\boldsymbol{k}}(\alpha)^{*}. \tag{C21}$$

This may be shortened, with Eq. (C10), as

$$K^{s(g)}\langle u_{m\boldsymbol{k}}|\hat{g}^{-1}(\boldsymbol{k})=\breve{g}(\boldsymbol{k})^{\dagger}_{mn}\langle u_{n,g\circ\boldsymbol{k}}|, \tag{C22}$$

which implies

$$\breve{g}(\boldsymbol{k})_{ml}K^{s(g)}\langle u_{l\boldsymbol{k}}|=\langle u_{m,g\circ\boldsymbol{k}}|\hat{g}(\boldsymbol{k}). \tag{C23}$$

This identity, with Eq. (C19), will be used to derive how the current operator transforms under symmetry in the next subsection.

2. Symmetry constraint on the orbital moment

We detail the derivation of the symmetry constraint of the multiband orbital moment in Eq. (126); we assume the reader is familiar with the outline of the proof sketched in Sec. VI B. As an intermediate step, let us derive Eq. (125), which describes the symmetry constraint on the current operator.

Proof of Eq. (125). The current operator transforms as

$$\hat{g}\hat{\boldsymbol{\Pi}}\hat{g}^{-1}=\hat{g}(-i)[\hat{\boldsymbol{r}},\hat{H}]\hat{g}^{-1}=(-1)^{s(g)}(-i)[\breve{g}^{-1}(\hat{\boldsymbol{r}}-\boldsymbol{\delta}),\hat{H}]=(-1)^{s(g)}\breve{g}^{-1}\hat{\boldsymbol{\Pi}}. \tag{C24}$$

Combining this with Eq. (26), we see that the operator, defined by

$$\hat{\boldsymbol{\Pi}}(\boldsymbol{k})=e^{-i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}\hat{\boldsymbol{\Pi}}e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{r}}}, \tag{C25}$$

transforms as

$$\hat{g}(\boldsymbol{k})\hat{\boldsymbol{\Pi}}(\boldsymbol{k})\hat{g}^{-1}(\boldsymbol{k})=(-1)^{s(g)}\breve{g}^{-1}\hat{\boldsymbol{\Pi}}(g\circ\boldsymbol{k}). \tag{C26}$$

The matrix elements of the velocity operator thus satisfy the following symmetry constraint:

$$\boldsymbol{\Pi}(g\circ\boldsymbol{k})_{mn}=(-1)^{s(g)}\breve{g}\left\langle u_{m,g\circ\boldsymbol{k}}\right|\hat{g}(\boldsymbol{k})\hat{\boldsymbol{\Pi}}(\boldsymbol{k})\hat{g}^{-1}(\boldsymbol{k})\left|u_{n,g\circ\boldsymbol{k}}\right\rangle. \tag{C27}$$

Inserting Eqs. (C19) and (C23) into this expression,

$$\begin{aligned}\boldsymbol{\Pi}(\,g\circ\boldsymbol{k}\,)_{mn}&=(-1)^{s(g)}\breve{g}\breve{g}(\boldsymbol{k})_{ml}K^{s(g)}\langle u_{l\boldsymbol{k}}|\hat{g}^{-1}(\boldsymbol{k})\hat{g}(\boldsymbol{k})\hat{\boldsymbol{\Pi}}(\boldsymbol{k})\hat{g}^{-1}(\boldsymbol{k})\hat{g}(\boldsymbol{k})\left|u_{a\boldsymbol{k}}\right\rangle K^{s(g)}\breve{g}^{-1}(\boldsymbol{k})_{an}\\&=(-1)^{s(g)}\breve{g}K^{s(g)}[\breve{g}^{*}\boldsymbol{\Pi}\breve{g}^{t}]_{mn}K^{s(g)}\big|_{\boldsymbol{k}}.\end{aligned} \tag{C28}$$

■

In the degenerate subspace projected by P, let us define $\varepsilon=\varepsilon_m$ for $m\in\{1,\ldots,D\}$; $\boldsymbol{k}$ dependence is implicit in this and the following notations. Combining Eq. (C28) with Eq. (67),

$$(-i)\epsilon_{abc}\sum_{\bar{n}}\frac{\Pi^{b}_{m\bar{n}}\Pi^{c}_{\bar{n}l}}{\varepsilon-\varepsilon_{\bar{n}}}\bigg|_{g\circ\boldsymbol{k}}=(-i)\epsilon_{abc}\breve{g}_{bd}\breve{g}_{ce}\sum_{\bar{n}}\frac{K^{s(g)}[\breve{g}^{*}\Pi^{d}\breve{g}^{T}]_{m\bar{n}}[\breve{g}^{*}\Pi^{e}\breve{g}^{T}]_{\bar{n}l}K^{s(g)}\big|_{\boldsymbol{k}}}{(\varepsilon-\varepsilon_{\bar{n}})\big|_{g\circ\boldsymbol{k}}}. \tag{C29}$$

Applying the Levi-Civita identity (for an orthogonal matrix satisfying $R^{T}=R^{-1}$)

$$\det[R^{T}]\,\epsilon_{lmn}=\epsilon_{abc}\,R^{T}_{la}\,R^{T}_{mb}\,R^{T}_{nc}\Rightarrow R_{al}\det[R]\,\epsilon_{lmn}=\epsilon_{abc}\,R_{bm}\,R_{cn}, \tag{C30}$$

and the reality of $\breve{g}$, we derive that the left-hand side of Eq. (C29) equals

$$(-1)^{s(g)}\det[\breve{g}]\breve{g}_{ab}K^{s(g)}(-i)\epsilon_{bde}\sum_{\bar{n}}\frac{[\breve{g}^{*}\Pi^{d}\breve{g}^{T}]_{m\bar{n}}[\breve{g}^{*}\Pi^{e}\breve{g}^{T}]_{\bar{n}l}\big|_{\boldsymbol{k}}}{(\varepsilon-\varepsilon_{\bar{n}})\big|_{g\circ\boldsymbol{k}}}K^{s(g)}. \tag{C31}$$

Let us introduce new labels $\bar{n} := (a',a'')$, such that a' labels the *distinct* energy eigenvalues, and a'' labels an arbitrarily chosen basis in the finite-dimensional subspace corresponding to energy $\varepsilon_{a'}$. We see in this labeling that $\varepsilon_{n\boldsymbol{k}}$ does not depend on a'', so we may shorten $\varepsilon_{\bar{n}\boldsymbol{k}} \to \varepsilon_{a'\boldsymbol{k}}$. Moreover, since the symmetry commutes with the Hamiltonian, $\varepsilon_{a'\boldsymbol{k}} = \varepsilon_{a',g\circ\boldsymbol{k}}$. Therefore, Eq. (C31) simplifies to

$$(-1)^{s(g)}\det[\breve{g}]\breve{g}_{ab}K^{s(g)}(-i)\epsilon_{bde}\sum_{a';\varepsilon_{a'}\neq\varepsilon}\frac{1}{(\varepsilon-\varepsilon_{a'})}\sum_{a''}[\breve{g}^*\Pi^d\breve{g}^T]_{m,(a',a'')}[\breve{g}^*\Pi^e\breve{g}^T]_{(a',a''),l}\Big|_{\boldsymbol{k}}K^{s(g)}. \tag{C32}$$

Since $\breve{g}$ is block-diagonal in the index a', the above equation may be expressed as

$$(-1)^{s(g)}\det[\breve{g}]\breve{g}_{ab}K^{s(g)}(-i)\epsilon_{bde}\sum_{a';\varepsilon_{a'}\neq\varepsilon}\frac{1}{(\varepsilon-\varepsilon_{a'})}\sum_{a'',b'',c''}[\breve{g}^*\Pi^d]_{m,(a',b'')}\breve{g}^T_{(a',b''),(a',a'')}\breve{g}^*_{(a',a''),(a',c'')}[\Pi^e\breve{g}^T]_{(a',c''),l}\Big|_{\boldsymbol{k}}K^{s(g)}.$$

Since each block diagonal of $\breve{g}$, corresponding to an energy subspace, is unitary, the above equation reduces to

$$\begin{aligned}&(-1)^{s(g)}\det[\breve{g}]\breve{g}_{ab}K^{s(g)}(-i)\epsilon_{bde}\sum_{a';\varepsilon_{a'}\neq\varepsilon}\frac{1}{(\varepsilon-\varepsilon_{a'})}\sum_{b'',c''}[\breve{g}^*\Pi^d]_{m,(a',b'')}\delta_{b'',c''}[\Pi^e\breve{g}^T]_{(a',c''),l}\Big|_{\boldsymbol{k}}K^{s(g)}\\&=(-1)^{s(g)}\det[\breve{g}]\breve{g}_{ab}K^{s(g)}(-i)\epsilon_{bde}\sum_{a';\varepsilon_{a'}\neq\varepsilon}\frac{1}{(\varepsilon-\varepsilon_{a'})}\sum_{a''}[\breve{g}^*\Pi^d]_{m,(a',a'')}[\Pi^e\breve{g}^T]_{(a',a''),l}\Big|_{\boldsymbol{k}}K^{s(g)}.\end{aligned} \tag{C33}$$

Restoring the usual labeling, we conclude that the left-hand side of Eq. (C29) equals

$$(-1)^{s(g)}\det[\breve{g}]\breve{g}_{ab}K^{s(g)}(-i)\epsilon_{bcd}\sum_{\bar{n}}\frac{[\breve{g}^*\Pi]^c_{m\bar{n}}[\Pi\breve{g}^T]^d_{\bar{n}l}}{\varepsilon-\varepsilon_{\bar{n}}}\Big|_{\boldsymbol{k}}K^{s(g)}, \tag{C34}$$

from which follows Eq. (126).

3. Appendix to symmetry of the first-order effective Hamiltonian

Let us analyze the symmetry constraints on the (a) Roth, (b) Zeeman, and (c) Berry terms in the first-order effective Hamiltonian [recall their definitions in Eq. (64)], in that order. The final goal is to derive Eq. (128).

(a) For Bloch electrons immersed in a field parallel to $\vec{z}$, $H_1^R = -B^z M^z$. For symmetries of semiclassical orbits (defined precisely in Sec. VI A), Eqs. (116) and (117) inform us that $[\breve{g}\boldsymbol{M}]^z = (-1)^{t(g)}M^z = (-1)^{u(g)}\det[\breve{g}]M^z$, and therefore Eq. (126) particularizes to

$$M^z|_{g\circ\boldsymbol{k}} = (-1)^{s(g)+u(g)}\,\breve{g}\,K^{s(g)}\,M^z\,K^{s(g)}\,\breve{g}^{-1}|_{\boldsymbol{k}}, \tag{C35}$$

with $u(g)\in\mathbb{Z}_2$ defined in Eq. (117).

(b) For symmetries in spin-orbit-coupled systems, we would like to demonstrate that $H_1^Z \propto B\sigma^z$ is constrained similarly to Eq. (C35):

$$\sigma^z|_{g\circ\boldsymbol{k}} = (-1)^{s(g)+u(g)}\,\breve{g}\,K^{s(g)}\,\sigma^z\,K^{s(g)}\,\breve{g}^{-1}|_{\boldsymbol{k}}, \tag{C36}$$

where $(\hbar/2)\sigma^z_{mn}(\boldsymbol{k}) = (\hbar/2)\langle u_{m\boldsymbol{k}}|\hat{\sigma}_z|u_{n\boldsymbol{k}}\rangle$ is the spin-half matrix defined in Eq. (17). We already know how the cell-periodic functions transform under symmetry [cf. Eq. (29)], so what remains is to determine how $\hat{\sigma}_z$ transforms under a symmetry of the orbit. For this purpose, the decomposition in Eq. (121) is useful in deriving

$$\hat{g}^{-1}\hat{\sigma}^z\hat{g} = (-1)^{s(g)+u(g)}\hat{\sigma}^z. \tag{C37}$$

Indeed, among the factors written on the right-hand side of Eq. (121), only time reversal (if present) and $\mathfrak{r}_x$ (if present) flips the z component of spin. Combining Eq. (C37) with Eq. (29), we then obtain Eq. (C36).

(c) The Berry term $H_1^B = l^{-2}\epsilon_{\alpha\beta}\mathfrak{X}^\beta v^\alpha$. Combining Eq. (130) with the constraint on the band velocity in Eq. (118),

$$\begin{aligned}\epsilon_{\alpha\beta}\mathfrak{X}^\beta v^\alpha|_{g\circ\boldsymbol{k}} - \epsilon_{\alpha\beta}\delta^\beta v^\alpha|_{g\circ\boldsymbol{k}} &= (-1)^s\epsilon_{\alpha\beta}\breve{g}_{\alpha\mu}\breve{g}_{\beta\nu}\big(\breve{g}K^s\mathfrak{X}^\nu K^s\breve{g}^{-1} + i(-1)^s\,\breve{g}\nabla^\nu_{\boldsymbol{k}}\breve{g}^{-1}\big)v^\mu\big|_{\boldsymbol{k}}\\&= (-1)^{s+u}\epsilon_{\alpha\beta}\big(\breve{g}^{-1}K^s\mathfrak{X}^\beta K^s\breve{g}^{-1} + i(-1)^s\,\breve{g}\nabla^\beta_{\boldsymbol{k}}\breve{g}^{-1}\big)v^\alpha\big|_{\boldsymbol{k}}.\end{aligned}$$

The net result of (a)–(c) is Eq. (128).

4. Topological obstruction to symmetry covariance of H_1

Supposing H_0 is g-symmetric, does a basis (for the cell-periodic functions) exist where H_1 transforms covariantly under g, for all $\boldsymbol{k}$ in the Brillouin torus? This section is a self-constrained exposition on the possible obstructions to symmetry covariance in topologically nontrivial band subspaces. The existence of a topological obstruction is suggested by the observation in paragraph (ii) of in Sec. VI C: the source of non-covariance is the Berry term. The Berry curvature is a measure of the "twisting" of the filled-band wave functions in $\boldsymbol{k}$

space; it is known that topologically nontrivial band structures exist whose curvature cannot be made to vanish [170].

We support this claim with a few case studies in the following subsections; a recurrent theme in these case studies is that the effective Hamiltonian of a symmetry-protected topological phase transforms anomalously (i.e., non-covariantly) under the symmetry in question. Our last case study has only the U(1) symmetry of charge conservation, and we will show that the effective Hamiltonian for a nontrivial Chern band transforms anomalously under a gauge transformation.

Before beginning properly, let us introduce a terminology. Supposing the second and third terms in Eq. (128) were absent; we say that H_1 transforms covariantly under the symmetry (resp. antisymmetry) g if $(-)^{s(g)+u(g)} = +1$ (resp. -1). In simple words, g is referred to as an antisymmetry of H_1 if it inverts the sign of H_1.

a. Wigner-Dyson class AII

Let us exemplify this claim with gapped band subspaces in the symmetry class AII in two [86,171–174] or three [175–178] spatial dimensions—spin-orbit-coupled, with the time-reversal symmetry satisfying $\hat{T}^2 = -I$; no assumption is made presently about the spatial symmetries; however we will elaborate on their roles in the next two sections (Appendices C 4 b and C 4 c). We have used the word "gapped" liberally to describe band subspaces which are energetically separated from all other bands at each wave vector in the Brillouin torus; this would include indirect-gap systems with nonvanishing Fermi lines or surfaces. It is well known that gapped band subspaces, in either 2D or 3D, are classified by a strong $\mathbb{Z}_2$ invariant [179]; we shall refer to the nontrivial phase (in both 2D and 3D) as the $\mathbb{Z}_2$-topological band.

In this context, we would like to define the effective Hamiltonian $H_1(\boldsymbol{k})$ over the entire torus; it is minimally a two-band Hamiltonian due to Kramers degeneracy at $\boldsymbol{k}^{(i)}$. Let us ask if H_1 may transform covariantly under the antisymmetry T; the non-covariant term on the right-hand side of Eq. (128) vanishes if either (i) $\boldsymbol{v}(\boldsymbol{k})$ can be made to vanish everywhere, i.e., the band(s) in P have a flat dispersion, or (ii) the sewing matrix corresponding to T, as defined by

$$\breve{T}(\boldsymbol{k})_{mn} = \langle u_{m,-\boldsymbol{k}}|\hat{T}|u_{n\boldsymbol{k}}\rangle K, \tag{C38}$$

can be made to be independent of $\boldsymbol{k}$. We disregard the implausibly fine-tuned scenario where the non-covariant term vanishes without satisfaction of (i) or (ii). For a trivial band subspace, we argue that an adiabatic continuation exists to a lattice of inert atoms, where both $\boldsymbol{v}(\boldsymbol{k}) = 0$ and $\breve{T}$ reduces to a $\boldsymbol{k}$-independent matrix which represents time reversal in the basis of Löwdin orbitals. Let us then consider (i) and (ii) in the context of a topological band subspace.

(i) For a $\mathbb{Z}_2$-topological band, $\boldsymbol{v}(\boldsymbol{k})$ cannot everywhere be zero if the associated tight-binding Hamiltonian has local (strictly short-ranged) hoppings [180]. In fact, the impossibility of a strictly short-ranged, flat-band Hamiltonian is more generally true for all of the strong topological band subspaces in larger-than-one spatial dimensions; this was first proven for class A in 2D [181], and then extended to the tenfold symmetry classes [180]. This rigorous result suggests that if a strictly-flat-band Hamiltonian exists for a $\mathbb{Z}_2$-topological band, it is likely to be a highly optimized scenario [182] which is challenging to realize in both theoretical and experimental laboratories. We know only of one model [183] (of a strong 2D topological insulator) with exactly flat bands [184]; the hopping elements here decay as a Gaussian. We henceforth assume that $\boldsymbol{v}(\boldsymbol{k})$ is a nonconstant function, which is the case of interest in almost all applications.

(ii) Given that the band is not flat, we are led to investigate the momentum dependence of the sewing matrix for a topological band. One expression for the strong $\mathbb{Z}_2$ invariant, in both 2D and 3D, involves the even-dimensional sewing matrix $\breve{T}$ defined in Eq. (C38). Since this matrix is skew-symmetric at any inversion-invariant wave vector $\boldsymbol{k}^{(i)}$, we might evaluate the quantity

$$\delta_i = \sqrt{\det[\breve{T}(\boldsymbol{k}^{(i)})]}/\mathrm{Pf}[\breve{T}(\boldsymbol{k}^{(i)})] = \pm 1, \tag{C39}$$

with $\mathrm{Pf}[\cdot]$ denoting the Pfaffian of $[\cdot]$. The product of δ_i over all $\boldsymbol{k}^{(i)}$ (numbering four in 2D, and eight in 3D) is the strong $\mathbb{Z}_2$ invariant, which equals $+1$ (-1) in correspondence with the trivial (topological) phase [175]; this definition implicitly assumes the continuity of the cell-periodic functions over the Brillouin torus. If the sewing matrix were constant over this torus, an immediate implication is $\prod_i \delta_i = +1$; alternatively stated, for the $\mathbb{Z}_2$-topological band, there is an obstruction to defining a constant sewing matrix; consequently, a non-flat-band H_1 must transform non-covariantly under the antisymmetry T.

b. Class AII with spatial inversion symmetry

Even for $\mathbb{Z}_2$-topological band subspaces, it is generically not true that a topological obstruction exists for all symmetries of the system. To exemplify this claim, let us consider a $\mathbb{Z}_2$-topological band (in 2D or 3D) with spatial inversion symmetry i. The spacetime inversion $T\mathrm{i}$ $(s = 1, u = 0)$ acts as an antisymmetry on the first-order effective Hamiltonian. For simplicity, we assume that bands are twofold degenerate everywhere on the Brillouin torus, and hence H_1 is a two-band Hamiltonian. We then ask if H_1 transforms covariantly under the antisymmetry $T\mathrm{i}$, or equivalently, if the corresponding sewing matrix (denoted as $\breve{T}_\mathrm{i}$) can be made constant over the torus. An algorithm for this has been proposed in Ref. [176], which is plausibly valid in the $\mathbb{Z}_2$-nontrivial phase (i.e., with $\prod_i \delta_i = -1$). Assuming such a gauge is found, $T\mathrm{i}$ symmetry then imposes the covariant antisymmetry condition:

$$H_1(\boldsymbol{k}) = -\breve{T}_\mathrm{i} H_1^*(\boldsymbol{k}) \breve{T}_\mathrm{i}^{-1}, \quad \nabla_{\boldsymbol{k}} \breve{T}_\mathrm{i} = 0, \tag{C40}$$

which follows from Eq. (128) for a constant sewing matrix. Evaluating the trace on both sides of Eq. (C40), we further deduce that H_1 is traceless:

$$\mathrm{Tr}[H_1(\boldsymbol{k})] = 0 \Rightarrow H_1(\boldsymbol{k}) = -\mu_B \tau_i B_j \xi_{ij}(\boldsymbol{k}); \tag{C41}$$

τ_i here are Pauli matrices describing an effective spin, which is generally distinct from the free spin due to spin-orbit coupling. A heuristic argument for the tracelessness of H_1 in Eq. (C41) already exists in the literature [25,31,74]. Here, we have clarified that H_1 is traceless only in a special basis where the

sewing matrix for $T\mathfrak{i}$ is constant; consequently, the unitary generated by H_1 [cf. Eq. (74)] has unit determinant, and the eigenphases of this unitary satisfy $\lambda_1 = -\lambda_2 \mod 2\pi$. We remind the reader that λ_a enter the multiband quantization conditions in Eq. (73), and also the condition for dHvA oscillations in Eq. (115).

The reader may be unsatisfied that the above conclusions relied on the existence of a special gauge. A more general proof of $\lambda_1 = -\lambda_2$ is provided in Sec. VI D [see the paragraph surrounding Eq. (146)].

c. Topological band subspaces protected by crystalline symmetries

One next case study demonstrates that a topological obstruction may exist for band subspaces, where the obstruction is protected solely by crystalline symmetries. 3D insulators having an improper spatial symmetry (i.e., $\det[\check{g}] = -1$) have a quantized magnetoelectric response [69]; i.e., the θ angle occurring in the axion Lagrangian [185] is symmetry-fixed to 0 or π. For inversion-symmetric (i) bands, $\theta/2\pi$ may be expressed as half the winding number of $\check{\mathfrak{i}}$ (the sewing matrix for $\mathfrak{i}$) [68,186]:

$$\frac{\theta}{2\pi} = -\frac{1}{48\pi^2}\int d^3\boldsymbol{k}\epsilon_{\alpha\beta\gamma}\mathrm{Tr}[(\check{\mathfrak{i}}\nabla_\alpha\check{\mathfrak{i}}^{-1})(\check{\mathfrak{i}}\nabla_\beta\check{\mathfrak{i}}^{-1})(\check{\mathfrak{i}}\nabla_\gamma\check{\mathfrak{i}}^{-1})]. \tag{C42}$$

$\theta = \pi$ thus implies a topological obstruction against H_1 transforming covariantly under $\mathfrak{i}$.

d. Wigner-Dyson class A

Having described the anomalous symmetry transformation of H_1 for symmetry-protected topological phases, we might ask if there is an analogous topological obstruction for charge-conserving band subspaces having no other symmetries; they fall into Wigner-Dyson class A. Even though the Bloch Hamiltonian is completely unconstrained, one always has, at the basic level, a "gauge symmetry," which reflects the ambiguity in how we label our bands in P; a gauge transformation such as in Eq. (22) might be viewed as a "do-nothing" symmetry operation.

Class-A band subspaces in 2D are classified by the TKNN invariant [117], or equivalently, the first Chern number (C_1). We would like to show that a nonzero C_1 necessarily implies that H_1 is not gauge-covariant. Indeed, it was already noted in Eq. (61) that the Berry term $l^{-2}\epsilon_{\alpha\beta}\mathfrak{X}^\beta v^\alpha$ generally results in a loss of covariance for any band subspace, trivial or nontrivial; the Roth and Zeeman terms are gauge-invariant (resp. -covariant) in the one-band (resp. multiband) case. We are led to ask whether H_1^B can be made to vanish by basis transformations within P. For a band with a generic dispersion, this amounts to asking whether there exists a gauge where $\mathfrak{X}(\boldsymbol{k}) = 0$ at each $\boldsymbol{k}$; this gauge does not exist if the Berry curvature $\mathcal{F}^z(\boldsymbol{k}) := \nabla_{\boldsymbol{k}} \times \mathrm{Tr}[\mathfrak{X}] \neq 0$. Since the net Berry curvature for a Chern band is nonzero, we conclude that H_1 (for a generically dispersing Chern band) must transform non-covariantly [187].

APPENDIX D: APPENDIX TO SEC. VII: "INTRABAND BREAKDOWN"

1. Derivation of the intraband scattering matrix

a. Review of connection formula in the lowest order

Let us pedagogically review the derivation of the connection formula in the lowest order in l^{-2}, with the eventual goal of generalizing the formula to the next order (to be carried out in the next subsection). The lowest-order problem was first studied by Azbel [11,109] and has reappeared in similar contexts [130,131,188], as more generally reviewed in Ref. [111].

The Hamiltonian in the breakdown region is approximated by the Peierls-Onsager Hamiltonian, which is in Weyl correspondence with Eq. (171):

$$H_0(\boldsymbol{K}) = \frac{K_x^2}{2m_1} - \frac{K_y^2}{2m_2}, \quad [H_0(\boldsymbol{K}) - E]f_{kE} = 0. \tag{D1}$$

We have further defined f_{kE} as the eigenfunction corresponding to eigenvalue E. Working in the Landau electromagnetic gauge [recall Eq. (77)], we perform a gauge transformation

$$f = e^{iq_xk_yl^2}\bar{f}; \tag{D2}$$

the resultant differential equation for $\bar{f}$ becomes independent of k_x, and is equivalent to the time-independent Schrödinger equation for a particle in an inverted parabolic potential, with coordinate k_y, as was first studied by Kemble [132]. Introducing the dimensionless variable

$$z = e^{-i\pi/4}(k_yl)\left(\frac{4m_1}{m_2}\right)^{1/4}, \tag{D3}$$

we obtain a Weber differential equation [189]

$$\left(\partial_z^2 - \tfrac{1}{4}z^2 + i\mu\right)\bar{f} = 0, \tag{D4}$$

with μ defined in Eq. (174). From Eq. (12.2.2) in Ref. [189], the solutions are linear combinations of two independent parabolic cylinder functions (PCFs):

$$\bar{f}(z) = \bar{c}_{\nearrow E}U(-i\mu,z) + \bar{c}_{\swarrow E}U(-i\mu,-z), \tag{D5}$$

with newly introduced coefficients $\bar{c}_{\nu E}$ that are to be determined. In the limit $z \gg 1$, the PCFs may be matched with the Zilberman functions [Eq. (179) without the H_1 correction]. We will assume some conditions on the zero-field band parameters, such that beyond-quadratic terms in $H_0(\boldsymbol{K})$ can still be neglected in this matching region. Employing the asymptotic expansion in 12.9.1 and 12.9.3 of [189],

(i) $k_y \to +\infty$,

$$\bar{f} \to \left[c^{(0)}_{\nearrow E}e^{\pi\mu/2} - i\,c^{(0)}_{\swarrow E}e^{-\pi\mu/2}\right] \times e^{i\mu\log|\mu|-i\mu}\frac{\Gamma(1/2-i\mu)}{\sqrt{2\pi}}\varpi + c^{(0)}_{\swarrow E}\varpi^*,$$

(ii) $k_y \to -\infty$,

$$\bar{f} \to \left[-i\,c^{(0)}_{\nearrow E}e^{-\pi\mu/2} + c^{(0)}_{\swarrow E}e^{\pi\mu/2}\right] \times e^{i\mu\log|\mu|-i\mu}\frac{\Gamma(1/2-i\mu)}{\sqrt{2\pi}}\varpi + c^{(0)}_{\nearrow E}\varpi^*, \tag{D6}$$

where we have introduced the coefficients $\{c^{(0)}_{\nu E}\}$ which differ from $\{\bar{c}_{\nu E}\}$ only by a ν-independent proportionality constant;

we have additionally defined

$$\varpi(k_y,E) = e^{-z^2/4+i\mu/2}\left[\frac{2|k_y|}{b}\right]^{i\mu-1/2},$$
$$\varpi^*(k_y,E) = e^{z^2/4-i\mu/2}\left[\frac{2|k_y|}{b}\right]^{-i\mu-1/2}, \quad \text{(D7)}$$

which may be identified with the Zilberman functions in the limit $|k_y| \gg |b|$ [recall the definition of b in Eq. (173)]:

$$g^{\nwarrow +}_{\boldsymbol{k}E} \overset{k_y\gg|b|}{\propto} e^{ik_xk_yl^2}\varpi(k_y,E),$$
$$g^{\searrow -}_{\boldsymbol{k}E} \overset{k_y\ll-|b|}{\propto} e^{ik_xk_yl^2}\varpi(k_y,E),$$
$$g^{\swarrow +}_{\boldsymbol{k}E} \overset{k_y\gg|b|}{\propto} e^{ik_xk_yl^2}\varpi^*(k_y,E),$$
$$g^{\nearrow -}_{\boldsymbol{k}E} \overset{k_y\ll-|b|}{\propto} e^{ik_xk_yl^2}\varpi^*(k_y,E). \quad \text{(D8)}$$

We emphasize that these identifications are made for uniquely defined Zilberman functions, for which the lower limits of the classical action integrals are specified as in Eq. (179) (with $H_1 = 0$). Following the discussion surrounding Eqs. (175)–(180), we may then identify

$$c^{(0)}_{\nwarrow E} = \left[c^{(0)}_{\nearrow E}\, e^{\pi\mu/2} - i\, c^{(0)}_{\swarrow E}\, e^{-\pi\mu/2}\right] \times e^{i\mu\log|\mu|-i\mu}\,\frac{\Gamma(1/2-i\mu)}{\sqrt{2\pi}},$$
$$c^{(0)}_{\searrow E} = \left[-i\, c^{(0)}_{\nearrow E}\, e^{-\pi\mu/2} + c^{(0)}_{\swarrow E}\, e^{\pi\mu/2}\right] \times e^{i\mu\log|\mu|-i\mu}\,\frac{\Gamma(1/2-i\mu)}{\sqrt{2\pi}}, \quad \text{(D9)}$$

which can be expressed as a matrix equation relating incoming to outgoing states

$$\begin{pmatrix} c^{(0)}_{\nwarrow E} \\ c^{(0)}_{\searrow E}\end{pmatrix} = \mathbb{S}^{(0)}(E,k_z)\begin{pmatrix} c^{(0)}_{\nearrow E} \\ c^{(0)}_{\swarrow E}\end{pmatrix}, \quad \text{(D10)}$$

with the lowest-order scattering matrix defined in Eq. (182). To summarize the results of this review, the eigenfunctions of Eq. (D1) in the limit $k_y \to \pm\infty$ are

$$f^{\pm}_{\boldsymbol{k}E} = e^{ik_xk_yl^2}\sum_{\nu}{}^{\pm} c^{(0)}_{\nu E}\,\frac{1}{\sqrt{|v^x_\nu|}}\,e^{-il^2\int_{k^\nu_{y0}(E)}^{k_y} k^\nu_x(z,E)dz}, \quad \text{(D11)}$$

where the superscript on $f^\pm$ corresponds to the sign in $k_y \to \pm\infty$; k^ν_{y0} is the coordinate of closest approach to the saddle point for the edge ν; $\sum^\pm_\nu$ runs over $\nwarrow$ and $\swarrow$ for f^+ (the two edges above the breakdown interval), and over $\nearrow$ and $\searrow$ for f^-; the various $c^{(0)}_{\nu E}$ are related as in Eqs. (D9), (D10), and (182).

b. Derivation of first-order-corrected connection formula

It is useful to estimate the size of the region in $\boldsymbol{k}$ space ($\Delta k_x \Delta k_y$), in the vicinity of the saddle point, where the Zilberman-Fischbeck wave functions are invalid; equivalently, this is where the asymptotic limits of the PCFs would not apply; we have called this the breakdown region. This is the region where z, the dimensionless variable entering the Weber differential equation [cf. Eq. (D3)], is of order one. Further assuming $m_1/m_2 = O(1)$, we obtain $\Delta k_y = O(l^{-1})$. Utilizing the hyperbolic asymptotes $k_y = \pm(b/a)k_x$ and assuming $(b/a) = O(1)$, we estimate $\Delta k_x = O(l^{-1})$; note that $\Delta k_x \Delta k_y = O(l^{-2})$.

Let us derive a first-order-corrected effective Hamiltonian ($\mathcal{H} = H_0 + H_1$) in the breakdown region. We first expand the symbol H_1 around the saddle point as in Eq. (194). The terms (δH_1) which we neglect to write explicitly are bounded by their values at the boundary of the breakdown region as $\delta H_1(\Delta\boldsymbol{k}) = O(l^{-4})$, with our estimates of $\Delta\boldsymbol{k}$ in the above paragraph. In other words, the explicit terms in Eq. (194), when evaluated on the boundary, are larger in magnitude than $O(l^{-4})$ and therefore expected to be relevant in the limit of small field. When these explicit terms are added to H_0, the result is an effective Hamiltonian that is identical in form to Eq. (D1):

$$\mathcal{H} = H_0(\boldsymbol{K}) + H_1(\boldsymbol{K}) = H_0(\boldsymbol{Q}) + H_1(\boldsymbol{0}) + O(l^{-4}), \quad \text{(D12)}$$

but shifted by an energy constant $H_1(\boldsymbol{0})$, and with shifted momentum variables

$$q_x = k_x + m_1H_{1x}, \quad q_y = k_y - m_2H_{1y} \quad \leftrightarrow$$
$$Q_x = K_x + mH_{1x}, \quad Q_y = K_y - m_2H_{1y}. \quad \text{(D13)}$$

It is useful to know which of the Roth, Berry, or Zeeman terms contribute to the effective Hamiltonian; let us individually expand H_1^R, H_1^B, and H_1^Z as in Eq. (194), keeping only the linear terms, which we define by $H^R_{1j}k_j$, etc. For example, the Berry term is expanded as

$$l^2H_1^B(\boldsymbol{k}) = \mathfrak{X}^y(\boldsymbol{k})v^x(\boldsymbol{k}) - \mathfrak{X}^x(\boldsymbol{k})v^y(\boldsymbol{k}) = \mathfrak{X}^y(\boldsymbol{0})\frac{k_x}{m_1} + \mathfrak{X}^x(\boldsymbol{0})\frac{k_y}{m_2} + \cdots := l^2\left(H^B_{1x}k_x + H^B_{1y}k_y\right) + \cdots, \quad \text{(D14)}$$

and vanishes when evaluated at the saddle point, where the band velocity $\boldsymbol{v}^\perp$ vanishes. Therefore the shift in the energy constant is only contributed by the gauge-invariant Roth and Zeeman terms:

$$H_1(\boldsymbol{0}) = H_1^R(\boldsymbol{0}) + H_1^Z(\boldsymbol{0}). \quad \text{(D15)}$$

We further deduce from Eq. (D14) that the shifts in the momentum variables k_x and k_y are, respectively,

$$m_1H_{1x} = l^{-2}\mathfrak{X}^y(\boldsymbol{0}) + m_1\left(H^R_{1x} + H^Z_{1x}\right),$$
$$m_2H_{1y} = l^{-2}\mathfrak{X}^x(\boldsymbol{0}) + m_2\left(H^R_{1y} + H^Z_{1y}\right). \quad \text{(D16)}$$

The similarity of $\mathcal{H}$ with the inverted-harmonic-oscillator Hamiltonian implies that it may be solved with the same techniques, with some small modifications. We assume here the reader has some familiarity with the "same techniques," which we have reviewed in the previous subsection (Appendix D 1 a) and will presently extend.

Let us then define the eigenfunction of $\mathcal{H}$, with the $O(l^{-4})$ correction henceforth truncated, as

$$\begin{aligned} 0 &= [\mathcal{H}(\boldsymbol{K}) - E] f_{\boldsymbol{k}E} = [H_0(\boldsymbol{Q}) + H_1(\boldsymbol{0}) - E] f_{\boldsymbol{k}E} \\ &= [H_0(\boldsymbol{Q}) - \tilde{E}] f_{\boldsymbol{k}E}, \\ \tilde{E} &:= E - H_1(\boldsymbol{0}). \end{aligned} \tag{D17}$$

Performing a gauge transformation,

$$f = e^{iq_x k_y l^2} \bar{f}, \tag{D18}$$

we see that $\bar{f}$ satisfies the Weber differential equation in the modified variable q_y and with modified eigenvalue $\tilde{E}$. Let us define $f^{\pm}_{\boldsymbol{k}E}$ to be the asymptotic limits of f in the limit $k_y \to \pm\infty$. Utilizing results from our review in Appendix D 1 a, especially Eq. (D11), we obtain

$$\begin{aligned} & e^{-iq_x k_y l^2} f^{\pm}_{\boldsymbol{k}E} \\ &= \sum_{\nu}{}^{\pm} \left\{ c^{(0)}_{\nu E} \frac{1}{\sqrt{|v^x_\nu|}} e^{-il^2 \int_{k^\nu_{y0}(E)}^{k_y} k^\nu_x(z,E) dz} \right\}_{(k_y,E)\to(q_y,\tilde{E})}, \end{aligned} \tag{D19}$$

with $\{c^{(0)}_{\nu E}\}$ related as in Eqs. (D9), (D10), and (182); $\boldsymbol{k}^\nu_0$ is the wave vector of closest approach to the saddle point for the edge ν; $\sum^{\pm}_\nu$ runs over $\nwarrow$ and $\swarrow$ for f^+ (the two edges above the breakdown interval), and over $\nearrow$ and $\searrow$ for f^-. Upon substituting $k_y \to q_y = k_y - m_2 H_{1y}$ in the curly brackets of Eq. (D19), Eq. (D19) is expressible as

$$\begin{aligned} f^{\pm}_{\boldsymbol{k},E} = \sum_{\nu}{}^{\pm} & c^{(0)}_{\nu\tilde{E}} e^{im_1 l^2 H_{1x} k^\nu_{y0}(\tilde{E}) + im_2 l^2 H_{1y} k^\nu_{x0}(\tilde{E})} \frac{e^{ik_x k_y l^2}}{\sqrt{|v^x_\nu|}} \\ & \times e^{-il^2 \int_{k^\nu_{y0}(E)}^{k_y} (k^\nu_x - \frac{H^\nu_1 - H_1(\boldsymbol{0})}{v^x_\nu}) dz} \Big|_{E\to\tilde{E}} + O(l^{-2}), \end{aligned} \tag{D20}$$

as we prove at the end of this subsection. We may identify the last line of Eq. (D20) as the Zilberman-Fischbeck function defined in Eq. (177). Further defining

$$\begin{aligned} c_{\nu E} &:= c^{(0)}_{\nu\tilde{E}}\ \upsilon_{\nu E}, \\ \upsilon_{\nu E} &:= e^{im_1 l^2 H_{1x} k^\nu_{y0}(\tilde{E}) + im_2 l^2 H_{1y} k^\nu_{x0}(\tilde{E})}, \end{aligned} \tag{D21}$$

we cast Eq. (D20) in the simple form

$$f^{\pm}_{\boldsymbol{k},E} = \sum_{\nu}{}^{\pm} c_{\nu E} \tilde{g}^\nu_{\boldsymbol{k}E}, \tag{D22}$$

which may be identified with Eqs. (175)–(180). We are finally ready to derive the scattering matrix defined in Eq. (181) and expressed in Eqs. (187) and (190). Combining Eq. (D21) with (D10) and (182),

$$\begin{aligned} \mathbb{S}(E,l^2) &= \begin{pmatrix} \upsilon_{\nwarrow E} & 0 \\ 0 & \upsilon_{\swarrow E} \end{pmatrix} \mathbb{S}^{(0)}(\tilde{E},l^2) \begin{pmatrix} \upsilon^*_{\nearrow E} & 0 \\ 0 & \upsilon^*_{\searrow E} \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{T}(\tilde{\mu}) \upsilon_{\nwarrow E} \upsilon^*_{\nearrow E} & \mathcal{R}(\tilde{\mu}) \upsilon_{\swarrow E} \upsilon^*_{\searrow E} \\ \mathcal{R}(\tilde{\mu}) \upsilon_{\nwarrow E} \upsilon^*_{\nearrow E} & \mathcal{T}(\tilde{\mu}) \upsilon_{\swarrow E} \upsilon^*_{\searrow E} \end{pmatrix}. \end{aligned} \tag{D23}$$

Inserting the integral expression for υ [from Eq. (D21)], and further applying the definition of $\boldsymbol{k}^\nu_0$, we obtain Eq. (190). If we neglect the Roth and Zeeman corrections, we find that the Berry term is sufficient to restore gauge covariance:

$$\begin{aligned} & \mathbb{S}(E,l^2) \\ & \overset{H_1 = H_1^B}{=} \begin{pmatrix} \mathcal{T}(\mu) e^{i\mathfrak{X}^y(\boldsymbol{0}) 2b(E)} & \mathcal{R}(\mu) e^{-i\mathfrak{X}^x(\boldsymbol{0}) 2a(E)} \\ \mathcal{R}(\mu) e^{i\mathfrak{X}^x(\boldsymbol{0}) 2a(E)} & \mathcal{T}(\mu) e^{-i\mathfrak{X}^y(\boldsymbol{0}) 2b(E)} \end{pmatrix} + O(l^{-2}) \\ & = \begin{pmatrix} \mathcal{T}(\mu) e^{i\int_{-b}^{b} \mathfrak{X}^y(0,k_y) dk_y} & \mathcal{R}(\mu) e^{-i\int_{-a}^{a} \mathfrak{X}^x(k_x,0) dk_x} \\ \mathcal{R}(\mu) e^{i\int_{-a}^{a} \mathfrak{X}^x(k_x,0) dk_x} & \mathcal{T}(\mu) e^{-i\int_{-b}^{b} \mathfrak{X}^y(0,k_y) dk_y} \end{pmatrix} \\ & \quad + O\left(l^{-2}, \left(\frac{b}{G}\right)^2, \left(\frac{a}{G}\right)^2\right). \end{aligned} \tag{D24}$$

In the first equality, we have made use of the expansion of H_1^B in Eq. (D14); the second equality follows from

$$\int_{-b}^{b} \mathfrak{X}^y(0,k_y) dk_y = 2\mathfrak{X}^y(\boldsymbol{0}) b + O((b/G)^2), \tag{D25}$$

where the correction is of order $(b/G)^2$, with G a typical reciprocal period.

Up to $O(l^{-2},(b/G)^2,(a/G)^2)$, the $O(1)$ phases in Eq. (190) may be creatively interpreted as the Roth-Berry-Zeeman phase averaged over all possible tunneling trajectories in the classically forbidden region. For example, the phase acquired for the tunneling trajectory in the $\vec{y}$ direction may be expressed as

$$e^{i2m_1 H_{1x} b(\tilde{E}) l^2} \approx \exp\left[i \int_{-b}^{b} \left\{ \frac{\overline{\tilde{H}_1}}{v^x} \right\}_{k_y} dk_y \right], \tag{D26}$$

where $\{\bar{\cdot}\}_{k_y}$ denotes the k_x average of the quantity $\cdot$ over a fixed-k_y cross section of the forbidden region:

$$\begin{aligned} \left\{ \frac{\tilde{H}_1}{v^x} \right\}_{k_y} &= \frac{l}{2} \int_{-1/l}^{1/l} \frac{H_1(\boldsymbol{k}) - H_1(\boldsymbol{0})}{v^x(\boldsymbol{k})} dk_x \\ &\approx \frac{m_1 l}{2} \int_{-1/l}^{1/l} \frac{H_{1x} k_x + H_{1y} k_y}{k_x} dk_x = m_1 H_{1x}. \end{aligned} \tag{D27}$$

In the last equality, we have used the Cauchy principal value for the integral $\int dk_x/k_x$.

Proof of identification of Eq. (D19) with Eq. (D20). From the exponent in the second line of Eq. (D20),

$$\begin{aligned} & \int_{k^\nu_{y0}(\tilde{E})}^{k_y} \frac{H^\nu_1 - H_1(\boldsymbol{0})}{v^x_\nu} dt \\ & = m_1 H_{1x} \left(k_y - k^\nu_{y0}(\tilde{E}) \right) \\ & \quad + m_2 H_{1y} \left(k^\nu_x(k_y,\tilde{E}) - k^\nu_{x0}(\tilde{E}) \right) + O(l^{-4}), \end{aligned} \tag{D28}$$

where, as a reminder, we have $k^\nu_x(k_y,E)$ as the k_x coordinate of the section s_ν at wave vector k_y and energy E. In deriving Eq. (D28), we employed $H^\nu_1(\boldsymbol{k}) - H_1(\boldsymbol{0}) = k_x H_{1x} + k_y H_{1y} + \cdots$ from Eq. (193), and the identity $dk_y/v^x = -dk_x/v^y$ along a constant-energy contour. The uncertainty $O(l^{-4})$ in Eq. (D28) is estimated by evaluating the neglected terms at the boundary of the breakdown region, where $\boldsymbol{k} = O(l^{-1})$. Substituting

Eq. (D28) into Eq. (D20), we obtain

$$f^{\pm}_{k,E} = e^{ik_xk_yl^2}\sum_{\nu}{}^{\pm} c^{(0)}_{\nu\tilde{E}}\, e^{il^2m_1H_{1x}k_y+il^2m_2H_{1y}k^{\nu}_x(k_y,\tilde{E})} \times \frac{1}{\sqrt{|v^{\nu}_x|}}\exp\left\{-il^2\int_{k^{\nu}_{y0}(E)}^{k_y}k^{\nu}_x dz\right\}\Bigg|_{E\to\tilde{E}} + O(l^{-2}). \tag{D29}$$

To complete the identification of this expression with Eq. (D19), we apply the following three observations:

$$\text{(i)}\quad e^{ik_xk_yl^2}e^{il^2m_1H_{1x}k_y} = e^{iq_xk_yl^2}, \tag{D30}$$

from the fundamental theorem of calculus,

$$\text{(ii)}\quad m_2H_{1y}k^{\nu}_x(k_y,\tilde{E}) + \int_{k^{\nu}_{y0}(\tilde{E})}^{k_y}k^{\nu}_x(z,\tilde{E})dz = \int_{k^{\nu}_{y0}(\tilde{E})}^{q_y}k^{\nu}_x(z,E)dz + O(l^{-4}), \tag{D31}$$

and finally, (iii) bearing in mind that the expressions are to be identified with an uncertainty of $O(l^{-2})$, we might directly replace $|v^x_\nu(q_y,\tilde{E})| \approx |v^x_\nu(k_y,\tilde{E})|$ in the square-root prefactor.

c. *Equivalence of two Zilberman-Fischbeck functions*

We would like to prove in the semiclassical region (sm) that

$$\text{For } \boldsymbol{k}\in sm,\quad g^{\nu}_{\boldsymbol{k}E} = \tilde{g}^{\nu}_{\boldsymbol{k}\tilde{E}} + O(l^{-2}). \tag{D32}$$

We need the following three identities: (i) In the semiclassical region where the Zilberman-Fischbeck functions are valid, we may assume $k^{\nu}_x = O(1)$ and therefore

$$2m_1H_1(\boldsymbol{0})/k^{\nu}_x(k_y,E)^2 = O(l^{-2}); \tag{D33}$$

combining this assumption with Eq. (80), we derive

$$k^{\nu}_x(k_y,\tilde{E}) = k^{\nu}_x(k_y,E) - H_1(\boldsymbol{0})/v^x_\nu + O(l^{-4}). \tag{D34}$$

(ii) The same assumption in Eq. (D33) implies, with $v^x_\nu = k^{\nu}_x/m_1$, that

$$v^x_\nu(k_y,\tilde{E}) = v^x_\nu(k_y,E) - H_1(\boldsymbol{0})/m_1v^x_\nu + O(l^{-4}). \tag{D35}$$

(iii) Lastly, applying the fundamental theorem of calculus,

$$\begin{aligned}&\int_{k^{\nu}_{y0}(\tilde{E})}^{k_y}\left[k^{\nu}_x(z,E) - \frac{H^{\nu}_1}{v^x_\nu}\right]dz\\ &= \int_{k^{\nu}_{y0}(E)}^{k_y}\left[k^{\nu}_x(z,E) - \frac{H^{\nu}_1}{v^x_\nu}\right]dz\\ &\quad + \left[k^{\nu}_{y0}(\tilde{E}) - k^{\nu}_{y0}(E)\right]k^{\nu}_{x0}(E) + O(l^{-4})\\ &= \int_{k^{\nu}_{y0}(E)}^{k_y}\left[k^{\nu}_x(z,E) - \frac{H^{\nu}_1}{v^x_\nu}\right]dz + O((a/G)l^{-2},l^{-4}).\end{aligned}$$

In the last equality, we applied that the coordinate of closest approach $k_{x0} = \pm a(E)$ (the hyperbolic parameter) for $E > 0$ and is otherwise zero. Substituting (i)–(iii) into Eq. (177), we derive Eq. (D32) as desired.

APPENDIX E: APPENDIX TO SEC. VIII: "EFFECTIVE HAMILTONIAN FOR GENERAL BAND TOUCHINGS"

1. Calculus with Weyl-symmetrized operators

Here we collect several identities which are useful in the calculus of Weyl-symmetrized operators.

We are interested in kinetic quasimomentum operators with the noncommutative relation:

$$[K_x,K_y] = il^{-2}. \tag{E1}$$

It immediately follows that

$$[f(K_x),k_y] = il^{-2}f'(K_x), \tag{E2}$$

with f' denoting a derivative with respect to K_x.

We are very often interested in symmetrized functions of $\boldsymbol{K}$. Besides our definition of symmetrization with the Fourier formula in Eq. (50), a more elementary definition exists for polynomials [73]: given a monomial $k^m_xk^n_y$, its symmetrized form is obtained from extracting all terms with m powers of K_x and n powers of K_y in the noncommutative binomial expansion

$$\frac{m!n!}{(m+n)!}(K_x+K_y)^{m+n}. \tag{E3}$$

One may verify that this symmetrization preserves the structure of products:

$$[(sk_x+tk_y+u)^v] = (sK_x+tK_y+u)^v,\quad s,t,u\in\mathbb{C},\ v\in\mathbb{Z}, \tag{E4}$$

which implies that the exponential structure is also preserved:

$$e^{i\boldsymbol{K}\cdot\boldsymbol{R}} = \sum_{n=0}^{\infty}\frac{(i\boldsymbol{K}\cdot\boldsymbol{R})^n}{n!} = \sum_{n=0}^{\infty}\left[\frac{(i\boldsymbol{k}\cdot\boldsymbol{R})^n}{n!}\right] = [e^{i\boldsymbol{k}\cdot\boldsymbol{r}}]. \tag{E5}$$

This identify underlies the Fourier definition of symmetrization in Eq. (50).

For any function of K_x:

$$[k_y,f(K_x)] = (1/2)[k_yf(K_x)+f(K_x)k_y] := (1/2)\{k_y,f(K_x)\}, \tag{E6}$$

as may be proven by Taylor-expanding f and symmetrizing individual terms (e.g., $[k_yK^n_x]$) with the rule in Eq. (E3).

Symmetrization of a symbol commutes with addition:

$$[f(\boldsymbol{k})] + [g(\boldsymbol{k})] = [f(\boldsymbol{k}) + g(\boldsymbol{k})]. \tag{E7}$$

Like many basic identities, it may be proven by Fourier analysis:

$$\begin{aligned}&\int d\boldsymbol{R}\,\check{f}(\boldsymbol{R})e^{i\boldsymbol{K}\cdot\boldsymbol{R}} + \int d\boldsymbol{R}'\check{g}(\boldsymbol{R}')e^{i\boldsymbol{K}\cdot\boldsymbol{R}'}\\ &= \int d\boldsymbol{R}(\check{f}(\boldsymbol{R})e^{i\boldsymbol{K}\cdot\boldsymbol{R}} + \check{g}(\boldsymbol{R})e^{i\boldsymbol{K}\cdot\boldsymbol{R}})\\ &= \int d\boldsymbol{R}(\check{f}(\boldsymbol{R}) + \check{g}(\boldsymbol{R}))e^{i\boldsymbol{K}\cdot\boldsymbol{R}}.\end{aligned}$$

The product rule for two symmetrized operators is described in Eq. (229); its nontriviality originates from the noncommutivity of Eq. (E1). We review the proof of the multiplication rule by

Roth [24], which combines Fourier analysis, and the Baker-Campbell-Hausdorff identity $e^A e^B = e^{A+B} e^{[A,B]/2}$:

$$\begin{aligned}
&\int d\boldsymbol{r} \int d\boldsymbol{r}' \check{A}(\boldsymbol{r})\check{B}(\boldsymbol{r}') e^{-i\boldsymbol{K}\cdot\boldsymbol{r}} e^{-i\boldsymbol{K}\cdot\boldsymbol{r}'} \\
&= \int d\boldsymbol{r} \int d\boldsymbol{r}' \check{A}(\boldsymbol{r})\check{B}(\boldsymbol{r}') e^{-i\boldsymbol{K}\cdot(\boldsymbol{r}+\boldsymbol{r}')} e^{-il^{-2}\epsilon_{\alpha\beta} r_\alpha r'_\beta/2} \\
&= \left[\int d\boldsymbol{r} \int d\boldsymbol{r}' \check{A}(\boldsymbol{r})\check{B}(\boldsymbol{r}') e^{-i\boldsymbol{k}\cdot(\boldsymbol{r}+\boldsymbol{r}')} e^{-il^{-2}\epsilon_{\alpha\beta} r_\alpha r'_\beta/2} \right] \\
&= \left[e^{(i/2)l^{-2}\epsilon_{\alpha\beta}\nabla^\alpha_{\boldsymbol{k}}\nabla^\beta_{\boldsymbol{k}'}} \int d\boldsymbol{r} \int d\boldsymbol{r}' \check{A}(\boldsymbol{r})\check{B}(\boldsymbol{r}') \right. \\
&\quad \left. \times e^{-i\boldsymbol{k}\cdot\boldsymbol{r}} e^{-i\boldsymbol{k}'\cdot\boldsymbol{r}'} \Big|_{\boldsymbol{k}=\boldsymbol{k}'} \right].
\end{aligned} \tag{E8}$$

An application of this product rule to a commutator of two symmetrized operators leads to

$$\begin{aligned}
&[A(\boldsymbol{K}), B(\boldsymbol{K})] \\
&= [[A(\boldsymbol{k}), B(\boldsymbol{k})]] + \frac{i}{2l^2}\epsilon_{\alpha\beta}\left[\left\{\nabla^\alpha_{\boldsymbol{k}} A, \nabla^\beta_{\boldsymbol{k}} B\right\}\right] + O(l^{-4}),
\end{aligned} \tag{E9}$$

where $[[a,b]] = [ab - ba]$ and $[\{a,b\}] = [ab + ba]$.

a. Symmetrized operators which are independent of k_y

A particularization of the Roth product rule [cf. Eq. (229)] for functions independent of k_y is

$$A(K_x)B(K_x) = [A(k_x)B(k_x)]_{k_x \to K_x}. \tag{E10}$$

An operator acting in $\boldsymbol{r}$ space (or more generally, an operator acting in both $\boldsymbol{r}$ and k_x space) commutes with the operation $[\cdot]_{k_x\to K_x}$, i.e.,

$$\hat{F}(\hat{\boldsymbol{r}}, \nabla_{\boldsymbol{r}}) A(K_x, \boldsymbol{r}) = [\hat{F}(\hat{\boldsymbol{r}}, \nabla_{\boldsymbol{r}}) A(k_x, \boldsymbol{r})]_{k_x \to K_x}, \tag{E11}$$

$$\hat{G}(K_x, \hat{\boldsymbol{r}}, \nabla_{\boldsymbol{r}}) A(K_x, \boldsymbol{r}) = [\hat{G}(k_x, \hat{\boldsymbol{r}}, \nabla_{\boldsymbol{r}}) A(k_x, \boldsymbol{r})]_{k_x \to K_x}, \tag{E12}$$

which may also be proven from Fourier analysis.

2. Relating our ansatz to Slutskin's function

To lowest order in l^{-2}, our ansatz for the wave function takes the form

$$\begin{aligned}
\Psi(\boldsymbol{r}) &= \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k}} \alpha(\boldsymbol{k}, \boldsymbol{r}), \\
\alpha(\boldsymbol{k}, \boldsymbol{r}) &:= \sum_n e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{nK_x0}(\boldsymbol{r}) f_{n\boldsymbol{k}},
\end{aligned} \tag{E13}$$

with $\sum_{\boldsymbol{k}}$ shorthand for a continuous integral over the Brillouin torus. We would like our ansatz to be independent of the choice of unit cell in $\boldsymbol{k}$ space, i.e.,

$$\alpha(\boldsymbol{k}, \boldsymbol{r}) = \alpha(\boldsymbol{k} + \boldsymbol{G}, \boldsymbol{r}), \tag{E14}$$

for any reciprocal vector $\boldsymbol{G}$. This is ensured if we impose the following boundary conditions on the wave function in the $(K_x,0)$ representation:

$$f_{n\boldsymbol{k}} = f_{n\boldsymbol{k}+\boldsymbol{G}_x}, \tag{E15}$$

$$f_{m\boldsymbol{k}} = \sum_n \tilde{S}_{mn}(K_x, 0; \boldsymbol{G}_y) f_{n\boldsymbol{k}+\boldsymbol{G}_y}, \tag{E16}$$

$$\tilde{S}_{mn}(K_x, 0; \boldsymbol{G}) := \int d\boldsymbol{\tau} \left[u^*_{mk_x0}(\boldsymbol{\tau})\right] e^{i\boldsymbol{G}\cdot\boldsymbol{\tau}} u_{nK_x0}(\boldsymbol{\tau}). \tag{E17}$$

Here, $\tilde{S}$ is formally an infinite-dimensional matrix, $\int d\boldsymbol{\tau}$ denotes an integration over the real-space unit cell, and $\boldsymbol{G}_x$ and $\boldsymbol{G}_y$ are the primitive reciprocal vectors of a rectangular lattice:

$$\begin{aligned}
\boldsymbol{G}_x &:= 2\pi\vec{x}/a_x, \quad G_x := 2\pi/a_x, \\
\boldsymbol{G}_y &:= 2\pi\vec{y}/a_y, \quad G_y := 2\pi/a_y;
\end{aligned} \tag{E18}$$

the choice of a rectangular lattice is merely for notational simplicity. $\alpha(\boldsymbol{k},\boldsymbol{r}) = \alpha(\boldsymbol{k} + \boldsymbol{G}_x, \boldsymbol{r})$ follows from the periodicity in k_x of both (i) the wave function [Eq. (E15)], and (ii) the operator that acts on the wave function:

$$e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{nK_x0} = e^{i(\boldsymbol{k}+\boldsymbol{G}_x)\cdot\boldsymbol{r}} u_{n,K_x+G_x,0} \tag{E19}$$

[cf. Eq. (213)]. The same operator is, however, not periodic in k_y, and therefore the corresponding boundary condition on the wave function is more complicated. To verify that this boundary condition produces the desired periodicity, $\alpha(\boldsymbol{k},\boldsymbol{r}) = \alpha(\boldsymbol{k} + \boldsymbol{G}_y, \boldsymbol{r})$, apply the operation $\sum_m e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{mK_x0}(\boldsymbol{r})$ on both sides of Eq. (E16) and apply the completeness relation in Eq. (216).

Our discussion about boundary conditions may seem more formal than practical, since in many applications we would only be interested in $f_{n\boldsymbol{k}}$ for $\boldsymbol{k}$ in the vicinity of a point; the area of interest is typically much smaller than the Brillouin torus. On the other hand, assuming such formalities, we would show that our ansatz is equivalent to an expansion in Slutskin's basis functions [12] (denoted $\chi_{n\boldsymbol{k}}$):

$$\Psi(\boldsymbol{r}) = \frac{1}{\sqrt{N}} \sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{nK_x0} f_{n\boldsymbol{k}} = \frac{1}{\sqrt{N}} \sum_{n\boldsymbol{k}} f_{n\boldsymbol{k}} \chi_{n\boldsymbol{k}}, \tag{E20}$$

$$\chi_{n\boldsymbol{k}}(\boldsymbol{r}) := u_{n,k_x+y/l^2,0}(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}}. \tag{E21}$$

While $e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{nK_x0}$ is a differential operator acting on $f_{n\boldsymbol{k}}$, $\chi_{n\boldsymbol{k}}$ acts on $f_{n\boldsymbol{k}}$ by multiplication, and therefore has a more intuitive interpretation as a wave function over real space.

The first step to proving Eq. (E20) is to equivalently express Slutskin's function as

$$\begin{aligned}
&u_{n,k_x+y/l^2,0}(\boldsymbol{r}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \\
&\quad = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{R}} e^{-i[k_x - (i/l^2)(\partial/\partial k_y)](r_x - R_x)} W_n(\boldsymbol{r} - \boldsymbol{R}) e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \\
&\quad = u_{n,K^*_x,0} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}, \\
&K^*_x := k_x - \frac{i}{l^2}\frac{\partial}{\partial k_y},
\end{aligned} \tag{E22}$$

with help from the identity Eq. (213). What remains is to prove

$$\Psi(\boldsymbol{r}) = \sum_{n\boldsymbol{k}} f_{n\boldsymbol{k}} u_{nK^*_x0} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} = \sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{nK_x0} f_{n\boldsymbol{k}}, \tag{E23}$$

which is analogous to an integration by parts; for notational simplicity, we shall no longer write out normalization factors. As an intermediate step, we will further identify the above quantity as equal to an expansion

$$\Psi(\boldsymbol{r}) = \sum_{R_x,k_y} h(x-R_x,y,k_y),$$
$$h(x-R_x,y,k_y) := \sum_n h_n(x-R_x,y,k_y)\check{f}_{nR_xk_y}, \tag{E24}$$

in the basis functions

$$h_n(x-R_x,y,k_y) := e^{i\{k_y-(x-R_x)/l^2\}y}\sum_{R_y} W_n(\boldsymbol{r}-\boldsymbol{R}), \tag{E25}$$

with expansion coefficients

$$\check{f}_{nR_xk_y} := \sum_{k_x} e^{ik_xR_x} f_{n,\boldsymbol{k}}. \tag{E26}$$

We may identify these expansion coefficients as the Fourier coefficients of the periodic function $f_{n\boldsymbol{k}}$ [cf. Eq. (E15)]. h_n may be viewed as the magnetic analog of a "hybrid" function, which is spatially extended in $\vec{y}$ (as a Luttinger-Kohn function) but exponentially localized in $\vec{x}$ (as a Wannier function). Indeed, setting $l^{-2}=0$ in Eq. (E25),

$$h_n(x-R_x,y,k_y) \overset{l^{-2}=0}{=} e^{ik_yy}\sum_{R_y} W_n(\boldsymbol{r}-\boldsymbol{R}) = \sum_{k_x} e^{-ik_xR_x}\chi^{(0)}_{n\boldsymbol{k}}, \tag{E27}$$

with the Luttinger-Kohn function defined as

$$\chi^{(0)}_{n\boldsymbol{k}}(\boldsymbol{r}) = e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{nk_x0}(\boldsymbol{r}). \tag{E28}$$

It is known that the Luttinger-Kohn functions form a complete orthonormal basis in which any function can be expanded [72]; we thus expect for small fields that $\{h_n\}$ forms a linearly independent basis, though we avoid assuming orthogonality. Furthermore, we insist that the expansion Eq. (E24) is independent of the choice of unit cell in the Brillouin circle parametrized by k_y, i.e., for each R_x,

$$h(x-R_x,y,k_y) = h(x-R_x,y,k_y+G_y); \tag{E29}$$

this imposes a boundary condition on the wave function $\check{f}_{nR_xk_y}$, in close analogy with Eqs. (E14)–(E16). We may exploit this periodicity to express Eq. (E24) as

$$\Psi(\boldsymbol{r}) = \sum_{R_x}\sum_{k_y} h(x-R_x,y,k_y)$$
$$= \sum_{R_x}\sum_{k_y} h(x-R_x,y,k_y+(x-R_x)/l^2)$$
$$= \sum_{R_xk_y} e^{ik_yy}\sum_{nR_y} W_n(\boldsymbol{r}-\boldsymbol{R})\sum_{k_x} e^{ik_xR_x} f_{nk_xk_y+(x-R_x)/l^2}.$$

This quantity is equal to the right-hand side of Eq. (E23), as we now demonstrate:

$$\sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}u_{nK_x0}f_{n\boldsymbol{k}}$$
$$= \sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\sum_{\boldsymbol{R}} W_n(\boldsymbol{r}-\boldsymbol{R})e^{-iK_x(x-R_x)}f_{n\boldsymbol{k}}$$
$$= \sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\sum_{\boldsymbol{R}} W_n(\boldsymbol{r}-\boldsymbol{R})e^{-i(k_x+il^{-2}\partial_y)(x-R_x)}f_{n\boldsymbol{k}}$$
$$= \sum_{nk_y} e^{ik_yy}\sum_{\boldsymbol{R}} W_n(\boldsymbol{r}-\boldsymbol{R})\sum_{k_x} e^{ik_xR_x} f_{nk_xk_y+(x-R_x)/l^2}. \tag{E30}$$

The left-hand side of Eq. (E23) may be expressed as

$$\sum_{n\boldsymbol{k}} f_{n\boldsymbol{k}}u_{nK_x^*}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$$
$$= \sum_{n\boldsymbol{k}} f_{n\boldsymbol{k}}\sum_{\boldsymbol{R}} W_n(\boldsymbol{r}-\boldsymbol{R})e^{-i(k_x-il^{-2}\partial_y)(x-R_x)}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$$
$$= \sum_{n\boldsymbol{k}} f_{n\boldsymbol{k}}\sum_{\boldsymbol{R}} W_n(\boldsymbol{r}-\boldsymbol{R})e^{-ik_x(x-R_x)}e^{-iy(x-R_x)/l^2}e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$$
$$= \sum_{k_yR_x} e^{i\{k_y-(x-R_x)/l^2\}y}\sum_{nR_y} W_n(\boldsymbol{r}-\boldsymbol{R})\sum_{k_x} e^{ik_xR_x} f_{n\boldsymbol{k}},$$

which may be identified with Eq. (E24).

3. Alternative derivation of the infinite-band effective Hamiltonian

We offer a derivation of Eq. (226) and its equivalent, symmetrized form in Eq. (245); these are effective-Hamiltonian equations which formally act on the wave functions over all bands. Equation (245) was previously derived in Sec. VIII C utilizing the Roth product rule of two symmetrized operators [cf. Eq. (229)]; the following, alternative derivation does not rely on this rule.

From Eq. (240),

$$\sum_n \tilde{\mathcal{H}}_{mn}(\boldsymbol{K})f_{n\boldsymbol{k}}$$
$$= \int d\boldsymbol{\tau}\sum_n u^\dagger_{mK_x0}(\boldsymbol{\tau})\hat{H}_0(\boldsymbol{K})u_{nK_x0}(\boldsymbol{\tau})f_{n\boldsymbol{k}}$$
$$= \int d\boldsymbol{\tau}\sum_n u^\dagger_{mK_x0}(\boldsymbol{\tau})\left\{[\hat{H}_0(k_x,0)] + k_y\hat{\Pi}_y + \frac{k_y^2}{2m}\right\} \times u_{nK_x0}(\boldsymbol{\tau})f_{n\boldsymbol{k}}. \tag{E31}$$

To derive the last equality in Eq. (E31), we need the following identity:

$$\hat{H}_0(\boldsymbol{K}) = [\hat{H}_0(\boldsymbol{k})] = \left[\hat{H}_0(k_x,0) + \hat{\Pi}^yk_y + \frac{k_y^2}{2m}\right] = [\hat{H}_0(k_x,0)] + \hat{\Pi}^yk_y + \frac{k_y^2}{2m}. \tag{E32}$$

The second equality follows from Eq. (11), and the last equality assumed the Landau gauge for the kinetic quasimomentum operators: $K_x = k_x + il^{-2}\partial_y, K_y = k_y$.

We separately consider each of the three terms in the last line of Eq. (E31). The first term is simply evaluated as

$$\int d\boldsymbol{\tau} u^{\dagger}_{mK_x0}(\boldsymbol{\tau})\hat{H}_0(K_x,0)u_{nK_x0}(\boldsymbol{\tau}) = \left[\int d\boldsymbol{\tau} u^{*}_{mk_x0}(\boldsymbol{\tau})\hat{H}_0(k_x,0)u_{nk_x0}(\boldsymbol{\tau})\right] = \tilde{H}_0(K_x,0)_{mn}. \tag{E33}$$

Here we have made use of the basic identities Eq. (E10) and

$$\hat{H}_0(K_x,0)u_{nK_x0}(\boldsymbol{\tau}) = \left[\hat{H}_0(k_x,0)u_{nk_x0}(\boldsymbol{\tau})\right]; \tag{E34}$$

the latter follows from Eq. (E12). It should be emphasized that the right-hand side of Eq. (E33) corresponds to the symbol $\langle u_{mk_x0}|\hat{H}_0(k_x,0)|u_{nk_x0}\rangle$, with cell-periodic functions which are smooth with respect to k_x; this assumption of smoothness is justified in Sec. VIII A.

For the second and third terms, a few basic identities for noncommuting operators [cf. Eq. (E2)] are helpful:

$$\left[u^{\dagger}_{mK_x0},k_y\right] = il^{-2}\left[\partial_{k_x}u^{*}_{mk_x0}\right], \tag{E35}$$

$$\left[u^{\dagger}_{mK_x0},k_y^2\right] = 2il^{-2}k_y\left[\partial_{k_x}u^{*}_{mk_x0}\right] - l^{-4}\left[\partial^2_{k_x}u^{*}_{mk_x0}\right]. \tag{E36}$$

We remind the reader that $[\cdot,\cdot]$ is a commutator, while $[\cdot]$ is a Weyl symmetrization of $\cdot$. We also need the identity

$$i\int d\boldsymbol{\tau}\left[\partial_{k_x}u^{*}_{mk_x0}(\boldsymbol{\tau})\right]u_{nK_x0}(\boldsymbol{\tau}) = \left[i\int d\boldsymbol{\tau}\,\partial_{k_x}u^{*}_{mk_x0}(\boldsymbol{\tau})u_{nk_x0}(\boldsymbol{\tau})\right] = -\tilde{\mathfrak{X}}^{x}_{mn}(K_x,0),$$

with $\tilde{\mathfrak{X}}$ defined in Eq. (15). Employing the above identity, Eq. (E11), and Eq. (E35), the second term is evaluated as

$$\int d\boldsymbol{\tau}\sum_n u^{\dagger}_{mK_x0}(\boldsymbol{\tau})(k_y\tilde{\Pi}^{y})u_{nK_x0}(\boldsymbol{\tau})f_{n\boldsymbol{k}} = \sum_{n,o}\left[k_y\delta_{m,o} - l^{-2}\tilde{\mathfrak{X}}^{x}_{mo}(K_x,0)\right]\tilde{\Pi}^{y}_{on}(K_x,0)f_{n\boldsymbol{k}}. \tag{E37}$$

Here, it was also necessary to insert a complete set of cell-periodic operators [cf. Eq. (216)].

The third term in Eq. (E31) is evaluated with aid from Eq. (E36) and the orthonormality condition in Eq. (217):

$$\int d\boldsymbol{\tau} u^{\dagger}_{mK_x0}\frac{k_y^2}{2m}u_{nK_x0}(\boldsymbol{\tau}) = \frac{k_y^2}{2m}\delta_{mn} - \frac{k_y}{ml^2}\tilde{\mathfrak{X}}^{x}_{mn}(K_x,0) - \frac{1}{2ml^4}\left[\left\langle\partial^2_{k_x}u_{mk_x0}\middle|u_{nk_x0}\right\rangle\right]. \tag{E38}$$

We remind the reader of our Dirac notation:

$$\left\langle\partial^2_{k_x}u_{mk_x0}\middle|u_{nk_x0}\right\rangle := \int d\boldsymbol{\tau}\left[\left\{\partial^2_{k_x}u^{*}_{mk_x0}(\boldsymbol{\tau})\right\}u_{nk_x0}(\boldsymbol{\tau})\right]. \tag{E39}$$

Applying the identity

$$\langle u_m|u_n\rangle = \delta_{mn} \Rightarrow \langle\partial^2 u_m|u_n\rangle + 2\langle\partial u_m|\partial u_n\rangle + \langle u_m|\partial^2 u_n\rangle = 0, \tag{E40}$$

we may express the symbol of the last term in Eq. (E38) as proportional to

$$\left\langle\partial^2_{k_x}u_{mk_x0}\middle|u_{nk_x0}\right\rangle = i\,\partial_{k_x}\tilde{\mathfrak{X}}^{x}_{mn} - \sum_o\tilde{\mathfrak{X}}^{x}_{mo}\tilde{\mathfrak{X}}^{x}_{on}\Big|_{k_x,0}. \tag{E41}$$

Inserting Eqs. (E33), (E37), (E38), and (E41) into Eq. (E31), we finally obtain

$$\frac{1}{N}\int d\boldsymbol{r}u^{\dagger}_{mK_x0}(\boldsymbol{r})e^{-i\boldsymbol{k}\cdot\boldsymbol{r}}\hat{H}\Psi(\boldsymbol{r}) = \sum_n\left[\left(\tilde{H}_0 + k_y\tilde{\Pi}^{y} - l^{-2}\tilde{\mathfrak{X}}^{x}\tilde{\Pi}^{y} + \frac{k_y^2}{2m} - \frac{k_y}{ml^2}\tilde{\mathfrak{X}}^{x}\right)_{mn} - \frac{1}{2ml^4}\left(i\,\partial_{k_x}\tilde{\mathfrak{X}}^{x}_{mn} - \sum\nolimits_o\tilde{\mathfrak{X}}^{x}_{mo}\tilde{\mathfrak{X}}^{x}_{on}\right)\right]_{K_x,0}f_{n\boldsymbol{k}}, \tag{E42}$$

from which we may identify the effective Hamiltonian acting on $f_{n\boldsymbol{k}}$ as that of Eq. (226).

We may symmetrize the above Hamiltonian with respect to $\boldsymbol{K}$ to obtain Eq. (245). The identity in Eq. (E6) is useful for this purpose. Let us tackle Eq. (E42) term by term:

$$k_y\tilde{\Pi}^{y} = \frac{1}{2}\{k_y,\tilde{\Pi}^{y}\} - \frac{i}{2l^2}\partial_{K_x}\tilde{\Pi}^{y} = \frac{1}{2}\{k_y,\tilde{\Pi}^{y}\} + \frac{1}{2l^2}[\tilde{\mathfrak{X}}^{x},\tilde{\Pi}^{y}]. \tag{E43}$$

Therefore, the sum of following two terms is symmetric:

$$k_y\tilde{\Pi}^{y} - l^{-2}\tilde{\mathfrak{X}}^{x}\tilde{\Pi}^{y} = \frac{1}{2}\{k_y,\tilde{\Pi}^{y}\} - \frac{1}{2l^2}\{\tilde{\mathfrak{X}}^{x},\tilde{\Pi}^{y}\}. \tag{E44}$$

Consider another term in Eq. (E42):

$$-\frac{k_y}{ml^2}\tilde{\mathfrak{X}}^{x} = -\frac{1}{2ml^2}(\{k_y,\tilde{\mathfrak{X}}^{x}\} + [k_y,\tilde{\mathfrak{X}}^{x}]) = -\frac{1}{2ml^2}\{k_y,\tilde{\mathfrak{X}}^{x}\} + \frac{1}{2ml^4}i\,\partial_{K_x}\tilde{\mathfrak{X}}^{x}. \tag{E45}$$

The last term here cancels a term in Eq. (E42). Finally, note that the $(\tilde{\mathfrak{X}}^{x})^2$ is already symmetric, trivially.

4. Comparison with the effective Hamiltonian in the representation of field-modified Bloch functions

We have claimed that the effective Hamiltonian of Eq. (230) validly describes any band dispersion; when particularized to the case of (i) a single nondegenerate band, or (ii) a subspace of degenerate bands, we may make an instructive comparison with the effective Hamiltonians derived by Roth [24] (reviewed in Secs. IV A and IV B).

In both cases (i) and (ii), the full velocity matrix $\tilde{\boldsymbol{\Pi}}$ and its diagonal component $\tilde{\boldsymbol{v}}$ [recall their definitions in Eqs. (13) and (16)] satisfy $(\tilde{\boldsymbol{\Pi}} - \tilde{\boldsymbol{v}})_{mn} = 0$, or equivalently $\boldsymbol{\Pi} = \boldsymbol{v}$; this follows from Eqs. (14) and (16). This property and the diagonality of $\tilde{\boldsymbol{v}}$ imply that

$$(\mathring{\mathfrak{X}}^{\beta}\tilde{\Pi}^{\alpha})_{mn} = \sum_{\bar{l}}\tilde{\mathfrak{X}}^{\beta}_{m\bar{l}}\tilde{\Pi}^{\alpha}_{\bar{l}n} = \sum_{\bar{l}}\tilde{\mathfrak{X}}^{\beta}_{m\bar{l}}(\tilde{\Pi}^{\alpha} - \tilde{v}^{\alpha})_{\bar{l}n} + \sum_{l}\tilde{\mathfrak{X}}^{\beta}_{ml}(\tilde{\Pi}^{\alpha} - \tilde{v}^{\alpha})_{ln}$$

$$= [\tilde{\mathfrak{X}}^{\beta}(\tilde{\Pi}^{\alpha} - \tilde{v}^{\alpha})]_{mn},$$
$$(\tilde{\Upsilon}^{y}\tilde{\Pi}^{x})_{mn} = [\tilde{\Upsilon}^{y}(\tilde{\Pi}^{x} - \tilde{v}^{x})]_{mn}. \tag{E46}$$

Furthermore, the assumption of nondegeneracy in the band energies (for at least a local region in $\boldsymbol{k}$) implies the existence of energy functions ($\varepsilon_{n\boldsymbol{k}}$) and cell-periodic functions ($u_{n\boldsymbol{k}}$) which are both smooth with respect to $\boldsymbol{k}$. In such a smooth energy basis, $\mathfrak{X}^{y}(\boldsymbol{k})$ is well defined, and its off-block-diagonal component $\mathring{\mathfrak{X}}^{y}(\boldsymbol{k})$ satisfies

$$i\mathring{\mathfrak{X}}^{y}_{m\bar{n}}(\boldsymbol{k}) = -\frac{\tilde{\Pi}^{y}_{m\bar{n}}(\boldsymbol{k})}{\varepsilon_{m\boldsymbol{k}} - \varepsilon_{\bar{n}\boldsymbol{k}}}, \tag{E47}$$

which is, for $\boldsymbol{k} = (k_x,0)$, also the defining relation for $\tilde{\Upsilon}(k_x)$ [cf. Eq. (235)]. When this identification is made in Eq. (232), as well as those in Eq. (E46), we obtain

$$\mathcal{H}_1^R = \frac{1}{2l^2}[\epsilon^{\alpha\beta}\{\tilde{\mathfrak{X}}^{\beta},(\tilde{\Pi}^{\alpha} - \tilde{v}^{\alpha})\}]_{K_x,0}, \tag{E48}$$

which is almost identical to the original Roth term [cf. Eq. (59) and H_1^R in Eq. (64)]; the sole difference is that $\mathcal{H}_1^R$ is independent of k_y. This difference originates from the different representations for the wave functions: $\Psi = \sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{nK_x0} f_{n\boldsymbol{k}}$ in the basis of field-modified Luttinger-Kohn functions, and $\Psi = \sum_{n\boldsymbol{k}} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} u_{n\boldsymbol{K}} f_{n\boldsymbol{k}}$ for field-modified Bloch functions. Finally, $\boldsymbol{\Pi} = \boldsymbol{v}$ also implies that

$$\mathcal{H}_1^B = -\frac{1}{2l^2}\{\mathfrak{X}^{x},v^{y}\}_{K_x,0}, \tag{E49}$$

which may be compared to the original Berry term [cf. Eq. (58) and H_1^B in Eq. (64)]. Since the cell-periodic function is independent [resp. dependent] of k_y in the $(K_x,0)$ representation [resp. (K_x,k_y) representation], the Berry term proportional to $\mathfrak{X}^{y}$ is absent in Eq. (E48), but present in Eqs. (58) and (64).

APPENDIX F: APPENDIX TO SEC. IX: "INTERBAND BREAKDOWN"

1. Connection to Weber's differential equation and Landau-Zener dynamics

Our aim is to derive Weber's differential equation from the effective Hamiltonian equation [Eq. (275)]. To begin, let us elaborate on the basis of field-modified Luttinger-Kohn functions in which Eq. (275) is represented. We have presupposed a basis where $\tilde{u}_{nk_x0}$ are energy bands along $k_y = 0$; this fixes the basis up to U(1) × U(1) gauge transformations; i.e., each energy band may be multiplied by a k_x-dependent phase. This arbitrariness is partially removed by insisting that the diagonal elements of $\mathfrak{X}^{x}(k_x,0)$ (a 2×2 matrix) vanish; this is the parallel-transport condition within each band. The off-diagonal elements of $\mathfrak{X}^{x}(k_x,0)$ vanish because they represent a coupling between distinct representations of g_2 [cf. Eq. (270)] [125]. The vanishing of $\mathfrak{X}^{x}(\boldsymbol{0})$ (as a 2×2 matrix) justifies the neglect of the third $O(l^{-2})$ term in Eq. (237), from which we have derived Eq. (275).

We remove the k_x dependence of Eq. (275) by the transformation

$$\tilde{f}_{n\boldsymbol{k}} = e^{ik_xk_yl^2}\phi_{n\boldsymbol{k}}, \tag{F1}$$

$$0 = e^{-ik_xk_yl^2}(\mathcal{H}_0(\boldsymbol{K}) - E) * \tilde{f}_{\boldsymbol{k}}$$
$$= ([\varepsilon_{\boldsymbol{0}} - E]I + k_y\Pi^{y}) * \phi + \frac{i}{l^2}\Pi^{x}\frac{\partial}{\partial k_y} * \phi. \tag{F2}$$

Here, we introduce $\varepsilon_{\boldsymbol{0}}$ as the energy at the II-Dirac point ($\varepsilon_{\boldsymbol{0}} = 0$ in the main text); Π^{x} is a diagonal matrix with elements $\Pi^{x}_{11} := u + v$ and $\Pi^{x}_{22} := u - v$. Assuming that $u^2 > v^2$, one can find a nonunitary transformation to a two-component wave function $\bar{f}$ which satisfies a differential equation that has been well-studied in the Landau-Zener scattering problem. Each component of $\bar{f}$ satisfies Weber's differential equation, which is solved by parabolic cylinder functions (PCFs). The transformation has the form

$$\tilde{f}_{n\boldsymbol{k}} = \alpha(\boldsymbol{k},E)\,\beta(k_y)\sum_{\bar{m}=1}^{2}\bar{T}_{n\bar{m}}\bar{f}_{\bar{m}}(k_y), \tag{F3}$$

$$\alpha(\boldsymbol{k},E) = \exp\left[i\left(k_x - \frac{1}{2}\mathrm{sgn}\big[\Pi^{x}_{11}\big]\,\mathrm{Tr}[(\tilde{\Pi}^{x})^{-1}]\,E\right)k_yl^2\right]$$
$$= \exp\left[i(k_x - k_{xc})k_yl^2\right],$$
$$\beta(k_y) = \exp\left[i\frac{1}{2}\mathrm{sgn}\big[\Pi^{x}_{11}\big]\,\mathrm{Tr}[(\tilde{\Pi}^{x})^{-1}]\,\varepsilon_{\boldsymbol{0}}k_yl^2\right.$$
$$\left. + \frac{1}{4}\mathrm{sgn}\big[\Pi^{x}_{11}\big]\left(\frac{\Pi^{y}_{11}}{|\Pi^{x}_{11}|} + \frac{\Pi^{y}_{22}}{|\Pi^{x}_{22}|}\right)k_y^2l^2\right],$$
$$\bar{T} = (\tilde{\Pi}^{x})^{-1/2}V, \quad \text{with} \quad \tilde{\Pi}^{x} := \mathrm{sgn}\big[\Pi^{x}_{11}\big]\Pi^{x}, \tag{F4}$$

and $V \in \mathrm{SU}(2)$. Note that k_{xc} in the second line is the coordinate of the hyperbolic center, $\tilde{\Pi}^{x}$ (defined above) is positive-definite, and $\bar{T}$ is independent of $\{k_x,k_y,E\}$. This transformation was first derived in Ref. [12], with the assumption that Π^{y} is real owing to spacetime-inversion symmetry [190]. The more general proof that is presented here demonstrates that solubility by PCFs does not require this symmetry.

Proof of transformation to Weber's differential equation.

(i) In the nonunitarily transformed basis

$$\bar{\phi} = \tilde{v}_x^{1/2}\phi, \tag{F5}$$

the eigenvalue equation [Eq. (F2)] takes the form

$$0 = ([\varepsilon_{\bar{\boldsymbol{k}}} - E](\tilde{\Pi}^{x})^{-1} + k_y(\tilde{\Pi}^{x})^{-1/2}\Pi_y(\tilde{\Pi}^{x})^{-1/2}) * \bar{\phi}$$
$$+ \mathrm{sgn}\big[\Pi^{x}_{11}\big]\frac{i}{l^2}\frac{\partial}{\partial k_y} * \bar{\phi}. \tag{F6}$$

(ii) We can remove the terms proportional to identity by

$$\tilde{\phi} = \exp\left\{-\mathrm{sgn}\big[\Pi^{x}_{11}\big]\left[\frac{i}{2}l^2(\varepsilon_{\bar{\boldsymbol{k}}} - E)\mathrm{Tr}[(\tilde{\Pi}^{x})^{-1}]k_y\right.\right.$$
$$\left.\left. + \frac{i}{4}l^2\left(\frac{\Pi^{y}_{11}}{|\Pi^{x}_{11}|} + \frac{\Pi^{y}_{22}}{|\Pi^{x}_{22}|}\right)k_y^2\right]\right\}(\tilde{\Pi}^{x})^{1/2}\phi. \tag{F7}$$

In our model,

$$\frac{E}{2}\mathrm{Tr}[(\tilde{\Pi}^{x})^{-1}] = \frac{E\big(\Pi^{x}_{11} + \Pi^{x}_{22}\big)}{2\Pi^{x}_{11}\Pi^{x}_{22}} = \frac{Eu}{u^2 - v^2} = k_{xc}. \tag{F8}$$

The operator on $\tilde{\phi}$ has the generic form

$$m\sigma_3 + k_y \boldsymbol{v}\cdot\boldsymbol{\sigma} + \mathrm{sgn}\left[\Pi^x_{11}\right]\frac{i}{l^2}\frac{\partial}{\partial k_y}. \tag{F9}$$

In the next steps, we will find a basis where the coefficient of σ_3 is linear in k_y, and that of σ_1 is independent of k_y. Indeed, given any two three-vectors $\boldsymbol{a}$ and $\boldsymbol{v}$ [in our context $\boldsymbol{a} = (0,0,m)$], we can always find a basis where

$$\boldsymbol{a}\cdot\boldsymbol{\sigma} + k_y \boldsymbol{v}\cdot\boldsymbol{\sigma} \tag{F10}$$

is transformed to

$$a_1'\sigma_1 + a_3'\sigma_3 + k_y|v|\sigma_3. \tag{F11}$$

This follows from the homomorphism between SU(2) and SO(3) [114]. From a geometrical perspective, we are looking for a plane in $\mathbb{R}^3$ that is spanned by two vectors $\boldsymbol{a}$ and $\boldsymbol{v}$; we parametrize this plane by (x,z), such that $\vec{z} = \boldsymbol{v}/|v|$. Let us show this explicitly:

(iii) We rotate to a basis where the matrix multiplying k_y is diagonal:

$$\boldsymbol{m}\cdot\boldsymbol{\sigma} + k_y|v|\sigma_3 + \mathrm{sgn}\left[\Pi^x_{11}\right]\frac{i}{l^2}\frac{\partial}{\partial k_y}. \tag{F12}$$

(iv) Shifting the origin of k_y to absorb the m_3 term,

$$m_1\sigma_1 + m_2\sigma_2 + k_y|v|\sigma_3 + \mathrm{sgn}\left[\Pi^x_{11}\right]\frac{i}{l^2}\frac{\partial}{\partial k_y}. \tag{F13}$$

(v) Performing a rotation with $\exp i\sigma_3\theta$,

$$\sqrt{m_1^2+m_2^2}\sigma_1 + k_y|v|\sigma_3 + \mathrm{sgn}\left[\Pi^x_{11}\right]\frac{i}{l^2}\frac{\partial}{\partial k_y}. \tag{F14}$$

Henceforth assuming $\Pi^x_{11} > 0$, we obtain the first-order matrix differential equation in Eq. (277), which is expressed with the hyperbolic parameters $\bar{a}$ and $\bar{b}$ defined in Eq. (268).

The general procedure outlined above, when applied to our minimal model, leads to the particular forms $V = e^{-i\sigma_2\pi/4}\sigma_3$ and $\bar{T}$ of Eq. (276).

The case of $\bar{a} = 0$ was previously discussed in Sec. IX C 2. Henceforth assuming $\bar{a} \neq 0$, and changing variables as

$$z = 2\sqrt{-i}\frac{\sqrt{\bar{\mu}}}{\bar{b}}k_y, \tag{F15}$$

with $\bar{\mu}$ defined in Eq. (271), each component of $\bar{f}$ now satisfies a second-order differential equation:

$$\left[\partial_z^2 - \frac{z^2}{4} + \frac{1}{2} + i\bar{\mu}\right]\bar{f}_{\bar{1}} = 0,$$
$$\left[\partial_z^2 - \frac{z^2}{4} + \frac{1}{2} + (i\bar{\mu}-1)\right]\bar{f}_{\bar{2}} = 0. \tag{F16}$$

The above equations may be identified with Weber's differential equation:

$$\left[\partial_z^2 - \frac{z^2}{4} + \frac{1}{2} + \nu\right]\psi = 0, \tag{F17}$$

which is solved generally by parabolic cylinder functions (or Weber-Hermite functions)

$$\psi = p_1 D_\nu(z) + p_2 \bar{D}_\nu(z), \quad \text{where} \quad \bar{D}_\nu(z) = D_\nu(-z). \tag{F18}$$

$D_\nu(z)$ is an entire function of both ν and z [189] and satisfies the recurrence relation

$$\partial D_\nu|_z + \frac{z}{2}D_\nu(z) = \nu D_{\nu-1}(z). \tag{F19}$$

Note that [189] employs a different notation for the PCF: $D_\nu(z) = U(-1/2-\nu,z)$. $\bar{f}_{\bar{1}}$ and $\bar{f}_{\bar{2}}$ are related as

$$\bar{f}_{\bar{2}} = -\frac{\sqrt{i}}{\sqrt{\bar{\mu}}}\left(\partial + \frac{z}{2}\right)\bar{f}_{\bar{1}}, \tag{F20}$$

which implies, via the recurrence relation, that

$$\bar{f}_{\bar{1}}(z) = p_1 D_{i\bar{\mu}}(z) + p_2 \bar{D}_{i\bar{\mu}}(z),$$
$$\bar{f}_{\bar{2}}(z) = \frac{\sqrt{\bar{\mu}}}{\sqrt{i}}[p_1 D_{i\bar{\mu}-1}(z) - p_2 \bar{D}_{i\bar{\mu}-1}(z)]. \tag{F21}$$

Combining these equations with the asymptotic limits of the PCFs (Eqs. (12.9.1) and (12.9.3) of [188]), we obtain the leading-order terms for $\bar{f}$ in the limits $k_y \to \pm\infty$ (denoted by $\pm$ in the argument):

$$\begin{aligned}
\text{for } \bar{b} > 0, \quad \bar{f}_{\bar{1}}(+) &\approx \varphi(|k_y/\bar{b}|)[p_1 Z + p_2 Z^{-3}],\\
\bar{f}_{\bar{1}}(-) &\approx \varphi(|k_y/\bar{b}|)[p_1 Z^{-3} + p_2 Z],\\
\bar{f}_{\bar{2}}(+) &\approx -p_2\,\varphi^*(|k_y/\bar{b}|)\,\mathcal{G}\,Z^{-1},\\
\bar{f}_{\bar{2}}(-) &\approx p_1\,\varphi^*(|k_y/\bar{b}|)\,\mathcal{G}\,Z^{-1},
\end{aligned} \tag{F22}$$

where we have introduced the variables

$$\varphi(|k_y/\bar{b}|) := \exp i\bar{\mu}\left[\frac{k_y^2}{\bar{b}^2} + \ln\left|\frac{2k_y}{\bar{b}}\right| + \frac{1}{2}\ln\bar{\mu}\right], \tag{F23}$$

$$Z := e^{\pi\bar{\mu}/4}, \quad \text{and}$$
$$\mathcal{G} := \frac{\sqrt{-i}\sqrt{2\pi\bar{\mu}}}{\Gamma(1-i\bar{\mu})} = \sqrt{Z^4 - Z^{-4}}e^{i\arg\Gamma(1+i\bar{\mu})-i\pi/4}. \tag{F24}$$

2. Derivation of connection formula for $|E| > 0$

The goal of this section is to derive the connection formula [cf. Eq. (284)] for interband breakdown at finite energy away from the II-Dirac point.

Let us assume that the semiclassical interval [where we apply the (K_x,k_y) representation] overlaps with the breakdown interval [the $(K_x,0)$ representation]; this overlap region is an interval in k_y satisfying

$$|k_y| \gg l^{-1}, |\bar{b}|, l^{-1}\sqrt{\frac{\bar{b}}{\bar{a}}} \quad \text{and} \quad |k_y| \ll G_y, \tag{F25}$$

with k_y originating from the II-Dirac point, and G_y the reciprocal period. In this overlap region, will apply the general

transformation [Eq. (273)] that relates wave functions in the two representations. Combining Eq. (273) with the WKB form of the (K_x,k_y) wave function in Eq. (283), we obtain that

$$\text{for } k_y>0,\quad \tilde{f}_{l\boldsymbol{k}}=\sum_{n=1}^{2}c_n^{+}w_{n\boldsymbol{k}}M_{ln}(k_y)+O\big(l^{-2},\tfrac{k_y}{G_y}\big),$$
$$\text{for } k_y<0,\quad \tilde{f}_{l\boldsymbol{k}}=\sum_{n=1}^{2}c_n^{-}w_{n\boldsymbol{k}}M_{ln}(k_y)+O\big(l^{-2},\tfrac{k_y}{G_y}\big), \tag{F26}$$

with $w_{n\boldsymbol{k}}$ a Zilberman-Fischbeck function defined in Eqs. (281) and (282), and the overlap matrix defined as

$$M_{ln}(k_y)=\big\langle \tilde{u}_{l,k_x^n(k_y),0}\big|u_{n,k_x^n(k_y),k_y}\big\rangle. \tag{F27}$$

Here, $\tilde{u}_{l,k_x,0}$ and u_{n,k_x,k_y} are classical symbols of operators occurring in the basis functions of the $(K_x,0)$ and (K_x,k_y) representations [cf. Eqs. (A3) and (272)]. Take care in defining M that the band index n appears in both bra and ket.

To make progress, we would need the asymptotic forms of the quantities $w,M,\tilde{f}$ as $k_y\to\pm\infty$:

$$\{w_{n\boldsymbol{k}}^{\pm},M^{\pm}(E),\tilde{f}_{l\boldsymbol{k}}^{\pm}\}:=\lim_{k_y\to\pm\infty}\{w_{n\boldsymbol{k}},M(k_y,E),\tilde{f}_{l\boldsymbol{k}}\}. \tag{F28}$$

We consider them in turn:

$$w_{n\boldsymbol{k}}^{\pm}=\frac{1}{\sqrt{|\bar{v}_x|}}\alpha(\boldsymbol{k},E)[\varphi(|k_y/b|)\lambda^{-1/2}]^{\mp(-1)^n}\mathcal{W}_{n\pm}(k_y), \tag{F29}$$
$$\lambda:=e^{i\bar{\mu}(\ln\bar{\mu}-1)},\quad \lambda^{\pm1/2}=e^{\pm i\bar{\mu}/2(\ln\bar{\mu}-1)}, \tag{F30}$$

with α defined in Eq. (F4) and φ in Eq. (F23); $\mathcal{W}_{n\pm}$ is the single-band Wilson line defined in Eq. (282). For both $n\in\{1,2\}$, $|v_x^n|$ approaches the same value (denoted as $|\bar{v}_x|$) as $k_y\to\pm\infty$. This asymptotic form of the overlap matrix M is derived in Appendix F 3 to be

$$\begin{aligned}&\text{for } E>0,\\ &M_{l1}^{+}=e^{-i\theta_1}\mathcal{W}_{1+}^{-1}\bar{T}_{l\bar{1}},\quad M_{l2}^{+}=-e^{-i\theta_2}\mathcal{W}_{2+}^{-1}\bar{T}_{l\bar{2}},\\ &M_{l1}^{-}=e^{-i\theta_1}\mathcal{W}_{1-}^{-1}\bar{T}_{l\bar{2}},\quad M_{l2}^{-}=e^{-i\theta_2}\mathcal{W}_{2-}^{-1}\bar{T}_{l\bar{1}},\\ &\text{for } E<0,\\ &M_{l1}^{+}=e^{-i\theta_1}\mathcal{W}_{1+}^{-1}\bar{T}_{l\bar{1}},\quad M_{l2}^{+}=e^{-i\theta_2}\mathcal{W}_{2+}^{-1}\bar{T}_{l\bar{2}},\\ &M_{l1}^{-}=-e^{-i\theta_1}\mathcal{W}_{1-}^{-1}\bar{T}_{l\bar{2}},\quad M_{l2}^{-}=e^{-i\theta_2}\mathcal{W}_{2-}^{-1}\bar{T}_{l\bar{1}},\end{aligned} \tag{F31}$$

with $\bar{T}$ a $\boldsymbol{k}$-independent, nonunitary transformation matrix defined in Eq. (276); θ_n are band-dependent phases that should be present on principle (see discussion in Appendix F 3) but do not ultimately affect the quantization condition.

Inserting Eqs. (F29)–(F31) on the right-hand side of Eq. (F26), the cancellation of the Wilson lines leads to

$$\begin{aligned}\text{for } E>0,\quad &\sum_n c_n^{+}w_{n\boldsymbol{k}}^{+}M_{ln}^{+}=\frac{\alpha(\boldsymbol{k})}{\sqrt{|\bar{v}_x|}}\{c_1^{+}(e^{-i\theta_1}\bar{T}_{l\bar{1}})(\lambda^{-1/2}\varphi)+c_2^{+}(-e^{-i\theta_2}\bar{T}_{l\bar{2}})(\lambda^{1/2}\varphi^{*})\},\\ &\sum_n c_n^{-}w_{n\boldsymbol{k}}^{-}M_{ln}^{-}=\frac{\alpha(\boldsymbol{k})}{\sqrt{|\bar{v}_x|}}\{c_1^{-}(e^{-i\theta_1}\bar{T}_{l\bar{2}})(\lambda^{1/2}\varphi^{*})+c_2^{-}(e^{-i\theta_2}\bar{T}_{l\bar{1}})(\lambda^{-1/2}\varphi)\};\\ \text{for } E<0,\quad &\sum_n c_n^{+}w_{n\boldsymbol{k}}^{+}M_{ln}^{+}=\frac{\alpha(\boldsymbol{k})}{\sqrt{|\bar{v}_x|}}\{c_1^{+}(e^{-i\theta_1}\bar{T}_{l\bar{1}})(\lambda^{-1/2}\varphi)+c_2^{+}(e^{-i\theta_2}\bar{T}_{l\bar{2}})(\lambda^{1/2}\varphi^{*})\},\\ &\sum_n c_n^{-}w_{n\boldsymbol{k}}^{-}M_{ln}^{-}=\frac{\alpha(\boldsymbol{k})}{\sqrt{|\bar{v}_x|}}\{c_1^{-}(-e^{-i\theta_1}\bar{T}_{l\bar{2}})(\lambda^{1/2}\varphi^{*})+c_2^{-}(e^{-i\theta_2}\bar{T}_{l\bar{1}})(\lambda^{-1/2}\varphi)\}.\end{aligned} \tag{F32}$$

From Eq. (F3), and applying that $\beta=0$ in our minimal model (since Π^y is off-diagonal and $\varepsilon_{\boldsymbol{0}}=0$),

$$\tilde{f}_{l\boldsymbol{k}}^{\pm}=\alpha(\boldsymbol{k})\sum_{\bar{m}=\bar{1}}^{\bar{2}}\bar{T}_{l\bar{m}}\bar{f}_{\bar{m}}^{\pm}(k_y), \tag{F33}$$

where $\bar{m}$ labels the diabatic basis vectors, and $\bar{f}$ satisfies the matrix differential equation in Eq. (277). Let us relate $\bar{f}$ above and below the Dirac point. Both $\bar{a}$ and $\bar{b}$ change sign across $E=0$, but $\bar{a}/\bar{b}$ does not. We will exploit a symmetry of Eq. (277): if $\bar{f}$ is a solution for $\bar{a}>0$, $\bar{r}(k_y)=\bar{f}(-k_y)$ is a solution for $\bar{a}<0$. Therefore, the asymptotic forms above and below the Dirac point are related by $\bar{r}_{\bar{i}}^{\pm}(k_y)=\bar{f}_{\bar{i}}^{\mp}(-k_y)$. Further employing that $\bar{f}$ only depends on k_y through $\varphi(|k_y/\bar{b}|)$,

$$\bar{r}_{\bar{i}}^{\pm}(|k_y/\bar{b}|)=\bar{f}_{\bar{i}}^{\mp}(|k_y/\bar{b}|). \tag{F34}$$

Utilizing the asymptotic forms of $\bar{f}$ in Eqs. (F22)–(F24) for $E>0$ (recall that the sign of E and $\bar{b}$ are identical with our assumption that $u,v,w>0$), in combination with Eq. (F34), we derive

$$\begin{aligned}\text{for } E>0,\quad &\tilde{f}_{l\boldsymbol{k}}^{+}=\alpha(\boldsymbol{k})\{\bar{T}_{l\bar{1}}\varphi(p_1Z+p_2Z^{-3})+\bar{T}_{l\bar{2}}\varphi^{*}(-p_2\mathcal{G}Z^{-1})\},\\ &\tilde{f}_{l\boldsymbol{k}}^{-}=\alpha(\boldsymbol{k})\{\bar{T}_{l\bar{1}}\varphi(p_1Z^{-3}+p_2Z)+\bar{T}_{l\bar{2}}\varphi^{*}(p_1\mathcal{G}Z^{-1})\};\\ \text{for } E<0,\quad &\tilde{f}_{l\boldsymbol{k}}^{-}=\alpha(\boldsymbol{k})\{\bar{T}_{l\bar{1}}\varphi(p_1Z+p_2Z^{-3})+\bar{T}_{l\bar{2}}\varphi^{*}(-p_2\mathcal{G}Z^{-1})\},\\ &\tilde{f}_{l\boldsymbol{k}}^{+}=\alpha(\boldsymbol{k})\{\bar{T}_{l\bar{1}}\varphi(p_1Z^{-3}+p_2Z)+\bar{T}_{l\bar{2}}\varphi^{*}(p_1\mathcal{G}Z^{-1})\}.\end{aligned} \tag{F35}$$

Comparing this with Eq. (F32), we identify

$$\frac{1}{\sqrt{|\bar{v}_x|}}\begin{pmatrix} c_1^- e^{-i\theta_1} \\ c_2^- e^{-i\theta_2}\end{pmatrix} = \begin{pmatrix} \lambda^{-1/2}\,\mathcal{G}\,Z^{-1} & 0 \\ \lambda^{1/2}\,Z^{-3} & \lambda^{1/2}\,Z\end{pmatrix}\sigma_1^{(1-\mathrm{sgn}[E])/2}\begin{pmatrix} p_1 \\ p_2\end{pmatrix},$$
$$\frac{1}{\sqrt{|\bar{v}_x|}}\begin{pmatrix} c_1^+ e^{-i\theta_1} \\ c_2^+ e^{-i\theta_2}\end{pmatrix} = \begin{pmatrix} \lambda^{1/2}\,Z & \lambda^{1/2}\,Z^{-3} \\ 0 & \lambda^{-1/2}\,\mathcal{G}\,Z^{-1}\end{pmatrix}\sigma_1^{(1-\mathrm{sgn}[E])/2}\begin{pmatrix} p_1 \\ p_2\end{pmatrix}, \tag{F36}$$

with $\sigma_1^1 := \sigma_1$ a Pauli matrix, and $\sigma_1^0 := I$ the 2×2 identity. Since the expressions for positive and negative energy differ only in the relabeling of dummy variables $p_1 \leftrightarrow p_2$, the scattering matrix is independent of the sign of the energy. Removing the p_1 and p_2 from our equations, we finally relate c^- to c^+ by the scattering matrix in Eq. (284).

3. Asymptotic form of overlap matrix M

The overlap matrix M defined in Eq. (F27) may be viewed as a basis transformation between the Luttinger-Kohn and crystal-momentum representations, as we have reviewed in Sec. III E. M is determined, with an accuracy of $O(k_y/G_y)$, from the following eigenvalue equation:

$$\left[-E + k_x^n(k_y,E)\,\Pi^x(\mathbf{0}) + k_y\Pi^y(\mathbf{0})\right]_{ml} M_{ln}(k_y,E) = 0. \tag{F37}$$

It is assumed that $u_{nk_x^n(k_y,E)k_y}$ [occurring in Eq. (F27)] are energy eigenfunctions of $\hat{H}_0(\boldsymbol{k})$ with eigenvalue $\varepsilon_{nk_x^nk_y} = E$. The goal of this section is to derive the asymptotic form of M [denoted by $M^\pm$ in Eq. (F31)] as $k_y \to \pm\infty$.

Before a detailed proof of Eq. (F31), we argue for the form of $M^\pm$:

(i) In the limit $k_y \to \pm\infty$, the adiabatic basis of energy bands (labeled by n) coincides with the diabatic basis (labeled by $\bar{n}$) up to a phase, as we have argued in the caption of Fig. 8. Each column of $M^\pm$ is therefore proportional to a column of the matrix $\bar{T}_{n\bar{n}}$ defined in Eq. (276), which transforms between adiabatic and diabatic bases.

(ii) What remains is to argue for the proportionality phase factors. If we ignore $\mathcal{W}$ and $e^{i\theta_j}$, there remains a -1 phase factor which reflects the π Berry phase acquired in the 2π rotation of pseudospin-half. Indeed, we might view $\bar{a}\tau_1 + \frac{\bar{a}}{\bar{b}}k_y\tau_3$ as the Hamiltonian of a pseudospin coupled to a pseudo-magnetic field, and label the diabatic basis $|\pm1\rangle$ according to its eigenvalue under τ_3. The diabatic basis coincides, modulo a phase factor, with the adiabatic basis in the two limits $k_y \gg |\bar{b}|$ and $k_y \ll -|\bar{b}|$. As k_y is varied from $+\infty$ to $-\infty$, the adiabatic basis is parallel-transported along $k_x^n(k_y,E)$ as $|+\rangle \to e^{i\phi_{+-}}|-\rangle$ and $|-\rangle \to e^{i\phi_{-+}}|+\rangle$. The product of the two phases $e^{i(\phi_{+-}+\phi_{-+})} = -1$ is independent of phase redefinitions of the diabatic basis: $|\pm\rangle \to |\pm\rangle e^{i\varphi_\pm}$. To explain this independence, we may view the combined parallel transport of $|+\rangle \to e^{i\phi_{+-}}|-\rangle \to e^{i(\phi_{+-}+\phi_{-+})}|+\rangle$ as the adiabatic rotation of a pseudospin-half by 2π within a plane. $e^{i(\phi_{+-}+\phi_{-+})} = -1$ may then be identified with the Berry phase, which is half the solid angle [20] subtended by the rotation.

(iii) In the Bloch problem, there is an intrinsic ambiguity in the definition of nondegenerate energy bands $|u_{n\boldsymbol{k}}\rangle$, which may arbitrarily be redefined by a $\boldsymbol{k}$-dependent phase. This phase ambiguity may be separated into two contributions: (ii-a) the single-band Berry connection $\mathfrak{X}_n(\boldsymbol{k})$ encodes the phase relationship between infinitesimally separated wave vectors $\boldsymbol{k}$ and $\boldsymbol{k}+\delta\boldsymbol{k}$. (ii-b) In addition, there remains, for each band labeled by $n = 1,2$, a global phase ambiguity encoded by $e^{i\theta_n}$, which explains their presence in Eq. (F31). That is, given a fixed connection, there remains a gauge freedom in redefining each band by a $\boldsymbol{k}$-independent phase. The ambiguity described in (ii-a) is expressed in Eq. (F31) as the integral of the connection along the constant-energy band contour in Eq. (282). We remark that there is no sense in which $\mathcal{W}_{n\pm}$ asymptotically converges to a unique phase factor; reassuringly, the final expression for the quantization condition involves only closed-loop integrals of $\mathfrak{X}_n$.

Proof of Eq. (F31). Let us make the argument of (i) precise. One may verify that

$$0 = \left[-k_{xc} + \bar{a}\tau_1 + \bar{k}_x^n + \frac{\bar{a}}{\bar{b}}k_y\tau_3\right]_{ml}[\bar{T}^{-1}M]_{ln}. \tag{F38}$$

For large $|k_y/\bar{b}| \gg 1$, we may neglect $\bar{a}\tau_1$ relative to $\bar{a}k_y\tau_3/\bar{b}$:

$$0 \approx \left[-k_{xc} + \bar{k}_x^n(k_y,E) + \frac{\bar{a}}{\bar{b}}k_y\tau_3\right]_{ml}[\bar{T}^{-1}M]_{ln} \tag{F39}$$

which determines the columns of $M^\pm$ up to a phase (denoted as γ below):

$$M_{l1}^+ = e^{i\gamma_{1+}}\bar{T}_{l\bar{1}}, \quad M_{l2}^- = e^{i\gamma_{2-}}\bar{T}_{l\bar{1}},$$
$$M_{l2}^- = e^{i\gamma_{2+}}\bar{T}_{l\bar{2}}, \quad M_{l1}^- = e^{i\gamma_{1-}}\bar{T}_{l\bar{2}}. \tag{F40}$$

The first line follows from setting $[\bar{T}^{-1}M]_{ln} \propto \delta_{l,1}$ in Eq. (F39), which leads to $0 \approx -k_{xc} + \bar{k}_x^n + \bar{a}k_y/\bar{b}$; for $k_y \gg |\bar{b}|$, this corresponds to the left band contour ($n = 1$), and for $k_y \ll -|\bar{b}|$ to the right ($n = 2$).

Our next step is to derive the phases $\gamma_{n\pm}$. As an intermediate step, we determine $\gamma_{n\pm}$ in a special basis for the energy bands which we denote as

$$\left\{\breve{u}_{n\bar{k}_x(k_y,E),k_y}\middle|n \in \{1,2\}, \quad k_y \in \mathbb{R}\right\}. \tag{F41}$$

Up to a relabeling of band indices, $\breve{u}$ and $\tilde{u}$ are defined to be continuous where their domains (in $\boldsymbol{k}$ space) overlap—at the two hyberbolic vertices. Explicitly, if we define the 2×2 overlap matrix,

$$\breve{M}_{ln}(k_y,E) = \left\langle \tilde{u}_{l,\bar{k}_x^n(k_y,E),0}\middle|\breve{u}_{n,\bar{k}_x^n(k_y,E),k_y}\right\rangle,$$
$$\text{then } \breve{M}(0,E) = \sigma_1^{(1-\mathrm{sgn}[E])/2}. \tag{F42}$$

The reason for this dependence on the sign of E: we have defined $\tilde{u}_1$ to have a larger velocity $(\partial\varepsilon/\partial k_x)$ than $\tilde{u}_2$, so $\tilde{u}_1$ corresponds to the left hyperbolic vertex ($\breve{u}_1$) at positive energy, and to the right hyperbolic vertex ($\breve{u}_2$) at negative energy. We further insist that $\breve{u}$ satisfies the parallel-transport condition, i.e., for any segment of a hyperbolic arm, $0 = \exp(-\int\langle\breve{u}_n|\nabla_{\boldsymbol{k}}\breve{u}_n\rangle\cdot d\boldsymbol{k})$. This condition, combined with the

reality of the pseudospin Hamiltonian ($\bar{a}\tau_1 + \frac{\bar{a}}{\bar{b}}k_y\tau_3$), ensures the reality of $\breve{M}$ for all k_y. It is simple to find a real function that interpolates between the known values of $\breve{M}$ at $k_y = 0$ [cf. Eq. (F42)] and at $\pm\infty$ (each column of $\breve{M}$ must be proportional to a column of the real matrix $\bar{T}$). The result is

$$\begin{aligned}
\text{for } E > 0, \quad & \breve{M}_{l1}(+) = \bar{T}_{l\bar{1}}, \quad \breve{M}_{l2}(+) = -\bar{T}_{l\bar{2}}, \\
& \breve{M}_{l1}(-) = \bar{T}_{l\bar{2}}, \quad \breve{M}_{l2}(-) = \bar{T}_{l\bar{1}}; \\
\text{for } E < 0, \quad & \breve{M}_{l1}(+) = \bar{T}_{l\bar{1}}, \quad \breve{M}_{l2}(+) = \bar{T}_{l\bar{2}}, \\
& \breve{M}_{l1}(-) = -\bar{T}_{l\bar{2}}, \quad \breve{M}_{l2}(-) = \bar{T}_{l\bar{1}}.
\end{aligned} \tag{F43}$$

Given that Eq. (F43) holds for the special basis $\breve{u}$, it follows that Eq. (F31) holds for any basis

$$\left\{u_{n\bar{k}_x^n(k_y),k_y}\middle| k_y \in \mathbb{R}\right\} \tag{F44}$$

that is differentiable with respect to k_y. Indeed, we may define the phase mismatch between u_n and $\breve{u}_n$ at the hyperbolic vertex $(\bar{k}_x^n(0),0)$ as

$$u_{n\bar{k}_x^n(0),0} = e^{-i\theta_n}\breve{u}_{n\bar{k}_x^n(0),0}. \tag{F45}$$

Finally, the Wilson line [$\mathcal{W}_{n\pm}$, as defined in Eq. (282)] accounts for the additional phase mismatch between u_n and $\breve{u}_n$, which originates from u_n not satisfying the parallel-transport condition. To recapitulate,

$$\breve{u}_{n,\boldsymbol{k}_x^n,k_y\to\pm\infty} = e^{i\theta_n}\,\mathcal{W}_{n\pm}\,u_{n,\boldsymbol{k}_x^n,k_y\to\pm\infty}, \tag{F46}$$

which may be substituted into Eq. (F43) to derive Eq. (F31).

4. Perturbative treatment of quasirandom spectrum

a. Case study of interband breakdown: Single II-Dirac graph with $\bar{\mu} \approx 0$

From Eq. (294), we identify

$$\begin{aligned}
f(E,B;\tau) &= f(E,B;0) + \delta\tau(E,B) f_1(E,B), \\
f_0(E,B) &:= f(E,B;0) = \cos\left[\frac{\Omega_1+\Omega_2}{2}\bigg|_{E,l^2}\right], \\
f_1(E,B) &= -\cos\left[\frac{\Omega_1-\Omega_2}{2}\bigg|_{E,l^2} + \omega(\bar{\mu})\right], \\
\delta\tau(E,B) &= \sqrt{1-\rho^2}, \quad \rho(\bar{\mu}) = e^{-\pi\bar{\mu}}, \quad \tau_0 = 0,
\end{aligned} \tag{F47}$$

with $\bar{\mu} \propto E^2/B$ and ω defined in Eqs. (271) and (285), respectively.

In the semiclassical limit $\bar{\mu} \to 0$, the Landau fan is determined by Eq. (295), and the first-order correction to the Landau fan is given in Eq. (300). Just as in Eq. (300), we will employ the shorthand $O' = \partial O/\partial E$ throughout this Appendix. Further assuming that $(S_1 + S_2)$ is slowly varying on the scale of $E_n^0 - E_0^0$ [i.e., $(S_1+S_2)' = O(1)$], and restricting ourselves to $n = O(1)$, Eqs. (295) and (300) particularize to

$$\begin{aligned}
E_n^0(B) &= E_0^0 + \frac{2n\pi}{l^2(S_1+S_2)'|_{E_0^0}} + O(l^{-4}), \\
\delta E_n^1(B) &= \frac{2\sqrt{\pi}(-1)^{n+1}}{l(S_1+S_2)'|_{E_n^0}} \frac{v}{\sqrt{w}(u^2-v^2)^{3/4}} \\
&\quad \times \left\{E_0^0 + \frac{2n\pi}{l^2(S_1+S_2)'|_{E_0^0}}\right\} \\
&\quad \times \sin\left[\omega + \frac{l^2(S_1-S_2)}{2}\right]\bigg|_{E_n^0} + O(\bar{\mu}^{3/2}l^{-2}, l^{-5}).
\end{aligned} \tag{F48}$$

The validity of the last expression rests on a double constraint on the field: it cannot be too large, as reflected in the $O(l^{-5})$ uncertainty; on the other hand for any nonzero energy, the field also cannot be too small, since $\bar{\mu}^{3/2}l^{-2} \propto E^3 l$. With these caveats in mind, we observe that the amplitude of δE_n^1 goes as $B^{1/2}$ at weak field; for $n \neq 0$, this crosses over to a $B^{3/2}$ dependence at intermediate field. Equation (300) and the second line of Eq. (F48) are derived at the end of this section (Appendix F 4 b).

b. Derivation of first-order correction Eq. (300)

As defined in Eq. (291), the domain of Ω_j does not include $E = 0$. Due to the continuity of the quantization condition Eq. (294) across $E = 0$ (as we had argued in Sec. IX D 1), we may as well extend the domain by $\Omega_j(0) := \Omega_j(0^+)$, with 0^+ an infinitesimally small positive quantity; our results would be unchanged if we had instead chosen $\Omega_j(0) := \Omega_j(0^-)$. We may then express the extended functions concisely as

$$\begin{aligned}
\frac{\Omega_1-\Omega_2}{2} &= \frac{l^2(S_1-S_2)}{2} - \frac{\pi}{2} + \pi\Theta^+(E), \\
\frac{\Omega_1+\Omega_2}{2} &= \frac{l^2(S_1+S_2)}{2} + \frac{\pi}{2},
\end{aligned} \tag{F49}$$

with the step function defined by

$$\Theta^+(x) = \begin{cases} 0 & \text{for } x \geqslant 0, \\ 1 & \text{for } x < 0. \end{cases} \tag{F50}$$

Inserting Eq. (F49) into Eq. (F47),

$$f_0 = -\sin\left[\frac{l^2(S_1+S_2)}{2}\right], \tag{F51}$$

$$\left.\frac{\partial f_0}{\partial E}\right|_{E_n^0} = \left.\frac{(-1)^{n+1}}{2} l^2(S_1+S_2)'\right|_{E_n^0}, \tag{F52}$$

$$\begin{aligned}
f_1 &= -\cos\left[\omega + \frac{l^2(S_1-S_2)}{2} - \frac{\pi}{2} + \pi\Theta^+(E)\right] \\
&= (-1)^{\Theta^+(E)+1}\sin\left[\omega + \frac{l^2(S_1-S_2)}{2}\right].
\end{aligned} \tag{F53}$$

In the second equality, we applied that $\sin[\Omega_1/2 + \Omega_2/2] = (-1)^n$ when evaluated at E_n^0, as deducible from Eq. (295). Inserting Eqs. (F52) and (F53) into Eq. (299), the first-order

correction in energy is then

$$\delta E_n^1 = \left.\frac{-\delta\tau f_1}{\frac{\partial f_0}{\partial E}}\right|_{E_n^0(B)} = 2(-1)^{n+1+\Theta^+}\frac{\delta\tau}{l^2(S_1+S_2)'}\sin\left[\omega+\frac{l^2(S_1-S_2)}{2}\right]\Bigg|_{E_n^0}. \tag{F54}$$

Since we are in a parameter regime where $\bar{\mu}$ is small, $\delta\tau$ [defined in Eq. (F47)] is approximated by

$$\delta\tau = \sqrt{2\pi\bar{\mu}} + O(\bar{\mu}^{3/2}) = \sqrt{\pi}\frac{v}{\sqrt{w}(u^2-v^2)^{3/4}}\,l\,|E| + O(\bar{\mu}^{3/2}), \tag{F55}$$

where we have utilized the definition of $\bar{\mu}$ in Eq. (271). Since

$$(-1)^{\Theta^+(E)}|E| = E = \mathrm{sgn}[E]|E|, \tag{F56}$$

we may just as well replace $(-1)^{\Theta^+}$ in Eq. (F54) by $\mathrm{sgn}[E]$, and finally obtain Eq. (300) as desired. For Landau levels indexed by $n = O(1)$, we may substitute Eqs. (F55) and (F48) into Eq. (300), and derive the second line of Eq. (F48).

[1] R. Peierls, Z. Phys. **80**, 763 (1933).
[2] L. Onsager, The Lond., Edinb., Dublin Philos. Mag. J. Sci. **43**, 1006 (1952).
[3] L. M. Lifshitz and A. Kosevich, Dokl. Akad. Nauk SSSR **96**, 963 (1954).
[4] L. M. Lifshitz and A. Kosevich, JETP **2**, 636 (1956).
[5] J. M. Luttinger, Phys. Rev. **84**, 814 (1951).
[6] D. Shoenberg, *Magnetic Oscillations in Metals* (Cambridge University Press, Cambridge, UK, 1984).
[7] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Thomson Learning, Boston, MA, 1976).
[8] T. Champel and V. P. Mineev, Philos. Mag., Part B **81**, 55 (2001).
[9] W. J. de Haas and P. M. van Alphen, Proc. Netherlands Roy. Acad. Sci. **33**, 1106 (1930).
[10] L. W. Shubnikov and W. J. de Haas, Proc. Netherlands Roy. Acad. Sci. **33**, 130 (1930).
[11] M. Y. Azbel, JETP **12**, 891 (1961).
[12] A. Slutskin, JETP **26**, 474 (1968).
[13] M. H. Cohen and L. M. Falicov, Phys. Rev. Lett. **7**, 231 (1961).
[14] E. I. Blount, Phys. Rev. **126**, 1636 (1962).
[15] A. B. Pippard, Proc. R. Soc. London, Ser. A **270**, 1 (1962).
[16] A. B. Pippard, Proc. R. Soc. London, Ser. A **256**, 317 (1964).
[17] W. G. Chambers, Phys. Rev. **149**, 493 (1966).
[18] M. Kaganov and A. Slutskin, Phys. Rep. **98**, 189 (1983).
[19] Y. N. Proshin and N. K. Useinov, Phys. Usp. **38**, 39 (1995).
[20] M. V. Berry, Proc. R. Soc. London A **392**, 45 (1984).
[21] G. P. Mikitik and Y. V. Sharlai, Phys. Rev. Lett. **82**, 2147 (1999).
[22] M.-C. Chang and Q. Niu, Phys. Rev. B **53**, 7010 (1996).
[23] W. Kohn, Phys. Rev. **115**, 1460 (1959).
[24] L. Roth, J. Phys. Chem. Solids **23**, 433 (1962).
[25] L. M. Roth, Phys. Rev. **145**, 434 (1966).
[26] M. Wilkinson and R. J. Kay, Phys. Rev. Lett. **76**, 1896 (1996).
[27] R. G. Littlejohn and W. G. Flynn, Phys. Rev. Lett. **66**, 2839 (1991).
[28] R. G. Littlejohn and W. G. Flynn, Phys. Rev. A **44**, 5239 (1991).
[29] C. Emmrich and A. Weinstein, Commun. Math. Phys. **176**, 701 (1996).
[30] G. Panati, H. Spohn, and S. Teufel, Phys. Rev. Lett. **88**, 250405 (2002).
[31] G. P. Mikitik and Y. V. Sharlai, J. Exp. Theor. Phys. **87**, 747 (1998).
[32] A. Alexandradinata, C. Wang, W. Duan, and L. Glazman, Phys. Rev. X **8**, 011027 (2018).
[33] G. Sundaram and Q. Niu, Phys. Rev. B **59**, 14915 (1999).
[34] M. Chang and Q. Niu, J. Phys.: Condens. Matter **20**, 193202 (2008).
[35] D. Culcer, Y. Yao, and Q. Niu, Phys. Rev. B **72**, 085110 (2005).
[36] A. Alexandradinata and L. Glazman, Phys. Rev. Lett. **119**, 256601 (2017).
[37] H. Nielsen and M. Ninomiya, Phys. Lett. B **130**, 389 (1983).
[38] X. Wan, A. M. Turner, A. Vishwanath, and S. Y. Savrasov, Phys. Rev. B **83**, 205101 (2011).
[39] A. A. Burkov, M. D. Hook, and L. Balents, Phys. Rev. B **84**, 235126 (2011).
[40] B. Bradlyn, J. Cano, Z. Wang, M. G. Vergniory, C. Felser, R. J. Cava, and B. A. Bernevig, Science **353**, aaf5037 (2016).
[41] S. Murakami, New J. Phys. **9**, 356 (2007).
[42] Z. Wang, Y. Sun, X.-Q. Chen, C. Franchini, G. Xu, H. Weng, X. Dai, and Z. Fang, Phys. Rev. B **85**, 195320 (2012).
[43] S. M. Young and C. L. Kane, Phys. Rev. Lett. **115**, 126803 (2015).
[44] T. Zhang, Z. Song, A. Alexandradinata, H. Weng, C. Fang, L. Lu, and Z. Fang, Phys. Rev. Lett. **120**, 016401 (2018).
[45] M. Serbyn and L. Fu, Phys. Rev. B **90**, 035402 (2014).
[46] T. E. O'Brien, M. Diez, and C. W. J. Beenakker, Phys. Rev. Lett. **116**, 236401 (2016).
[47] M. Koshino, Phys. Rev. B **94**, 035202 (2016).
[48] G. H. Wannier and D. R. Fredkin, Phys. Rev. **125**, 1910 (1962).
[49] J. Zak, Phys. Rev. **168**, 686 (1968).
[50] G. Nenciu, Rev. Mod. Phys. **63**, 91 (1991).
[51] G. Zil'berman, JETP **5**, 208 (1957).
[52] H. J. Fischbeck, Phys. Status Solidi B **38**, 11 (1970).
[53] W. Yao, D. Xiao, and Q. Niu, Phys. Rev. B **77**, 235406 (2008).
[54] Z. Wang, A. Alexandradinata, R. J. Cava, and B. A. Bernevig, Nature (London) **532**, 189 (2016).
[55] T. H. Hsieh, H. Lin, J. Liu, W. Duan, A. Bansil, and L. Fu, Nat. Commun. **3**, 982 (2012).
[56] A. A. Soluyanov, D. Gresch, Z. Wang, Q. Wu, M. Troyer, X. Dai, and B. A. Bernevig, Nature (London) **527**, 495 (2015).
[57] H. Isobe and N. Nagaosa, Phys. Rev. Lett. **116**, 116803 (2016).
[58] E. J. Bergholtz, Z. Liu, M. Trescher, R. Moessner, and M. Udagawa, Phys. Rev. Lett. **114**, 016806 (2015).

[59] L. Muechler, A. Alexandradinata, T. Neupert, and R. Car, Phys. Rev. X **6**, 041069 (2016).
[60] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, M. I. Katsnelson, I. V. Grigorieva, S. V. Dubonos, and A. A. Firsov, Nature (London) **438**, 197 (2005).
[61] J. Zak, Phys. Rev. Lett. **54**, 1075 (1985).
[62] C. Fang, M. J. Gilbert, X. Dai, and B. A. Bernevig, Phys. Rev. Lett. **108**, 266802 (2012).
[63] E. I. Blount, in *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic Press, New York, 1962), Vol. 13.
[64] J. Zak, Phys. Rev. Lett. **62**, 2747 (1989).
[65] F. Wilczek and A. Zee, Phys. Rev. Lett. **52**, 2111 (1984).
[66] M. Lax, *Symmetry Principles in Solid State and Molecular Physics* (Wiley-Interscience, Hoboken, NJ, 1974).
[67] K. Shiozaki, M. Sato, and K. Gomi, Phys. Rev. B **95**, 235425 (2017).
[68] T. L. Hughes, E. Prodan, and B. A. Bernevig, Phys. Rev. B **83**, 245132 (2011).
[69] C. Fang, M. J. Gilbert, and B. A. Bernevig, Phys. Rev. B **86**, 115112 (2012).
[70] A. Alexandradinata, Z. Wang, and B. A. Bernevig, Phys. Rev. X **6**, 021008 (2016).
[71] See related footnote just after Ref. [119].
[72] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).
[73] A useful set of introductory notes can be found at https://terrytao.wordpress.com/2012/10/07/some-notes-on-weyl-quantisation.
[74] E. M. Lifshitz and L. P. Pitaevskii, *Statistical Physics, Part 2* (Elsevier, Amsterdam, Netherlands, 1980).
[75] J. Zak, Phys. Rev. Lett. **67**, 2565 (1991).
[76] P. Misra, Phys. Lett. A **33**, 339 (1970).
[77] J. Zak, Phys. Lett. A **117**, 367 (1986).
[78] P. G. Harper, J. Phys.: Condens. Matter **3**, 3047 (1991).
[79] E. Brown, Phys. Rev. **133**, A1038 (1964).
[80] D. R. Hofstadter, Phys. Rev. B **14**, 2239 (1976).
[81] M. Wilkinson, J. Phys. A: Math. Gen. **20**, 4337 (1987).
[82] M. Wilkinson, J. Phys. A: Math. Gen. **27**, 8123 (1994).
[83] S. Freund and S. Teufel, Anal. PDE **9**, 773 (2016).
[84] H. J. Schellnhuber and G. M. Obermair, Phys. Rev. Lett. **45**, 276 (1980).
[85] L. Michel and J. Zak, Phys. Rev. B **59**, 5998 (1999).
[86] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).
[87] J. Zak, Phys. Rev. B **23**, 2824 (1981).
[88] R. A. Evarestov and V. P. Smirnov, Phys. Status Solidi **122**, 231 (1984).
[89] H. Bacry, Commun. Math. Phys. **153**, 359 (1993).
[90] B. Bradlyn, L. Elcoro, J. Cano, M. G. Vergniory, Z. Wang, C. Felser, M. I. Aroyo, and B. A. Bernevig, Nature (London) **547**, 298 (2017).
[91] F. A. Butler and E. Brown, Phys. Rev. **166**, 630 (1968).
[92] G. Panati, H. Spohn, and S. Teufel, Commun. Math. Phys. **242**, 547 (2003).
[93] R. Rammal and J. Bellissard, J. Phys. (France) **51**, 1803 (1990).
[94] O. Gat and J. E. Avron, Phys. Rev. Lett. **91**, 186801 (2003).
[95] J. B. Keller, Ann. Phys. **4**, 180 (1958).
[96] R. K. Pathria, *Statistical Mechanics*, 2nd ed. (Elsevier, Amsterdam, Netherlands, 1996).
[97] D. Xiao, J. Shi, and Q. Niu, Phys. Rev. Lett. **95**, 137204 (2005).
[98] C. Duval, Z. Horváth, P. A. Horváthy, L. Martina, and P. C. Stichel, Phys. Rev. Lett. **96**, 099701 (2006).
[99] C. Duval, Z. Horváth, P. A. Horváthy, L. Martina, and P. C. Stichel, Mod. Phys. Lett. B **20**, 373 (2006).
[100] K. Yabana and H. Horiuchi, Prog. Theor. Phys. **75**, 592 (1986).
[101] L. A. Fal'kovskii, JETP **49**, 609 (165).
[102] B. M. Gorbovitskii and V. I. Perel, Zh. Eksp. Teor. Fiz. **85**, 1812 (1983).
[103] V. Schweikhard, I. Coddington, P. Engels, V. P. Mogendorff, and E. A. Cornell, Phys. Rev. Lett. **92**, 040404 (2004).
[104] G. Juzeliūnas and P. Öhberg, Phys. Rev. Lett. **93**, 033602 (2004).
[105] G. Juzeliūnas, P. Öhberg, J. Ruseckas, and A. Klein, Phys. Rev. A **71**, 053614 (2005).
[106] D. Jaksch and P. Zoller, New J. Phys. **5**, 56 (2003).
[107] A. Jacob, P. Öhberg, G. Juzeliūnas, and L. Santos, New J. Phys. **10**, 045022 (2008).
[108] W. Kohn, Proc. Phys. Soc. **72**, 1147 (1958).
[109] M. Y. Azbel, JETP **46**, 929 (1964).
[110] W. Y. Hsu and L. M. Falicov, Phys. Rev. B **13**, 1595 (1976).
[111] M. V. Berry and K. E. Mount, Rep. Prog. Phys. **35**, 315 (1972).
[112] S. Y. Zhou, G.-H. Gweon, A. V. Fedorov, P. N. First, W. A. de Heer, D.-H. Lee, F. Guinea, A. H. Castro Neto, and A. Lanzara, Nat. Mater. **6**, 770 (2007).
[113] G. Giovannetti, P. A. Khomyakov, G. Brocks, P. J. Kelly, and J. van den Brink, Phys. Rev. B **76**, 073103 (2007).
[114] M. Tinkham, *Group Theory and Quantum Mechanics* (Dover, Mineola, NY, 2003).
[115] M. H. Cohen and E. I. Blount, Philos. Mag. **5**, 115 (1960).
[116] J. J. Sakurai, *Modern Quantum Mechanics* (Pearson Education, London, 2005).
[117] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).
[118] A. A. Soluyanov and D. Vanderbilt, Phys. Rev. B **83**, 035108 (2011).
[119] S. Weinberg, *The Quantum Theory of Solids, Volume 1: Foundations* (Cambridge University Press, Cambridge, UK, 2005).
[120] The last four rows do not describe a standard representation of a point group, owing to the nontrivial action of g in $\boldsymbol{k}$ space. To map the equation to a more standard representation, we may collect $\{\breve{g}_i\}_{i=1}^{N}$ as blocks in a single permutation matrix representing an N cycle. This larger matrix indeed represents g in P_g. With slight abuse of language, we will anyway refer to $\breve{g}_i$ as a representation of P_g.
[121] H. Hiller, Am. Math. Mon. **93**, 765 (1986).
[122] B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).
[123] A. Alexandradinata, X. Dai, and B. A. Bernevig, Phys. Rev. B **89**, 155114 (2014).
[124] A. Alexandradinata and B. A. Bernevig, Phys. Rev. B **93**, 205104 (2016).
[125] J. Höller and A. Alexandradinata, arXiv:1708.02943.
[126] Lecture notes at http://math.ucla.edu/~sharifi/groupcoh.pdf.
[127] G. Li, B. Yan, Z. Wang, and K. Held, Phys. Rev. B **95**, 035102 (2017).
[128] D. Gresch, G. Autès, O. V. Yazyev, M. Troyer, D. Vanderbilt, B. A. Bernevig, and A. A. Soluyanov, Phys. Rev. B **95**, 075146 (2017).
[129] A. M. Kosevich, Low Temp. Phys. **30**, 97 (2004).
[130] M. Wilkinson, Proc. R. Soc. London, Ser. A **391**, 305 (1984).
[131] L. C. Davis and S. H. Liu, Phys. Rev. **158**, 689 (1967).
[132] E. C. Kemble, Phys. Rev. **48**, 549 (1935).

[133] H. Weng, C. Fang, Z. Fang, B. A. Bernevig, and X. Dai, Phys. Rev. X **5**, 011029 (2015).

[134] B. Q. Lv, H. M. Weng, B. B. Fu, X. P. Wang, H. Miao, J. Ma, P. Richard, X. C. Huang, L. X. Zhao, G. F. Chen, Z. Fang, X. Dai, T. Qian, and H. Ding, Phys. Rev. X **5**, 031013 (2015).

[135] S.-Y. Xu, I. Belopolski, N. Alidoust, M. Neupane, G. Bian, C. Zhang, R. Sankar, G. Chang, Z. Yuan, C.-C. Lee, S.-M. Huang, H. Zheng, J. Ma, D. S. Sanchez, B. Wang, A. Bansil, F. Chou, P. P. Shibayev, H. Lin, S. Jia, and M. Z. Hasan, Science **349**, 613 (2015).

[136] S.-Y. Xu, C. Liu, N. Alidoust, M. Neupane, D. Qian, I. Belopolski, J. D. Denlinger, Y. J. Wang, H. Lin, L. A. Wray, G. Landolt, B. Slomski, J. H. Dil, A. Marcinkova, E. Morosan, Q. Gibson, R. Sankar, F. C. Chou, R. J. Cava, A. Bansil, and M. Z. Hasan, Nat. Commun. **3**, 1192 (2012).

[137] Y. Tanaka *et al.*, Nat. Phys. **8**, 800 (2012).

[138] C. Herring, Phys. Rev. **52**, 365 (1937).

[139] C. Herring, Phys. Rev. **52**, 361 (1937).

[140] P. Hořava, Phys. Rev. Lett. **95**, 016405 (2005).

[141] There exists a related notion of co-dimension that is defined differently in the literature [39,140].

[142] D. T. Son and B. Z. Spivak, Phys. Rev. B **88**, 104412 (2013).

[143] J. Xiong, S. K. Kushwaha, T. Liang, J. W. Krizan, M. Hirschberger, W. Wang, R. Cava, and N. Ong, Science **350**, 413 (2015).

[144] A. Bouhon and A. M. Black-Schaffer, Phys. Rev. B **95**, 241101 (2017).

[145] L. Michel and J. Zak, Europhys. Lett. **50**, 519 (2000).

[146] If the line node is protected by a mirror symmetry, we have assumed here that the field is not aligned orthogonal to the mirror plane.

[147] B. A. Foreman, J. Phys.: Condens. Matter **12**, R435 (2000).

[148] If the band is not representable by Wannier functions, one may employ a Fourier-integral representation of u_{nk_x0} instead [cf. Eqs. (49) and (50)].

[149] P. Löwdin, J. Chem. Phys. **19**, 1396 (1951).

[150] R. Winkler, *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*, Springer Tracts in Modern Physics (Springer, New York, 2003).

[151] J. E. Moyal, Math. Proc. Cambridge Philos. Soc. **45**, 99 (1949).

[152] To define a local circulation for each hyperbolic arm, draw a line that closes the arm into a circle (in the sense of homotopy equivalence), e.g., $\curvearrowleft\to\circlearrowleft$, $\curvearrowright\to\circlearrowright$. We then assign the orientation by interpreting the circle as a clock face.

[153] They are the symmetries of MTe_2 in the sense of a group isomorphism.

[154] In contrast, the Dirac cone of graphene is symmetric under $\mathfrak{c}_{3z}$ and cannot be tilted.

[155] The third $O(l^{-2})$ term may be dropped explicitly if we work in a basis where $\mathfrak{X}^x = 0$, as further elaborated in Appendix F 1.

[156] C. Wittig, J. Phys. Chem. B **109**, 8428 (2005).

[157] E. O. Kane and E. I. Blount, in *Tunneling Phenomenon in Solids*, edited by E. Burstein and S. Lundqvist (Plenum Press, New York, 1969).

[158] W. Kohn, Phys. Rev. **115**, 809 (1959).

[159] Y. Xu, F. Zhang, and C. Zhang, Phys. Rev. Lett. **115**, 265304 (2015).

[160] H. Huang, S. Zhou, and W. Duan, Phys. Rev. B **94**, 121117 (2016).

[161] T.-R. Chang, S.-Y. Xu, D. S. Sanchez, W.-F. Tsai, S.-M. Huang, G. Chang, C.-H. Hsu, G. Bian, I. Belopolski, Z.-M. Yu, S. A. Yang, T. Neupert, H.-T. Jeng, H. Lin, and M. Z. Hasan, Phys. Rev. Lett. **119**, 026404 (2017).

[162] Y. Gao and Q. Niu, Proc. Natl. Acad. Sci. USA **114**, 7295 (2017).

[163] T. Thonhauser, D. Ceresoli, D. Vanderbilt, and R. Resta, Phys. Rev. Lett. **95**, 137205 (2005).

[164] D. Ceresoli, T. Thonhauser, D. Vanderbilt, and R. Resta, Phys. Rev. B **74**, 024408 (2006).

[165] J. Shi, G. Vignale, D. Xiao, and Q. Niu, Phys. Rev. Lett. **99**, 197202 (2007).

[166] K.-T. Chen and P. A. Lee, Phys. Rev. B **84**, 205137 (2011).

[167] D. J. Griffiths, *Introduction to Quantum Mechanics* (Pearson Prentice Hall, Upper Saddle River, NJ, 2005).

[168] P. Lowdin, J. Chem. Phys. **18**, 365 (1950).

[169] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).

[170] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).

[171] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).

[172] B. A. Bernevig, T. L. Hughes, and S. C. Zhang, Science **314**, 1757 (2006).

[173] R. Roy, Phys. Rev. B **79**, 195321 (2009).

[174] L. Fu and C. L. Kane, Phys. Rev. B **74**, 195312 (2006).

[175] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).

[176] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).

[177] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306 (2007).

[178] R. Roy, Phys. Rev. B **79**, 195322 (2009).

[179] A. Kitaev, AIP Conf. Proc. **1134**, 22 (2009).

[180] N. Read, Phys. Rev. B **95**, 115309 (2017).

[181] L. Chen, T. Mazaheri, A. Seidel, and X. Tang, J. Phys. A: Math. Theor. **47**, 152001 (2014).

[182] C. H. Lee, D. P. Arovas, and R. Thomale, Phys. Rev. B **93**, 155155 (2016).

[183] E. Kapit and E. Mueller, Phys. Rev. Lett. **105**, 215303 (2010).

[184] The model described in this reference describes a Chern insulator, but one can combine Chern insulators with opposite chiralities to obtain a class-AII strong topological insulator.

[185] F. Wilczek, Phys. Rev. Lett. **58**, 1799 (1987).

[186] A. M. Turner, Y. Zhang, R. S. K. Mong, and A. Vishwanath, Phys. Rev. B **85**, 165120 (2012).

[187] The problem actually goes deeper: H_1 cannot even be defined continuously over the torus. For a related discussion, see Ref. [83].

[188] J. M. Robbins, S. C. Creagh, and R. G. Littlejohn, Phys. Rev. A **41**, 6052 (1990).

[189] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST Handbook of Mathematical Functions*, 1st ed. (Cambridge University Press, New York, 2010).

[190] See Appendix 1 of Ref. [12].