Effective quantum theories for Bloch dynamics in inhomogeneous systems with nontrivial band structure

Christian Wickles[*] and Wolfgang Belzig[†]

Universität Konstanz, Fachbereich Physik, 78457 Konstanz, Germany

Starting from a general N-band Hamiltonian with weak spatial and temporal variations, we derive a low-energy effective theory for transport within one or several overlapping bands. To this end, we use the Wigner representation that allows us to systematically construct the unitary transformation that brings the Hamiltonian into band-diagonal form. We address the issue of gauge invariance and discuss the necessity of using kinetic variables in order to obtain a low-energy effective description that is consistent with the original theory. Essentially, our analysis is a semiclassical one and quantum corrections appear as Berry curvatures in addition to quantities that are related to the appearance of persistent currents. We develop a transport framework, which is manifestly gauge invariant, and it is based on a quantum Boltzmann formulation along with suitable definitions of current density operators such that Liouville's theorem is satisfied. Finally, we incorporate the effects of an external electromagnetic field into our theory.

I. INTRODUCTION

When performing a semiclassical analysis, one naturally encounters Berry phases[1] and meanwhile, the importance of these so-called geometrical phases in condensed matter physics is beyond question.[2–5] For example, not long ago, it was realized that the electric polarizability can be defined in terms of a Berry curvature, for the first time a consistent formulation of this subject.[6] Furthermore, research in the field of the anomalous Hall effect (AHE) has shown that the intrinsic contribution is related to a Berry curvature, which is a quantum-mechanical property of a perfect crystal.[7–9] Also, magnetic monopoles appearing in the definition of a momentum space effective magnetic field give important modifications to universal conductance fluctuations.[5] Finally, topological interference effects arise from spin Berry phases in single molecular magnets.[10]

There are different ways to obtain semiclassical transport equations (see Refs. 3 and 4, and references therein), like wave-packet analysis,[11–13] or the systematic diagonalization method developed by Gosselin and coworkers,[14–16] which, however, does not include the possibility of time-dependent perturbations. Furthermore, there are various works treating semiclassical quantum transport equations that incorporate Berry phase effects: the scenario of a general two-band model is considered by Wong and Tserkovnyak,[17] and spin-orbit coupled systems[18] as well as a non-Abelian gauge-field formulation[19] is investigated. The Löwdin partitioning, or quasidegenerate perturbation theory, used in the book of Winkler[20] to derive effective models for certain bands in spin-orbit coupled semiconductors can also be related to a semiclassical treatment, however, there focus is only put on the Hamiltonian, not on the dynamical variables or other aspects of the system.

In this work, we present a self-contained derivation of the semiclassical dynamics, which is based on the Wigner representation[21]—or phase-space representation of quantum mechanics—which is a natural starting point for a semiclassical analysis. One big advantage is that one can obtain corrections systematically to arbitrary order in $\hbar$. Also, a requantization of the effective theory is not necessary, which

is a big drawback of the wave-packet analysis, which derives a Lagrangian from the equations of motion for the wave-packet center of mass coordinates, and it is not always clear what the canonical conjugate variables are. The relation between canonical and kinetic pairs of conjugate variables, however, emerges naturally in our formalism. We adopt a 4-component vector notation, which allows us to incorporate spatial inhomogeneity as well as temporal variation on an equal basis. In the course of our treatment, we will find how fictitious electric and magnetic fields (real space and its momentum space pendants) appear in effective theories, and we complete our work by developing a low-energy effective quantum transport theory, which is manifestly gauge invariant and consistent with a description in the original frame. Finally, we address the interesting question of how an external electromagnetic field modifies the formalism.

Approaches related to ours have been considered before, albeit from a somewhat different perspective and mainly focused on stationary situations. The early work of Blount employs an approach (termed *mixed representation* therein) similar to the phase-space method used in our work.[22] Even though the concept of Berry phase was not known at that time, Blount's paper shares many results with the present work, aside from using different terminology. Furthermore, Littlejohn *et al.* elaborate in some detail the subtle issues related to the multicomponent WKB analysis.[23,24] Their proposed idea of reducing the problem of multi-component WKB analysis to a single-component one is strongly related to the idea presented in this article. More specifically, analyzing the general spin-orbit problem,[24] they perform an approximate diagonalization to find the quantities restricted to the subspace of a single polarization, and subsequently employ the results of scalar WKB theory, at the expense of a noncanonical treatment. Within their work, they too come to the conclusion that without appreciation of the role of Berry's phase, a proper semiclassical description of a multiband system seems difficult, if not impossible.

We have several scenarios in mind of applying our formalism to studying the electron dynamics in the presence of an arbitrary inhomogeneous and time-dependent ferromagnetic exchange field, which exhibits many interesting phenomena.[25–27] One can make various generalizations like

adding a spin-orbit coupling term, which would give rise to an even much broader range of new effects such as the anomalous Hall effect, and the latter would be additionally modified by the inhomogeneous magnetization. Studying spin-transport phenomena and/or adding thermal gradients delivers a whole new range of possibilities.[28] Furthermore, there is a recently discovered class of materials called topological insulators, in which the quasiparticle momentum is intimately coupled to the spin-degree of freedom and thus behaves similar to a relativistic Dirac fermion.[29,30] Now, adding an additional coupling between spatial and spin degrees-of-freedom leads to a physically rich system already subject to various studies.[31–35] Considering the effects of mechanical rotation of these systems[36,37] is another interesting application for our formalism.

Approaches that diagonalize band space in order to obtain an effective low-energy description have been performed in the case of quasifree electrons in ferromagnetic metals with spatiotemporally varying exchange field. There, one can come up with a position-dependent unitary matrix $U(\boldsymbol{r})$ that diagonalizes spin space, which maps the problem on that of a free particle in an (fictitious) electromagnetic field.[38] The converse situation of coupling between spin and momentum as given by the spin-orbit interaction can equally be treated by a momentum-dependent unitary matrix $U(\boldsymbol{p})$, for example, the Foldy-Wouthuysen transformation in the case of the relativistic Dirac equation.[39]

It is clear that the full quantum-mechanical problem of surface Dirac fermions coupled to a general inhomogeneous and time-dependent magnetization texture will, in general, be very complex due to the locking of spin and momentum as well as coupling between spin and spatial degrees of freedom. If one wants to diagonalize spin space, the unitary transformation has to involve the pair of canonical operators $\boldsymbol{r}$ and $\boldsymbol{p}$, which is rather difficult due to noncommutation of $\boldsymbol{r}$ and $\boldsymbol{p}$. Our approach relies on the fact that, if $\boldsymbol{r}$ and $\boldsymbol{p}$ are classical variables, such a unitary transformation is much more simple to find, and then, we resort to quantum corrections, which are of the order $\hbar$. Despite the fact that we are performing formally an expansion in $\hbar$, it does not necessarily need to be restricted to the semiclassical regime. In fact, as pointed out later, our actual expansion parameter might be a different one, depending on the physical system and the regime under investigation. For example, in the case of the Dirac theory, as we will discuss thoroughly in Sec. VI, the actual scale relevant for our expansion is the Compton wavelength $\lambda_c = \frac{\hbar}{mc}$ so the resulting Pauli equation still correctly describes the quantum regime for scalar potentials smooth on the scale of λ_c.

The outline of this work is as follows. In Sec. II, we will introduce the unitary transformation that performs a rotation within band space such that the Hamiltonian becomes band diagonal. Since this transformation is not uniquely defined, we will discuss the implications of this additional gauge degree of freedom. This motivates a description in terms of kinetic variables, which leads to the appearance of Berry curvatures, which is discussed in some detail in Sec. III. We will also investigate how observables change in the course of the diagonalization and we discuss the electronic spectrum as well as energy corrections appearing therein. Subsequently, in Sec. IV, we develop a manifestly gauge invariant description of the physical system restricted to a certain band, i.e., we

seek a projected theory without the necessity to refer back to the original Hamiltonian. To this end, we find equations of motion for the quasiprobability density, essentially a quantum Boltzmann equation applicable to a nonequilibrium scenario. We also find current densities that obey a conservation law corresponding to Liouville's theorem in classical mechanics. Our quantum-mechanical equations of motion are formally similar to the equations of motion for the center-of-mass motion of a wave packet and, in fact, the latter is just a special case of our formulation. Finally, in Sec. V, we illustrate how to treat the external electromagnetic field, which can be done in the spirit of a hierarchy of effective theories. After concluding the discussion of general systems, in Sec. VI, we apply the apparatus developed to the Dirac equation and readily find a relativistic version of the Pauli Hamiltonian, thereby gaining some interesting insights into the structure of the Dirac equation. For a concise summary of this work, the reader might want to go to Sec. VI where all central results are being referenced and find immediate application.

II. THE QUEST FOR A BAND-DIAGONAL HAMILTONIAN

We now consider a Hamiltonian that consists of N bands and which is almost diagonal in momentum space $\boldsymbol{p}$, but has some additional spatial- and/or temporal variation imprinted on it. In the usual quantum representation, it is expressed in terms of the canonical pair of operators $[\hat{r}_i, \hat{p}_j] = i\hbar\delta_{ij}$ and additionally carries the $N \times N$ dimensional matrix structure. Prominent examples that fall into this category are the Dirac equation to be studied in more detail in Sec. VI as well as the aforementioned system of surface Dirac fermions coupled to a spatially dependent magnetization.

For practical reasons, we resort to a description in terms of single-particle Greens functions, or more specifically the inverse thereof, $\Xi \equiv i\hbar\partial_t \mathbb{1}_N - \mathcal{H}(\hat{\boldsymbol{r}},\hat{\boldsymbol{p}},t)$. One reason is that only Ξ represents the complete equation of motion, i.e., all kinetic equations involve this operator, and not $\mathcal{H}$ alone, for example, we can generally write quantum kinetic equations in the compact form

$$[\Xi, \mathcal{D}] = 0, \tag{1}$$

where $\mathcal{D}$ represents any dynamical variable like the usual retarded or Keldysh Greens functions, density of states, or the density matrix for which Eq. (1) reduces to the well-known von Neumann equation

$$i\hbar\partial_t \rho = [\mathcal{H}, \rho]. \tag{2}$$

The properties of $\mathcal{D}$ (the type of Greens function, etc.) enter through appropriate boundary conditions in our parameter space. For the example just given, we need to fix our density matrix at some initial time $\rho(t_0) = \rho_0$.

We now transform everything into the Wigner representation,[40] so that our Hamiltonian and our observables are a function of the variables $\boldsymbol{r}, \boldsymbol{p}, t$ and energy ϵ so that (see Appendix A for more details)

$$\Xi(\boldsymbol{r},t,\boldsymbol{p},\epsilon) \equiv \epsilon\mathbb{1}_N - \mathcal{H}(\boldsymbol{r},t,\boldsymbol{p}). \tag{3}$$

Multiplication of operators is now performed by virtue of the Moyal product,[41]

$$* \equiv \exp \frac{i\hbar}{2}\Lambda, \tag{4}$$

where the differential operator Λ is given by

$$\begin{aligned}\Lambda &= \overleftarrow{\partial_r}\,\overrightarrow{\partial_p} - \overleftarrow{\partial_t}\,\overrightarrow{\partial_\epsilon} - \overleftarrow{\partial_p}\,\overrightarrow{\partial_r} + \overleftarrow{\partial_\epsilon}\,\overrightarrow{\partial_t}\\ &= -\overleftarrow{\partial_x}\,\overrightarrow{\partial_\pi} + \overleftarrow{\partial_\pi}\,\overrightarrow{\partial_x}\end{aligned} \tag{5}$$

and we introduced the compact 4-component vector notation $\boldsymbol{x} = (t,\boldsymbol{r})$ and $\boldsymbol{\pi} = (\epsilon,-\boldsymbol{p})$. Note that we always use $\boldsymbol{x}$, $\partial_{\boldsymbol{x}} = (\partial_t,\partial_{\boldsymbol{r}})$ and $\mathcal{A}_{\boldsymbol{x}} = (\mathcal{A}_t,\mathcal{A}_{\boldsymbol{r}})$ in contravariant notation, whereas we implicitly assume covariant notation for the symbols $\boldsymbol{\pi}$, $\partial_{\boldsymbol{\pi}} = (\partial_\epsilon,-\partial_{\boldsymbol{p}})$ and $\mathcal{A}_{\boldsymbol{\pi}} = (\mathcal{A}_\epsilon,-\mathcal{A}_{\boldsymbol{p}})$. Since contraction will always be between pairs of $\boldsymbol{x}$ and $\boldsymbol{\pi}$, there is no need to indicate covariant vectors. This provides us a symmetric and compact notation in the following treatment. The kinetic equation (1) is straightforwardly transformed into the Wigner picture:

$$[\Xi \overset{*}{,} \mathcal{D}] = 0. \tag{6}$$

We are now looking for a unitary matrix $\mathcal{U}(\boldsymbol{x},\boldsymbol{\pi})$ that transforms our initial Hamiltonian $\Xi(\boldsymbol{x},\boldsymbol{\pi})$ into a band-diagonal Hamiltonian, denoted by $\bar{\Xi}(\boldsymbol{x},\boldsymbol{\pi})$ in the following, i.e.,

$$\mathcal{U} * \mathcal{U}^\dagger = \mathcal{U}^\dagger * \mathcal{U} = \mathbb{1}, \tag{7}$$

$$\mathcal{U} * \Xi * \mathcal{U}^\dagger = \bar{\Xi}. \tag{8}$$

Let us note that upon this transformation, $\mathcal{U} * \epsilon * \mathcal{U}^\dagger$ can acquire off-diagonal elements when $\mathcal{U}$ depends on time, thus requiring the diagonalization of the combination $\Xi = \epsilon \mathbb{1}_N - \mathcal{H}$ rather than $\mathcal{H}$ alone. This is one major difference to previous semiclassical schemes.[15] For diagonalizations that do not require explicit dependency on the time parameter, treating $\mathcal{H}$ and Ξ is equivalent, and in the following, we will use the term Hamiltonian equally for both objects. Note that another advantage of using Ξ instead of $\mathcal{H}$ is the form invariance of the kinetic equation (6) under unitary transformations.

In the classical limit, the operators $\hat{r}_i$ and $\hat{p}_j$ commute, while in the Wigner representation, the Moyal product (4) becomes trivial as $\hbar \to 0$; after all, in this formulation, the Moyal product encodes the noncommutativity of the canonical variables. Then, we essentially have to diagonalize an $N \times N$ matrix $\mathcal{H}$, whose elements are functions of $\boldsymbol{r}$, t, and $\boldsymbol{p}$. We call the unitary matrix associated with this rotation in band space $U_0(\boldsymbol{r},\boldsymbol{p},t)$, so that

$$\mathcal{U}_0 \Xi \mathcal{U}_0^\dagger = \bar{\Xi}_0 \Leftrightarrow \mathcal{U}_0 \mathcal{H} \mathcal{U}_0^\dagger = \bar{\mathcal{H}}_0, \tag{9}$$

and the diagonal elements of $\bar{\Xi}_0$ constitute the classical energies of the N bands described by our initial Hamiltonian. At any rate, we assume from now on that we know the diagonalization matrix $\mathcal{U}_0$ analytically. Note that all our matrices parametrically depend on $\boldsymbol{x}$ and $\boldsymbol{\pi}$, so we essentially diagonalize locally at every point in $2 \times (3+1)$-dimensional parameter space (see Fig. 1 for an illustration), which becomes meaningful in the semiclassical limit, as position $\boldsymbol{r}$, momentum $\boldsymbol{p}$, time t, and energy ϵ are well defined in this limit. Also note that $U_0(\boldsymbol{r},\boldsymbol{p},t)$ is not uniquely defined, which brings up the problem of gauge invariance as discussed in detail later.
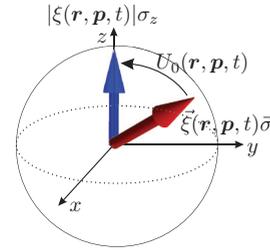

FIG. 1. (Color online) Illustration of the band-diagonalization scheme: Bloch-sphere representation for a simple two-band model $\mathcal{H} = \bar{\bar{\xi}}(\boldsymbol{r},\boldsymbol{p},t) \cdot \vec{\boldsymbol{\sigma}} = \xi_x \sigma_x + \xi_y \sigma_y + \xi_z \sigma_z$ where the vector is rotated by $U_0(\boldsymbol{r},\boldsymbol{p},t)$ such that it aligns along the z axis at each point in phase space.

Furthermore, if the band degrees-of-freedom couple to only either $\boldsymbol{x}$ or $\boldsymbol{\pi}$, then we already found the exact expression $\mathcal{U} = \mathcal{U}_0$. For example, treating metallic ferromagnets with inhomogeneous magnetization and neglecting spin-orbit interactions,[27] it is just necessary to only diagonalize $\boldsymbol{m}(\boldsymbol{r})\boldsymbol{\sigma}$, since the energy dispersion is diagonal in spin space. Then, the effect of $\mathcal{U}$ is that $\boldsymbol{p}$ acquires an additional vector potential: $\boldsymbol{p} \to \boldsymbol{p} - i\hbar\mathcal{U}(\partial_{\boldsymbol{r}}\mathcal{U}^\dagger)$.

The general situation, where the degrees-of-freedom couple to both spatial and momentum coordinates simultaneously, is much more involved and chances that we can come up with an exact solution for $\mathcal{U}(\boldsymbol{x},\boldsymbol{\pi})$ are slim. Therefore we adopt a gradient expansion approach where we expand the Moyal product $*$ in powers of $\hbar$ and calculate corrections systematically order by order. The fact that we are performing formally an expansion in $\hbar$ does not necessarily imply that we are restricted to the semiclassical regime. Our actual expansion parameter might be a different one, depending on the physical system and the regime under investigation. Typically, the expansion parameter is the ratio of the energy associated with band-dependent spatiotemporal variations and the interband energy distance Δ, as illustrated in Fig. 2, and therefore, the larger the band separations (or weaker coupling), the better the approximation becomes. To emphasize this fact, we will introduce the notation of $\hbar$ in this paper to indicate that the expansion is not necessarily a semiclassical one.

We assume that we already solved the zeroth-order problem that yields the unitary matrix $\mathcal{U}_0$, so that now, we can introduce
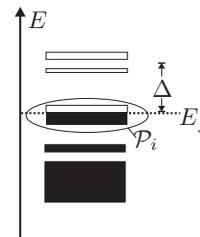
FIG. 2. Illustration of the bands described by the Hamiltonian $\mathcal{H}$ and Δ is a typical energy scale for interband distances. The projection operator on the ith band is denoted by $\mathcal{P}_i$. For low-energy processes around the Fermi energy E_f, only one or few overlapping bands are relevant and thus contribute to, e.g., transport properties. Corrections due to the influence of other bands enter as Berry curvatures and are $\propto 1/\Delta$.

gradient corrections to this matrix by writing

$$\mathcal{U} = (\mathbb{1} + \bar{\mathcal{U}}_1 + \bar{\mathcal{U}}_2 + \cdots)\mathcal{U}_0, \tag{10}$$

which describes the matrix that diagonalizes our Hamiltonian Ξ to arbitrary order in $\hbar$. $\bar{\mathcal{U}}_1$ is $\propto \hbar$ and will be specified explicitly later. As a result of the expansion of the Moyal product, we obtain the covariant derivatives

$$\bar{\partial}_i = \mathcal{U}_0 \partial_i \mathcal{U}_0^\dagger = \partial_i - i\mathcal{A}_i, \tag{11}$$

which acquire the Berry connections

$$\mathcal{A}_i \equiv i\mathcal{U}_0(\partial_i \mathcal{U}_0^\dagger), \tag{12}$$

where ∂_i is a placeholder for any of the possible derivatives $\partial_t, \partial_\epsilon, \partial_r$, and ∂_p. Berry phases, or often called geometric phases, are omnipresent in modern physics, and describe phases picked up along a trajectory in curved geometries. Note that $\mathcal{A}_i$ is an $N \times N$ matrix, which is Hermitian:

$$\mathcal{A}_i^\dagger = -i(\mathcal{U}_0 \partial_i \mathcal{U}_0^\dagger)^\dagger = -i(\partial_i \mathcal{U}_0)\mathcal{U}_0^\dagger = i\mathcal{U}_0(\partial_i \mathcal{U}_0^\dagger) = \mathcal{A}_i, \tag{13}$$

where the last step is due to $0 = \partial_i(\mathcal{U}_0 \mathcal{U}_0^\dagger) = (\partial_i \mathcal{U}_0)\mathcal{U}_0^\dagger + \mathcal{U}_0(\partial_i \mathcal{U}_0^\dagger)$ and the diagonal elements of $\mathcal{A}_i$ describe the usual Berry connections arising from transport within a certain band. The off-diagonal elements mix contributions from two bands and thus describe effects due to interband coupling during transport in a certain band. As we will see later, these interband transitions will give rise to important corrections, for example, the correction to the energy that appears to leading order in $\hbar$. Transforming higher-order derivatives is straightforward, except for the additional freedom of exchanging partial derivatives, which leads to a general relation between Berry connection matrices $\mathcal{A}_i$. Starting from

$$\mathcal{U}_0 \partial_j \partial_i \mathcal{U}_0^\dagger = \bar{\partial}_j \bar{\partial}_i = (\partial_j - i\mathcal{A}_j)(\partial_i - i\mathcal{A}_i), \tag{14}$$

and likewise,

$$\mathcal{U}_0 \partial_i \partial_j \mathcal{U}_0^\dagger = \bar{\partial}_i \bar{\partial}_j = (\partial_i - i\mathcal{A}_i)(\partial_j - i\mathcal{A}_j), \tag{15}$$

and, if we assume symmetry of second derivatives $\partial_j \partial_i = \partial_i \partial_j$, the following relation should hold:

$$\partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i = i[\mathcal{A}_i, \mathcal{A}_j]. \tag{16}$$

In fact, this identity can be directly shown by using the definition of $\mathcal{A}$, see Eq. (12), and the exchange of partial derivatives.

Now we have everything at hand to systematically calculate $\bar{\Xi} = \bar{\Xi}_0 + \bar{\Xi}_1 + \bar{\Xi}_2 \ldots$ to arbitrary order in $\hbar$ and to first order, we explicitly obtain the following result for the transformed expression:

$$\bar{\Xi}_1 = \bar{\mathcal{U}}_1 \bar{\Xi}_0 + \bar{\Xi}_0 \bar{\mathcal{U}}_1^\dagger - \frac{\hbar}{2}\{\mathcal{A}_\pi, \partial_x \bar{\Xi}_0\} + \frac{\hbar}{2}\{\mathcal{A}_x, \partial_\pi \bar{\Xi}_0\}$$
$$+ \frac{i\hbar}{2}(\mathcal{A}_\pi \bar{\Xi}_0 \mathcal{A}_x - \mathcal{A}_x \bar{\Xi}_0 \mathcal{A}_\pi). \tag{17}$$

We obtain an additional constraint for $\bar{\mathcal{U}}_n$ from the condition of unitarity (7), which to first order in $\hbar$ reads

$$\bar{\mathcal{U}}_1 + \bar{\mathcal{U}}_1^\dagger + \frac{i\hbar}{2}[\mathcal{A}_\pi, \mathcal{A}_x] \overset{!}{=} 0, \tag{18}$$

and is obtained by substituting $\mathbb{1}$ for $\bar{\Xi}_0$ into the transformation relation (17). Since this relation fixes the Hermitian part of $\bar{\mathcal{U}}_1$,

we can make the ansatz

$$\bar{\mathcal{U}}_1 = -\frac{i\hbar}{4}[\mathcal{A}_\pi, \mathcal{A}_x] + \mathcal{Y}_1,$$

with $\mathcal{Y}_1 = -\mathcal{Y}_1^\dagger$ assumed to be antihermitian, such that condition (18) is satisfied. Plugging $\bar{\mathcal{U}}_1$ back into Eq. (17) yields

$$\bar{\Xi}_1 = [\mathcal{Y}_1, \bar{\Xi}_0] - \frac{\hbar}{2}\left\{\mathcal{A}_\pi, \left[\partial_x - \frac{i}{2}\mathcal{A}_x, \bar{\Xi}_0\right]\right\}$$
$$+ \frac{\hbar}{2}\left\{\mathcal{A}_x, \left[\partial_\pi - \frac{i}{2}\mathcal{A}_\pi, \bar{\Xi}_0\right]\right\} + O(\hbar^2). \tag{19}$$

We observe that $[\mathcal{Y}_1, \bar{\Xi}_0]$ is completely off-diagonal and $\mathcal{Y}_1$ is well suited to absorb the off-diagonal part of the last two terms on the right-hand side of Eq. (19), in the following denoted as $\mathcal{R}_o$. After all, the objective is to diagonalize $\bar{\Xi}$, so we want the off-diagonal part of $\bar{\Xi}_1$ to vanish, which is achieved by the condition

$$[\mathcal{Y}_1, \bar{\Xi}_0] = \mathcal{R}_o. \tag{20}$$

The solution can be readily given as

$$\Rightarrow (\mathcal{Y}_1)_{ij} = \frac{(\mathcal{R}_o)_{ij}}{\epsilon_i - \epsilon_j} \tag{21}$$

in the case of a completely diagonalized $(\bar{\mathcal{H}}_0)_{ij} = \epsilon_i \delta_{ij}$ and which is anti-Hermitian as desired. Here, we explicitly see that corrections due to $\mathcal{Y}$ are inversely proportional to the separation between bands, and thus are of the same order as corrections due to Berry phases. In fact, the commutator on the left side with the Hamiltonian $\bar{\mathcal{H}}_0$ is exactly what also appears in the von Neumann equation suggesting further that this term describes corrections due to interband dynamics. From Eq. (20), we also see that the expansion fails near mode conversion points, so that explicit semiclassical diagonalization is not possible. Instead, one can only perform block diagonalization and, within the subspace of relevant bands, on has to resort to other methods not subject of this work.

We remark that the difference between the diagonalization of Ξ and $\mathcal{H}$ is merely the modification of $\mathcal{Y}_1$ by that it acquires an additional contribution from the off-diagonal elements of the Berry connection $\mathcal{A}_t$. However, this term is crucial when our transformation $\mathcal{U}$ is time dependent and only then will the formalism yield consistent results. Note that the definition (20) does not fix the imaginary diagonal part of $\mathcal{Y}_1$, but without loss of generality, we can set this part to zero. A nonzero imaginary diagonal part corresponds to the linear order expansion of the gauge phase factors $e^{i\chi(x,\pi)}$ [see Eq. (23)]. Essentially, the freedom of choice here can be reduced to the problem of gauge invariance, to be discussed in the next section.

Yet, we also want to capture the situation of degenerate or overlapping bands, so the desired form of $\bar{\Xi}$ is in general block diagonal. To formally express this matter, we introduce the projectors $\mathcal{P}_i$ that define the bands (see Fig. 2) and that, in the end, we want use for our effective theory. Of course, the set of $\mathcal{P}_i$ has to be specified together with $\mathcal{U}_0$, since the unitary transformation has a freedom of how we distribute the bands amongst the entries of our matrix. For example, a natural choice would be to sort the bands with respect to their energies. Then, introducing the projected Berry

connection $\mathcal{A}_{\boldsymbol{\pi}}^{(d)} \equiv \mathcal{P}_d \mathcal{A}_{\boldsymbol{\pi}} \mathcal{P}_d \equiv \sum_i \mathcal{P}_i \mathcal{A}_{\boldsymbol{\pi}} \mathcal{P}_i$, we can write for the Hamiltonian that is diagonalized up to first order in $\hbar$,

$$\bar{\Xi} = \bar{\Xi}_0 - \hbar \mathcal{A}_{\boldsymbol{\pi}}^{(d)} \partial_{\boldsymbol{x}} \bar{\Xi}_0 + \hbar \mathcal{A}_{\boldsymbol{x}}^{(d)} \partial_{\boldsymbol{\pi}} \bar{\Xi}_0$$
$$+ \frac{i\hbar}{4} \mathcal{P}_d(\{\mathcal{A}_{\boldsymbol{\pi}}, [\mathcal{A}_{\boldsymbol{x}}, \bar{\Xi}_0]\} - \{\mathcal{A}_{\boldsymbol{x}}, [\mathcal{A}_{\boldsymbol{\pi}}, \bar{\Xi}_0]\}) \mathcal{P}_d. \quad (22)$$

As compared to expression (19), we can drop two anti-commutators, since after truncation, $\mathcal{A}^{(d)}$ and $\bar{\Xi}_0$ commute. As discussed more thoroughly below, this is even the case when either $\bar{\Xi}_0$ is block diagonal and/or the projected Berry connections are non-Abelian.

Apart from the obvious term $\bar{\Xi}_0 = \epsilon - \bar{\mathcal{H}}_0$, which can be understood as the classical energy, the last term of Eq. (22) can be thought of as correction to the energy due to interband transitions and corresponds to the energy associated with persistent circulating currents, as for example, the magnetic Zeeman energy in the case of the Dirac equation (see discussion in Sec. VI). The two middle terms in Eq. (22) appear to be Berry phase corrections to this energy, however, they are not unique in the sense that they depend on the specific form of U_0. The meaning will become more apparent later, but before, we address the question of gauge invariance of the effective Hamiltonian (22).

A. Gauge invariance

We mentioned previously that there is an additional degree of freedom in the choice of unitary transformations $\mathcal{U}_0$ and $\mathcal{U}_0'$, which all yield the same diagonal Hamiltonian, $\bar{\Xi} = \mathcal{U}_0 \Xi \mathcal{U}_0^\dagger = \mathcal{U}_0' \Xi \mathcal{U}_0'^\dagger$. These different unitary transformations are related by local phase factors in $2 \times (3+1)$-dimensional parameter space, so that we can formally connect two unitary transformations $\mathcal{U}$ and $\mathcal{U}'$ by the (block-)diagonal $N \times N$ phase matrix:

$$\Phi(\boldsymbol{x}, \boldsymbol{\pi}) = \begin{pmatrix} e^{i\chi_1(\boldsymbol{x}, \boldsymbol{\pi})} & 0 & 0 & \\ 0 & e^{i\chi_2(\boldsymbol{x}, \boldsymbol{\pi})} & 0 & \cdots \\ 0 & 0 & e^{i\chi_3(\boldsymbol{x}, \boldsymbol{\pi})} & \\ & \vdots & & \ddots \end{pmatrix}, \quad (23)$$

where $\chi_n(\boldsymbol{x}, \boldsymbol{\pi})$ are arbitrary functions in our parameter space, which are appropriately termed gauge fields. In fact, $\mathcal{U}_0' = \Phi \mathcal{U}_0$ describes a gauge transformation and can be thought of as a local phase transformation in an extended phase space that includes time and energy. This gauge transformation changes the Berry connection matrices according to

$$\mathcal{A}_k' = i\mathcal{U}_0' \partial_k \mathcal{U}_0'^\dagger = \Phi \mathcal{A}_k \Phi^\dagger + i\Phi \partial_k \Phi^\dagger = \Phi \mathcal{A}_k \Phi^\dagger + \mathcal{X}_k, \quad (24)$$

where the field $\mathcal{X}_k = i\Phi \partial_k \Phi^\dagger = \partial_k \mathrm{diag}(\chi_1, \chi_2, \ldots, \chi_N)$ is a (block-)diagonal matrix containing the partial derivatives of the phases. $\Phi \mathcal{A}_k \Phi^\dagger$ modifies the off-diagonal elements by giving them additional phase factors, while the change in diagonal elements is due to $\mathcal{X}_k$. In principle, we have to distinguish three different cases here, the first one being that $\bar{\Xi}_0$ is completely diagonalized with well-separated bands, which corresponds to the situation just described. However, when we keep part of $\bar{\Xi}_0$ block-diagonal because bands are overlapping or degenerate and we can distinguish the bands, i.e., they have a physical meaning that we want to retain (for

example, we have spin-degenerate bands but want to describe spin-dependent physics) then we have only a U(1) gauge freedom within this subblock. Third and last, if we have M degenerate bands, i.e., there is an M-dimensional subblock in $\bar{\Xi}_0$ that is proportional to the unit matrix, and furthermore, we cannot or do not want to distinguish between the degenerate bands, we have the additional degree of freedom to rotate within this degenerate space giving us an additional SU(M) gauge invariance. Contrary to the first two cases, this last one describes a situation with the effective description of this M-dimensional subspace being a non-Abelian gauge theory with the symmetry group U(1) × SU(M) and consequently, $\mathcal{X}_k$ constitutes a non-Abelian field. We do not differentiate between these cases explicitly in the following because they are straightforwardly treated in our formulas, thus requiring us only to comment in situations where special care is required.

According to the preceding discussion, the projected Berry connection matrix is only modified due to $\mathcal{X}$,

$$\mathcal{A}_{\boldsymbol{x}}'^{(d)} = \mathcal{P}_d \mathcal{A}_{\boldsymbol{x}}' \mathcal{P}_d = \mathcal{A}_{\boldsymbol{x}}^{(d)} + \mathcal{X}_{\boldsymbol{x}}, \quad (25)$$

and the alternative transformation due to $\mathcal{U}_0'$ leads to the Hamiltonian

$$\bar{\Xi}' = \mathcal{U}' * \Xi * \mathcal{U}'^\dagger = \bar{\Xi}_0 - \frac{\hbar}{2}\left\{\mathcal{A}_{\boldsymbol{\pi}}'^{(d)}, \partial_{\boldsymbol{x}} \bar{\Xi}_0\right\} + \frac{\hbar}{2}\left\{\mathcal{A}_{\boldsymbol{x}}'^{(d)}, \partial_{\boldsymbol{\pi}} \bar{\Xi}_0\right\}$$
$$+ \frac{i\hbar}{4} \mathcal{P}_d(\{\mathcal{A}_{\boldsymbol{\pi}}, [\mathcal{A}_{\boldsymbol{x}}, \bar{\Xi}_0]\} - \{\mathcal{A}_{\boldsymbol{x}}, [\mathcal{A}_{\boldsymbol{\pi}}, \bar{\Xi}_0]\}) \mathcal{P}_d + O(\hbar^2). \quad (26)$$

The last term, representing the interband transition corrections to the energy, does not change, since the additional terms due to $\mathcal{X}_{\boldsymbol{x}}, \mathcal{X}_{\boldsymbol{\pi}}$ are projected out by $\mathcal{P}_d$ and thus, as a consequence, are absorbed by $\mathcal{Y}_1'$ viz. $\bar{\mathcal{U}}_1'$, which of course does not need to coincide with $\bar{\mathcal{U}}_1$.

However, the other two terms linear in $\hbar$ do explicitly depend on the gauge, and therefore change the effective Hamiltonian. Clearly, this shows that the effective Hamiltonian alone is an incomplete description as it directly depends on this additional degree of freedom. Therefore, in order to make any sense out of this, we have to identify our physical observables, because in the end, the physical results derived from our effective theory should not depend on a specific gauge.

III. CANONICAL VERSUS KINETIC VARIABLES AND GAUGE INVARIANT DESCRIPTION

The previous section showed us that there is still an ingredient missing in our effective theories. In order to investigate this matter, let us study the dynamics of our system, and construct the effective theory such that the results obtained within this description are consistent with what one would obtain in the original frame. The question is now, whether one can find a manifest gauge invariant formulation and, as we will explain in the following, it is indeed possible.

A. Parameter transformation to kinetic variables

Let us consider the operator $\mathcal{S}$ that describes some physical observable of our system and, in performing the rotation that brings our Hamiltonian $\mathcal{H}$ into diagonal form, it transforms

our observable $\mathcal{S}$ along with it. The observable in the rotated frame $\bar{\mathcal{S}}$ is then related to the original operator $\mathcal{S}$ by virtue of relation (48). We now consider the projected system, i.e., we have in mind to develop an effective, yet exhaustive description of the physics taking place within a certain band that is sufficiently well separated from all other bands in order to treat this band independently to a good approximation. For the moment, let us assume that our observable $\mathcal{S}$ is band diagonal, i.e., it is a scalar function of $\boldsymbol{x}$ and $\boldsymbol{\pi}$, so that the last term in relation (48) vanishes and we have

$$\bar{\mathcal{S}} = \bar{\mathcal{S}}_0 - \frac{\hbar}{2}\left\{\mathcal{A}_{\boldsymbol{\pi}}^{(d)}, \partial_x \bar{\mathcal{S}}_0\right\} + \frac{\hbar}{2}\left\{\mathcal{A}_x^{(d)}, \partial_{\boldsymbol{\pi}} \bar{\mathcal{S}}_0\right\} + O(\hbar^2), \quad (27)$$

where $\bar{\mathcal{S}}_0 = \mathcal{U}_0 \mathcal{S} \mathcal{U}_0^\dagger$. Later, we will lift this restriction and consider a matrix $\mathcal{S}$ with general structure in band space, so that we will also get this additional term, giving rise to important contributions. However, for the moment, Eq. (27) is nothing but a Taylor expansion of $\bar{\mathcal{S}}_0$ to first order in $\hbar$,

$$\bar{\mathcal{S}} = \bar{\mathcal{S}}_0\big[\boldsymbol{x} - \hbar\mathcal{A}_{\boldsymbol{\pi}}^{(d)}, \boldsymbol{\pi} + \hbar\mathcal{A}_x^{(d)}\big] + O(\hbar^2), \quad (28)$$

suggesting the parameter transformation to band projected kinetic variables $\boldsymbol{X}$ and $\boldsymbol{\Pi}$,

$$\boldsymbol{X}^{(d)} = \boldsymbol{x} - \hbar\mathcal{A}_{\boldsymbol{\pi}}^{(d)}, \quad (29)$$

$$\boldsymbol{\Pi}^{(d)} = \boldsymbol{\pi} + \hbar\mathcal{A}_x^{(d)}, \quad (30)$$

so that we can write

$$\bar{\mathcal{S}} = \bar{\mathcal{S}}_0(\boldsymbol{X}, \boldsymbol{\Pi}) + O(\hbar^2). \quad (31)$$

We note that to leading order in $\hbar$, we can equally write $\mathcal{A}(\boldsymbol{x}, \boldsymbol{\pi}) = \mathcal{A}(\boldsymbol{X}, \boldsymbol{\Pi}) + O(\hbar)$, since Berry connection terms are already linear in $\hbar$.

$\boldsymbol{P} = \boldsymbol{p} - \hbar\mathcal{A}_r$ is of course well known in the Hamilton formulation of particles in an electromagnetic field. In an analogous way, the position $\boldsymbol{R} = \boldsymbol{r} + \hbar\mathcal{A}_p$ acquires an additional Berry connection with its corresponding Berry curvature, or momentum space magnetic field that gives rise to the so-called anomalous velocity term. Furthermore, $E = \epsilon + \hbar\mathcal{A}_t$ attains a contribution that gives rise to an electromotive force and appears in the form of an effective electric field (see, for example, Ref. 42). The same is true for the electromagnetic field, where the electric field can be also rewritten in terms of a time-dependent phase, effectively changing the gauge. Finally, for reasons of symmetry, one would also have $T = t - \hbar\mathcal{A}_\epsilon$ but, at least in noninteracting Hamiltonian systems, $\mathcal{A}_\epsilon$ is zero, since the energy dependency in Ξ is trivial. However, the situation is different if one considers an interacting system and uses an effective noninteracting quasiparticle description, because the self-energy that includes these interaction in general carries a nontrivial energy dependence inherited by the single-particle Greens function and thus by Ξ.[43]

Let us have a look at the commutator relations between the kinetic variables,[14,16]

$$\big[R_i^{(d)} \overset{*}{,} P_j^{(d)}\big] = i\hbar\big(\delta_{ij} + \underline{\underline{\Theta}}_{ji}^{rp}\big),$$
$$\big[R_i^{(d)} \overset{*}{,} R_j^{(d)}\big] = i\hbar\epsilon_{ijk}\mathcal{B}_k^{(p)}, \quad (32)$$
$$\big[P_i^{(d)} \overset{*}{,} P_j^{(d)}\big] = i\hbar\epsilon_{ijk}\mathcal{B}_k^{(r)},$$

where we introduced

$$\mathcal{B}^{(r)} = \hbar\left(\partial_r \times \mathcal{A}_r^{(d)}\right) - i\hbar\left(\mathcal{A}_r^{(d)} \times \mathcal{A}_r^{(d)}\right), \quad (33)$$

$$\mathcal{B}^{(p)} = \hbar\left(\partial_p \times \mathcal{A}_p^{(d)}\right) - i\hbar\left(\mathcal{A}_p^{(d)} \times \mathcal{A}_p^{(d)}\right), \quad (34)$$

which can be considered as a generalized magnetic field in real space and reciprocal (or momentum) space. Such non-Abelian Berry curvatures have been treated in Ref. 19. Furthermore,

$$\underline{\underline{\Theta}}_{ij}^{rp} = \hbar\left(\frac{\partial \mathcal{A}_{p_j}^{(d)}}{\partial r_i} - \frac{\partial \mathcal{A}_{r_i}^{(d)}}{\partial p_j} - i\big[\mathcal{A}_{r_i}^{(d)}, \mathcal{A}_{p_j}^{(d)}\big]\right). \quad (35)$$

As we will also encounter later, the dimensionless tensor $\underline{\underline{\Theta}}^{rp}$ describes the change in the metric of the phase space due to the parameter transformation from canonical to kinetic variables (29). We can make this more apparent by relating it to the change in differentials

$$dR_i = \big(\delta_{ij} + \hbar\partial_{r_j}\mathcal{A}_{p_i}\big)dr_j, \quad (36)$$

$$dP_i = \big(\delta_{ij} - \hbar\partial_{p_j}\mathcal{A}_{r_i}\big)dp_j, \quad (37)$$

so that

$$d\boldsymbol{R}\cdot d\boldsymbol{P} = d\boldsymbol{r}\left(\underline{1} + \underline{\underline{\Theta}}^{rp}\right)d\boldsymbol{p} + O(\hbar^2). \quad (38)$$

A more compact way to write these Berry curvatures is to use the covariant derivative (11), projected onto our band-diagonal space $\bar{\partial}_x^{(d)} = \mathcal{P}_d\bar{\partial}_x\mathcal{P}_d$, for example,

$$\underline{\underline{\Theta}}_{ij}^{rp} = \hbar\big(\bar{\partial}_{t_i}^{(d)}\mathcal{A}_{p_j}^{(d)} - \bar{\partial}_{p_j}^{(d)}\mathcal{A}_{r_i}^{(d)}\big). \quad (39)$$

It is well known that the Berry curvatures are invariant with respect to gauge transformations, and the commutator is essential as it provides the full SU(M) gauge invariance in the non-Abelian case.

In accordance with the effective magnetic fields (33) and (34), we introduced the effective electric fields

$$\mathcal{E}^{(r)} = \hbar\big(\partial_r\mathcal{A}_t^{(d)} - \partial_t\mathcal{A}_r^{(d)} - i\big[\mathcal{A}_r^{(d)}, \mathcal{A}_t^{(d)}\big]\big), \quad (40)$$

$$\mathcal{E}^{(p)} = \hbar\big(\partial_p\mathcal{A}_t^{(d)} - \partial_t\mathcal{A}_p^{(d)} - i\big[\mathcal{A}_p^{(d)}, \mathcal{A}_t^{(d)}\big]\big), \quad (41)$$

which shows us indeed that $\mathcal{A}_t$ appears in the role of a generalized electric potential, however, it can also depend on momentum $\boldsymbol{P}$. In the Abelian case (for non-Abelian fields, it works if we take $\text{Tr}_M\mathcal{E}$ and $\text{Tr}_M\mathcal{B}$ or when we take the covariant derivatives (11) along with the full matrix structure of the Berry connections), these fictitious fields obey homogeneous Maxwell equations

$$\boldsymbol{\partial}_{r/p}\cdot\mathcal{B}^{(r/p)} = 0, \quad \boldsymbol{\partial}_{r/p}\times\mathcal{E}^{(r/p)} + \partial_t\mathcal{B}^{(r/p)} = 0, \quad (42)$$

however, in order to determine these fields independently as in classical electrodynamics, one would need two additional inhomogeneous equations containing (effective) source terms as inhomogeneities. Note that, in general, $\mathcal{B}^{(r/p)}$ and $\mathcal{E}^{(r/p)}$ depend on $\boldsymbol{r}$ and $\boldsymbol{p}$ simultaneously. As in Eq. (32), the effective magnetic fields can be also defined in terms of commutator relations

$$\mathcal{E}^{(r)} = \frac{1}{i\hbar}[E^{(d)} \overset{*}{,} P^{(d)}], \quad \mathcal{E}^{(p)} = -\frac{1}{i\hbar}[E^{(d)} \overset{*}{,} R^{(d)}]. \quad (43)$$

There exists a sum rule for the fictitious fields

$$\text{Tr}_N\mathcal{B}^{(p)} = 0, \quad \text{Tr}_N\mathcal{B}^{(r)} = 0, \quad \text{Tr}_N\underline{\underline{\Theta}}^{rp} = 0,$$
$$\text{Tr}_N\mathcal{E}^{(p)} = 0, \quad \text{Tr}_N\mathcal{E}^{(r)} = 0, \quad (44)$$

which can be found by taking the trace over all bands (we denote this sum over bands as Tr_N here and throughout this work) and making use of the identity (16). This means that all the effective forces for each band balance each other in total, or in other words, if all bands are completely filled (and thus, the density operator ρ is proportional to the unit matrix $\mathbb{1}_N$), the system does not experience any net force.

The basis of all calculations within the Wigner representation of quantum theory is the Moyal bracket $[\bar{\mathcal{S}} \overset{*}{,} \bar{\mathcal{T}}]$ between two operators $\bar{\mathcal{S}}$ and $\bar{\mathcal{T}}$. Now, we want to rewrite this in terms of kinetic variables only, which is achieved by transforming the derivatives in the Moyal product to act on kinetic variables. For concise notation, we introduce the tensor of Berry curvatures as [α, β denote indices with respect to $(t, \boldsymbol{r}, \boldsymbol{p})$],

$$\underline{\Theta}_{\alpha,\beta} = \hbar\big(\partial_\alpha \mathcal{A}_\beta^{(d)} - \partial_\beta \mathcal{A}_\alpha^{(d)} - i\big[\mathcal{A}_\alpha^{(d)}, \mathcal{A}_\beta^{(d)}\big]\big), \qquad (45)$$

or explicitly, in terms of the fictitious fields used previously:

$$\underline{\Theta} = \begin{pmatrix} 0 & -\mathcal{E}^{(r)} & -\mathcal{E}^{(p)} \\ \mathcal{E}^{(r)} & \epsilon_{ijk}\mathcal{B}_k^{(r)} & \underline{\Theta}^{rp} \\ \mathcal{E}^{(p)} & -(\underline{\Theta}^{rp})^T & \epsilon_{ijk}\mathcal{B}_k^{(p)} \end{pmatrix}, \qquad (46)$$

where we recognize the top-left part of $\underline{\Theta}$ as being essentially the electromagnetic field tensor.

Neglecting terms of order $O(\hbar^2)$, we find the explicit form of the Moyal product after the transformation to kinetic variables

$$* \to \exp\left[\frac{i\hbar}{2}(-\overleftarrow{\partial}_X \overrightarrow{\partial}_\Pi + \overleftarrow{\partial}_\Pi \overrightarrow{\partial}_X) + \frac{i\hbar}{2}(\overleftarrow{\partial}_\Pi \overleftarrow{\partial}_R)\underline{\Theta}\begin{pmatrix} \overrightarrow{\partial}_\Pi \\ \overrightarrow{\partial}_R \end{pmatrix}\right]. \qquad (47)$$

This is a central result of this section, since it shows that expressing all quantities in terms of kinetic variables allows us to deal solely with manifest gauge invariant expressions. Essentially, the bottom line of this parameter transformation is that it changes the metric of the Moyal product by the appearance of the Berry curvatures $\underline{\Theta}$. Immediate consequences of result (47) are the equations of motions to be discussed in detail in Sec. IV.

B. Transformation of general operators

Let us briefly comment on observables with nontrivial band matrix structure $\mathcal{O}(\boldsymbol{x}, \boldsymbol{\pi})$, and their transformation into the rotated frame, which is performed analogously to the transformation (19),

$$\bar{\mathcal{O}} = \mathcal{U} * \mathcal{O} * \mathcal{U}^\dagger = \bar{\mathcal{O}}_0 + [\mathcal{Y}_1, \bar{\mathcal{O}}_0] - \frac{\hbar}{2}\left\{\mathcal{A}_\pi, \left[\partial_x - \frac{i}{2}\mathcal{A}_x, \bar{\mathcal{O}}_0\right]\right\} + \frac{\hbar}{2}\left\{\mathcal{A}_x, \left[\partial_\pi - \frac{i}{2}\mathcal{A}_\pi, \bar{\mathcal{O}}_0\right]\right\} + O(\hbar^2), \qquad (48)$$

and the back transformation is given by

$$\mathcal{U}_0 \mathcal{O} \mathcal{U}_0^\dagger = \mathcal{U}_0(\mathcal{U}^\dagger * \bar{\mathcal{O}} * \mathcal{U})\mathcal{U}_0^\dagger = \bar{\mathcal{O}} - \hbar[\mathcal{Y}_1, \bar{\mathcal{O}}] + \frac{\hbar}{2}\left\{\mathcal{A}_\pi, \left[\partial_x - \frac{i}{2}\mathcal{A}_x, \bar{\mathcal{O}}\right]\right\} - \frac{\hbar}{2}\left\{\mathcal{A}_x, \left[\partial_\pi - \frac{i}{2}\mathcal{A}_\pi, \bar{\mathcal{O}}\right]\right\} + O(\hbar^2), \qquad (49)$$

which can be readily checked by plugging Eq. (48) into the back transformation (49) and dropping terms of order $\hbar^2$. Now, treating the diagonal and off-diagonal part of $\bar{\mathcal{O}}_0 = \mathcal{U}_0 \mathcal{O} \mathcal{U}_0^\dagger = \bar{\mathcal{O}}_0^{(d)} + \bar{\mathcal{O}}_0^{(o)}$ separately, we find

$$\bar{\mathcal{O}}_0^{(d)}(\boldsymbol{X}, \boldsymbol{\Pi}) + \frac{i\hbar}{4}\mathcal{P}_d\big(\{\mathcal{A}_\pi, [\mathcal{A}_x, \bar{\mathcal{O}}_0^{(d)}]\} - \frac{i\hbar}{4}\{\mathcal{A}_x, [\mathcal{A}_\pi, \bar{\mathcal{O}}_0^{(d)}]\}\big)\mathcal{P}_d + O(\hbar^2), \qquad (50)$$

and the contribution arising from the off-diagonal part of $\bar{\mathcal{O}}_0^{(o)}$,

$$\mathcal{P}_d\left([\mathcal{Y}_1, \bar{\mathcal{O}}_0^{(o)}] - \frac{\hbar}{2}\left\{\mathcal{A}_\pi, \left[\partial_x - \frac{i}{2}\mathcal{A}_x, \bar{\mathcal{O}}_0^{(o)}\right]\right\} + \frac{\hbar}{2}\left\{\mathcal{A}_x, \left[\partial_\pi - \frac{i}{2}\mathcal{A}_\pi, \bar{\mathcal{O}}_0^{(o)}\right]\right\}\right)\mathcal{P}_d + O(\hbar^2), \qquad (51)$$

which are both independently gauge invariant. While the gauge invariance of the former is straightforward to show, the later requires more work and we have to take into account that $\mathcal{Y}_1$ is modified under a gauge transformation as

$$\mathcal{Y}_1 \to \mathcal{Y}_1 - \frac{i\hbar}{4}(\{\chi_x, \mathcal{A}_\pi\} - \{\chi_\pi, \mathcal{A}_x\}), \qquad (52)$$

along with $\partial_\alpha \bar{\mathcal{O}} \to \partial_\alpha \bar{\mathcal{O}} + i[\chi_\alpha, \bar{\mathcal{O}}]$ and $\mathcal{A}_\alpha \to \mathcal{A}_\alpha + \chi_\alpha$.

C. Expectation values in the rotated frame

Let us now go back to the initial question of the dynamics of our system within the effective theory by studying the expectation values of physical observables, which can be

obtained in the Wigner representation by the integration over the complete phase space,

$$\langle \mathcal{S} \rangle = \int d^d r \int \frac{d^d p}{(2\pi\hbar)^d} \text{Tr}_N \{\rho(\boldsymbol{x}, \boldsymbol{\pi}) * \mathcal{S}(\boldsymbol{x}, \boldsymbol{\pi})\}, \qquad (53)$$

and the trace is with respect to the matrix structure. Note that the factor $1/(2\pi\hbar)^d = 1/h^d$ describes the proper quantization of the phase space volume and thus is directly obtained by transforming quantum averages into the Wigner representation.

If we assume the integration over the whole phase space to be unbounded and any surface contribution from the integrand at infinity to vanish, we can perform partial integrations to show that all the partial derivatives in the Moyal product $*$

cancel each other, so that we can equally write

$$\langle \mathcal{S} \rangle = \int \mathrm{d}^d r \int \frac{\mathrm{d}^d p}{(2\pi\hbar)^d} \, \mathrm{Tr}_N \{ \rho(\boldsymbol{x},\boldsymbol{\pi}) \mathcal{S}(\boldsymbol{x},\boldsymbol{\pi}) \}. \quad (54)$$

By using the cyclic property of the trace and by partial integration, we can easily show that

$$\langle \mathcal{O} \rangle = \int \frac{\mathrm{d}^d r \, \mathrm{d}^d p}{(2\pi\hbar)^d} \, \mathrm{Tr}_N \{ \mathcal{U} * \mathcal{O} * \mathcal{U}^\dagger \} = \langle \bar{\mathcal{O}} \rangle, \quad (55)$$

and we can write

$$\langle \mathcal{S} \rangle = \langle \bar{\mathcal{S}} \rangle = \int \mathrm{d}^d r \int \frac{\mathrm{d}^d p}{(2\pi\hbar)^d} \, \mathrm{Tr}_N \{ \bar{\rho}(\boldsymbol{x},\boldsymbol{\pi}) \bar{\mathcal{S}}(\boldsymbol{x},\boldsymbol{\pi}) \}. \quad (56)$$

In equilibrium, the density operator is given by

$$\rho(\boldsymbol{x},\boldsymbol{\pi}) = f_\mathrm{D}(\epsilon) \, \delta(\bar{\Xi}) = f_\mathrm{D}(\epsilon) \, \delta[\epsilon - \bar{\mathcal{H}}(\boldsymbol{x},\boldsymbol{\pi})], \quad (57)$$

or, if we do not want our results to be energy resolved, we use directly

$$\rho(\boldsymbol{x},\boldsymbol{p}) = \int \mathrm{d}\epsilon \, f_\mathrm{D}(\epsilon)\delta[\epsilon - \bar{\mathcal{H}}(\boldsymbol{x},\boldsymbol{\pi})], \quad (58)$$

which gives us the density in phase space and, as we will discuss later, however, it is to be interpreted as a quasiprobability. An explicitly time-dependent Hamiltonian $\mathcal{H}(\boldsymbol{x},\boldsymbol{\pi})$ has to be treated using Eq. (97) instead. If a band is completely filled, it becomes 1 at the diagonal element corresponding to that band. If energies are well separated, and we assume excitations localized in energy space, we can assume any off-diagonal elements in the density operator to vanish. In fact, those off-diagonal entries correspond to coherent excitations that are split amongst several bands. If the band splitting is sufficiently large, this leads to rapid oscillations subject to decoherence. In the end, the projection operation defined by $\mathcal{P}_\mathrm{d}$ is essentially enforced by the diagonal representation of the Hamiltonian $\bar{\Xi}$, and by the density matrix ρ, which gives us only those states that have coherences within bands (or degenerate/overlapping bands so that, again, the energy argument applies). In particular, this is certainly true for low-energy transport, where physics takes place only in the vicinity of the Fermi level.

In order to proceed, let us transform the integration variables to kinetic ones, and in doing so, we also have to take into account how the volume element in phase space changes, which is given by the determinant of the Jacobian,

$$D^{-1} \equiv \det \frac{\partial(\boldsymbol{R},\boldsymbol{P})}{\partial(\boldsymbol{r},\boldsymbol{p})} = \det(\underline{1} + \underline{\underline{\Theta}}^{rp})$$

$$= 1 + \mathrm{Tr}\underline{\underline{\Theta}}^{rp} + O(\hbar^2), \quad (59)$$

or $D(\boldsymbol{X},\boldsymbol{\Pi}) = 1 - \mathrm{Tr}\underline{\underline{\Theta}}^{rp} + O(\hbar^2)$. In non-Abelian situations, the nontrivial matrix structure of the Berry curvature will be inherited by D, which will be accounted for by performing the integration before taking the trace, thus yielding

$$\langle \mathcal{S} \rangle = \mathrm{Tr}_N \left\{ \int \frac{\mathrm{d}^d R \, \mathrm{d}^d P}{(2\pi\hbar)^d} \, D(\boldsymbol{X},\boldsymbol{\Pi}) \, \bar{\rho}(\boldsymbol{X},\boldsymbol{\Pi}) \bar{\mathcal{S}}(\boldsymbol{X},\boldsymbol{\Pi}) \right\}. \quad (60)$$

D describes, for example, charge accumulation in the case of a topological insulator with a magnetization structure induced by ferromagnetic exchange. This effect of the Berry curvature D on the density of states has been already discovered by Xiao and coworkers.[44]

Let us look at this in another way by using Eq. (49) to transform an observable $\bar{\mathcal{O}}$ back into the original frame. In addition, we let $\mathcal{S}$ be a general observable that can posses an arbitrary matrix structure, so that contrary to relation (27), the additional dipole term becomes relevant. In the end, we want to establish the connection with Eq. (60), so we are interested in expectation values or phase-space densities (which then are quasiprobability distributions as discussed later),

$$s(\boldsymbol{x},\boldsymbol{\pi}) = \mathrm{Tr}_N \big(\tfrac{1}{2} \{ \rho \overset{*}{,} \mathcal{S} \} \big) = \mathrm{Tr}_N \mathcal{O}. \quad (61)$$

We identify $\bar{\mathcal{O}} = \frac{1}{2} \{ \bar{\rho} \overset{*}{,} \bar{\mathcal{S}} \}$ and plug in the back transformation (49), so that

$$s(\boldsymbol{x},\boldsymbol{\pi}) = \mathrm{Tr}_N \{ \bar{\mathcal{O}} + \hbar \mathcal{A}_\pi \partial_x \bar{\mathcal{O}} - \hbar \mathcal{A}_x \partial_\pi \bar{\mathcal{O}} \}$$
$$+ i\hbar \mathrm{Tr} \{ \bar{\mathcal{O}} [\mathcal{A}_x, \mathcal{A}_\pi] \} + O(\hbar^2).$$

According to the discussion above, it is reasonable to assume that our observable is given as a function of the kinetic variables, i.e., $\bar{\mathcal{O}}(\boldsymbol{X},\boldsymbol{\Pi})$ and it is instructive to treat the interband and the intraband contributions separately by splitting $\bar{\mathcal{O}} = \mathcal{P}_\mathrm{d} \bar{\mathcal{O}} \mathcal{P}_\mathrm{d} + \bar{\mathcal{O}}^{(\mathrm{o})} \equiv \bar{\mathcal{O}}^{(\mathrm{d})} + \bar{\mathcal{O}}^{(\mathrm{o})}$ and likewise for $\bar{\mathcal{S}} = \bar{\mathcal{S}}^{(\mathrm{d})} + \bar{\mathcal{S}}^{(\mathrm{o})}$, so that the contribution from the diagonal part becomes[45]

$$s^{(\mathrm{d})}(\boldsymbol{x},\boldsymbol{\pi}) = \mathrm{Tr}_N \left\{ D(\boldsymbol{x},\boldsymbol{\pi}) \, \bar{\mathcal{O}}^{(\mathrm{d})} \big(\boldsymbol{X} + \hbar \mathcal{A}_\pi^{(\mathrm{d})}, \boldsymbol{\Pi} - \hbar \mathcal{A}_x^{(\mathrm{d})} \big) \right\}$$
$$+ O(\hbar^2), \quad (62)$$

which basically undoes the variable transformation so that we go back to the canonical pair of variables and can write

$$s^{(\mathrm{d})}(\boldsymbol{x},\boldsymbol{\pi}) = \mathrm{Tr}_N \{ \bar{\rho}(\boldsymbol{x},\boldsymbol{\pi}) \, D(\boldsymbol{x},\boldsymbol{\pi}) \, \bar{\mathcal{S}}^{(\mathrm{d})}(\boldsymbol{x},\boldsymbol{\pi}) \} + O(\hbar^2). \quad (63)$$

In addition, we rewrote the last term with the help of identity (16), $i[\mathcal{A}_x, \mathcal{A}_\pi] = \partial_x \mathcal{A}_\pi - \partial_\pi \mathcal{A}_x$ and, according to our previous discussion, we have $\Theta^{\epsilon t} = 0$, as $\mathcal{U}_0$ was assumed to not explicitly depend on the energy parameter so we can replace this term with $\mathrm{Tr}\underline{\underline{\Theta}}^{rp}$. This contribution gives rise to the correction factor $D(\boldsymbol{x},\boldsymbol{\pi})$ that we already encountered before, and thus, the last result is consistent with relation (60).

The implications of the diagonal part of $\bar{\mathcal{S}}$ can be summarized as undoing the parameter transformation together with the appearance of the correction factor $D(\boldsymbol{x},\boldsymbol{p})$, which locally changes the density. However, it is not always possible to ignore the off-diagonal part of the observable $\bar{\mathcal{S}}$, one prominent example will be the current density [see Eq. (85)]. With a series of straightforward manipulations involving the cyclic property of the trace along with identity (16), we eventually arrive at

$$s^{(\mathrm{o})}(\boldsymbol{x},\boldsymbol{\pi}) = \partial_x \mathrm{Tr}_N \bar{\rho} \, \frac{\hbar}{2} \{ \mathcal{A}_\pi, \bar{\mathcal{S}}^{(\mathrm{o})} \}$$
$$- \partial_\pi \mathrm{Tr}_N \bar{\rho} \, \frac{\hbar}{2} \{ \mathcal{A}_x, \bar{\mathcal{S}}^{(\mathrm{o})} \} + O(\hbar^2), \quad (64)$$

so that both contributions to the expectation value, Eqs. (63) and (64) together read

$$s = \mathrm{Tr}_N \bar{\rho} \, D \bar{\mathcal{S}}^{(\mathrm{d})} + \partial_x \mathrm{Tr}_N \bar{\rho} \, \frac{\hbar}{2} \{ \mathcal{A}_\pi, \bar{\mathcal{S}}^{(\mathrm{o})} \}$$
$$- \partial_\pi \mathrm{Tr}_N \bar{\rho} \, \frac{\hbar}{2} \{ \mathcal{A}_x, \bar{\mathcal{S}}^{(\mathrm{o})} \} + O(\hbar^2). \quad (65)$$

The importance of these last two terms will become clearer in Sec. IV A when discussing the kinetic equations of the effective theory.

To summarize this section, we have seen that when we use the kinetic terms X and $\mathbf{\Pi}$ as basic quantities for our observables, we end up with expressions that are manifest gauge invariant [c.f. Eqs. (47) and (60)]. In fact, these kinetic variables appear consistently in virtually all equations of physical relevance, and moreover, it is exactly these quantities that we obtain, if we transform the canonical variables into the rotated frame,

$$X = \mathcal{U} * x * \mathcal{U}^\dagger = x - i\hbar\mathcal{U} * \partial_{\boldsymbol{\pi}}\mathcal{U}^\dagger,$$
$$\mathbf{\Pi} = \mathcal{U} * \boldsymbol{\pi} * \mathcal{U}^\dagger = \boldsymbol{\pi} + i\hbar\mathcal{U} * \partial_{\boldsymbol{x}}\mathcal{U}^\dagger.$$

D. Is the diagonalization transformation canonical?

Before continuing, we would like to point out that without the projection, X and $\mathbf{\Pi}$ still obey the canonical commutation relations, which can be easily seen by noting that the set of unprojected Berry curvatures vanishes according to identity (16) [for example, Eqs. (34) or (41) with $\mathcal{A}^{(d)}$ replaced by the $N \times N$ Berry connection $\mathcal{A}$]. This is actually not surprising, since then our unitary matrix $\mathcal{U}_0$ is connected by $SU(N)$ gauge invariance to the identity transformation, which clearly has vanishing Berry curvature, or in other words, $\mathcal{U}_0$ itself is a $SU(N)$ gauge transformation, which naturally keeps the full $SU(N)$ Berry curvature invariant. This implies that our unitary transformation is a canonical one, however, since X and $\mathbf{\Pi}$ now posses a complicated matrix structure in the N-dimensional band space, they no longer commute with nontrivial matrices within this band space. For example, in a simple particle-hole symmetric two-band model,

$$\mathcal{H} = \boldsymbol{E}(\boldsymbol{p}) \cdot \boldsymbol{\sigma} + V(\boldsymbol{r})\mathbb{1}_2, \tag{66}$$

our band-diagonalized Hamiltonian has the form

$$\bar{\mathcal{H}} = E(\boldsymbol{p})\sigma_z + V(\boldsymbol{R})\mathbb{1}_2, \tag{67}$$

while $\boldsymbol{R} = \boldsymbol{r} + \hbar\mathcal{A}_{\boldsymbol{p}}$ acquires a 2×2 Berry connection matrix, so that $[\boldsymbol{R}_i, \boldsymbol{R}_j] = 0 = [\boldsymbol{P}_i, \boldsymbol{P}_j]$ and $[\boldsymbol{R}_i, \boldsymbol{P}_j] = i\hbar\delta_{ij}$. Instead, the commutator $[\boldsymbol{R}, \sigma_z] = \hbar[\mathcal{A}_{\boldsymbol{p}}, \sigma_z] \neq 0$ now encodes the complicated dynamics of interband scattering, making the problem as a whole not easier tractable so, in the general case, the only way out is the truncation scheme. And only due to the restriction of the Berry connection matrices into a certain subspace do the corresponding Berry curvatures yield a nonvanishing value.

E. Electronic spectrum and magnetic dipole energy

In course of the preceding discussion, we have seen that rewriting the Hamiltonian in terms of kinetic variables would render it gauge invariant. However, in order to calculate the electronic spectrum, one essentially has to fully diagonalize it, and which has to be done in terms of canonical variables. But it turns out that this remaining gauge-dependence of the Hamiltonian in the canonical representation would only affect the wave functions or quantities that build upon them like the retarded Greens function or the density operator. These objects will acquire local phase factors that depend on $\boldsymbol{x}$ and $\boldsymbol{p}$, and we will illustrate this point later by explicitly studying the situation in the case of the Dirac equation. At this time, we

can conclude that the electronic spectrum of the system is also gauge independent.

Let us briefly discuss the last term of the transformation equations like Eq. (22), and that we ignored by now,

$$\bar{\mathcal{H}}_\mathcal{M} = \frac{i\hbar}{4}\mathcal{P}_\mathrm{d}(\{\mathcal{A}_{\boldsymbol{r}}, [\mathcal{A}_{\boldsymbol{p}}, \bar{\mathcal{H}}_0] - \{\mathcal{A}_{\boldsymbol{p}}, [\mathcal{A}_{\boldsymbol{r}}, \bar{\mathcal{H}}_0]\}\})\mathcal{P}_\mathrm{d}, \tag{68}$$

which gives rise to corrections due to virtual transitions to other bands. The term "magnetic" comes from the fact that $\bar{\mathcal{H}}_\mathcal{M}$ corresponds to the energy of a magnetic dipole in an external magnetic field, where the magnetic dipole is an intrinsic property of the band. For example, in the case of the Dirac equation, this term becomes essentially the magnetic Zeeman term, though in other scenarios, one obtains generalizations thereof, and even in the absence of external magnetic fields, $\bar{\mathcal{H}}_\mathcal{M}$ can be nonzero.

Since we will encounter terms like in $\bar{\mathcal{H}}_\mathcal{M}$ later, let us define the quantity [α, β are any combinations of $(t, \boldsymbol{r}, \boldsymbol{p})$]

$$\underline{\underline{\Omega}}_{\alpha,\beta} = \frac{i\hbar}{2}\mathcal{P}_\mathrm{d}\{\mathcal{A}_\alpha, [\mathcal{A}_\beta, \bar{\mathcal{H}}_0]\}\mathcal{P}_\mathrm{d}, \tag{69}$$

which is an antisymmetric tensor in $(t, \boldsymbol{r}, \boldsymbol{p})$ space, and in addition, it is a band-diagonal matrix, or block diagonal according to the structure defined by the projector $\mathcal{P}_\mathrm{d}$. The explicit structure of $\underline{\underline{\Omega}}$ is

$$\underline{\underline{\Omega}} = \begin{pmatrix} 0 & -\boldsymbol{\Omega}^{tr} & -\boldsymbol{\Omega}^{tp} \\ \boldsymbol{\Omega}^{tr} & \epsilon_{ijk}\Omega_k^{rr} & \underline{\underline{\Omega}}^{rp} \\ \boldsymbol{\Omega}^{tp} & -(\underline{\underline{\Omega}}^{rp})^T & \epsilon_{ijk}\Omega_k^{pp} \end{pmatrix}, \tag{70}$$

which has been defined in analogy to the Berry curvatures $\underline{\underline{\Theta}}$ given in Eq. (46).

Now, we can express the energy term as

$$\bar{\mathcal{H}}_\mathcal{M} = \mathrm{Tr}\underline{\underline{\Omega}}^{rp}, \tag{71}$$

where the trace is only with respect to coordinates, and not band indices, and we write explicitly

$$\underline{\underline{\Omega}}_{ij}^{rp} = \frac{i\hbar}{2}\mathcal{P}_\mathrm{d}\{\mathcal{A}_{r_i}, [\mathcal{A}_{p_j}, \bar{\mathcal{H}}_0]\}\mathcal{P}_\mathrm{d}. \tag{72}$$

As will become more apparent later, terms involving the $\underline{\underline{\Omega}}$ tensor are related to circular currents and give rise to important terms that should not be ignored.

For the sake of completeness, and in order to establish the link to other treatments in literature (for example, Ref. 3, and references therein), let us express our quantities $\underline{\underline{\Theta}}$ and $\underline{\underline{\Omega}}$ in terms of Bloch functions $(\mathcal{A}_{\boldsymbol{p}})_{ij} = \langle u_i | i\partial_{\boldsymbol{p}} | u_j \rangle$, with the Bloch band indices i and j. We start by introducing the gauge-invariant transition elements

$$\Gamma_{ij}^{(\alpha,\beta)} \equiv -2\hbar\Im\{(\mathcal{A}_\alpha)_{ij}(\mathcal{A}_\beta)_{ji}\}$$
$$= -2\hbar\Im\{\langle u_i | i\partial_\alpha | u_j \rangle \langle u_j | i\partial_\beta | u_i \rangle\}, \tag{73}$$

which are anti-Hermitian, $\Gamma_{ij}^{(\alpha,\beta)} = -\Gamma_{ji}^{(\alpha,\beta)}$ due to the Hermiticity of $\mathcal{A}$ and, as the name suggests, they describe corrections due to virtual transitions between band i and j. Then the Berry curvatures projected onto band i can be expressed as

$$\underline{\underline{\Theta}}_{\alpha\beta} = \hbar(\partial_\alpha\mathcal{A}_\beta^{(d)} - \partial_\beta\mathcal{A}_\alpha^{(d)})_{ii} = \sum_j \Gamma_{ij}^{(\alpha,\beta)}, \tag{74}$$

and one can interpret the Berry connection as being corrections to the kinetic variables x, π due to virtual transitions into all other bands j. The sum rule (44) is then a direct consequence of the anti-Hermiticity of Γ. In the same spirit, the dipole terms projected on Bloch band i read

$$\underline{\underline{\Omega}}_{\alpha\beta} = \frac{1}{2} \sum_j (\epsilon_i - \epsilon_j) \Gamma_{ij}^{(\alpha,\beta)}, \tag{75}$$

where the energies are $\epsilon_i \equiv (\bar{\mathcal{H}}_0)_{ii}$.

As might be apparent from the above expressions, there exists a direct link between $\underline{\underline{\Theta}}$ and $\underline{\underline{\Omega}}$ in the case of a two-level system, and which we establish in the following. First, a general diagonal 2-level Hamiltonian $\bar{\mathcal{H}}_0$ can be divided into a symmetric part $\propto \mathbb{1}_2$, which consequently commutes out in expression (69), and an antisymmetric part $\propto \sigma_z$ responsible for a finite contribution. After some simple algebra, we find $\text{Tr}_2 \sigma_z \underline{\underline{\Omega}} = 0$, which implies that $\underline{\underline{\Omega}}$ is band diagonal, and the two bands get the same contribution, in particular the magnetic dipole energy is the same for the two bands. The diagonal part is obtained by taking the trace and using its cyclic properties,

$$\begin{aligned}
\text{Tr}_2 \underline{\underline{\Omega}}_{\alpha,\beta} &= i\hbar \text{Tr}_2 \bar{\mathcal{H}}_0 [\mathcal{A}_\alpha, \mathcal{A}_\beta] \\
&= \hbar \text{Tr}_2 \bar{\mathcal{H}}_0 \left(\partial_\alpha \mathcal{A}_\beta^{(\mathrm{d})} - \partial_\beta \mathcal{A}_\alpha^{(\mathrm{d})} \right) = \text{Tr}_2 (\bar{\mathcal{H}}_0 \underline{\underline{\Theta}}_{\alpha,\beta}), \quad (76)
\end{aligned}$$

where in the last two steps, we used the identity (16) along with definition (74) to replace the commutator. We can finally cast the result into the form

$$\underline{\underline{\Omega}} = \tfrac{1}{2} \text{Tr}_2 (\bar{\mathcal{H}}_0 \underline{\underline{\Theta}}) \mathbb{1}_2, \tag{77}$$

where in this expression, the diagonal contribution of $\bar{\mathcal{H}}_0$ drops out because of the sum rule $\text{Tr}_2 \underline{\underline{\Theta}} = 0$.

In particular, we can express the magnetic dipole energy as

$$\bar{\mathcal{H}}_{\mathcal{M}} = \tfrac{1}{2} \text{Tr}_2 (\bar{\mathcal{H}}_0 \text{Tr} \underline{\underline{\Theta}}^{rp}) \mathbb{1}_2, \tag{78}$$

where $\text{Tr} \underline{\underline{\Theta}}^{rp} \equiv \sum_i \underline{\underline{\Theta}}_{ii}^{rp}$, whereas Tr_2 denotes the trace with respect to the two-level band space.

IV. EQUATIONS OF MOTION

We are now interested in the dynamics of the systems, especially, in a consistent effective descriptions of the physics within a certain band.

A. Quasiprobability distributions and currents

Within the Wigner framework, the basic quantity is the quasiprobability distribution $\rho(r, p)$, while the usual momentum and position probability distributions can be generally defined as marginals:

$$\rho(r) = \int \frac{\mathrm{d}^d p}{(2\pi\hbar)^d} \rho(p, r), \quad \rho(p) = \int \frac{\mathrm{d}^d r}{(2\pi\hbar)^d} \rho(p, r),$$

which are linked to the probability interpretation of quantum mechanics, i.e., $\rho(r) = |\psi(r)|^2$ and $\rho(p) = |\psi(p)|^2$. In the rotated frame, however, one has to express them in terms of kinetic variables instead, in order to obtain gauge invariant results. Now the goal, which we want to pursue in the following, is to find proper quantities in the rotated frame that lead to a consistent physical description when projected onto a certain band.

Since the velocity operator is given by

$$\frac{\mathrm{d} R}{\mathrm{d} t} = \frac{1}{i\hbar} [\bar{\Xi} \overset{\star}{,} R], \tag{79}$$

and inspired by Eq. (65) for the expectation values in the rotated frame, we define the quasiprobability density $\mathfrak{n}$, current density j, and force density q as follows:

$$\mathfrak{n}(R, P, t) \equiv \text{Tr}_N (\bar{\rho}\, D), \tag{80}$$

$$\begin{aligned}
j(R, P, t) \equiv \text{Tr}_N \left(\bar{\rho}\, D \frac{\mathrm{d} R}{\mathrm{d} t} \right) &- \nabla_R \times \text{Tr}_N \bar{\rho} \Omega^{pp} \\
&- \nabla_P \cdot \text{Tr}_N \bar{\rho} \underline{\underline{\Omega}}^{rp}, \quad (81)
\end{aligned}$$

$$\begin{aligned}
q(R, P, t) \equiv \text{Tr}_N \left(\bar{\rho}\, D \frac{\mathrm{d} P}{\mathrm{d} t} \right) &- \nabla_P \times \text{Tr}_N \bar{\rho} \Omega^{rr} \\
&+ \nabla_R \cdot \text{Tr}_N \bar{\rho} (\underline{\underline{\Omega}}^{rp})^T, \quad (82)
\end{aligned}$$

where we used that $\underline{\underline{\Omega}}^{pr} = -(\underline{\underline{\Omega}}^{rp})^T$. These densities obey a conservation law in phase space, or Liouville's theorem that states

$$\partial_t \mathfrak{n}(R, P, t) + \nabla_R\, j(R, P, t) + \nabla_P\, q(R, P, t) = 0. \tag{83}$$

We substitute the definitions (80)–(82) into Liouville's theorem and establish the identity (83) after some algebra, using the kinetic equation (95) for $\bar{\rho}$ and the equality of second partial derivatives and dropping terms of order $O(\hbar^2)$. Furthermore, we need to use the following identities between Berry curvatures:

$$\begin{aligned}
\partial_t D - \nabla_R \mathcal{E}^{(p)} + \nabla_P \mathcal{E}^{(r)} &= 0, \\
\partial_{R_i} D + \partial_{R_k} \underline{\underline{\Theta}}_{ik}^{rp} - (\nabla_P \times \mathcal{B}^{(r)})_i &= 0, \\
-\partial_{P_i} D - \partial_{P_k} \underline{\underline{\Theta}}_{ki}^{rp} - (\nabla_R \times \mathcal{B}^{(p)})_i &= 0,
\end{aligned} \tag{84}$$

which can be readily shown by plugging in the definitions (33)–(35), (40), and (41). This result strongly emphasizes the importance of including the correction factor $D(R, P, t)$ into the expectation values.

Here, it is important to realize that the current matrices have an off-diagonal structure that one has to take into account in the light of expression (65) and which eventually leads to the last two terms in the current densities (81) and (82), and constitute divergence-free, or circular currents. Explicitly, for the 4-component current, this off-diagonal part takes the form

$$\left(\frac{\mathrm{d} X}{\mathrm{d} t} \right)_{\mathrm{o}} = -i[\bar{\mathcal{H}}_0, \mathcal{A}_\pi] \tag{85}$$

and is completely expressed in terms of quantities of the rotated frame and arises due to the off-diagonal elements of the Berry connection matrix.

Just like the Wigner function is a quasiprobability and can be given physical sense only after taking expectation values, the same applies to $\mathfrak{n}$, j, and q. In particular, physical meaning can be given only to quantities like $n(R)$ or $j(R)$. This is related to Heisenberg's uncertainty, which states that momentum and position uncertainty have to be larger than Plank's constant $\hbar$, i.e., $\Delta P \Delta R \gtrsim \hbar$. The same is true for the conjugate variables time and energy.

Liouville's theorem constitutes a conservation law for the quasiprobability densities, and one can, for example, integrate

Eq. (83) over all momenta in order to find a continuity equation for the probability densities of the particle and current densities,

$$\partial_t n(\boldsymbol{R},t) + \nabla_R \boldsymbol{j}(\boldsymbol{R},t) = -\oint_{\partial V} \mathrm{d}^{d-1} \boldsymbol{P} \, \boldsymbol{q}(\boldsymbol{R},\boldsymbol{P},t), \quad (86)$$

where the closed surface integral on the right-hand side has been obtained by virtue of Gauss theorem, and it vanishes if we assume that there is no net momentum current flow through the surface at infinity. Explicitly, the particle density is

$$n(\boldsymbol{R},t) = \int \frac{\mathrm{d}^d P}{(2\pi\hbar)^d} \, \mathrm{Tr}_N \left(\bar{\rho} \, D \right), \quad (87)$$

and likewise for the current density,

$$\boldsymbol{j}(\boldsymbol{R},t) = \int \frac{\mathrm{d}^d P}{(2\pi\hbar)^d} \, \mathrm{Tr}_N \left(\bar{\rho} \, D \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t} \right) - \nabla_R \times \int \frac{\mathrm{d}^d P}{(2\pi\hbar)^d}$$
$$\times \mathrm{Tr}_N \bar{\rho} \boldsymbol{\Omega}^{pp} - \oint \frac{\mathrm{d}^{d-1} \boldsymbol{S}_P}{(2\pi\hbar)^d} \mathrm{Tr}_N \bar{\rho} \underline{\underline{\Omega}}^{rp}, \quad (88)$$

while the previous discussion has shown that these definitions are in fact meaningful physical quantities and constitute the central result of this section. We will see interesting implications of the additional dipole term in the current when studying the examples of the Dirac equation.

For the sake of completeness, we use the result (65) to define densities corresponding to the physical quantity $\mathcal{S}$ in the diagonalized frame:

$$S(\boldsymbol{R},t) \equiv \int \frac{\mathrm{d}^d P}{(2\pi\hbar)^d} \, \mathrm{Tr}_N \left(\bar{\rho} \, D\bar{\mathcal{S}} - \nabla_R \, \bar{\rho} \frac{\hbar}{2} \{ \mathcal{A}_P, \bar{\mathcal{S}}^{(o)} \} \right)$$
$$+ \oint \frac{\mathrm{d}^{d-1} \boldsymbol{P}}{(2\pi\hbar)^d} \mathrm{Tr}_N \bar{\rho} \frac{\hbar}{2} \{ \mathcal{A}_R, \bar{\mathcal{S}}^{(o)} \}, \quad (89)$$

and likewise for $S(\boldsymbol{P})$. This expression is consistent with the definition of a density operator $\delta(\boldsymbol{r} - \boldsymbol{r}_0) S(\boldsymbol{r},\boldsymbol{p})$ which, when transformed into the rotated frame yields the given result. Furthermore, all the given expressions can still explicitly depend on energy via the gauge invariant parameter E, which, however, does not affect the discussion here.

B. Operator equations of motion

In order to calculate the densities (81) and (82), we need to evaluate the equations of motion for kinetic position $\boldsymbol{R}$ and momentum $\boldsymbol{P}$. Using this result and Eq. (47), we can now immediately write down the equations of motion:

$$\frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t} = \frac{1}{i\hbar}[\bar{\Xi} \overset{*}{,} \boldsymbol{R}] = \frac{\partial\bar{\mathcal{H}}}{\partial\boldsymbol{P}}(\underline{\underline{1}} + \underline{\underline{\Theta}}^{rp}) - \mathcal{E}^{(p)} + \frac{\partial\bar{\mathcal{H}}}{\partial\boldsymbol{R}} \times \mathcal{B}^{(p)}, \quad (90)$$

$$\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = \frac{1}{i\hbar}[\bar{\Xi} \overset{*}{,} \boldsymbol{P}] = -(\underline{\underline{1}} + \underline{\underline{\Theta}}^{rp})\frac{\partial\bar{\mathcal{H}}}{\partial\boldsymbol{R}} + \mathcal{E}^{(r)} + \frac{\partial\bar{\mathcal{H}}}{\partial\boldsymbol{P}} \times \mathcal{B}^{(r)}, \quad (91)$$

or, to leading order in $\hbar$,

$$\frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t}(\underline{\underline{1}} - \underline{\underline{\Theta}}^{rp}) = \frac{\partial\bar{\mathcal{H}}}{\partial\boldsymbol{P}} - \mathcal{E}^{(p)} - \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} \times \mathcal{B}^{(p)}, \quad (92)$$

$$(\underline{\underline{1}} - \underline{\underline{\Theta}}^{rp})\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = -\frac{\partial\bar{\mathcal{H}}}{\partial\boldsymbol{R}} + \mathcal{E}^{(r)} + \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t} \times \mathcal{B}^{(r)}, \quad (93)$$

where we see that the effect of $\underline{\underline{\Theta}}^{rp}$ is related to a change of phase-space in the course of the diagonalization transformation. The beauty of this result is the symmetry in which effective magnetic and electric fields appear in these equations.

We are now performing a quantum average of the kinetic equations (92) and (93) with respect to a density matrix that is peaked around a certain value $\boldsymbol{P}_c$ and $\boldsymbol{R}_c$, which in effect corresponds to a Gaussian wave packet that is well beyond the limits of the Heisenberg uncertainty. Then our equations of motion (92) and (93) essentially become classical ones with all kinetic variables replaced by P_c and R_c. These equations of motion for the center of mass coordinates of a wave packet have been directly obtained by Sundaram and coworkers.[11] The circular current terms in Eqs. (81) and (82) vanish for the wave packet, which is to be expected physically, since this description reduces the electron to a point particle with coordinates $\boldsymbol{P}_c$ and $\boldsymbol{R}_c$, and which does not posses any internal motion described by the circular currents. Mathematically, after integration over the whole phase space, one can rewrite them in terms of surface integrals, which vanish for the well-localized wave packet.

C. Nonequilibrium description

Now, it is just a matter of finding the density matrix $\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t)$ in order to explicitly calculate anything within this framework, in particular, when one is interested in nonequilibrium phenomena. One way to proceed is within Keldysh formalism and to consider the kinetic equation for the lesser Greens function,[40]

$$[\Xi \overset{*}{,} \mathcal{G}^<] = 0, \quad (94)$$

which, after transformation into the rotated frame, can be evaluated to leading order correction in $\hbar$ by using Eq. (47), which by virtue of Eqs. (90) and (91) can be recast into

$$\left[\partial_t + \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t}\nabla_R + \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\nabla_P \right.$$
$$\left. + \left(\frac{\partial\bar{\mathcal{H}}}{\partial t} + \mathcal{E}^{(r)}\frac{\partial\bar{\mathcal{H}}}{\partial\boldsymbol{P}} - \mathcal{E}^{(p)}\frac{\partial\bar{\mathcal{H}}}{\partial\boldsymbol{R}} \right)\partial_E \right]\bar{\mathcal{G}}^< = 0. \quad (95)$$

When we integrate the lesser Greens function over energy, we obtain the density matrix

$$\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t) = \int \frac{\mathrm{d}E}{2\pi} \, \bar{\mathcal{G}}^<(X,\boldsymbol{\Pi}), \quad (96)$$

and, since the terms proportional to $\partial_E \bar{\mathcal{G}}^<$ vanish after integrating Eq. (95) over all energies, we are left with the following differential equation for the density matrix,

$$\left(\partial_t + \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t}\nabla_R + \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\nabla_P \right)\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t) = 0, \quad (97)$$

which is nothing but a Boltzmann equation including quantum corrections in terms of Berry curvatures. In the end, it is quite analogous to the quantum Boltzmann equation along with the gradient expansion we used Ref. 46. Of course, the major difference is the absence of the collision integral in the present formulation, however, it is possible to include it here as well, which is, however, left for future investigations. Of course, $\bar{\rho}$ can additionally include discrete quantum degrees of freedom

like the spin, which means that our rotated frame is block diagonal and so is $\bar{\rho}$.

We now show how to obtain the distribution function, when it is known at some initial time, $\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t=0)$. In general, these problems can be formulated in terms of an inhomogeneous Boltzmann equation with a source term $S(\boldsymbol{R},\boldsymbol{P},t)$,

$$\left(\partial_t + \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t}\nabla_R + \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\nabla_P\right)\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t) = S(\boldsymbol{R},\boldsymbol{P},t). \quad (98)$$

To proceed in the usual way, we define the Greens function $\mathcal{G}_\mathrm{c}$,

$$\left(\partial_t + \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t}\nabla_R + \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}\nabla_P\right)\mathcal{G}_\mathrm{c}(\boldsymbol{R},\boldsymbol{P},t,\boldsymbol{R}',\boldsymbol{P}',t')$$
$$= \delta(t-t')\delta(\boldsymbol{R}-\boldsymbol{R}')\delta(\boldsymbol{P}-\boldsymbol{P}'), \quad (99)$$

for which we find the general solution

$$\mathcal{G}_\mathrm{c}(\boldsymbol{R},\boldsymbol{P},t,\boldsymbol{R}',\boldsymbol{P}',t') = \Theta(t-t')\,\delta[\boldsymbol{R}-\boldsymbol{R}_\mathrm{c}(t)]\,\delta[\boldsymbol{P}-\boldsymbol{P}_\mathrm{c}(t)], \quad (100)$$

where Θ is the Heaviside step function, and $\boldsymbol{R}_\mathrm{c}(t)$ and $\boldsymbol{P}_\mathrm{c}(t)$ describe the classical orbits, which obey the equations of motion:

$$\dot{\boldsymbol{R}}_\mathrm{c} = \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t}, \quad \boldsymbol{R}_\mathrm{c}(t') = \boldsymbol{R}',$$

$$\dot{\boldsymbol{P}}_\mathrm{c} = \frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t}, \quad \boldsymbol{P}_\mathrm{c}(t') = \boldsymbol{P}'.$$

This solution describes ballistic trajectories given by the kinetic equations (90) and (91) and which including the effects of the Berry curvatures, while the particle being initially at phase-space coordinates $(\boldsymbol{R}',\boldsymbol{P}',t')$. For example, in the absence of any spatially dependent potential, one simply has straight lines described by the Greens function $\mathcal{G}_\mathrm{c} = \Theta(t-t_0)\,\delta(\boldsymbol{R}-\boldsymbol{V}(t-t_0))\,\delta(\boldsymbol{P}-\boldsymbol{P}')$, where $\boldsymbol{V} = \frac{\partial \bar{\mathcal{H}}_0}{\partial \boldsymbol{P}}$ is the group velocity. This is analogous to other quasiclassical equations like the Eilenberger equation, which can be also described in terms of classical trajectories.[40,47]

Now, the solution is readily given by

$$\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t) = \int_{-\infty}^{+\infty} \mathrm{d}t' \int \mathrm{d}^d R' \int \mathrm{d}^d P'$$
$$\times \mathcal{G}_\mathrm{c}(\boldsymbol{R},\boldsymbol{P},t,\boldsymbol{R}',\boldsymbol{P}',t')\,S(\boldsymbol{R}',\boldsymbol{P}',t'), \quad (101)$$

where, due to the delta functions in $\mathcal{G}_\mathrm{c}$, we get contributions only from trajectories which end at $(\boldsymbol{R},\boldsymbol{P})$ at time t. Since the trajectory is well defined by the given pair $(\boldsymbol{R},\boldsymbol{P})$ already, we can find all corresponding points $(\boldsymbol{R}',\boldsymbol{P}')$ by going back in time so we can eventually rewrite the above integral as

$$\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t) = \int_{-\infty}^{t} \mathrm{d}t'\,S[\boldsymbol{R}_\mathrm{c}(t'),\boldsymbol{P}_\mathrm{c}(t'),t'], \quad (102)$$

where now, the classical orbits are defined such that its position in phase space at time t is $\boldsymbol{R}_\mathrm{c}(t) = \boldsymbol{R}$ and $\boldsymbol{P}_\mathrm{c}(t) = \boldsymbol{P}$.

Let us now go back to the original initial value problem, which we can easily solve by using the source term $\delta(t)\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t=0)$, so that the general solution reads

$$\bar{\rho}(\boldsymbol{R},\boldsymbol{P},t) = \Theta(t)\,\bar{\rho}[\boldsymbol{R}_\mathrm{c}'(0),\boldsymbol{P}_\mathrm{c}'(0),0], \quad (103)$$

which has the desired properties for $t \geqslant 0$. This indeed corresponds to a motion in phase space as an incompressible fluid, as stated by Liouville's theorem.

D. Polarization

Another way to interpret the kinetic variables is obtained by considering the electrical polarization which we pursue according to the pioneering work of Vanderbilt and coworkers.[6] They derive the polarization in terms of adiabatic transport, where some perturbation that leads to a polarization in the crystal is adiabatically turned on. The current due to adiabatic transport is[48]

$$\boldsymbol{j} = \int \frac{\mathrm{d}^d r\,\mathrm{d}^d p}{(2\pi\hbar)^d} \mathrm{Tr}_N \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t} = -\int_\mathrm{BZ} \frac{\mathrm{d}^d p}{(2\pi\hbar)^d} \mathrm{Tr}_N \mathcal{E}^{(p)}, \quad (104)$$

where integration is only within the Brillouin zone. The electric polarization is then obtained by integrating over time,

$$\mathfrak{P}(\boldsymbol{R}) = \int \mathrm{d}t\,\boldsymbol{j}(\boldsymbol{R},t). \quad (105)$$

The definition as an adiabatic transport process is necessary in order to obtain a truly gauge invariant result.[6] In the periodic gauge, however,[3] it is nevertheless possible to construct

$$\mathfrak{P} = \int \frac{\mathrm{d}^d p}{(2\pi\hbar)^d}\,\mathcal{A}_{\boldsymbol{p}}, \quad (106)$$

but in the general case, it is required to consider the full Berry connection structure (including in particular $\mathcal{A}_t$) in order to obtain the correct result for the polarization. In the same spirit, one could define a polarization in momentum space, which can be thought of as a Doppler shift in the rotated frame or also termed anomalous velocity.

In this sense, one could loosely interpret these Berry connections as shifts that the canonical variables acquire and which depend on full phase space, albeit these shifts are not directly physically observable, only when one does an integration with respect to either the position or the momentum variable one obtains the observable electric polarization in real space or reciprocal space, respectively. This is analogous to the quasiprobability distribution $\rho(\boldsymbol{r},\boldsymbol{p})$ in the Wigner picture, which can be given only physical interpretation as probability density when integrated either over whole momentum space or real space. However, if we try to transfer this idea, we still will have a shift that depends on the frame, i.e., it is still not gauge invariant. Only for a very specific gauge, one obtains the physical polarization.

E. Bulk-boundary correspondence

It is interesting to note that the circular current term

$$-\nabla_{\boldsymbol{R}} \times \int \frac{\mathrm{d}^d P}{(2\pi\hbar)^d} \mathrm{Tr}_N \bar{\rho}\boldsymbol{\Omega}^{pp} \quad (107)$$

appearing in the current density (88) is a manifestation of the bulk-boundary correspondence.[49] In the homogeneous bulk, these circular currents compensate each other and yield zero, but at the edge of the system, there will be residual currents flowing on the surface. Due to the circular nature of these edge currents, the associated edge states are also of topological

nature. Therefore a nontrivial (bulk) value of $\boldsymbol{\Omega}^{pp}$ indicates a topological state of matter.

V. HIERARCHY OF EFFECTIVE THEORIES: INCLUSION OF THE ELECTROMAGNETIC FIELD

In many cases, one starts from an effective theory which requires dealing with gauge invariant momentum and/or position operators in the original frame. This might happen if one uses an effective theory derived from a more comprehensive theory, or more commonly, one simply wishes to include an external electromagnetic field. Qualitatively, one expects no new concepts or phenomena to emerge in effective theories at a lower level in the hierarchy, as one could also go directly from the topmost theory to the lowest in a single step.

Here, our focus lies on the combined effect of the electromagnetic field and the Berry connections emerging from the diagonalization of a band Hamiltonian. In this case, we are starting from a theory which is gauge invariant in real space, thus formulated in terms of the minimally coupled momentum $\boldsymbol{p} - q\boldsymbol{A}(x)$. Since all our derivations so far are only linear in $\hbar$, we treat the electromagnetic vector potential $q\boldsymbol{A}$ (q is the charge of the particle, so we usually have $q = -e$) on a different level than $\hbar\mathcal{A}$, thus formally taking $\hbar$ and q as independent expansion parameters. To begin with, we have the usual canonical pair $\boldsymbol{x}$ and $\boldsymbol{\pi}$ and the kinetic pair $\boldsymbol{x}$ and $\boldsymbol{\pi} + q\boldsymbol{A}(\boldsymbol{x})$, where $\boldsymbol{A} = (-\phi, \boldsymbol{A_r})$ is the usual 4-component vector potential of the electromagnetic field. In the following, we denote $\boldsymbol{B}(\boldsymbol{r},t) = \nabla_r \times \boldsymbol{A_r}$ and $\boldsymbol{E}(\boldsymbol{r},t) = -\partial_t \boldsymbol{A_r} + \nabla_r A_0$ (not to be confused with the energy E) as the external electric and magnetic fields.

The Hamiltonian is specified as $\mathcal{H}[\boldsymbol{x}, \boldsymbol{p} - q\boldsymbol{A_r}(\boldsymbol{x})]$ and is diagonalized by an appropriate unitary matrix $\mathcal{U}[\boldsymbol{x}, \boldsymbol{p} - q\boldsymbol{A_r}(\boldsymbol{x})]$, so that in the diagonalized frame, we arrive at the kinetic pair of variables

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{x} - \hbar\mathcal{A}_{\boldsymbol{\pi}}(\boldsymbol{x}, \boldsymbol{p} - q\boldsymbol{A_r}), \\ \boldsymbol{\Pi} &= \boldsymbol{\pi} + q\boldsymbol{A}(X) + \hbar\mathcal{A}_{\boldsymbol{x}}(\boldsymbol{x}, \boldsymbol{p} - q\boldsymbol{A_r}). \end{aligned} \tag{108}$$

Since derivatives now also act on $\boldsymbol{A_r}$, we have

$$\mathcal{A}_{x_k} = \mathcal{A}_{X_k} - q\big(\partial_{X_k} A_l\big)\mathcal{A}_{P_l}, \tag{109}$$

so that we can write more explicitly

$$\boldsymbol{R} = \boldsymbol{r} + \hbar\mathcal{A}_{\boldsymbol{P}}, \tag{110}$$

$$E = \epsilon - q\phi + \hbar\mathcal{A}_{\boldsymbol{R}} + \hbar q\boldsymbol{E}\cdot\mathcal{A}_{\boldsymbol{P}}, \tag{111}$$

$$\boldsymbol{P} = \boldsymbol{p} - q\boldsymbol{A_r}(\boldsymbol{x}) - \hbar\mathcal{A}_{\boldsymbol{R}} - \hbar q\boldsymbol{B}\times\mathcal{A}_{\boldsymbol{P}}. \tag{112}$$

Note that $\mathcal{A}_{\boldsymbol{X}}$, as well as $\mathcal{A}_{\boldsymbol{P}}$ still carry the full gauge invariant momentum $\boldsymbol{p} - q\boldsymbol{A_r}$ and for simplicity, all Berry connections involved are assumed to be Abelian.

The fictitious fields in the presence of an electromagnetic field can be derived by plugging these kinetic variables into the commutator relations (32) and (43) and keeping all terms up to order $O(\hbar)$ and quadratic in the fields $O(q^2)$. In order to obtain manifest gauge invariant results, we express everything in terms of kinetic variables by using $\partial_{x_j} = \partial_{X_j} - q(\partial_{X_j} A_l)\partial_{P_l}$ and $\boldsymbol{E}(X) = \boldsymbol{E}(x) + (\mathcal{A}_{\boldsymbol{P}}\partial_r)\boldsymbol{E}(x)$. Denoting primed quantities as the Berry curvatures describing the combined effect

of external electromagnetic and the diagonalization, we can summarize the results as

$$\begin{aligned} \boldsymbol{\mathcal{B}}'^{(r)} &= q\boldsymbol{B}(\boldsymbol{R}) + \boldsymbol{\mathcal{B}}^{(r)} + q[\text{Tr}(\underline{\underline{\Theta}}^{rp})\boldsymbol{B} - \boldsymbol{B}\underline{\underline{\Theta}}^{rp}] \\ &\quad + q^2(\boldsymbol{\mathcal{B}}^{(p)}\cdot\boldsymbol{B})\boldsymbol{B}, \end{aligned} \tag{113}$$

$$\boldsymbol{\mathcal{B}}'^{(p)} = \boldsymbol{\mathcal{B}}^{(p)}, \tag{114}$$

$$\begin{aligned} \boldsymbol{\mathcal{E}}'^{(r)} &= q\boldsymbol{E}(\boldsymbol{X}) + \boldsymbol{\mathcal{E}}^{(r)} + q\underline{\underline{\Theta}}^{rp}\boldsymbol{E} + q\boldsymbol{B}\times\boldsymbol{\mathcal{E}}^{(p)} \\ &\quad + q^2\boldsymbol{B}\times(\boldsymbol{E}\times\boldsymbol{\mathcal{B}}^{(p)}), \end{aligned} \tag{115}$$

$$\boldsymbol{\mathcal{E}}'^{(p)} = \boldsymbol{\mathcal{E}}^{(p)} + q\boldsymbol{E}\times\boldsymbol{\mathcal{B}}^{(p)}, \tag{116}$$

$$\underline{\underline{\Theta}}'^{rp}_{ij} = \underline{\underline{\Theta}}^{rp}_{ij} + q\big[(\boldsymbol{\mathcal{B}}^{(p)}\cdot\boldsymbol{B})\delta_{ij} - \mathcal{B}^{(p)}_i B_j\big]. \tag{117}$$

Here, unprimed quantities $\boldsymbol{\mathcal{B}}$, $\boldsymbol{\mathcal{E}}$, $\underline{\underline{\Theta}}$ correspond to the Berry curvatures of the system in the diagonalized frame in absence of the external electromagnetic field $q\boldsymbol{A}$.

It is important to note that these results have to be substituted into the equations of motion (90) and (91) and *not* (92) and (93) since the latter has been obtained by dropping terms beyond leading order corrections in Berry curvatures, whereas our results for the fictitious fields are in fact higher order in Berry connections ($q^2\hbar$ actually already corresponds to terms of third order in the expansion of the Moyal product).

The term $\boldsymbol{E} \times \boldsymbol{\mathcal{B}}^{(p)}$ of $\boldsymbol{\mathcal{E}}'^{(p)}$ is the anomalous velocity term that is, for example, responsible for the quantum Hall effect, since the Berry curvature $\boldsymbol{\mathcal{B}}^{(p)}$ becomes nontrivial when the system is in the quantum Hall state. Similarly, in systems with spin-orbit interactions, this term constitutes the intrinsic contribution to the anomalous Hall effect. The reciprocal effect thereof is described by the term $q\boldsymbol{B} \times \boldsymbol{\mathcal{E}}^{(p)}$, where a external magnetic field transforms a momentum space electric field into a real-space one.

$\underline{\underline{\Theta}}'^{rp}$ gets modified by a magnetic field term that resembles a dipole interaction between a real-space and the momentum space magnetic field. In fact, as we have seen in result (78), and as we will see explicitly at the end of this section, $\underline{\underline{\Theta}}'^{rp}$ is directly related to the magnetic dipole energy in the case of a two-level model, and thus, it is the magnetic field term in $\underline{\underline{\Theta}}'^{rp}$, which will give rise to the Zeeman energy in the Hamiltonian. Furthermore, the term quadratic in the fields can be absorbed by $\underline{\underline{\Theta}}'^{rp}$, i.e., fictitious electric field can be recast into

$$\boldsymbol{\mathcal{E}}'^{(r)} = q\boldsymbol{E} + \boldsymbol{\mathcal{E}}^{(r)} + q\underline{\underline{\Theta}}'^{rp}\boldsymbol{E} + q\boldsymbol{B}\times\boldsymbol{\mathcal{E}}^{(p)}. \tag{118}$$

From this expression, we see that the external electric field enters effectively as $q(\mathbb{1} + \underline{\underline{\Theta}}'^{rp})\boldsymbol{E}$, and the electric field acting on the system is thus renormalized. An analogous renormalization of the external magnetic field appears also in the result for $\boldsymbol{\mathcal{B}}'^{(r)}$.

The modification of the transformation rule (48) for arbitrary observables is easily found by virtue of Eq. (109),

$$\begin{aligned} \bar{\mathcal{O}}' &= \bar{\mathcal{O}} + [\mathcal{Y}'_1 - \mathcal{Y}_1, \bar{\mathcal{O}}_0] + \frac{\hbar}{2}q\boldsymbol{E}\{\mathcal{A}_{\boldsymbol{P}}, \partial_E\bar{\mathcal{O}}_0\} \\ &\quad - \epsilon_{ijk}qB_k\frac{\hbar}{2}\left\{\mathcal{A}_{P_i}, \left[\partial_{P_j} - \frac{i}{2}\mathcal{A}_{P_j}, \bar{\mathcal{O}}_0\right]\right\}, \end{aligned} \tag{119}$$

where again, primed symbols represent quantities in presence of an external magnetic field. In particular, evaluating the above

expression for the Hamiltonian, we find for the magnetization energy in the rotated frame $\bar{\mathcal{H}}_{\mathcal{M}} = \mathrm{Tr}\underline{\underline{\Omega}}'^{rp}$ [see Eq. (71)], where

$$\underline{\underline{\Omega}}'^{rp}_{ij} = \underline{\underline{\Omega}}^{rp}_{ij} + q[(\boldsymbol{\Omega}^{pp}\cdot\boldsymbol{B})\delta_{ij} - \Omega^{pp}_i B_j]. \qquad (120)$$

Notice the resemblance between this expression and Eq. (117); in fact, it turns out we obtain $\underline{\underline{\Omega}}'$ by substituting the Berry curvatures with the analog quantities of $\underline{\underline{\Omega}}$ in results (113)–(117). In two dimensions, $\underline{\underline{\Theta}}'^{rp}_{33}$ is zero,[50] and thus $\mathrm{Tr}\underline{\underline{\Theta}}'^{rp} = \mathrm{Tr}\underline{\underline{\Theta}}^{rp} + q\,\boldsymbol{B}\cdot\boldsymbol{\mathcal{B}}^{(p)}$, i.e., a factor 2 less than in the 3D case, and likewise, in 1D $\mathrm{Tr}\underline{\underline{\Theta}}'^{rp} = \mathrm{Tr}\underline{\underline{\Theta}}^{rp}$. Nevertheless, we see that the magnetic field couples in the shape of a dipole term $\boldsymbol{B}\cdot\boldsymbol{\mathcal{B}}^{(p)}$ to the momentum space magnetic field $\boldsymbol{\mathcal{B}}^{(p)}$. Since the magnetic dipole energy is already of order $\hbar$, we can equally write $\boldsymbol{B}(\boldsymbol{X})$ and $\boldsymbol{B}(\boldsymbol{x})$, since substituting canonical and kinetic variables affects only higher orders in $\hbar$.

VI. THE DIRAC EQUATION: AN ALTERNATIVE PERSPECTIVE ON DERIVING THE PAULI HAMILTONIAN

The goal of this section is illustrate our formalism, while at the same time serving as a concise summary of the framework, essentially making use of all the central results. The Dirac equation and its descendants like the Pauli Hamiltonian are the most fundamental equations in condensed matter physics, nevertheless, we present some interesting insight into the physics revealed by this rather simple equation.

We consider the Dirac Hamiltonian in the presence of a scalar potential and minimally coupled to the magnetic field,[51]

$$H = c\boldsymbol{\alpha}(\boldsymbol{p} - q\boldsymbol{A}) + mc^2\beta + V(\boldsymbol{x})\mathbb{1}_4, \qquad (121)$$

where m is the rest mass and $\boldsymbol{p}$ the momentum of the electron. As

$$\boldsymbol{\alpha} = \begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & -\mathbb{1}_2 \end{pmatrix}, \qquad (122)$$

H acts on the 4-dimensional Dirac Spinor and thus can be considered as a 4-band model.

The Foldy-Wouthuysen transformation,[39] which brings the Dirac equation into diagonal form, is described by the unitary matrix

$$\mathcal{U}_0 = \frac{(E_P + mc^2)\mathbb{1}_4 + c\beta\boldsymbol{\alpha}\boldsymbol{P}}{\sqrt{E_P(E_P + mc^2)}}, \qquad (123)$$

where $E_P = c\sqrt{P^2 + m^2c^2}$ is the relativistic energy of the electron with gauge invariant momentum P. Performing, as outlined before [cf. Eq. (22)], the diagonalization with respect to $\mathcal{U}_0$, we arrive at

$$\begin{aligned} \bar{\mathcal{H}} &= \mathcal{U} * H * \mathcal{U}^\dagger \\ &= E_P\sigma_0\tau_z + \big(V(\boldsymbol{r})\sigma_0 + \hbar(\partial_{\boldsymbol{r}} V)\mathcal{A}^{(\mathrm{d})}_{\boldsymbol{P}} + qE_P\boldsymbol{B}\cdot\boldsymbol{\mathcal{B}}^{(p)}\big)\tau_0. \end{aligned} \qquad (124)$$

The 2×2-matrices τ_i denote the Pauli matrices in electron-positron space, while σ_i describes the usual spin degree of freedom. The magnetic dipole term is given according to Eq. (120) in terms of an interaction term between magnetic field $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ and the fictitious momentum space magnetic field $\boldsymbol{\mathcal{B}}^{(p)}$, to be specified below. We note that in this case, all the correction terms are equal for both electron and positron

bands, albeit the term positron becomes only meaningful when all negative energy states are completely occupied.

The full matrix structure of the Berry connection, split into diagonal and off-diagonal parts, i.e., $\mathcal{A}_{\boldsymbol{P}} = \mathcal{A}^{(\mathrm{d})}_{\boldsymbol{P}}\tau_0 + \mathcal{A}^{(\mathrm{o})}_{\boldsymbol{P}}\tau_y$, reads

$$\mathcal{A}^{(\mathrm{d})}_{\boldsymbol{P}} = \frac{c^2\boldsymbol{P}\times\boldsymbol{\sigma}}{2E_P(E_P + mc^2)} \approx \frac{\lambda_c}{4\hbar}\frac{\boldsymbol{P}\times\boldsymbol{\sigma}}{mc}, \qquad (125)$$

$$\mathcal{A}^{(\mathrm{o})}_{\boldsymbol{P}} = \frac{c\boldsymbol{\sigma}}{2E_P} - \frac{\boldsymbol{P}(\boldsymbol{P}\boldsymbol{\sigma})\,c^3}{2E_P^2(E_P + mc^2)} \approx \frac{\lambda_c}{2\hbar}\boldsymbol{\sigma}, \qquad (126)$$

where in the last step, we are dropping terms of order $O(p/mc)^2$.

The kinetic variables for both positive and negative energy states read explicitly

$$\begin{aligned} \boldsymbol{P} &= \boldsymbol{p} - q\boldsymbol{A}(\boldsymbol{r}) - q\hbar\boldsymbol{B}\times\mathcal{A}^{(\mathrm{d})}_{\boldsymbol{P}}, \\ \boldsymbol{R} &= \boldsymbol{r} + \hbar\mathcal{A}^{(\mathrm{d})}_{\boldsymbol{P}}, \end{aligned}$$

so that the Berry curvature (34) is also equal for both bands, and yields

$$\boldsymbol{\mathcal{B}}^{(p)} = -\frac{\hbar mc^4}{2E_P^3}\boldsymbol{\sigma} - \frac{\boldsymbol{P}(\boldsymbol{P}\cdot\boldsymbol{\sigma})\,\hbar c^4}{2E_P^3(E_P + mc^2)} \approx -\frac{\lambda_c^2}{2\hbar}\boldsymbol{\sigma}, \qquad (127)$$

where again, in the last step we took the nonrelativistic limit. Of course, the general expressions are still valid for arbitrary velocities, for example, in the opposite, the ultra-relativistic limit, we find

$$\boldsymbol{\mathcal{B}}^{(p)} \overset{p\gg mc}{\to} -\frac{\boldsymbol{P}(\boldsymbol{P}\cdot\boldsymbol{\sigma})\hbar}{2P^4}.$$

Using the results (113)–(117), we immediately find $\boldsymbol{\mathcal{E}}^{(r)} = 0 = \boldsymbol{\mathcal{E}}^{(p)}$ and

$$\begin{aligned} \boldsymbol{\mathcal{B}}^{(r)} &= q\boldsymbol{B}(\boldsymbol{R}) + q^2(\boldsymbol{\mathcal{B}}^{(p)}\cdot\boldsymbol{B})\boldsymbol{B}, \\ \underline{\underline{\Theta}}^{rp}_{ij} &= q\big[(\boldsymbol{\mathcal{B}}^{(p)}\cdot\boldsymbol{B})\delta_{ij} - \mathcal{B}^{(p)}_i B_j\big], \end{aligned} \qquad (128)$$

where here, we used $\boldsymbol{E}(\boldsymbol{R}) = 0$, since we are already using the scalar potential $V(\boldsymbol{R})$.

In terms of kinetic variables, the Hamiltonian $\bar{\mathcal{H}}$ can now be rewritten (let us restrict ourselves to the positive energy branch, whose excitations correspond to electrons, so that $q = -e$), neglecting terms of order $\hbar^2$,

$$\bar{\mathcal{H}} = E_P + V(\boldsymbol{R}) + qE_P\boldsymbol{B}(\boldsymbol{R})\cdot\boldsymbol{\mathcal{B}}^{(p)}, \qquad (129)$$

so that in view of this, the spin-orbit interaction that appears in the rotated frame can be reinterpreted as being a result of the shift the kinetic position operator attains. However, we should keep in mind that this shift is gauge dependent and becomes physically meaningful only when integrated over the whole momentum space (which then is equivalent to the polarization) or in the form of the Berry curvature, viz the momentum space magnetic field appearing in the kinetic equations. In the nonrelativistic limit, the magnetic dipole energy is just the usual Zeeman term [using Eqs. (71) and (72)]

$$\bar{\mathcal{H}}_{\mathcal{M}} = -\boldsymbol{\mu}_s\cdot\boldsymbol{B}, \qquad (130)$$

where we introduced the magnetic moment of the electron spin, $\boldsymbol{\mu}_s = -g_s\mu_{\mathrm{B}}\boldsymbol{\sigma}/2 = -\mu_{\mathrm{B}}\boldsymbol{\sigma}$, and we assume a g factor of 2 within the validity of the Dirac theory without quantum

corrections from the radiative field, and $\mu_{\rm B} = \frac{e\hbar}{2m}$ is the Bohr magneton.

Using results (90) and (91) together with Eqs. (128) and (129), we obtain the operator equations of motion of a relativistic electron, which moves in an electromagnetic field that is smooth on the scale of the Compton wavelength λ_c. We do not explicitly write the expression here, however, our result is consistent with the work of Bliokh.[52]

For the sake of our discussion, we will nevertheless specify the equations of motion in absence of a magnetic field,

$$\frac{{\rm d}\boldsymbol{R}}{{\rm d}t} = \frac{\partial \bar{\mathcal{H}}}{\partial \boldsymbol{P}} - \frac{{\rm d}\boldsymbol{P}}{{\rm d}t} \times \boldsymbol{\mathcal{B}}^{(p)} = \frac{\partial \bar{\mathcal{H}}}{\partial \boldsymbol{P}} + \frac{\lambda_c^2}{2\hbar}\frac{{\rm d}\boldsymbol{P}}{{\rm d}t} \times \boldsymbol{\sigma}, \quad (131)$$

$$\frac{{\rm d}\boldsymbol{P}}{{\rm d}t} = -\frac{\partial \bar{\mathcal{H}}}{\partial \boldsymbol{R}} = -e\boldsymbol{E}(\boldsymbol{R}), \quad (132)$$

whose anomalous velocity term is a factor 2 larger than what one would obtain from the Hamiltonian (124), naively taking the canonical variable $\boldsymbol{r}$ to be the physical position operator. Essentially, the same conclusion has been reached in Refs. 3 and 22 for the case of the relativistic Dirac equation.

The particle density is, according to results (87) and (128), given by

$$n(\boldsymbol{R},t) = \int \frac{{\rm d}^3 P}{(2\pi\hbar)^3}\, {\rm Tr}_4 \bar{\rho}(1 - 2q\boldsymbol{B}\cdot\boldsymbol{\mathcal{B}}^{(p)}), \quad (133)$$

and let us take the vacuum state with all negative energies filled, so that $\bar{\rho} = \Theta(-\tau_z E_P)$ and the first term gives the vacuum charge. However, from the correction factor proportional to the magnetic field one might conclude that a sufficiently strong magnetic field leads to a charge accumulation with respect to the vacuum state. This kind of charge accumulation is also present in topological insulators.[33] Here, however, factors like $D = 1 - {\rm Tr}\underline{\underline{\Theta}}^{rp} = 1 - \frac{2\mathcal{H}_M}{mc^2}$ are relevant only for enormous magnetic fields of $\approx 10^{10}T$, so we can safely drop all corrections due to $\underline{\underline{\Theta}}^{rp}$.

It is also insightful to study the current density in the presence of the electromagnetic field, because it gives a direct meaning to the divergence-free current terms in Eq. (88). For the simplicity of our discussion, we restrict ourselves to the nonrelativistic limit, where $\boldsymbol{\mathcal{B}}^{(p)} = -\frac{\lambda_c^2}{2\hbar}\boldsymbol{\sigma} = \frac{\boldsymbol{\mu}_s}{emc^2}$ and $E_P = mc^2 + \boldsymbol{P}^2/2m$, and furthermore, we treat external fields $\boldsymbol{E}$ and $\boldsymbol{B}$ only to linear order. Since we want to study excitations of positive energy, we ignore the completely filled lower band.

Then, our charge density is

$$n(\boldsymbol{R},t) = -e \int \frac{{\rm d}^d P}{(2\pi\hbar)^d}\, \bar{\rho},$$

so that with the help of Eqs. (88) and (90) as well as the definition of the spin magnetic moment [c.f. Eq. (130)] the charge current density becomes

$$\boldsymbol{j}(\boldsymbol{R},t) = -e \int \frac{{\rm d}^d P}{(2\pi\hbar)^d}\, \frac{\boldsymbol{P}}{m}\bar{\rho} - \frac{e}{mc^2}n(\boldsymbol{R},t)\,\boldsymbol{E}\times\boldsymbol{\mu}_s$$
$$- \boldsymbol{\mu}_s \times \nabla_{\boldsymbol{R}} n(\boldsymbol{R},t), \quad (134)$$

where the surface term vanishes since we assume our electron momentum to be peaked around some mean value $\boldsymbol{P}_c$. The density $\bar{\rho}$ can be explicitly determined from Eq. (97), here, however, we only integrate this equation over all momenta

and arrive at the continuity equation

$$\partial_t n(\boldsymbol{R},t) + \nabla_{\boldsymbol{R}}\cdot\boldsymbol{j}(\boldsymbol{R},t) = 0. \quad (135)$$

The first term in Eq. (134) is the usual definition of a nonrelativistic current, and the second term is the anomalous velocity contribution due to the electric field. Similar to the electron wave packet considered in Eq. (131), using the nonrelativistic Pauli Hamiltonian without the distinction between canonical and kinetic variables would yield an anomalous velocity term that is just half of the correct value. However, the last term seems unusual at first as it constitutes a persistent current, which itself gives rise to a magnetic moment,

$$\frac{1}{2}\int {\rm d}^3 R\, \boldsymbol{R}\times\boldsymbol{j}_{\rm per} = -\frac{1}{2}\int {\rm d}^3 R\, \boldsymbol{R}\times(\boldsymbol{\mu}_s\times\nabla_{\boldsymbol{R}} n)$$
$$= \boldsymbol{\mu}_s \int {\rm d}^3 R\, n(\boldsymbol{R},t) = \mathcal{N}\boldsymbol{\mu}_s, \quad (136)$$

where $\mathcal{N} = 1$ is the number of electrons. This result suggests that the notion of electron spin and this internal persistent current are just different interpretations of the same effect. For example, a magnetic field couples to this circular current via the magnetic moment it generates, and gives rise to the Zeeman energy (130). This comes close to the original suggestion of an internal rotation of the electron by Uhlenbeck and Goudsmit,[53] yet unlike their idea, it is not the motion of a solid object, instead it is a genuine quantum phenomenon, and thus does not suffer the same deficiencies concerning rotation velocities of the electron that would have to be faster than the speed of light. Recently, an analogous conclusion has been reached in the approach by Chuu *et al.* using the picture of a self-rotating wave packet.[13]

In the same work, it has also been noted that the spin-orbit term is due to the difference between canonical position and physical position given by the wave-packet center of mass, which is consistent to what we found in Eq. (129). As a consequence, the spin-orbit term in the Pauli Hamiltonian is *not* gauge invariant, it gives the correct energy spectrum, but not the correct equations of motion. That is why it has not been realized for a long time, as the Pauli Hamiltonian has been mainly put to test with electronic spectra in atoms or solid matter. To explicitly show this point, we go into a different rotated frame so that the Berry connection is changed to $\mathcal{A}_{\boldsymbol{p}} \to \mathcal{A}_{\boldsymbol{p}} + \hbar\nabla_{\boldsymbol{p}}\chi(\boldsymbol{p})$, and we get an additional term in the Pauli Hamiltonian $\hbar(\nabla_{\boldsymbol{p}}\chi)(\partial_r V)$. Let us assume for the moment that we are working in momentum space and in the operator representation of quantum mechanics, where the Hamiltonian formally looks the same. Now, performing a local gauge transformation in momentum space by adding the phase factor $e^{i\chi(\boldsymbol{p})}$ to the wave function, we obtain $e^{-i\chi(\boldsymbol{p})}V(\boldsymbol{r})e^{i\chi(\boldsymbol{p})} = V(\boldsymbol{r} - \hbar\nabla_{\boldsymbol{p}}\chi) = V(\boldsymbol{r}) - \hbar(\nabla_{\boldsymbol{p}}\chi)(\partial_r V) + O(\hbar^2)$, which exactly cancels the additional term, and we are left with the original Hamiltonian. Usually, one enforces gauge-invariance in real space to obtain the electromagnetic field. This is a an example of gauge invariance in reciprocal space which leads to the phenomena of spin-orbit interaction, or the anomalous velocity.

In the above analysis, we needed to drop terms of order $\hbar^2$, though here, $\hbar$ is just a formal expansion parameter, the real relevant scale being the Compton wavelength $\lambda_c \equiv \frac{\hbar}{mc}$. For

example, here, we find that the order of $\hbar \mathcal{A}^{(\mathrm{d})}$ is $\frac{\hbar p}{4m^2c^2} = \lambda_c \frac{v}{c}$, which is to be compared with our position coordinate, where one realizes that corrections are indeed small unless the potential $V(\boldsymbol{r})$ varies strongly on the scale of the Compton wavelength λ_c. Consequently, the Hamiltonian (124) still describes the correct quantum behavior for sufficiently smooth potentials and thus, the term semiclassical expansion seems misleading. However, for the case of the Coulomb δ-like potential of the nucleus of an atom it is obviously no longer true, and this leads to corrections like the Darwin term, which is essential to understand atomic spectra.[51]

Sometimes, one interprets the terms appearing in the Pauli Hamiltonian as being a result of what one termed *Zitterbewegung*, and which one envisages as helical motion of the electron. Then the spin-orbit interaction and its physical consequences such as the anomalous velocity can be understood as a result of this motion. Obviously, we do not observe such an oscillatory motion here in the effective theory, when the excitation is confined to a single band. Essentially, due to the Wigner transformation, this rapid oscillation has been transformed into energy space, and instead, the effects like anomalous velocity now appear in form of interband corrections. We remark that in the case of massless Dirac fermions and if one considers excitations in the vicinity of the Dirac point, both positive and negative energy states have to be taken into account simultaneously so that excitations are not confined to positive or negative energies alone. In this scenario, as for example in graphene, one obtains those features termed *Zitterbewegung*.[54]

VII. CONCLUSIONS AND OUTLOOK

In this work, we first studied a general Hamiltonian that contains an additional matrix structure describing different bands, and the goal was to bring this Hamiltonian into a band-diagonal form, which has been achieved by performing a rotation in band space. The diagonal representation is very practical when one wants to study the low-energy response of the system because then, only one or few degenerate bands are relevant. Since in the very general case this diagonalization involves the pair of noncommuting position $\boldsymbol{r}$ and momentum $\boldsymbol{p}$ operators, we performed the diagonalization perturbatively in $\hbar$ by using the Wigner representation. We investigated the Hamiltonian and physical observables in the rotated frame and how Berry curvatures emerge naturally during this diagonalization procedure. Essentially, Berry connections describe corrections to the canonical variables like position and momentum due to interband scattering. This led to the distinction between canonical and kinetic variables, where the kinetic variables in the rotated frame are strongly linked to the canonical variables of the original theory. The canonical ones are, on the other side, responsible for the proper quantum structure due to their canonical commutation relations, which defines the quantization. Therefore both the canonical and kinetic variables are an important part of our effective description, and only then will it be consistent. We also established a link between gauge invariance in momentum space and the spin-orbit interaction, and the Hamiltonian expressed in terms of canonical variables is enough if one merely wants to know the energy spectrum of the system.

Having in mind a gauge-invariant description, we expressed the Hamiltonian and observables in terms of kinetic variables, which naturally leads to the appearance of Berry curvatures that describe various effects intrinsic to the band structure. In addition to Berry curvatures, we identified further gauge invariant objects, which are related to circular, or persistent currents, which itself interact with magnetic fields give for example rise to energy corrections, like the Zeeman interaction. For the Dirac equation, the electron is naturally delocalized in space and we found the appearance of an internal motion in the form of circular persistent currents giving rise to a magnetic dipole moment, and which exactly corresponds to the spin magnetic moment of the electron.

Using these kinetic variables, we formulated a Boltzmann transport equation that incorporates intrinsic effects expressed in terms of Berry curvatures. To this end, we consistently defined various quasiprobability densities that are connected by a conservation law: Liouville's theorem in phase space. These can be used to obtain physically meaningful densities and currents within the effective description. Furthermore, it is rather straightforward to include impurity scattering in the same manner as in Ref. 46, which is subject of future work. The results derived in this work provide a good starting point to transform the collision integral into the diagonalized frame, automatically incorporating the effect of an external electromagnetic field.

A related work by Wong and Tserkovnyak treated a general two-band system by studying a quantum kinetic equation approach in the rotated frame.[17] However, the rotation diagonalizes the Hamiltonian only to zeroth order in a gradient expansion, so that when studying gradient corrections the diagonalization is no longer exact. It is therefore no longer sufficient to consider only the diagonal part of the kinetic equation. To this end, it is necessary to take into account the off-diagonal elements, which is in contrast to our formulation where already the rotation transformation includes gradient corrections.

Despite performing a formal expansion in $\hbar$, similar to treating the semiclassical limit, we never abandon the quantum description. One major advantage compared to other semiclassical approaches is that there is no need for a re-quantization (see Appendix B for some details). In fact, our real expansion parameter might be a different one, like the Compton wave-length as we have seen in the case of the relativistic Dirac equation. Also, our approach is systematic in the sense that we can go to arbitrary order in $\hbar$ or interband coupling. In this respect, it is also very interesting to look at terms second order in $\hbar$, which yield important contributions, for example, the Darwin term in the case of the low-energy limit of the relativistic Dirac equation. Furthermore, new physical phenomena emerge at $O(\hbar^2)$, like the magnetoelectric coupling in insulators which received new attention recently, also due to the discovery of topological insulators.[55,56]

Then, one could study a variety of systems that involve both spin-orbit interaction (SOI) and an inhomogeneous and time-dependent magnetization. This is interesting for spin-orbit coupled semiconductors where one has different types such as the Rashba SOI, Dresselhaus SOI, or in the case of strong SOI in III-V ferromagnetic semiconductors, one can utilize the Luttinger Hamiltonian for hole transport. In addition, one can

have nonmagnetic and magnetic impurity scattering that adds additional complexity to the system and is naturally studied in terms of a collision integral within the Boltzmann approach. Thus the formalism developed in this work seems practical to attack these kinds of problems in order to study various (magneto-, spin-, or thermo-)transport and dynamical properties. Finally, it would be interesting to include many particle effects like electron-electron interactions in the investigation.

APPENDIX A: WIGNER REPRESENTATION

In the operator representation of quantum mechanics, we can write the canonical commutation relation (quantities with hats indicate operators in the original picture),

$$[\hat{\boldsymbol{\pi}}_\mu, \hat{\boldsymbol{x}}^\nu] = i\hbar \delta_\mu^{\ \nu}. \tag{A1}$$

In 4-component notation, the Wigner transformation explicitly reads

$$\int d^4 z\, e^{iz\pi/\hbar} A\left(\boldsymbol{x} + \frac{z}{2}, \boldsymbol{x} - \frac{z}{2}\right) = A(\boldsymbol{x}, \boldsymbol{\pi}), \tag{A2}$$

and the Moyal product obeys the axiom of associativity,

$$A * B * C = A * (B * C) = (A * B) * C, \tag{A3}$$

and also

$$(A * B)^\dagger = B^\dagger * A^\dagger, \tag{A4}$$

however, the Moyal product is noncommutative, i.e., $A * B \neq B * A$. In fact, the latter property encodes the noncommutativity of operators in quantum theory, which one can explicitly see by considering the commutator of two general observables dependent on $\boldsymbol{x}$ and $\boldsymbol{\pi}$ and which are band diagonal,

$$[A(\boldsymbol{x}, \boldsymbol{\pi})\, \overset{*}{,}\, B(\boldsymbol{x}, \boldsymbol{\pi})] = 2i\, A(\boldsymbol{x}, \boldsymbol{\pi})\, \sin(\hbar\Lambda/2)\, B(\boldsymbol{x}, \boldsymbol{\pi})$$

$$= i\hbar\, \{A, B\}_p + O(\hbar^3). \tag{A5}$$

It becomes essentially an extended Poisson bracket in the semiclassical limit, which reduces to the normal Poisson bracket when neither A nor B explicitly depend on ϵ, which is usually the case.

The Wigner representation has the very practical feature of automatically symmetrizing operators, when one transforms back into the operator formulation, for example, using the back transformation of Eq. (A2),

$$A\left(\boldsymbol{x} + \frac{z}{2}, \boldsymbol{x} - \frac{z}{2}\right) = \int \frac{d^4\pi}{(2\pi\hbar)^4}\, e^{-iz\pi/\hbar} A(\boldsymbol{x}, \boldsymbol{\pi}), \tag{A6}$$

one obtains

$$xp \rightarrow \tfrac{1}{2}\, (\hat{x}\hat{p} + \hat{p}\hat{x}), \tag{A7}$$

so what is usually required to be put in by hand in usual quantum mechanics in order to obtain Hermitian observables, comes out automatically. Another, more direct way to see this is by using identities of the form

$$\tfrac{1}{2}(x * p + p * x) = xp. \tag{A8}$$

The special operator ordering obtained in this manner is the so-called Wigner-Weyl ordering.

APPENDIX B: SOME NOTES ON REQUANTIZATION

In this section, we address the question of the quantum theoretical aspects of our semiclassical analysis. In particular, in literature, there is the question of how to properly quantize the theory in the Lagrangian formalism, since one needs to identify the canonical pair of variables, which is the starting point of the quantization. As mentioned previously, using wave packet analysis,[3] one obtains the same set of equations of motion for the center of mass coordinates of the wave packet as in Eqs. (92) and (93). Then one can find the Lagrangian, but one does not know anything about canonical variables, and in the general case, it is not always easy or possible to find the pair of canonical variables.[3]

However, in our framework, we did not encounter such problems, since firstly, we are working in the Hamilton formalism, where one knows the canonical variables and secondly, our theory is quantized at any time, albeit neglecting terms of order $O(\hbar^2)$. This quantization is encoded in the structure of the Moyal product, or more explicitly in the special structure of the commutator in Eq. (47). Furthermore, without major effort, we can go back to the operator representation of quantum mechanics by undoing the Wigner transformation.

Since many of our expressions [for example, Eqs. (19), (48), (29), etc.] are straightforwardly transformed back into the operator formulation of quantum mechanics, we formally obtain the same expression, except mixtures of $\boldsymbol{r}$ and $\boldsymbol{p}$ appear properly symmetrized (see Appendix A). In fact, we retain the full quantum theory, which gives correct results at least to order $\hbar$. One just has pay attention that the Wigner transformation is *only* with respect to the canonical pair of variables so kinetic variables have to be replaced accordingly.

---

*christian.wickles@uni-konstanz.de

†wolfgang.belzig@uni-konstanz.de

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

[2] G. P. Mikitik and Y. V. Sharlai, Phys. Rev. Lett. **82**, 2147 (1999).

[3] D. Xiao, M.-C. Chang, and Q. Niu, Rev. Mod. Phys. **82**, 1959 (2010).

[4] J. N. Fuchs, F. Piéchon, M. O. Goerbig, and G. Montambaux, Eur. Phys. J. B **77**, 351 (2010).

[5] K. M. D. Hals, A. K. Nguyen, X. Waintal, and A. Brataas, Phys. Rev. Lett. **105**, 207204 (2010).

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

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

[8] N. A. Sinitsyn, J. Phys.: Condens. Matter **20**, 023201 (2008).

[9] N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, and N. P. Ong, Rev. Mod. Phys. **82**, 1539 (2010).

[10] M. N. Leuenberger and D. Loss, Phys. Rev. B **63**, 054414 (2001).

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

[12]D. Culcer, Y. Yao, and Q. Niu, Phys. Rev. B **72**, 085110 (2005).

[13]C.-P. Chuu, M.-C. Chang, and Q. Niu, Solid State Commun. **150**, 533 (2010).

[14]P. Gosselin, A. Bérard, and H. Mohrbach, Eur. Phys. J. B **58**, 137 (2007).

[15]P. Gosselin, J. Hanssen, and H. Mohrbach, Phys. Rev. D **77**, 085008 (2008).

[16]P. Gosselin, H. Boumrar, and H. Mohrbach, Europhys. Lett. **84**, 50002 (2008).

[17]C. H. Wong and Y. Tserkovnyak, Phys. Rev. B **84**, 115209 (2011).

[18]D. Culcer and Q. Niu, Phys. Rev. B **74**, 035209 (2006).

[19]C. Gorini, P. Schwab, R. Raimondi, and A. L. Shelankov, Phys. Rev. B **82**, 195316 (2010).

[20]R. Winkler, *Spin-orbit Coupling Effects in Two-dimensional Electron and Hole Systems* (Springer-Verlag, Berlin, Heidelberg, 2003).

[21]E. Wigner, Phys. Rev. **40**, 749 (1932).

[22]E. I. Blount, in *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic Press, New York, 1962), Vol. 13, pp. 305–373.

[23]R. G. Littlejohn and W. G. Flynn, Phys. Rev. A **44**, 5239 (1991).

[24]R. G. Littlejohn and W. G. Flynn, Phys. Rev. A **45**, 7697 (1992).

[25]G. Tatara and H. Kohno, Phys. Rev. Lett. **92**, 086601 (2004).

[26]G. Tatara, H. Kohno, and J. Shibata, Phys. Rep. **468**, 213 (2008).

[27]C. H. Wong and Y. Tserkovnyak, Phys. Rev. B **80**, 184411 (2009).

[28]G. E. W. Bauer, E. Saitoh, and B. J. van Wees, Nat. Mater. **11**, 391 (2012).

[29]D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008).

[30]X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).

[31]X.-L. Qi, T. L. Hughes, and S.-C. Zhang, Nat. Phys. **4**, 273 (2008).

[32]T. Yokoyama, J. Zang, and N. Nagaosa, Phys. Rev. B **81**, 241410 (2010).

[33]K. Nomura and N. Nagaosa, Phys. Rev. B **82**, 161401 (2010).

[34]C. Wickles and W. Belzig, Phys. Rev. B **86**, 035151 (2012).

[35]Y. Tserkovnyak and D. Loss, Phys. Rev. Lett. **108**, 187201 (2012).

[36]M. Matsuo, J. Ieda, E. Saitoh, and S. Maekawa, Phys. Rev. Lett. **106**, 076601 (2011).

[37]M. Matsuo, J. Ieda, E. Saitoh, and S. Maekawa, Phys. Rev. B **84**, 104410 (2011).

[38]Y. Tserkovnyak and C. H. Wong, Phys. Rev. B **79**, 014402 (2009).

[39]L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).

[40]J. Rammer and H. Smith, Rev. Mod. Phys. **58**, 323 (1986).

[41]J. E. Moyal, Math. Proc. Cambridge Philos. Soc. **45**, 99 (1949).

[42]S. E. Barnes and S. Maekawa, Phys. Rev. Lett. **98**, 246601 (2007).

[43]R. Shindou and L. Balents, Phys. Rev. B **77**, 035110 (2008).

[44]D. Xiao, J. Shi, and Q. Niu, Phys. Rev. Lett. **95**, 137204 (2005).

[45]Since $\mathcal{O}^{(d)}$ effectively projects the Berry connection matrices $\mathcal{A}^{(d)}$, so that we have $[\bar{\mathcal{S}}^{(d)}, \mathcal{A}^{(d)}] = 0$, the Taylor expansion is mathematically well defined. In the case of Abelian gauge fields, the situation is trivial, while in the non-Abelian case, we need to make the assumption not to probe the internal, degenerate structure and accordingly, $\bar{\mathcal{S}}_0$ should be diagonal in the corresponding subspace in order to describe an observable of this kind.

[46]C. Wickles and W. Belzig, Phys. Rev. B **80**, 104435 (2009).

[47]A. Shelankov, J. Low Temp. Phys. **60**, 29 (1985).

[48]D. J. Thouless, Phys. Rev. B **27**, 6083 (1983).

[49]A. M. Essin and V. Gurarie, Phys. Rev. B **84**, 125132 (2011).

[50]However, even in lower-dimensional system, we are still using all three dimensions in our vector calculus and according to (117), $\Theta_{33}^{\prime rp} \neq 0$, so we have to pay attention to these subtleties for dimensions $\leqslant 2$.

[51]F. Schwabl, *Advanced Quantum Mechanics*, edited by R. Hilton and A. Lahee (Springer-Verlag, Berlin, Heidelberg, 2008).

[52]K. Y. Bliokh, Europhys. Lett. **72**, 7 (2005).

[53]G. E. Uhlenbeck and S. Goudsmit, Nature (London) **117**, 264 (1926).

[54]M. Katsnelson, Eur. Phys. J. B **51**, 157 (2006).

[55]A. M. Essin, A. M. Turner, J. E. Moore, and D. Vanderbilt, Phys. Rev. B **81**, 205104 (2010).

[56]S. Coh, D. Vanderbilt, A. Malashevich, and I. Souza, Phys. Rev. B **83**, 085108 (2011).