Berry phase, hyperorbits, and the Hofstadter spectrum

Ming-Che Chang and Qian Niu
Department of Physics, University of Texas at Austin, Austin, TX 78712

(March 16, 2004)

Abstract

We develop a semiclassical theory for the dynamics of electrons in a magnetic Bloch band, where the Berry phase plays an important role. This theory, together with the Boltzmann equation, provides a framework for studying transport problems in high magnetic fields. We also derive an Onsager-like formula for the quantization of cyclotron orbits, and we find a connection between the number of orbits and Hall conductivity. This connection is employed to explain the clustering structure of the Hofstadter spectrum. The advantage of this theory is its generality and conceptual simplicity.

PACS numbers: 72.10.–d 72.15.Gd 73.20.Dx

The theory of semiclassical dynamics of Bloch electrons in a weak electromagnetic field plays a fundamental role in our understanding of electronic spectral and transport properties in metals and semiconductors. The basic ingredients of this theory are the following pair of equations:

$$\dot{\mathbf{r}} = \frac{\partial \mathcal{E}_n(\mathbf{k})}{\hbar \partial \mathbf{k}}, \quad \hbar \dot{\mathbf{k}} = -e\mathbf{E} - e\dot{\mathbf{r}} \times \mathbf{B}, \tag{1}$$

where $\mathcal{E}_n(\mathbf{k})$ is the energy for the n-th band, and $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields [1]. The validity of these relations depends on the absence of interband tunneling, therefore (1) holds if the external fields are sufficiently weak.

A natural question is how should these semiclassical equations be modified for a magnetic Bloch band (MBB). Such a band is obtained when an electron is subject simultaneously to a periodic potential and a magnetic field (not necessarily weak), such that the magnetic flux per unit cell of the periodic potential (plaquette) is a rational multiple of the flux quantum h/e. For example, in a tight binding model Hofstadter showed that a Bloch band is broken into q subbands if the rational number is p/q [2]. At the opposite limit, when the magnetic field is much stronger than the periodic potential, a Landau level is broadened and split into p subbands [3]. These subbands at both limits are manifestations of the magnetic Bloch bands.

It is difficult to resolve these MBBs in a naturally occuring solid. For example, with a lattice constant $a = 5$ Å and a magnetic field $B = 3$ Tesla, the value of p/q is of order 10^{-2}. However, p/q can be a significant fraction of unity if we are using an artificial lattice with a much larger period. This is realized, for example, by imposing an optical interference pattern on top of a heterostructure to create a grid potential on the two dimensional electron gas beneath. Recent experiments have demonstrated that disorder can be reduced to such a degree that the effect of magnetic bands emerges in the transport properties of the electron gas [4].

Our goal is to derive the counterpart equations to Eq. (1) for a MBB, and use them to explore the dynamics of electrons under a weak electromagnetic perturbation. To simplify the discussion, assume the electrons are confined in a two-dimensional periodic potential, with a magnetic field perpendicular to it. Because the vector potential for a constant magnetic field is not periodic, the Hamiltonian H does not commute with usual translation operators. However, we can define *magnetic* translation operators $\tilde{T}(\mathbf{R})$ that commute with H [5]. Analogous to the Bloch states, we require the eigenstates of H to satisfy the relation $\tilde{T}(\mathbf{R})\Psi_n(\mathbf{k}) = \exp(i\mathbf{k}\cdot\mathbf{R})\Psi_n(\mathbf{k})$ [6]. This is not well-defined in general, because the magnetic translation operators for different displacements, $\mathbf{R}_1$ and $\mathbf{R}_2$, do not commute unless there is an integer number of flux quanta in the area $|\mathbf{R}_1 \times \mathbf{R}_2|$. Therefore, if $\phi = p/q$, we have to choose a unit cell consisting of q plaquettes. Correspondingly, the area of the magnetic Brillouin zone is reduced by a factor of q. Furthermore, because of the magnetic translation symmetry, the energy spectrum is exactly q-fold degenerate.

Electric perturbation and transport — If we write a magnetic Bloch state in the form $\Psi_n(\mathbf{k}_0) = \exp(i\mathbf{k}_0 \cdot \mathbf{r})u_n(\mathbf{k}_0)$, then u_n is modified by a weak and homogeneous electric field into

$$u_n(\mathbf{k}) - i\hbar \sum_{n'\neq n} \frac{u_{n'}(\mathbf{k})\langle u_{n'}(\mathbf{k})|\dot{u}_n(\mathbf{k})\rangle}{\mathcal{E}_n(\mathbf{k}) - \mathcal{E}_{n'}(\mathbf{k})}, \tag{2}$$

where $\mathbf{k} = \mathbf{k}(t) \equiv \mathbf{k}_0 - e\mathbf{E}t/\hbar$, and we have used a time dependent vector potential for the electric field. The above result is obtained by adiabatic perturbation theory, which is valid for weak fields. The average velocity in such a state can be easily evaluated as:

$$\dot{\mathbf{r}} = \frac{\partial \mathcal{E}_n(\mathbf{k})}{\hbar \partial \mathbf{k}} - \dot{\mathbf{k}} \times \hat{z}\ \Omega_n(\mathbf{k}), \quad \hbar\dot{\mathbf{k}} = -e\mathbf{E}, \tag{3}$$

where $\hat{z}$ is the unit vector along the direction of the magnetic field. The second term in the expression for $\dot{\mathbf{r}}$ comes from the first order non-adiabatic correction in the wave function, with [7]

$$\Omega_n(\mathbf{k}) = i\left(\langle\frac{\partial u_n}{\partial k_1}|\frac{\partial u_n}{\partial k_2}\rangle - \langle\frac{\partial u_n}{\partial k_2}|\frac{\partial u_n}{\partial k_1}\rangle\right). \tag{4}$$

The equations in Eq. (3) are the new set of semiclassical equations. Notice that the usual Lorentz force term for $\dot{\mathbf{k}}$ is absent because the magnetic field has already been included in the band structure. On the other hand, the velocity has an extra term involving $\Omega_n(\mathbf{k})$, which will be called the "curvature" of the Berry phase, because its integral over an area bounded by a path C in $\mathbf{k}$-space is the Berry phase $\Gamma_n(C)$ [8]. In physical terms, $\Omega_n(\mathbf{k})$ describes the contribution of state $\Psi_n(\mathbf{k})$ to the Hall conductivity in the absence of scattering. The derivation of Eq. (3) is based on a homogeneous field; nevertheless, it should still be valid when the field is slowly varying in space and time. The generalization of Eq. (3) to higher dimensions is straightfoward. In that case there will be more than one component of the curvature.

The combination of Eq. (3) with the Boltzmann equation,

$$\dot{\mathbf{r}} \cdot \frac{\partial f}{\partial \mathbf{r}} + \dot{\mathbf{k}} \cdot \frac{\partial f}{\partial \mathbf{k}} = \left(\frac{\partial f}{\partial t}\right)_{\text{coll}}, \tag{5}$$

offers a general framework for *semiclassical* transport in a MBB. The right hand side is the collision term due to impurity scatterings, etc. It has to be cautioned that the Boltzmann equation is valid only when the scattering broadening of a MBB is small compared with its bandwidth. This does not pose essential difficulty when p/q is a simple fraction. In the more general situation of a large or infinite q (irrational ϕ), we have to modify our approach in the following way: We divide the total magnetic field B into B_0 and δB, where B_0 relates to the band structure not destroyed by disorder, and δB is a small perturbation. Then the semiclassical dynamics in the MBBs of B_0, driven by E and δB (see Eqs. (7) and (9) in the next section), will be employed in the Boltzmann equation. Under such a circumstance, the scattering broadening is only required to be smaller than the bandwidth for B_0.

To demonstrate the use of Eq. (3) in transport problems involving an electric perturbation, we consider a homogeneous system in which f depends on $\mathbf{k}$ only, and use the relaxation time approximation. The current to first order in $\mathbf{E}$ is:

$$\mathbf{J}_n = \mathbf{E} \times \hat{z} \frac{e^2}{\hbar} \int \frac{d^2\mathbf{k}}{(2\pi)^2} f_0 \Omega_n(\mathbf{k}) + \left(\frac{e}{\hbar}\right)^2 \int \frac{d^2\mathbf{k}}{(2\pi)^2} \tau(\mathbf{k}) \left(-\frac{\partial f_0}{\partial \mathcal{E}}\right) \left(\mathbf{E} \cdot \frac{\partial \mathcal{E}_n}{\partial \mathbf{k}}\right) \frac{\partial \mathcal{E}_n}{\partial \mathbf{k}}, \tag{6}$$

where $\tau(\mathbf{k})$ is the relaxation time and f_0 is the unperturbed Fermi-Dirac distribution. The first term is new and is due to the Berry phase curvature, which is nonzero in general. In fact, in the simple case of a filled band ($f_0 = 1$), for which the second term is zero, this term reduces to the topological Chern number discovered by Thouless *et al* [7]. The second term is the usual Boltzmann transport formula [1]. It was used for the calculation of longitudinal conductivity of magnetic bands [9]. Our theory justifies this usage because the Berry phase term only contributes to the Hall conductivity.

Compared to the brute force, all purpose Kubo formula approach, the semiclassical dynamics, in conjunction with the Boltzmann transport theory, offers a simple and intuitive picture of the behavior of the physical system. The semiclassical dynamics also provides a useful tool in problems with spatially varying and/or time dependent fields, where quantum mechanical calculations are usually very involved. A detailed study will appear in a separate publication [10].

Magnetic perturbation and hyperorbits — The remaining part of this letter will focus on how the MBBs for a given magnetic field B_0 are perturbed by adding δB. In this case, the equations for the semiclassical dynamics become

$$\dot{\mathbf{r}} = \frac{\partial \mathcal{E}_n(\mathbf{k})}{\hbar \partial \mathbf{k}} - \dot{\mathbf{k}} \times \hat{z} \Omega_n(\mathbf{k}), \quad \hbar \dot{\mathbf{k}} = -e \dot{\mathbf{r}} \times \delta B \hat{z}. \tag{7}$$

This result can be derived, for instance, by considering a wave packet in a MBB and studying how its center of mass moves in $\mathbf{r}$-space and $\mathbf{k}$-space [10]. Notice that the wave vector $\mathbf{k}$, which is a good quantum number for B_0, is no longer conserved in the presence of δB, even when both B_0 and δB are uniform.

After $\dot{\mathbf{r}}$ is eliminated by combining both equations in Eq. (7), the equation for $\dot{\mathbf{k}}$ takes the following form:

$$\hbar \dot{\mathbf{k}} = -\frac{\partial \mathcal{E}_n / \partial \mathbf{k} \times \hat{z} \delta B e / \hbar}{1 + \Omega_n(\mathbf{k}) \delta B e / \hbar}. \tag{8}$$

It is not difficult to see that $\mathbf{k}$ moves along a constant energy contour in the magnetic band structure. The presence of $\Omega_n(\mathbf{k})$ changes the speed of motion, but it does not alter the shape of the orbit for a given energy. The cyclotron orbit in $\mathbf{r}$-space can be derived from $\dot{\mathbf{r}} = \dot{\mathbf{k}} \times \hat{z}(\hbar / e\delta B)$, which shows that the $\mathbf{r}$-orbit is simply the $\mathbf{k}$-orbit rotated by $\pi/2$ and scaled by the factor $\hbar / e\delta B$. Such "hyperorbits" were introduced by Pippard [11]. We emphasize that, the existence of hyperorbits is a quantum effect, and cannot be explained classically. One possible way to detect them is by using an electron focusing device with a configuration similar to a mass spectrometer [12]. In order to have a successful observation, the hyperorbit has to be within the ballistic range of electron transport.

Analogous to ordinary cyclotron orbits, these hyperorbits will drift in an external electric field. By adding a $-e\mathbf{E}$ term to the second equation in Eq. (7), we obtain [1]

$$\dot{\mathbf{r}} = \frac{\hbar}{e\delta B}\dot{\mathbf{k}} \times \hat{z} - \frac{\mathbf{E} \times \hat{z}}{\delta B}. \tag{9}$$

The time average of the first term for a *closed* orbit is zero, while the second term describes the drifting of the hyperorbit that results in a Hall current. This will be used later to calculate the Hall conductivity for magnetic subbands in the Hofstadter spectrum.

Quantization of hyperorbits — The dynamical equation for $\mathbf{k}$ in Eq. (8) can be cast into the Lagrangian formulation

$$L(\mathbf{k}, \dot{\mathbf{k}}) = \frac{\hbar^2}{e\delta B}(k_1\dot{k}_2 - k_2\dot{k}_1) - \mathcal{E}_n(\mathbf{k}) + \hbar \mathbf{A}_n \cdot \dot{\mathbf{k}}, \tag{10}$$

where $\mathbf{A}_n$ is the "vector potential" for the Berry phase that satisfies $\nabla \times \mathbf{A}_n(\mathbf{k}) = \Omega_n(\mathbf{k})\hat{z}$. Apart from an unimportant constant, the propagator for a completed closed orbit C in period T is given by $\exp(i/\hbar \int_0^T Ldt)$. For a semiclassical orbit, the amplitudes for paths that circle different times must add constructively. This leads to the following quantization rule for the area of a hyperorbit in the n-th MBB [13],

$$\frac{1}{2}\oint_{C_m} (\mathbf{k} \times d\mathbf{k}) \cdot \hat{z} = 2\pi \left(m + \frac{1}{2} - \frac{\Gamma_n(C_m)}{2\pi} \right) \frac{e\delta B}{\hbar}, \tag{11}$$

where m is a non-negative integer, and $\Gamma_n(C_m)$ is the Berry phase for orbit C_m [14].

By using the constraint that the area of the outer-most orbit be smaller than the area of the first magnetic Brillouin zone of size $(2\pi/a)^2/q$, we found the number of quantized orbits to be the integer part of $1/(q\delta\phi) + \Gamma_n(C_{max})/2\pi + 1/2$, where $\delta\phi = \delta B a^2 e/h$. $\Gamma_n(C_{max})/2\pi$ can be replaced by the Hall conductivity σ_n (in units of e^2/h), because the Berry phase for the orbit C_{max} is very close to $2\pi\sigma_n$. For an integer value of $1/(q\delta\phi)$, we then have

$$\text{number of orbits } = \left| \frac{1}{q\delta\phi} + \sigma_n \right|. \tag{12}$$

These orbits will be broadened into subbands by tunneling to orbits with the same energy in other magnetic Brillouin zones. Therefore, this naive-looking formula relates the Hall conductivity σ_n of a parent band to the number of daughter subbands under a perturbation $\delta\phi$. It is crucial in understanding the clustering pattern for the Hofstadter spectrum.

The Hofstadter spectrum — Consider a Bloch band subject to a magnetic flux that can be expanded as

$$\phi = \cfrac{1}{f_1 + \cfrac{1}{f_2 + \cfrac{1}{f_3 + \cdots}}}. \tag{13}$$

Its r-th order approximation will be written as $\phi_r = p_r/q_r$. At the first order, the Bloch band is broken into f_1 subbands; each subband is further fragmented by an extra magnetic field at the second order, and so on. As will be shown below, the new semiclassical dynamics offers a clear and intuitive picture about how each subband (parent) in the r-th order should split into daughter subbands [15].

Our major findings are summarized below: (a) Firstly, it is important to distinguish between "closed" and "open" subbands. We define a "closed" subband to be a subband broadened from a closed hyperorbit; similarly an "open" subband is derived from an open orbit. For a square or a triangular lattice, all subbands except one for every parent band are closed. (b) The closed subbands at the same order all have the same Hall conductivity

$$\sigma_r = (-1)^{r-1} q_{r-1}, \tag{14}$$

where the subscript in σ refers to the order, not to the band index of the subband. If a parent band has only one open daughter band (eg., for a square or a triangular lattice), the Hall conductivity for this open subband is $(-1)^{r-1}q_{r-1} + (-1)^r q_r$. (c) Because of the difference in the Hall conductivities, a closed band will break into f_{r+1} subbands at the next order, while an open band will break into $f_{r+1} + 1$ subbands.

We briefly describe the derivation for the Hall conductivity of a closed subband. It is determined by the response of the corresponding hyperorbit under an electric field. Using Eq. (9), we know that the drifting velocity of a closed orbit is $\langle \dot{\mathbf{r}} \rangle = -\mathbf{E} \times \hat{z}/\delta B$. It follows that the Hall conductivity for a closed subband at the r-th order is $\sigma_r = e\rho_r/\delta B_{r-1}$, where ρ_r is the electron density per unit area and $\delta B_{r-1} = h\delta\phi_{r-1}/(ea^2)$. Since the electrons are equally distributed among the q_r subbands at the same order, ρ_r is equal to $1/q_r$ times the electron density of the original Bloch band. Eq. (14) is obtained after the identity $p_r/q_r - p_{r-1}/q_{r-1} = (-1)^{r-1}/(q_r q_{r-1})$ is used to evaluate $\delta\phi_{r-1}$. The Hall conductivity for an open subband can be figured out quite easily by using the sum rule: $\sigma_{\text{parent}} = \sum \sigma_{\text{daughter}}$ [10] [16].

For a square lattice, an open daughter band is always located at the center of a parent band; therefore, we know which subband the conductivity σ^{open} belongs to. The Hall conductivity distribution obtained this way is exactly the same as that which is obtained by using the Diophantine equation with some subsidiary constraints [7]. We should emphasize that this is the first time the (seemingly) erratic behavior of the Hall conductivities for the Hofstadter spectrum is given a clear and direct physical meaning.

With the help of Eq. (12), we can determine how a parent band is splitted by $\delta\phi$. Substituting $\delta\phi_r = (-1)^r/(q_{r+1}q_r)$ into Eq. (12), and taking into account the q_r-fold degeneracy for the r-th order magnetic Brillouin zone, we then have the number of daughter bands for an r-th order parent band

$$\mathcal{D}_r = |(-1)^r q_{r+1} + \sigma_r|/q_r. \tag{15}$$

In conjunction with Eq. (14), we obtain $\mathcal{D}_r = f_{r+1}$ for a closed band. Similarly we have $\mathcal{D}_{r+1} = f_{r+1} + 1$ for an open band. Azbel conjectured that there are f_{r+1} daughter subbands for *every* parent band [17]. It is clear from our calculation that his conjecture is correct only for a closed parent band. An expression similar to Eq. (15) has been obtained by Wilkinson, but his evaluation of such an expression required external input for the Hall conductivity [18].

Various values of ϕ have been used to check the predicted splitting of subbands with the actual Hofstadter spectrum, and the results are found to agree very well. One exception occurs when there is a degeneracy among the subbands. In this case the individual σ_r, as well as $\mathcal{D}_r$, cannot be determined uniquely. For example, when $\phi = 1/4$, the central two subbands are degenerate for a square lattice and the Hall conductivities can be either $(1, 1, -3, 1)$ or $(1, -3, 1, 1)$ $(q_0 = 1)$. One of these two subbands will have this -3 $(\sigma^{\rm open})$ when the degeneracy is lifted by next nearest neighbor coupling [19]. As expected, we found the subband corresponding to a hyperorbit that is closer to the nesting energy contour acquires this $\sigma^{\rm open}$.

For a triangular lattice, the open orbit is near the zone boundary (exact location requires some calculation). Therefore, the distribution of Hall conductivities and band splitting are no longer symmetric in energy. This explains the asymmetry in the spectrum generated by numerical calculation [19]. For a lattice without 3-fold or 4-fold symmetry, more than one open orbit may exist in a range of energy. In this case, the total Hall conductivities for open orbits can be figured out by the sum rule. However, further analysis is required to get the detailed distribution within them.

In summary, we have demonstrated the new semiclassical dynamics for magnetic Bloch bands and its application to a variety of phenomena involving strong magnetic fields. It can be used to calculate the transport properties, to obtain the quantization rule for hyperorbits, and to get the Hall conductivity for a magnetic subband. We also showed that the complex structure of the Hofstadter spectrum can be explained as a logically consistent part of this theory.

ACKNOWLEDGMENTS

Q. N. wishes to thank F. H. Claro, M. Kohmoto, W. Kohn, and M. Marder for many helpful discussions. M. C. C. wishes to thank E. Demircan, G. Georgakis, and R. Janhke for their comments on the manuscript. This work is supported by the R. A. Welch foundation.

REFERENCES

[1] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, (W. B. Saunders Co., Philadelphia, 1976) Chapter 12.

[2] D. R. Hofstadter *Phys. Rev.* **14**, 2239 (1976).

[3] A. Rauh, *Phys. Status. Solidi.* **B65**, K131 (1974).

[4] R. R. Gerhardts, D. Weiss and U. Wulf, *Phys. Rev.* **B43**, 5192 (1991).

[5] J. Zak *Phys. Rev.* **136**, A776 (1964); **134**, A1602 (1964).

[6] M. C. Chang and Q. Niu, *Phys. Rev.* **B50**, 10843 (1994).

[7] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982); M. Kohmoto, *Ann. Phys.* **160**, 343 (1985); *J. Phys. Soc. Japan* **62**, 659 (1993). Another example of Berry phase for a periodic system is given by Zak, *Phys. Rev. Lett.* **62**, 2747 (1989).

[8] M. V. Berry, *Proc. Roy. Soc. London* **A392**, 45 (1984); B. Simon, *Phys. Rev. Lett.* **51**, 2167 (1983).

[9] R. B. S.Oakeshott and A. MacKinnon, *J. Phys. Cond. Matt.* **5**, 6991 (1993).

[10] M. C. Chang and Q. Niu, unpublished.

[11] A. B. Pippard, in *Physics of Solids in Intense Magnetic Fields*, edited by E. D. Haidemenakis, (Plenum, New York, 1969), p.370.

[12] See the paper by H. van Houten, *Phys. Rev.* **B 39**, 8556 (1989) and the references therein. The electron focusing device has also been used to map out the constant energy contours in **k** space, see J. J. Heremans, M. B. Santos, and M. Shayegan, *Surf. Sci.* **305**, 348 (1994).

[13] A factor of 1/2 has been inserted to the right hand side for the following reason: The orbit acuqires a phase shift of $-\pi/2$ whenever it passes a caustic turning point. In this case, there are two caustics for a cyclotron orbit. Therefore there is an additional phase lag of $-\pi$ contributes to the nondynamical phase. For more detail, see L. D. Landau and E. M. Lifshitz, The Classical Theory of Fields, 4th ed. (Pergmon, Oxford, 1987). Van Houten's paper in Ref. 12 also has a nice discussion on the additional phase shift.

[14] A similar formula for quantization of orbits in phase space has been derived by Wilkinson, in which it is not apparent that his Berry phase is related to the Hall conductivity. See M. Wilkinson, *J. Phys.* **A17**, 3459 (1984), also H. Kuratsuji and S. Iida, *Prog. Theo. Phys.* **74**, 439 (1985); A. Entelis and S. Levit, *Phys. Rev. Lett.* **69**, 3001 (1992).

[15] The scaling property of the Hofstadter spectrum guarantees that the semiclassical approximation, which is valid at the first order, remains valid at finer energy scales. See, for example, D. J. Thouless and Q. Niu, *J. Phys.* **A16**, 1911 (1983).

[16] J. E. Avron and B. Simon, *J. Phys. A: Math. Gen.* **18**, 2199 (1985).

[17] M. Y. Azbel, *Zh. Eskp. Teor. Fiz.* **46**, 939 (1964) [*Sov. Phys. -JETP* **19**, 634 (1964)].

[18] M. Wilkinson, *Proc. R. Soc. Lond.* **A391**, 305 (1984); *J. Phys.* **A20**, 4337 (1987); **27**, 8123 (1994) and references therein.

[19] Y. Hatsugai and M. Kohmoto, *Phys. Rev.* **42**, 8282 (1990), also see F. H. Claro and G. H. Wannier, *Phys. Rev.* **B19**, 6068 (1979).