Open Momentum Space Method for Hofstadter Butterfly and the Quantized Lorentz Susceptibility

Biao Lian,[1] Fang Xie,[1] and B. Andrei Bernevig[1]

[1]*Department of Physics, Princeton University, Princeton, New Jersey 08544, USA*
(Dated: April 26, 2021)

We develop a generic $\mathbf{k} \cdot \mathbf{p}$ open momentum space method for calculating the Hofstadter butterfly of both continuum (moiré) models and tight-binding models, where the quasimomentum is directly substituted by the Landau level (LL) operators. By taking a LL cutoff (and a reciprocal lattice cutoff for continuum models), one obtains the Hofstadter butterfly with in-gap spectral flows. For continuum models such as the moiré model for twisted bilayer graphene, our method gives a sparse Hamiltonian, making it much more efficient than existing methods. The spectral flows in the Hofstadter gaps can be understood as edge states on a momentum space boundary, from which one can determine the two integers (t_ν, s_ν) of a gap ν satisfying the Diophantine equation. The spectral flows can also be removed to obtain a clear Hofstadter butterfly. While t_ν is known as the Chern number, our theory identifies s_ν as a dual Chern number for the momentum space, which corresponds to a quantized Lorentz susceptibility $\gamma_{xy} = eBs_\nu$.

Two-dimensional (2D) lattice electrons in large magnetic fields are known to exhibit Hofstadter butterfly spectra [1]. Conventionally, the Hofstadter butterfly is calculated at rational fluxes per unit cell $\varphi = 2\pi p/q$ in a basis with translation symmetry of q unit cells, where p and q are coprime integers. The calculation often involves a complicated construction of the matrix elements. In particular, for continuum $\mathbf{k} \cdot \mathbf{p}$ models obtained from plane wave expansions such as the moiré model for twisted bilayer graphene (TBG) [2–4], the Hofstadter Hamiltonian matrix is infinite dimensional and dense [5–10], which requires a large cutoff for the spectrum to converge.

In contrast, the Landau levels (LLs) of a $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian at small magnetic fields can be calculated by simply substituting the quasimomentum $\mathbf{k} = (k_x, k_y)$ with $(\frac{a+a^\dagger}{\sqrt{2}\ell}, \frac{a-a^\dagger}{i\sqrt{2}\ell})$, where a and $a^\dagger$ are the LL lowering and raising operators, and ℓ is the magnetic length [11]. In this letter, we demonstrate that such a substitution with a LL cutoff (and a reciprocal lattice cutoff for continuum models) provides an efficient method for calculating the Hofstadter butterfly in large magnetic fields, which greatly simplifies the Hamiltonian matrix elements [12]. In particular, for continuum models, this method yields a sparse Hamiltonian, whose spectrum can be efficiently calculated by the shift-and-invert Lanczos method.

The method can be understood as an open momentum space calculation, where the smaller of the momentum-space LL wavefunction radius cutoff and reciprocal lattice radius cutoff plays the role of a momentum space boundary. As a result, the spectrum contains not only the Hofstadter butterfly, but also in-gap spectral flow levels [12, 13] which can be understood as "momentum space edge states". We show that the spectral flows of these edges allow us to determine the two integers (t_ν, s_ν) in a Hofstadter gap ν satisfying the Diophantine equation [14–16], where t_ν is the Chern number of the gap. Moreover, we show that s_ν can be interpreted as a dual Chern number for the momentum space, which yields a quan-

tized Lorentz susceptibility (Eq. (15)). Furthermore, by identifying and removing the momentum space edge states, one can obtain the Hofstadter butterfly without spectral flows. We demonstrate our method for both continuum models and tight binding models in a 2D periodic lattice. We shall denote the lattice Bravais vectors as $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$, and the reciprocal vectors as $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, which satisfy $\boldsymbol{g}_i \cdot \boldsymbol{d}_j = 2\pi\delta_{ij}$ $(i, j = 1, 2)$.

Continuum models. At zero magnetic field, a continuum model can be written in the real space basis $|\mathbf{r}, \alpha\rangle$ as [3–7, 17]

$$H^{\alpha\beta}(\mathbf{r}) = \epsilon^{\alpha\beta}(-i\nabla) + \sum_j V_j^{\alpha\beta} e^{i\mathbf{q}_j^{\alpha\beta} \cdot \mathbf{r}}, \qquad (1)$$

where $\mathbf{r} = (x, y)$ is the real space position, $-i\nabla = -i(\partial_x, \partial_y)$ is the canonical momentum, and we assume there are M intrinsic orbitals labeled by α, β. $\epsilon^{\alpha\beta}(-i\nabla)$ and $V_j^{\alpha\beta} e^{i\mathbf{q}_j^{\alpha\beta} \cdot \mathbf{r}}$ are the electron kinetic term in free space and the periodic lattice potential between orbitals β and α, respectively. If one denote $\mathbf{Q} \in \boldsymbol{g}_1\mathbb{Z} + \boldsymbol{g}_2\mathbb{Z}$ as the reciprocal lattice, and choose the momentum origin of orbital α at $\mathbf{p}_\alpha$, one can define a momentum lattice $\mathbf{Q}_\alpha = \mathbf{p}_\alpha + \mathbf{Q}$ for orbital α, and $\mathbf{q}_j^{\alpha\beta}$ in Eq. (1) must be the difference $\mathbf{Q}'_\alpha - \mathbf{Q}_\beta$ between some sites $\mathbf{Q}'_\alpha$ and $\mathbf{Q}_\beta$ (see supplementary material (SM) [18] Sec. S2A). Generically, one can always fix all $\mathbf{p}_\alpha = \mathbf{0}$; however, in certain models (e.g., the TBG model [2]) nonzero $\mathbf{p}_\alpha$ choices are preferred.

One can transform the zero-magnetic-field Hamiltonian (1) into the momentum eigenbasis $|\mathbf{k}, \mathbf{Q}_\alpha, \alpha\rangle = \int \mathrm{d}^2\mathbf{r} e^{i(\mathbf{k}+\mathbf{Q}_\alpha) \cdot \mathbf{r}}|\mathbf{r}, \alpha\rangle$, where $\mathbf{k}$ is in the first Brillouin zone (BZ). The momentum space Hamiltonian under basis $|\mathbf{k}, \mathbf{Q}_\alpha, \alpha\rangle$ then takes the form [2, 18]

$$H_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta}^{\alpha\beta}(\mathbf{k}) = \epsilon^{\alpha\beta}(\mathbf{k} + \mathbf{Q}_\beta)\delta_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta} + \sum_j V_j^{\alpha\beta}\delta_{\mathbf{Q}'_\alpha, \mathbf{Q}_\beta + \mathbf{q}_j^{\alpha\beta}}. \qquad (2)$$

When a uniform out-of-plane magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$ is added, $-i\nabla$ in Eq. (1) is replaced by the kinematic

momentum $\mathbf{\Pi} = -i\nabla - \mathbf{A}(\mathbf{r})$, where $\mathbf{A}(\mathbf{r})$ is the vector potential satisfying $\partial_x A_y - \partial_y A_x = B$. The kinetic momentum satisfies $[\Pi_x, \Pi_y] = i/\ell^2$, where $\ell = 1/\sqrt{B}$ is the magnetic length. We also define the guiding center $\mathbf{R} = \mathbf{r} - \ell^2 \hat{\mathbf{z}} \times \mathbf{\Pi}$, which satisfies $[R_x, R_y] = -i\ell^2$, and $[\mathbf{R}, \mathbf{\Pi}] = 0$.

The usual Hofstadter method for continuum models employs the Landau basis defined by eigenstates of R_x and $\mathbf{\Pi}^2$, which has complicated matrix elements [5–10]. Here we shall take a different basis, under which we prove the nonzero magnetic field Hamiltonian can be simply obtained by the zero-field momentum-space Hamiltonian (2) with the substitution of Eq. (5).

We define $R_{\hat{\boldsymbol{\tau}}} = \mathbf{R} \cdot \hat{\boldsymbol{\tau}}$ as the guiding center along unit vector $\hat{\boldsymbol{\tau}}$, where we choose $\frac{\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)}{\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)}$ irrational. We also define a set of (linearly dependent) LL operators $a_{\mathbf{Q}_\alpha} = \frac{\ell}{\sqrt{2}}[\Pi_x - Q_{\alpha,x} - k_{0,x} + i(\Pi_y - Q_{\alpha,y} - k_{0,y})]$ and their conjugates $a^\dagger_{\mathbf{Q}_\alpha}$ associated with momentum sites $\mathbf{Q}_\alpha$, where $\mathbf{k}_0 = (k_{0,x}, k_{0,y})$ is a freely chosen real vector which we call the *center momentum*. We then construct an orthonormal basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ for orbital α and reciprocal site $\mathbf{Q}_\alpha$ by requiring

$$R_{\hat{\boldsymbol{\tau}}}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = [\lambda - \ell^2 \hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \mathbf{Q}_\alpha)]|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle,$$
$$a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = n|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle. \quad (3)$$

Here $n \geq 0$ is an integer LL number, while λ is a real number chosen in the set $\lambda + \ell^2 \hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)\mathbb{Z} + \ell^2 \hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)\mathbb{Z}$ representing the set, or abstractly, $\lambda \in \mathbb{R}/[\ell^2 \hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)\mathbb{Z} + \ell^2 \hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)\mathbb{Z}]$ (see SM [18] Sec. S2B). It can then be proved that all the states $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ form a complete basis for the continuum model satisfying $\langle \lambda, \mathbf{Q}_\alpha, n, \alpha|\lambda', \mathbf{Q}'_\beta, n', \beta\rangle = \delta_{\lambda\lambda'}\delta_{\mathbf{Q}_\alpha, \mathbf{Q}'_\beta}\delta_{nn'}\delta_{\alpha\beta}$.

The above basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ is advantageous because the nonzero-magnetic-field Hamiltonian is diagonal in λ and independent of λ. In SM [18] Sec. S2B, we show the Hamiltonian in a fixed λ subspace is

$$H^{\lambda,\alpha\beta}_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta} = \epsilon^{\alpha\beta}(\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\beta} + \mathbf{k}_0 + \mathbf{Q}_\beta)\delta_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta} + \sum_j V_j \delta_{\mathbf{Q}'_\alpha, \mathbf{Q}_\beta + \mathbf{q}^{\alpha\beta}_j}, \quad (4)$$

where we have defined $\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha} = \frac{1}{\sqrt{2}\ell}(a_{\mathbf{Q}_\alpha} + a^\dagger_{\mathbf{Q}_\alpha}, -ia_{\mathbf{Q}_\alpha} + ia^\dagger_{\mathbf{Q}_\alpha})$. Without ambiguity, we can drop the subindex $\mathbf{Q}_\alpha$ and simplify $(a_{\mathbf{Q}_\alpha}, a^\dagger_{\mathbf{Q}_\alpha})$ as $(a, a^\dagger)$, which acts as $a|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = \sqrt{n}|\lambda, \mathbf{Q}_\alpha, n-1, \alpha\rangle$ and $a^\dagger|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = \sqrt{n+1}|\lambda, \mathbf{Q}_\alpha, n+1, \alpha\rangle$. The Hamiltonian (4) is then exactly the zero-field Hamiltonian $H^{\alpha\beta}_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta}(\mathbf{k})$ in Eq. (2) with the substitution

$$k_x \to \frac{a + a^\dagger}{\sqrt{2}\ell} + k_{0,x}, \quad k_y \to \frac{a - a^\dagger}{i\sqrt{2}\ell} + k_{0,y}, \quad (5)$$

as we claimed earlier. One then only need calculate the spectrum for a fixed λ. Different λ and λ' subspaces have identical spectra, but have eigenstates differing by displacement $\lambda - \lambda'$ in the $\hat{\boldsymbol{\tau}}$ direction ($R_{\hat{\boldsymbol{\tau}}}$ eigenvalue).
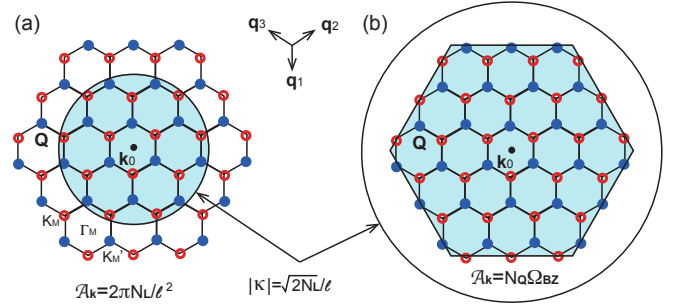

FIG. 1. (a) When $\varphi/2\pi < N_Q/N_L$, the momentum space (the shaded area) has a circular boundary of radius $\sqrt{2N_L}/\ell$. (b) When $\varphi/2\pi > N_Q/N_L$, the momentum space boundary is the reciprocal lattice boundary enclosing N_Q BZs (the shaded area).

To numerically calculate the spectrum of Hamiltonian (4), one can fix a center momentum $\mathbf{k}_0$, take a LL cutoff $n \leq N_L$, and take a cutoff of reciprocal lattice $\mathbf{Q}_\alpha$ at a boundary enclosing N_Q BZs. This yields a Hamiltonian of size MN_LN_Q for M intrinsic orbitals. If $\epsilon(\mathbf{k})$ only contains polynomials up to Δ-th power of $\mathbf{k}$, and the number of $\mathbf{q}^{\alpha\beta}_j$ is finite, $\langle \lambda, \mathbf{Q}'_\alpha, m, \alpha|H|\lambda, \mathbf{Q}_\beta, n, \beta\rangle$ will be zero for $|m-n| > \Delta$ or $|\mathbf{Q}'_\alpha - \mathbf{Q}_\beta| > \max(|\mathbf{q}^{\alpha\beta}_j|)$, so the Hamiltonian H is a sparse matrix. The low-energy eigenstates and spectrum can then be efficiently calculated by the Lanczos algorithm.

The cutoffs N_Q and N_L, however, lead to spectral flows in the Hofstadter gaps due to the absence of periodic boundary conditions [12, 13]. As an example, we calculate the Hofstadter butterfly of the TBG continuum model defined on a honeycomb momentum lattice $\mathbf{Q}_\alpha$ [2], which has a Dirac kinetic term $\epsilon(\mathbf{k}) = v_F\boldsymbol{\sigma}^* \cdot \mathbf{k}$, and 2×2 hopping matrices V_j between the nearest momentum sites, where $\boldsymbol{\sigma}^* = (\sigma_x, -\sigma_y)$ are the Pauli matrices (SM [18] Sec. S3). Fig. 2(a) shows the TBG spectrum at twist angle $\theta = 2.2°$ versus the flux per unit cell $\varphi = B|\boldsymbol{d}_1 \times \boldsymbol{d}_2|$, where we take $N_Q = 37$ and $N_L = 60$. Besides the Hofstadter butterfly, one can see numerous in-gap spectral flow levels.

The in-gap spectral flows are generically due to the presence of boundaries which host edge states [12, 13]. Here, as illustrated in Fig. 1(a) and (b), the cutoff N_Q sets a boundary of momentum radius $\sqrt{\frac{N_Q\Omega_{BZ}}{\pi}}$ enclosing N_Q BZs, where $\Omega_{BZ} = 4\pi^2/|\boldsymbol{d}_1 \times \boldsymbol{d}_2|$ is the BZ area; while the LL cutoff N_L yields a boundary $\sqrt{\langle \hat{\boldsymbol{\kappa}}^2_{\mathbf{Q}_\alpha}\rangle} \leq \frac{\sqrt{2N_L}}{\ell}$ for $\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha}$ in the Hamiltonian (4). The smaller value of $\sqrt{\frac{N_Q\Omega_{BZ}}{\pi}}$ and $\frac{\sqrt{2N_L}}{\ell}$ then serves as a momentum space boundary radius for Hamiltonian (4) (SM [18] Sec. S2C), which gives rise to edge states.

The momentum space edge state levels then generate the spectral flows versus the magnetic field B. This can be understood from the Diophantine equation [14–16, 18]

satisfied by the ν-th Hofstadter gap ($\nu \in \mathbb{Z}$) at $\varphi = 2\pi p/q$ flux per unit cell:

$$t_\nu p + s_\nu q = \nu , \qquad (6)$$

where (t_ν, s_ν) are two integer quantum numbers characterizing the gap. t_ν is the Chern number of the gap, while s_ν is referred to as the electromechanical quantum number in Ref. [5]. It is often rewritten as

$$t_\nu \frac{\varphi}{2\pi} + s_\nu = \rho , \qquad (7)$$

where $\rho = \nu/q$ is the number of occupied bulk states per unit cell in the gap [18–20]. Here it is more useful to rewrite it in a dual form

$$t_\nu + s_\nu \frac{2\pi}{\varphi} = \rho_K , \qquad (8)$$

where $\rho_K = 2\pi\rho/\varphi = \nu/p$. In SM [18] Sec. S2F, we show that ρ_K gives the number of occupied bulk states *per BZ* in the gap for the Hamiltonian (4) at a fixed λ. Furthermore, in Eq. (16), we show that s_ν plays the role of a dual Chern number for the momentum space. Eq. (8) then determines the in-gap spectral flows (Fig. 2(a)) in two different regimes as follows.

In the first regime $\varphi/2\pi < N_Q/N_L$, the momentum space boundary is a circle enclosing a φ-dependent area $\mathcal{A}_K = 2\pi N_L/\ell^2$ centered at $\mathbf{k}_0$ (Fig. 1(a)). In a gap, the total number of occupied states in the momentum area $\mathcal{A}_K$ is $\mathcal{N}_K = \rho_K \mathcal{A}_K / \Omega_{BZ} = N_L \rho_K \varphi/2\pi = N_L \rho$. Therefore, by Eq. (8) we have

$$\mathcal{N}_K = N_L(t_\nu \varphi/2\pi + s_\nu) . \qquad (9)$$

In a bulk gap, the number of occupied states $\mathcal{N}_K$ can only change by pumping edge states into (out of) the bulk. Therefore, the edge states necessarily produce in-gap spectral flows, where the rate of flowing levels is $d\mathcal{N}_K/d(\varphi/2\pi) = N_L t_\nu$ by Eq. (9). Fig. 2(c) shows a gap in this regime, where the midgap line (dashed line) crosses 16 levels as $\varphi/2\pi$ increases from 0.25 to 0.5, and $N_L = 60$. The flow rate is then $d\mathcal{N}_K/d(\varphi/2\pi) = 16/(0.5-0.25) = 64$, so we can identify the Chern number of the gap as the integer closest to $N_L^{-1} d\mathcal{N}_K/d(\varphi/2\pi) = 1.07$, namely, $t_\nu = 1$. Further, at $\varphi/2\pi = 0.5$, we counted there are $\mathcal{N}_K = 32$ levels from zero energy to the midgap energy (for TBG, $\mathcal{N}_K = 0$ is at zero energy (SM [18] Sec. S2G)), so we find the gap has $s_\nu = \mathcal{N}_K/N_L - t_\nu \varphi/2\pi = 0$ from Eq. (9).

In the second regime $\varphi/2\pi > N_Q/N_L$, the momentum space boundary is given by cutoff N_Q, which encloses a φ independent area $\mathcal{A}_K = N_Q \Omega_{BZ}$ (Fig. 1(b)). The number of occupied states $\mathcal{N}_K = \rho_K \mathcal{A}_K / \Omega_{BZ}$ in momentum area $\mathcal{A}_K$ in a gap is then

$$\mathcal{N}_K = N_Q(t_\nu + 2\pi s_\nu/\varphi) . \qquad (10)$$

This yields a spectral flow rate $d\mathcal{N}_K/d(2\pi/\varphi) = N_Q s_\nu$. Besides, for TBG which has a Dirac kinetic term, there
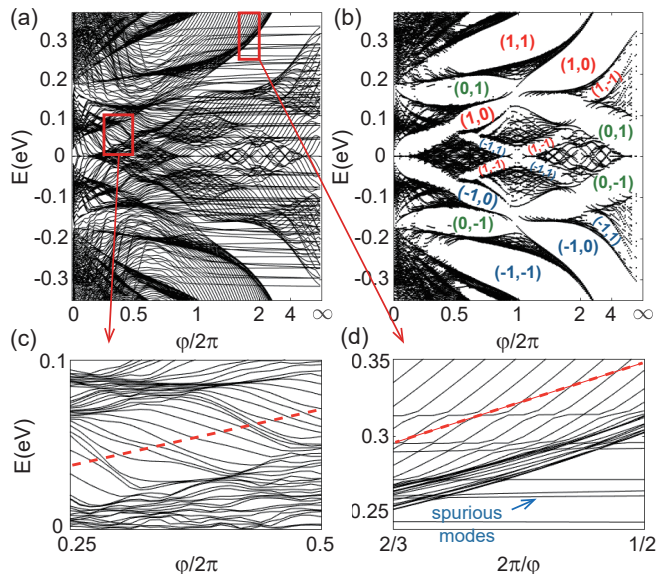

FIG. 2. (a) Hofstadter butterfly and spectral flow of $\theta = 2.2°$ TBG with $N_Q = 37$ and $N_L = 60$, and $\mathbf{k}_0$ at Γ point. The horizontal axis $\varphi/2\pi$ is linearly plotted in $[0, 1]$, and deformed into $2 - 2\pi/\varphi$ in $[1, \infty]$. (b) The Hofstadter butterfly after deleting the edge states (with $w = \min\{\ell^{-1}, 1.6\sqrt{\Omega_{BZ}}\}$, $P_c = 0.5$), and (t_ν, s_ν) in the gaps. (c) Zoom-in plot in the regime $\varphi/2\pi < N_Q/N_L$. (d) Zoom-in plot in the regime $\varphi/2\pi > N_Q/N_L$.

are $2N_Q$ horizontal levels at $\varphi/2\pi > N_Q/N_L$ in Fig. 2(a), which are spurious zero modes due to LL cutoff N_L (see SM [18] Sec. S3). These spurious levels should be excluded when counting $\mathcal{N}_K$. Fig. 2(d) shows a gap in this regime, where the midgap line crosses 6 levels (excluding the spurious modes) as $2\pi/\varphi$ decreases from 2/3 to 1/2, and $N_Q = 37$. The flow rate is then $d\mathcal{N}_K/d(2\pi/\varphi) = 6/(2/3-1/2) = 36$, thus s_ν can be identified as the integer closest to $N_Q^{-1} d\mathcal{N}_K/d(2\pi/\varphi) = 0.97$, namely, $s_\nu = 1$. Further, we counted there are $\mathcal{N}_K = 55$ levels (excluding the spurious modes) between midgap and zero energy at $2\pi/\varphi = 1/2$, thus the gap has a Chern number $t_\nu = \mathcal{N}_K/N_Q - 2\pi s_\nu/\varphi = 1$ from Eq. (10).

We note that models with a Dirac kinetic term $\epsilon(\mathbf{k}) = v_F \boldsymbol{\sigma}^* \cdot \mathbf{k}$ would have $\mathcal{N}_K = 0$ defined at half filling (zero energy for TBG), while models with a lower-bounded kinetic term (e.g., $\epsilon(\mathbf{k}) = k^2/2m_0$) would have $\mathcal{N}_K = 0$ below the lowest band (SM [18] Sec. S2G). More generically, if a gap persists below and above $\varphi/2\pi = N_Q/N_L$, one can identify t_ν and s_ν separately from the spectral flow rates at small and large φ, after which one can obtain $\mathcal{N}_K$ of the gap from Eq. (9) or (10).

The edge states and spurious modes can be easily removed from the spectrum. We define a boundary projector $P_{\kappa_b, w}$ onto basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ with $n > (\kappa_b - w)^2 \ell^2/2$ for some $w > 0$, where $\kappa_b \approx \min\{\frac{\sqrt{2N_L}}{\ell}, \sqrt{\frac{N_Q \Omega_{BZ}}{\pi}}\}$ is the radius of momentum space boundary. We can then

identify the eigenstates with $\langle P_{\kappa_b,w} \rangle > P_c$ above certain value $P_c \in [0,1]$ as momentum space edge states within distance w to the boundary, and delete them to obtain a bulk Hofstadter spectrum. For example, Fig. 2(b) is obtained by setting $w = \min\{\ell^{-1}, 1.6\sqrt{\Omega_{BZ}}\}$ and $P_c = 0.5$.

Tight-binding models. The substitution (5) can also be employed to calculate the Hofstadter butterfly of tight-binding models. Given the position $\boldsymbol{u}_\alpha$ of each Wannier orbital α in a unit cell in the continuum space, we denote orbital α at position $\boldsymbol{D} + \boldsymbol{u}_\alpha$ as $|\boldsymbol{D}, \alpha\rangle$, where $\boldsymbol{D} \in \boldsymbol{d}_1\mathbb{Z} + \boldsymbol{d}_2\mathbb{Z}$ is the lattice vector. The Hamiltonian under Peierls substitution [21–23] then takes the form

$$H = \sum_{j,\alpha,\beta} t_j^{\alpha\beta} T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta} ~, \tag{11}$$

where $\boldsymbol{D}_j \in \boldsymbol{d}_1\mathbb{Z} + \boldsymbol{d}_2\mathbb{Z}$, $t_j^{\alpha\beta}$ is the hopping from $|\boldsymbol{D}, \beta\rangle$ to $|\boldsymbol{D} + \boldsymbol{D}_j, \alpha\rangle$, and

$$T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta} = \sum_{\boldsymbol{D}} e^{i \int_{c_{\alpha\beta}} \mathbf{A}(\mathbf{r}) \cdot d\mathbf{r}} |\boldsymbol{D} + \boldsymbol{D}_j, \alpha\rangle \langle \boldsymbol{D}, \beta| \tag{12}$$

is the translation operator, with $c_{\alpha\beta}$ being the straight line segment from $\boldsymbol{D} + \boldsymbol{u}_\beta$ to $\boldsymbol{D} + \boldsymbol{D}_j + \boldsymbol{u}_\alpha$. At zero magnetic field, the Hamiltonian can be transformed into the momentum space basis $|\mathbf{k}, \alpha\rangle = \sum_{\boldsymbol{D}} e^{i\mathbf{k}\cdot(\boldsymbol{D}+\boldsymbol{u}_\alpha)}|\boldsymbol{D}, \alpha\rangle$ as

$$H^{\alpha\beta}(\mathbf{k}) = \sum_j t_j^{\alpha\beta} e^{-i\mathbf{k}\cdot(\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta)} ~. \tag{13}$$

At nonzero magnetic field, we define a basis as $\overline{|\lambda, n, \alpha\rangle} = \sum_{\boldsymbol{D}} |\boldsymbol{D}, \alpha\rangle\langle \boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \boldsymbol{0}, n, \alpha\rangle$, where $|\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha\rangle$ is the continuum space position eigenstate at position $\mathbf{r} = \boldsymbol{D} + \boldsymbol{u}_\alpha$, $|\lambda, \boldsymbol{0}, n, \alpha\rangle$ is the state defined in Eq. (3) in the continuum space at reciprocal site $\boldsymbol{0}$, and $\lambda \in \mathbb{R}/[\ell^2 \hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_1)\mathbb{Z} + \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_2)\mathbb{Z}]$. One can then show that $\overline{|\lambda, n, \alpha\rangle}$ forms a complete orthonormal basis of Hamiltonian (11) satisfying $\overline{\langle \lambda', n', \beta|}\overline{\lambda, n, \alpha\rangle} = \delta_{\lambda\lambda'}\delta_{n'n}\delta_{\beta\alpha}$ (SM [18] Sec. S4A). Furthermore, $T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta}$ is diagonal in λ and takes the λ independent form

$$T^{\lambda,\alpha\beta}_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta} = e^{-i(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0)\cdot(\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta)} \tag{14}$$

in a fixed λ subspace between basis $\overline{|\lambda, n, \beta\rangle}$ and $\overline{|\lambda, n', \alpha\rangle}$, where $\hat{\boldsymbol{\kappa}} = \frac{1}{\sqrt{2}\ell}(a + a^\dagger, -ia + ia^\dagger)$, with $a\overline{|\lambda, n, \alpha\rangle} = \sqrt{n}\overline{|\lambda, n-1, \alpha\rangle}$ and $a^\dagger\overline{|\lambda, n, \alpha\rangle} = \sqrt{n+1}\overline{|\lambda, n+1, \alpha\rangle}$ (SM [18] Sec. S4A). Therefore, the nonzero magnetic field tight-binding Hamiltonian (11) in a fixed λ is given by the zero-field momentum space Hamiltonian (13) with substitution (5). For nonstandard Peierls substitutions along nonstraight $c_{\alpha\beta}$ paths, $e^{-i(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0)\cdot(\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta)}$ in Eq. (14) becomes the path-ordered integral $\mathcal{P}e^{-i\int_{c_{\alpha\beta}}(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0)\cdot d\mathbf{r}}$ (SM [18] Sec. S4B).

The Hofstadter butterfly can then be numerically calculated with a LL cutoff, namely, $n \leq N_L$. Fig. 3(a) and (b) show the spectrum of the square lattice tight-binding model $H(\mathbf{k}) = -\cos k_x - \cos k_y$ [1] with cutoffs $N_L = 100$
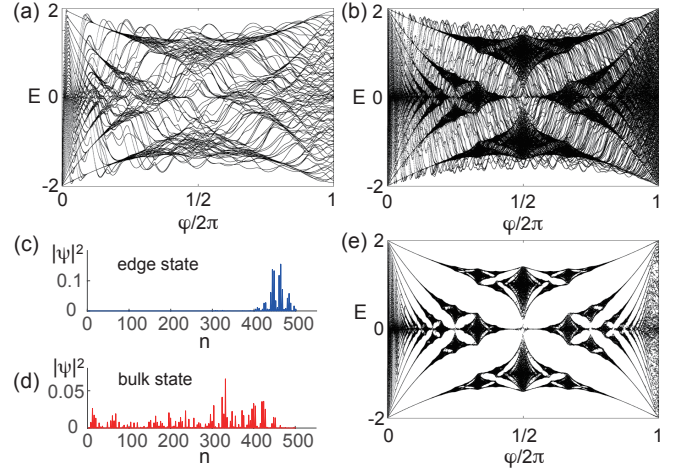

FIG. 3. The Hofstadter spectrum for tight-binding model $H(\mathbf{k}) = -\cos k_x - \cos k_y$ with $\mathbf{k}_0 = \boldsymbol{0}$ and LL cutoff (a) $N_L = 100$ and (b) $N_L = 500$. (c) Probability distribution of a typical momentum space edge state in (b) versus LL number n. (d) Probability distribution of a typical bulk state in (b). (e) The Hofstadter butterfly obtained by deleting the edge states in (b) (with $w = 3.5\ell^{-1}\sqrt{\varphi + 0.5}$, $P_c = 0.5$), which looks identical to that obtained by usual methods.

and $N_L = 500$, respectively, where we set $\mathbf{k}_0 = \boldsymbol{0}$. The spectrum exhibits both the Hofstadter butterfly and the spectral flows, which can again be understood as momentum space edge states. Since tight-binding models have no cutoff in the reciprocal lattice, the momentum space boundary is always at radius $\kappa_b = \frac{\sqrt{2N_L}}{\ell}$ given by N_L, and the spectral flows always satisfy Eq. (9).

One can define a boundary projector $P_{\kappa_b,w}$ onto basis $\overline{|\lambda, n, \alpha\rangle}$ with $n > (\kappa_b - w)^2\ell^2/2$ for certain $w > 0$, and identify the eigenstates with $\langle P_{\kappa_b,w}\rangle > P_c$ for some $P_c \in [0,1]$ as momentum space edge states. Fig. 3(c) and (d) show the probability of a typical edge state and bulk state versus LL number n, respectively. By deleting the edge states, one can obtain a high-quality Hofstadter butterfly without spectral flows, as shown in Fig. 3(e) (where $w = 3.5\ell^{-1}\sqrt{\varphi + 0.5}$ and $P_c = 0.5$).

Quantized Lorentz susceptibility. The Chern number t_ν is known to give a quantized Hall conductance via the Kubo formula $\sigma_{xy} = i\partial_\omega \int d\omega' \langle G_{\omega+\omega'}\hat{j}_x G_{\omega'}\hat{j}_y\rangle|_{\omega\to0} = t_\nu\frac{e^2}{h}$ [10], where G_ω is the Green's function at energy ω, and $\hat{\boldsymbol{j}} = (\hat{j}_x, \hat{j}_y)$ is the uniform current operator. The duality between Eqs. (7) and (8) suggests that s_ν behaves as a dual Chern number for the momentum space, thus s_ν should also give a quantized response. Indeed, by noting that the natural momentum-space dual of the current operator $\hat{\boldsymbol{j}}$ is the force operator $\hat{\boldsymbol{F}} = e\mathbf{B} \times \frac{d\mathbf{R}}{dt} = (\hat{F}_x, \hat{F}_y)$, we find s_ν leads to a quantized *Lorentz susceptibility* (SM [18] Sec. S6)

$$\gamma_{xy} = -i\frac{\partial}{\partial\omega}\int d\omega' \langle G_{\omega+\omega'}\hat{F}_x G_{\omega'}\hat{F}_y\rangle\Big|_{\omega\to0} = eBs_\nu. \tag{15}$$

It yields a Lorentz force per unit cell $F_x = \gamma_{xy} v_y$ on the system when the lattice is moving at velocity v_y. Furthermore, a formula similar to the Thouless-Kohmoto-Nightingale-den Nijs formula [10] at flux per unit cell $\varphi = 2\pi p/q$ can be derived for s_ν (SM [18] Sec. S6B2):

$$s_\nu = -i \sum_{n \in \text{occ}} \int_{\boldsymbol{d} \in \Omega_M} \frac{\mathrm{d}^2 \boldsymbol{d}}{2\pi} \hat{\mathbf{z}} \cdot \langle \partial_{\boldsymbol{d}} w_{n,\boldsymbol{d}} | \times | \partial_{\boldsymbol{d}} w_{n,\boldsymbol{d}} \rangle \; , \quad (16)$$

where Ω_M is a torus with periods $\boldsymbol{d}_1$ and $\boldsymbol{d}_2/p$ serving as a "dual magnetic BZ", $|w_{n,\boldsymbol{d}}\rangle = e^{i\ell^{-2}(\hat{\mathbf{z}} \times \mathbf{R}) \cdot \boldsymbol{d}} |\psi_{n,\ell^{-2}\hat{\mathbf{z}} \times \boldsymbol{d}}\rangle$ (see the explicit form in SM [18] Sec. S6C) is defined using the Bloch eigenstates $|\psi_{n,\mathbf{k}}\rangle$ of band n, and n runs over all occupied bands.

Discussion. It is worth noting that the cutoffs in our method affect the resolution but not the shape of the Hofstadter butterfly. Our method greatly simplifies the matrix element construction compared to usual methods [1, 5], and require neither rational flux per unit cell nor large magnetic unit cells, making it easy to calculate the Hofstadter spectra of complicated models [12, 24–26].

Moreover, it leads to a sparse Hamiltonian for continuum models. At small magnetic fields, our method reduces to the LL calculations of $\mathbf{k} \cdot \mathbf{p}$ Hamiltonians expanded at center momentum $\mathbf{k}_0$. The large magnetic field spectrum is insensitive to the choice of $\mathbf{k}_0$.

[1] Douglas R. Hofstadter, "Energy levels and wave functions of bloch electrons in rational and irrational magnetic fields," Phys. Rev. B **14**, 2239–2249 (1976).

[2] Rafi Bistritzer and Allan H. MacDonald, "Moiré bands in twisted double-layer graphene," Proceedings of the National Academy of Sciences **108**, 12233–12237 (2011), http://www.pnas.org/content/108/30/12233.full.pdf.

[3] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, "Graphene bilayer with a twist: Electronic structure," Phys. Rev. Lett. **99**, 256802 (2007).

[4] E. J. Mele, "Commensuration and interlayer coherence in twisted bilayer graphene," Phys. Rev. B **81**, 161405 (2010).

[5] R. Bistritzer and A. H. MacDonald, "Moiré butterflies in twisted bilayer graphene," Phys. Rev. B **84**, 035440 (2011).

[6] S. Janecek, M. Aichinger, and E. R. Hernández, "Two-dimensional bloch electrons in perpendicular magnetic fields: An exact calculation of the hofstadter butterfly spectrum," Phys. Rev. B **87**, 235429 (2013).

[7] Godfrey Gumbs, Andrii Iurov, Danhong Huang, and Liubov Zhemchuzhna, "Revealing hofstadter spectrum for graphene in a periodic potential," Phys. Rev. B **89**, 241407 (2014).

[8] Kasra Hejazi, Chunxiao Liu, and Leon Balents, "Landau levels in twisted bilayer graphene and semiclassical orbits," Phys. Rev. B **100**, 035115 (2019).

[9] Ya-Hui Zhang, Hoi Chun Po, and T. Senthil, "Landau level degeneracy in twisted bilayer graphene: Role of symmetry breaking," Phys. Rev. B **100**, 125104 (2019).

[10] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, "Quantized hall conductance in a two-dimensional periodic potential," Phys. Rev. Lett. **49**, 405–408 (1982).

[11] Roland Winkler, *Spin-orbit coupling effects in two-dimensional electron and hole systems*, Springer tracts in modern physics (Springer, Berlin, 2003).

[12] Biao Lian, Fang Xie, and B. Andrei Bernevig, "Landau level of fragile topology," Phys. Rev. B **102**, 041402 (2020).

[13] János K. Asbóth and Andrea Alberti, "Spectral flow and global topology of the hofstadter butterfly," Phys. Rev. Lett. **118**, 216801 (2017).

[14] F. H. Claro and G. H. Wannier, "Magnetic subband structure of electrons in hexagonal lattices," Phys. Rev. B **19**, 6068–6074 (1979).

[15] I Dana, Y Avron, and J Zak, "Quantised hall conductance in a perfect crystal," Journal of Physics C: Solid State Physics **18**, L679–L683 (1985).

[16] Indubala I Satija, *Butterfly in the Quantum World*, 2053-2571 (Morgan & Claypool Publishers, 2016).

[17] Toshikaze Kariyado and Ashvin Vishwanath, "Flat band in twisted bilayer bravais lattices," Phys. Rev. Research **1**, 033076 (2019).

[18] See Supplemental Material for details, which includes additional references [27–31].

[19] G. H. Wannier, "A result not dependent on rationality for bloch electrons in a magnetic field," physica status solidi (b) **88**, 757–765 (1978), https://onlinelibrary.wiley.com/doi/pdf/10.1002/pssb.2220880243.

[20] P Streda, "Theory of quantised hall conductivity in two dimensions," Journal of Physics C: Solid State Physics **15**, L717 (1982).

[21] J. M. Luttinger, "The effect of a magnetic field on electrons in a periodic potential," Phys. Rev. **84**, 814–817 (1951).

[22] Takafumi Kita and Masao Arai, "Theory of interacting bloch electrons in a magnetic field," Journal of

the Physical Society of Japan **74**, 2813–2830 (2005), https://doi.org/10.1143/JPSJ.74.2813.

[23] A. Alexandradinata and Leonid Glazman, "Semiclassical theory of landau levels and magnetic breakdown in topological metals," Phys. Rev. B **97**, 144422 (2018).

[24] Jonah Herzog-Arbeitman, Zhi-Da Song, Nicolas Regnault, and B. Andrei Bernevig, "Hofstadter topology: Noncrystalline topological materials at high flux," Phys. Rev. Lett. **125**, 236804 (2020).

[25] Xiaobo Lu, Biao Lian, Gaurav Chaudhary, Benjamin A. Piot, Giulio Romagnoli, Kenji Watanabe, Takashi Taniguchi, Martino Poggio, Allan H. MacDonald, B. Andrei Bernevig, and Dmitri K. Efetov, "Multiple flat bands and topological hofstadter butterfly in twisted bilayer graphene close to the second magic angle," (2020), arXiv:2006.13963 [cond-mat.mes-hall].

[26] G. William Burg, Biao Lian, Takashi Taniguchi, Kenji Watanabe, B. Andrei Bernevig, and Emanuel Tutuc, "Evidence of emergent symmetry and valley chern number in twisted double-bilayer graphene," (2020), arXiv:2006.14000 [cond-mat.mes-hall].

[27] Aaron L. Sharpe, Eli J. Fox, Arthur W. Barnard, Joe Finney, Kenji Watanabe, Takashi Taniguchi, M. A. Kastner, and David Goldhaber-Gordon, "Emergent ferromagnetism near three-quarters filling in twisted bilayer graphene," Science **365**, 605–608 (2019), https://science.sciencemag.org/content/365/6453/605.full.pdf.

[28] M. Serlin, C. L. Tschirhart, H. Polshyn, Y. Zhang, J. Zhu, K. Watanabe, T. Taniguchi, L. Balents, and A. F. Young, "Intrinsic quantized anomalous hall effect in a moiré heterostructure," Science **367**, 900–903 (2020), https://science.sciencemag.org/content/367/6480/900.full.pdf.

[29] Zhida Song, Zhijun Wang, Wujun Shi, Gang Li, Chen Fang, and B. Andrei Bernevig, "All magic angles in twisted bilayer graphene are topological," Phys. Rev. Lett. **123**, 036401 (2019).

[30] B. ANDREI BERNEVIG and Taylor L. Hughes, *Topological Insulators and Topological Superconductors*, student edition ed. (Princeton University Press, 2013).

[31] Robert Karplus and J. M. Luttinger, "Hall effect in ferromagnetics," Phys. Rev. **95**, 1154–1160 (1954).

SUPPLEMENTARY MATERIAL

CONTENTS

I. THE ALGEBRA IN A MAGNETIC FIELD

We first review the algebra obeyed by 2-dimensional (2D) electrons in a uniform static magnetic field B in the continuum real space (in the first quantized language). We denote $\mathbf{r} = (x, y)$ as the position operator, and $-i\nabla = (-i\hbar\partial_x, -i\hbar\partial_y)$ as the canonical momentum operator.

Assume the magnetic field B corresponds to a gauge field $\mathbf{A}(\mathbf{r}) = (A_x(\mathbf{r}), A_y(\mathbf{r}))$, which satisfies $\partial_x A_y - \partial_y A_x = B$. We can then define the magnetic length $\ell = \sqrt{\hbar/eB}$, where e is the electron charge, and $\hbar$ is the Planck constant. The kinematic momentum operator of an electron in a magnetic field is given by $\mathbf{\Pi} = (\Pi_x, \Pi_y)$, which satisfies

$$\Pi_x = -i\partial_x - eA_x \ , \qquad \Pi_y = -i\partial_y - eA_y \ , \qquad [\Pi_x, \Pi_y] = \frac{i\hbar^2}{\ell^2} \ , \tag{17}$$

or in vector form $\mathbf{\Pi} = -i\nabla - \mathbf{A}(\mathbf{r})$. In the absence of the gauge field, $\mathbf{\Pi} = -i\nabla$ is the same as the canonical momentum. We also define the real space guiding center coordinates $\mathbf{R} = (R_x, R_y)$, which satisfy

$$R_x = x + \frac{\ell^2}{\hbar}\Pi_y \ , \qquad R_y = y - \frac{\ell^2}{\hbar}\Pi_x \ , \qquad [R_x, R_y] = -i\ell^2 \ . \tag{18}$$

It can be written in the vector form as $\mathbf{R} = \mathbf{r} - \frac{\ell^2}{\hbar}\hat{\mathbf{z}} \times \mathbf{\Pi}$. Semiclassically, the guiding center is the central position of the cyclotron motion of an electron in magnetic field B. The kinematic momentum operator $\mathbf{\Pi}$ commutes with the guiding center operator $\mathbf{R}$, namely, $[\Pi_x, R_x] = [\Pi_x, R_y] = [\Pi_y, R_x] = [\Pi_y, R_y] = 0$.

For convenience, hereafter we set $e = \hbar = 1$, unless recovery of the original units is needed. Besides, we always understand $\mathbf{r}$ as the position operator instead of a vector parameter, except that $|\mathbf{r}\rangle$ stands for a state at position $\mathbf{r}$ (where $\mathbf{r}$ is a parameter).

II. BASIS COMPLETENESS AND MATRIX ELEMENTS FOR CONTINUUM MODELS

In this section, we give the detailed derivation of the basis we choose and the Hamiltonian matrix elements for continuum models at zero magnetic field and nonzero magnetic field.

A. Continuum model in real space

We consider a continuum model with M intrinsic orbitals per zero-magnetic-field unit cell. For example, in the one-valley one-spin twisted bilayer graphene (TBG) continuum model in Ref. [2] (see also Sec. III), the intrinsic orbitals are graphene sublattice and layer indices. In more generic examples with spin-orbit coupling, intrinsic orbitals also include spin, etc. We denote the lattice Bravais vectors as $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$, and the reciprocal vectors as $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$, which satisfy

$$\boldsymbol{g}_i \cdot \boldsymbol{d}_j = 2\pi\delta_{ij} \ , \qquad (i, j = 1, 2). \tag{19}$$

We denote the reciprocal lattice as

$$\mathbf{Q} = m_1\boldsymbol{g}_1 + m_2\boldsymbol{g_2} \ , \qquad (m_1, m_2 \in \mathbb{Z}). \tag{20}$$

In the absence of magnetic field, the continuum model Hamiltonian in a continuum space with a lattice potential is of the generic form

$$\mathbf{H} = \int \mathrm{d}^2\mathbf{r} c_\alpha^\dagger(\mathbf{r})\widetilde{H}^{\alpha\beta}(\mathbf{r})c_\beta(\mathbf{r}) \ , \tag{21}$$

where

$$\widetilde{H}^{\alpha\beta}(\mathbf{r}) = \widetilde{\epsilon}^{\alpha\beta}(-i\nabla) + \sum_j \widetilde{V}_j^{\alpha\beta}e^{i\mathbf{q}_j\cdot\mathbf{r}} \ . \tag{22}$$

Here α, β denote the M intrinsic orbitals, $c_\alpha(\mathbf{r})$, $c_\alpha^\dagger(\mathbf{r})$ are the electron annihilation and creation operators of orbital α at position $\mathbf{r}$, $\widetilde{\epsilon}^{\alpha\beta}(-i\nabla)$ is the kinetic term in free space, $\widetilde{V}_j^{\alpha\beta}$ is the momentum $\mathbf{q}_j$ component of the lattice potential. The momentum $\mathbf{q}_j$ of the lattice potential component $\widetilde{V}_j^{\alpha\beta}$ satisfy

$$\mathbf{q}_j \in \mathbf{Q} , \tag{23}$$

where $\mathbf{Q}$ is the set of reciprocal lattice sites in Eq. (20); thus the lattice potential is periodic with lattice Bravais vectors $\boldsymbol{d}_1, \boldsymbol{d}_2$. Besides, the Hamiltonian is Hermitian, namely, $\widetilde{\epsilon}^{\alpha\beta}(-i\nabla) = \widetilde{\epsilon}^{\beta\alpha}(-i\nabla)^\dagger$, and $V_j^{\alpha\beta} = V_{\bar{j}}^{\beta\alpha*}$ for momenta $\mathbf{q}_{\bar{j}} = -\mathbf{q}_j$.

Here we note that $\mathbf{H}$ denotes the second quantized Hamiltonian, and $\widetilde{H}^{\alpha\beta}(\mathbf{r})$ denotes the first quantized single-particle Hamiltonian. The basis of the first quantized Hamiltonian $\widetilde{H}^{\alpha\beta}(\mathbf{r})$ is given by

$$c_\alpha^\dagger(\mathbf{r})|0\rangle , \tag{24}$$

with $|0\rangle$ being the vacuum state. Since we do not consider interactions, we can work in the first quantized single-particle Hamiltonian hereafter.

B. Continuum model in real space with orbital-dependent momentum origin shifts

In writing the Hamiltonian (22) above, the momentum origins of all orbitals are chosen at the Γ point of the BZ of the lattice. In some continuum models, it is convenient to shift the momentum origins of different orbitals α to some desired momenta $\mathbf{p}_\alpha$. This is done by transforming the single-particle Hamiltonian (22) from the real space basis in Eq. (24) into a new real space basis defined by

$$|\mathbf{r}, \alpha\rangle = e^{-i\mathbf{p}_\alpha \cdot \mathbf{r}} c_\alpha^\dagger(\mathbf{r})|0\rangle , \tag{25}$$

where $\mathbf{p}_\alpha$ is an orbital α dependent momentum vector which can be chosen freely. Here we shall restrict the choices of $\mathbf{p}_\alpha$ so that

$$\mathbf{p}_\alpha = \mathbf{p}_\beta \quad \text{if} \quad \widetilde{\epsilon}^{\alpha\beta}(-i\nabla) \neq 0 , \tag{26}$$

which ensures the kinetic term under the new basis (25) to be a function of $-i\nabla$ only and is independent of $\mathbf{r}$ (see Eq. (27)). An example of models with such orbital-dependent momentum origin shifts is the TBG continuum model originally written down in Ref. [2], where the orbitals α of the upper layer have $\mathbf{p}_\alpha = k_\theta(0, 1)$, and the orbitals β of the lower layer have $\mathbf{p}_\beta = k_\theta(0, -1)$ (see Sec. III for definition of k_θ and more details). Condition (26) is also satisfied for the TBG continuum model, since the kinetic term $\widetilde{\epsilon}^{\alpha\beta}(-i\nabla)$ between an orbital α in the upper layer and an orbital β in the lower layer is zero.

Under the assumption (26), we find the first quantized single-particle Hamiltonian transforms under the new basis (25) into

$$H^{\alpha\beta}(\mathbf{r}) = \langle \mathbf{r}, \alpha|\mathbf{H}|\mathbf{r}, \beta\rangle = e^{i\mathbf{p}_\alpha \cdot \mathbf{r}} \widetilde{H}^{\alpha\beta}(\mathbf{r}) e^{-i\mathbf{p}_\beta \cdot \mathbf{r}} = \epsilon^{\alpha\beta}(-i\nabla) + \sum_j V_j^{\alpha\beta} e^{i\mathbf{q}_j^{\alpha\beta} \cdot \mathbf{r}} , \tag{27}$$

where we have defined

$$\epsilon^{\alpha\beta}(-i\nabla) = \widetilde{\epsilon}^{\alpha\beta}(-i\nabla - \mathbf{p}_\alpha) , \qquad V_j^{\alpha\beta} = \widetilde{V}_j^{\alpha\beta} , \qquad \mathbf{q}_j^{\alpha\beta} = \mathbf{q}_j + \mathbf{p}_\alpha - \mathbf{p}_\beta \in \mathbf{Q}_\alpha - \mathbf{Q}_\beta \quad (\mathbf{q}_j \in \mathbf{Q}) , \tag{28}$$

and the momentum lattice $\mathbf{Q}_\alpha$ for orbital α is defined as

$$\mathbf{Q}_\alpha = \mathbf{p}_\alpha + \mathbf{Q} = \mathbf{p}_\alpha + m_1\boldsymbol{g}_1 + m_2\boldsymbol{g}_2 , \qquad (m_1, m_2 \in \mathbb{Z}) , \tag{29}$$

where $\mathbf{Q}$ is the reciprocal lattice sites in Eq. (20). In particular, we note that with the constraint (26), the transformed kinetic term $\epsilon^{\alpha\beta}(-i\nabla)$ in Eq. (28) under the new real space basis (25) is still only a function of $-i\nabla$, and does not depend on $\mathbf{r}$. In contrast, if we choose vectors $\mathbf{p}_\alpha$ which do not satisfy the constraint (26), after the transformation of Eq. (27), we would have a kinetic term $\epsilon^{\alpha\beta}(-i\nabla, \mathbf{r}) = \widetilde{\epsilon}^{\alpha\beta}(-i\nabla - \mathbf{p}_\alpha)e^{i(\mathbf{p}_\alpha - \mathbf{p}_\beta)\cdot\mathbf{r}}$ that is position $\mathbf{r}$ dependent, which brings unnecessary complications. Therefore, we will impose the constraint (26).

We note that the single-particle Hamiltonian in Eq. (27) is generically invariant up to a unitary transformation under the translation of a Bravais lattice vector $\boldsymbol{d}_i$. Under the real space basis $|\mathbf{r}, \alpha\rangle$ in Eq. (25), if we define the

translation operator of distance $\boldsymbol{d}$ which satisfies $T_{\boldsymbol{d}}|\mathbf{r}, \alpha\rangle = |\mathbf{r}+\boldsymbol{d}, \alpha\rangle$, its real space representation in the first quantized language is given by $T_{\boldsymbol{d}} = e^{-\boldsymbol{d}\cdot\nabla}$, and we have

$$H^{\alpha\beta}(\mathbf{r} + \boldsymbol{d}_i) = T_{\boldsymbol{d}_i}^{-1} H^{\alpha\beta}(\mathbf{r}) T_{\boldsymbol{d}_i} = e^{\boldsymbol{d}\cdot\nabla} \left[\epsilon^{\alpha\beta}(-i\nabla) + \sum_j V_j^{\alpha\beta} e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}} \right] e^{-\boldsymbol{d}\cdot\nabla} = e^{i\mathbf{p}_\alpha\cdot\boldsymbol{d}_i} H^{\alpha\beta}(\mathbf{r}) e^{-i\mathbf{p}_\beta\cdot\boldsymbol{d}_i} \ , \qquad (30)$$

where $\boldsymbol{d}_i$ $(i = 1, 2)$ is a Bravais lattice vector, and we have used the assumption (26), and the relation (19). It is clear that if we set all $\mathbf{p}_\alpha = 0$, we would have the usual translational invariance $H^{\alpha\beta}(\mathbf{r} + \boldsymbol{d}_i) = H^{\alpha\beta}(\mathbf{r})$ which does not involve a unitary transformation.

For generality, we shall use the first quantized continuum model single-particle Hamiltonian in Eq. (27), which has orbital-dependent momentum origin shifts $\mathbf{p}_\alpha$, and the real space basis is defined in Eq. (25). We note that one can always choose all $\mathbf{p}_\alpha = 0$, in which case the form of the continuum model Hamiltonian reduces back to Eq. (22).

C. Transforming the model at zero magnetic field into momentum space

Eq. (27) gives the single-particle Hamiltonian in the absence of magnetic field. It can be written in the momentum space by Fourier transformation. To do this, we define the momentum space basis

$$|\mathbf{k}, \mathbf{Q}_\alpha, \alpha\rangle = \frac{1}{\sqrt{\Omega_{\text{tot}}}} \int \mathrm{d}^2\mathbf{r} e^{i(\mathbf{k}+\mathbf{Q}_\alpha)\cdot\mathbf{r}} |\mathbf{r}, \alpha\rangle \ , \qquad (31)$$

where $\mathbf{k}$ is the quasimomentum in the first Brillouin zone (BZ), $\mathbf{Q}_\alpha$ for intrinsic orbital α is defined in Eq. (29), and Ω_{tot} is the total area of the system. Under this momentum space basis $|\mathbf{k}, \mathbf{Q}_\alpha, \alpha\rangle$, the single-particle Hamiltonian $H^{\alpha\beta}(\mathbf{r})$ in Eq. (27) transforms into the following form diagonal in $\mathbf{k}$ (here $\mathbf{k}, \mathbf{k}'$ are in the first BZ):

$$H_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta}^{\alpha\beta}(\mathbf{k}', \mathbf{k}) = \langle \mathbf{k}', \mathbf{Q}'_\alpha, \alpha | \mathbf{H} | \mathbf{k}, \mathbf{Q}_\beta, \beta \rangle = \frac{1}{\Omega_{\text{tot}}} \int \mathrm{d}^2\mathbf{r}\mathrm{d}^2\mathbf{r}' e^{-i(\mathbf{k}'+\mathbf{Q}'_\alpha)\cdot\mathbf{r}} \langle \mathbf{r}, \alpha | \mathbf{H} | \mathbf{r}', \beta \rangle e^{i(\mathbf{k}+\mathbf{Q}_\beta)\cdot\mathbf{r}'}$$

$$= \frac{1}{\Omega_{\text{tot}}} \int \mathrm{d}^2\mathbf{r} e^{-i(\mathbf{k}'+\mathbf{Q}'_\alpha)\cdot\mathbf{r}} H^{\alpha\beta}(\mathbf{r}) e^{i(\mathbf{k}+\mathbf{Q}_\beta)\cdot\mathbf{r}} = \frac{1}{\Omega_{\text{tot}}} \int \mathrm{d}^2\mathbf{r} e^{-i(\mathbf{k}'+\mathbf{Q}'_\alpha)\cdot\mathbf{r}} \left[\epsilon^{\alpha\beta}(-i\nabla) + \sum_j V_j^{\alpha\beta} e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}} \right] e^{i(\mathbf{k}+\mathbf{Q}_\beta)\cdot\mathbf{r}} \qquad (32)$$

$$= \delta_{\mathbf{k},\mathbf{k}'} \left[\epsilon^{\alpha\beta}(\mathbf{k} + \mathbf{Q}_\beta) \delta_{\mathbf{Q}'_\alpha, \mathbf{Q}_\beta} + \sum_j V_j^{\alpha\beta} \delta_{\mathbf{Q}'_\alpha, \mathbf{Q}_\beta+\mathbf{q}_j^{\alpha\beta}} \right] = \delta_{\mathbf{k},\mathbf{k}'} H_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta}^{\alpha\beta}(\mathbf{k}) \ ,$$

where we have used the definition of $\mathbf{q}_j^{\alpha\beta}$ in Eq. (28), and the condition (26). The Hamiltonian diagonal in $\mathbf{k}$

$$H_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta}^{\alpha\beta}(\mathbf{k}) = \epsilon^{\alpha\beta}(\mathbf{k} + \mathbf{Q}_\beta) \delta_{\mathbf{Q}'_\alpha, \mathbf{Q}_\beta} + \sum_j V_j^{\alpha\beta} \delta_{\mathbf{Q}'_\alpha, \mathbf{Q}_\beta+\mathbf{q}_j^{\alpha\beta}} \qquad (33)$$

then gives Eq. (2) in the main text.

D. The Basis and Hamiltonian in nonzero magnetic field

Under *a uniform magnetic field* $\mathbf{B} = B\hat{\mathbf{z}}$, the canonical momentum $-i\nabla$ in Eq. (27) will be replaced by the kinetic momentum $\boldsymbol{\Pi} = -i\nabla - \mathbf{A}(\mathbf{r})$, where $\nabla \times \mathbf{A}(\mathbf{r}) = \mathbf{B}$. The (first quantized) single-particle Hamiltonian then reads

$$H^{\alpha\beta}(\mathbf{r}) = \epsilon^{\alpha\beta}(\boldsymbol{\Pi}) + \sum_j V_j^{\alpha\beta} e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}} \ . \qquad (34)$$

In the below, we will construct a basis under magnetic field, in which we show that the Hamiltonian in magnetic field is block diagonalized into blocks with identical matrix elements (different blocks differ by guiding center translations in the real space, see explanations below Eq. (55)), and each block is simply given by Eq. (33) (the momentum space Hamiltonian at zero magnetic field) with the substitution

$$\mathbf{k} \rightarrow \hat{\boldsymbol{\kappa}} + \mathbf{k}_0 \ , \qquad \hat{\boldsymbol{\kappa}} = \frac{1}{\sqrt{2}\ell}(a + a^\dagger, -ia + ia^\dagger) \ , \qquad (35)$$

where $\mathbf{k}_0$ is an arbitrarily chosen momentum vector, a and $a^\dagger$ are lowering and raising operators satisfying $[a, a^\dagger] = 1$ which we will define below, and ℓ is the magnetic length.

To begin, we would like to find a set of mutually commuting operators to define the quantum numbers of our basis. First, we define a set of Landau level (LL) lowering and raising operators associated with the momentum lattice sites $\mathbf{Q}_\alpha = (Q_{\alpha,x}, Q_{\alpha,y})$ of orbital α as

$$a_{\mathbf{Q}_\alpha} = \frac{\ell}{\sqrt{2}}[\Pi_x - Q_{\alpha,x} - k_{0,x} + i(\Pi_y - Q_{\alpha,y} - k_{0,y})] , \quad a^\dagger_{\mathbf{Q}_\alpha} = \frac{\ell}{\sqrt{2}}[\Pi_x - Q_{\alpha,x} - k_{0,x} - i(\Pi_y - Q_{\alpha,y} - k_{0,y})] , \quad (36)$$

where $\mathbf{k}_0 = (k_{0,x}, k_{0,y})$ is a freely chosen fixed momentum which we call the center momentum. They satisfy the commutation relation $[a_{\mathbf{Q}_\alpha}, a^\dagger_{\mathbf{Q}_\alpha}] = 1$. We note that for two different sites $\mathbf{Q}_\alpha$ and $\mathbf{Q}'_\alpha$, the operators $a_{\mathbf{Q}_\alpha}$ and $a_{\mathbf{Q}'_\alpha}$ only differ by a constant shift, thus are linearly dependent. We also note that $[a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}, a^\dagger_{\mathbf{Q}'_\alpha} a_{\mathbf{Q}'_\alpha}] \neq 0$ for $\mathbf{Q}_\alpha \neq \mathbf{Q}'_\alpha$. The eigenvalue of $a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}$ for a given $\mathbf{Q}$ runs over all the nonnegative integers.

Secondly, we define (recall that $\mathbf{R} = \mathbf{r} - \frac{\ell^2}{\hbar}\hat{\mathbf{z}} \times \mathbf{\Pi}$)

$$R_{\hat{\boldsymbol{\tau}}} = \mathbf{R} \cdot \hat{\boldsymbol{\tau}} = R_x \hat{\tau}_x + R_y \hat{\tau}_y \qquad (37)$$

as the guiding center along the $\hat{\boldsymbol{\tau}}$ direction, where $\hat{\boldsymbol{\tau}} = (\hat{\tau}_x, \hat{\tau}_y)$ is a unit vector so chosen that $\frac{\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)}{\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)}$ is an irrational number (the reason will be explained below Eq. (41)). In an infinite real space (which we assume is the case here), the eigenvalue of $R_{\hat{\boldsymbol{\tau}}}$ runs over all real numbers $\mathbb{R}$. This can be seen from the fact that one can shift the eigenvalue of $R_{\hat{\boldsymbol{\tau}}}$ by any value $b \in \mathbb{R}$ using the operator $e^{ib\mathbf{R}\cdot(\hat{\boldsymbol{\tau}}\times\hat{\mathbf{z}})/\ell^2}$, namely,

$$e^{ib\mathbf{R}\cdot(\hat{\boldsymbol{\tau}}\times\hat{\mathbf{z}})/\ell^2} R_{\hat{\boldsymbol{\tau}}} e^{-ib\mathbf{R}\cdot(\hat{\boldsymbol{\tau}}\times\hat{\mathbf{z}})/\ell^2} = R_{\hat{\boldsymbol{\tau}}} + b , \qquad (b \in \mathbb{R}) \qquad (38)$$

We note that once $R_{\hat{\boldsymbol{\tau}}}$ is diagonalized for a given direction $\hat{\boldsymbol{\tau}}$, one cannot further diagonalize the guiding center coordinate $R_{\hat{\boldsymbol{\tau}}'} = \mathbf{R} \cdot \hat{\boldsymbol{\tau}}'$ along any other direction $\hat{\boldsymbol{\tau}}' \neq \pm\hat{\boldsymbol{\tau}}$, since $[R_{\hat{\boldsymbol{\tau}}'}, R_{\hat{\boldsymbol{\tau}}}] = -i\ell^2 \hat{\mathbf{z}} \cdot (\hat{\boldsymbol{\tau}}' \times \hat{\boldsymbol{\tau}}) \neq 0$ unless $\hat{\boldsymbol{\tau}}' = \pm\hat{\boldsymbol{\tau}}$ (Note that $R_{\hat{\boldsymbol{\tau}}} = -R_{-\hat{\boldsymbol{\tau}}}$, so diagonalization of $R_{\hat{\boldsymbol{\tau}}}$ is sufficient).

Since $[\mathbf{R}, \mathbf{\Pi}] = 0$, it is easy to see that

$$[R_{\hat{\boldsymbol{\tau}}}, a_{\mathbf{Q}_\alpha}] = [R_{\hat{\boldsymbol{\tau}}}, a^\dagger_{\mathbf{Q}_\alpha}] = [R_{\hat{\boldsymbol{\tau}}}, a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}] = 0 \qquad (39)$$

for any $\mathbf{Q}_\alpha$. Therefore, $R_{\hat{\boldsymbol{\tau}}}$ and $a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}$ (for a given $\mathbf{Q}_\alpha$) can be diagonalized simultaneously. For a given $\mathbf{Q}_\alpha$, there is no other functions of $\mathbf{R}, \mathbf{\Pi}$ independent of $R_{\hat{\boldsymbol{\tau}}}$ and $a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}$ which commute with both $R_{\hat{\boldsymbol{\tau}}}$ and $a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}$, so $R_{\hat{\boldsymbol{\tau}}}$ and $a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}$ form the maximal commuting set of operators $\mathbf{R}, \mathbf{\Pi}$. We shall therefore use $R_{\hat{\boldsymbol{\tau}}}$ and $a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}$ to define the quantum numbers of our basis.

In the previous Hofstadter method for continuum models [5], the chosen basis makes the lattice potential term $e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}}$ in Eq. (34) a complicated dense matrix. We hope to find a different basis where the operator $e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}}$ have simple matrix elements, so that the Hamiltonian takes a much simpler form. Now we describe the construction of such a basis (in which the operator $e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}}$ has very simple matrix elements, see Eq. (48)). For each momentum site $\mathbf{Q}_\alpha$, we define a set of basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ satisfying

$$R_{\hat{\boldsymbol{\tau}}}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = [\lambda - \ell^2\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \mathbf{Q}_\alpha)]|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle , \qquad a^\dagger_{\mathbf{Q}_\alpha} a_{\mathbf{Q}_\alpha}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = n|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle , \qquad (40)$$

where $n \geq 0$ is a nonnegative integer (Landau level number), λ is a real number, and α is the intrinsic orbital index. Note that we have defined the eigenvalue of $R_{\hat{\boldsymbol{\tau}}}$ in a $\mathbf{Q}_\alpha$ dependent way (we could have defined the $R_{\hat{\boldsymbol{\tau}}}$ eigenvalue without $\ell^2\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \mathbf{Q}_\alpha)$, which would be equivalent to the basis used in usual method [5]). One advantage of such a definition is that, the operator $e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}}$ will not change the number λ when acting on the basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$, as we will prove in Eq. (48). Another advantage of such a definition in Eq. (40) is that, for a fixed number λ, the basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ of all quantum numbers α, $\mathbf{Q}_\alpha$ and n are orthonormal. To see this, consider two momentum sites $\mathbf{Q}_\alpha, \mathbf{Q}'_\alpha \in \mathbf{p}_\alpha + \boldsymbol{g}_1\mathbb{Z} + \boldsymbol{g}_2\mathbb{Z}$ (as defined in Eq. (29)), and assume $\mathbf{Q}_\alpha - \mathbf{Q}'_\alpha = m_1\boldsymbol{g}_1 + m_2\boldsymbol{g}_2$. If two states $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ and $|\lambda, \mathbf{Q}'_\alpha, n', \alpha\rangle$ have equal $R_{\hat{\boldsymbol{\tau}}}$ eigenvalues, namely,

$$\lambda - \ell^2\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \mathbf{Q}_\alpha) = \lambda - \ell^2\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \mathbf{Q}'_\alpha) \quad \rightarrow \quad m_1\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1) + m_2\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2) = 0 , \qquad (41)$$

we must have $m_1 = m_2 = 0$, namely, $\mathbf{Q}'_\alpha = \mathbf{Q}_\alpha$, since $\frac{\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)}{\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)}$ is an irrational number. Vice versa, if $\mathbf{Q}'_\alpha \neq \mathbf{Q}_\alpha$, the two states $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ and $|\lambda, \mathbf{Q}'_\alpha, n', \alpha\rangle$ will have different $R_{\hat{\boldsymbol{\tau}}}$ eigenvalues and will be orthogonal to each other. Therefore, the basis with a fixed number λ satisfies the orthonormal relation:

$$\langle\lambda, \mathbf{Q}'_\beta, n', \beta|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = \delta_{\beta\alpha}\delta_{\mathbf{Q}'_\beta\mathbf{Q}_\alpha}\delta_{n'n} . \qquad (42)$$

We note that Eq. (42) would not hold if we define $R_{\hat{\tau}}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = \lambda|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ instead, as two different $\mathbf{Q}$'s would have the same eigenvalue λ.

As we will show later below Eq. (48), the way of defining the basis in Eq. (40) will greatly simplify the matrix elements of Hamiltonian (34). However, before we move on, we note that the set of basis with a fixed number λ (satisfying Eq. (42)) is not a complete basis for the Hamiltonian (34), since for a fixed λ, the $R_{\hat{\tau}}$ eigenvalue $\lambda - \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \mathbf{Q}_\alpha)$ (of all momentum sites $\mathbf{Q}_\alpha$) does not run over the entire real number set $\mathbb{R}$ (the complete set of $R_{\hat{\tau}}$ eigenvalues is $\mathbb{R}$, see the argument above Eq. (38)). Therefore, we need to allow λ to run over a certain set to make the basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ a complete basis. We now prove this can be done by assuming λ runs over the following quotient set:

$$\lambda \in \Lambda_{\hat{\tau}} = \mathbb{R}/[\ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)\mathbb{Z} + \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)\mathbb{Z}] , \tag{43}$$

namely, each number λ labels a coset $\{\lambda + \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)\mathbb{Z} + \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)\mathbb{Z}\}$ of the subgroup $\ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)\mathbb{Z} + \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)\mathbb{Z}$ in the real number group $\mathbb{R}$. Two numbers λ and λ' label the same coset if they belongs to the same coset. For example, assume we have $\ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1) = 1$ and $\ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2) = \sqrt{2}$, then the quantum number $\lambda = 0$ will label the coset $\{\mathbb{Z} + \sqrt{2}\mathbb{Z}\}$. Accordingly, $\lambda = 0$ is identical to $\lambda = 1$ and $\lambda = \sqrt{2}$, etc. In contrast, $\lambda = 0$ and $\lambda = \sqrt{3}$ are not identical, which belong to different cosets. For definiteness, we will pick a fixed element λ in each coset to represent the coset, so each coset is represented by a definite number λ. For example, we can choose to use the definite number $\lambda = 0$ (instead of 1, $\sqrt{2}$, etc) to represent the coset $\{\mathbb{Z} + \sqrt{2}\mathbb{Z}\}$. In this way, we can represent each element (coset) in the set $\Lambda_{\hat{\tau}}$ in Eq. (43) by a definite number λ.

With the set of representative numbers λ given in Eq. (43), we now show that the basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ forms a complete orthonormal basis. Note that for each λ in Eq. (43), the $R_{\hat{\tau}}$ eigenvalue $\lambda - \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \mathbf{Q}_\alpha)$ of all $\mathbf{Q}_\alpha$ runs over all the numbers in the coset $\{\lambda - \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \mathbf{p}_\alpha) + \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)\mathbb{Z} + \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)\mathbb{Z}\}$, where $\mathbf{p}_\alpha$ is the momentum origin shift of orbital α defined in Eq. (25), and recall that $\mathbf{Q}_\alpha = \mathbf{p}_\alpha + \boldsymbol{g}_1\mathbb{Z} + \boldsymbol{g}_2\mathbb{Z}$ as defined in Eq. (29). Therefore, the $R_{\hat{\tau}}$ eigenvalue of $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ of all λ in Eq. (43) and all $\mathbf{Q}_\alpha$ runs over all the real numbers $\mathbb{R}$ (since λ is a continuous variable in the quotient set (43)). In particular, two orbital α states $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ and $|\lambda', \mathbf{Q}'_\alpha, n', \alpha\rangle$ will have their $R_{\hat{\tau}}$ eigenvalues in different cosets if $\lambda \neq \lambda'$ ($\lambda, \lambda' \in \Lambda_{\hat{\tau}}$ in Eq. (43), note that we have chosen a unique definite number λ in each coset to represent the coset in Eq. (43)). Therefore, we have the orthonormal relation

$$\langle \lambda', \mathbf{Q}'_\beta, n', \beta | \lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = \delta_{\beta\alpha}\delta_{\lambda\lambda'}\delta_{\mathbf{Q}'_\beta \mathbf{Q}_\alpha}\delta_{n'n} . \tag{44}$$

Thus, the basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ with λ defined in Eq. (43) forms a complete orthonormal basis of Hamiltonian (40).

We now show that the most important advantage of the basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ satisfying Eq. (44) is, the Hamiltonian (34) is diagonal in λ, and its matrix elements are independent of λ. To see this, we first note that from the commutation relations in Sec. (I), we have

$$e^{-i\mathbf{q}\cdot\mathbf{r}}R_{\hat{\tau}}e^{i\mathbf{q}\cdot\mathbf{r}} = R_{\hat{\tau}} - \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \mathbf{q}) , \qquad e^{-i\mathbf{q}\cdot\mathbf{r}}a_{\mathbf{Q}_\alpha}e^{i\mathbf{q}\cdot\mathbf{r}} = a_{\mathbf{Q}_\alpha} + \frac{\ell}{\sqrt{2}}(q_x + iq_y) = a_{\mathbf{Q}_\alpha - \mathbf{q}} , \tag{45}$$

where $\mathbf{r}$ is the position operator, and $\mathbf{q}$ is an arbitrary momentum vector. Therefore, we have

$$R_{\hat{\tau}}e^{i\mathbf{q}\cdot\mathbf{r}}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = e^{i\mathbf{q}\cdot\mathbf{r}}(R_{\hat{\tau}} - \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times \mathbf{q}))|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = [\lambda - \ell^2 \hat{\tau} \cdot (\hat{\mathbf{z}} \times (\mathbf{Q}_\alpha + \mathbf{q}))]e^{i\mathbf{q}\cdot\mathbf{r}}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle \tag{46}$$

$$a^\dagger_{\mathbf{Q}_\alpha + \mathbf{q}}a_{\mathbf{Q}_\alpha + \mathbf{q}}e^{i\mathbf{q}\cdot\mathbf{r}}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = e^{i\mathbf{q}\cdot\mathbf{r}}a^\dagger_{\mathbf{Q}_\alpha}a_{\mathbf{Q}_\alpha}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = ne^{i\mathbf{q}\cdot\mathbf{r}}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle , \tag{47}$$

which indicates

$$e^{i\mathbf{q}\cdot\mathbf{r}}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = |\lambda, \mathbf{Q}_\alpha + \mathbf{q}, n, \alpha\rangle , \tag{48}$$

where $\mathbf{r}$ is the position operator (understood as $\sum_\alpha \int \mathbf{r}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha|d^2\mathbf{r}$ when expanded in position basis), and it also plays the role of the generator of momentum translations.

We then examine the matrix elements of Hamiltonian (34). Using Eq. (48), we find the lattice potential term $V_j^{\alpha\beta}e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}}$ in the Hamiltonian (34) has matrix elements

$$\langle\lambda', \mathbf{Q}'_\alpha, m, \alpha|V_j^{\alpha\beta}e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}}|\lambda, \mathbf{Q}_\beta, n, \beta\rangle = V_j^{\alpha\beta}\langle\lambda', \mathbf{Q}'_\alpha, m, \alpha|\lambda, \mathbf{Q}_\beta + \mathbf{q}_j^{\alpha\beta}, n, \beta\rangle = V_j^{\alpha\beta}\delta_{\lambda'\lambda}\delta_{\mathbf{Q}'_\alpha, \mathbf{Q}_\beta + \mathbf{q}_j^{\alpha\beta}}\delta_{mn} , \tag{49}$$

where $\mathbf{r}$ on the left hand side is understood as the position operator (instead of a number). Therefore, the lattice hopping potential term $V_j^{\alpha\beta}e^{i\mathbf{q}_j^{\alpha\beta}\cdot\mathbf{r}}$ is diagonal in quantum numbers λ and n when acting on basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$, while changes the orbital from β to α, and shifts the reciprocal momentum $\mathbf{Q}_\beta$ to $\mathbf{Q}_\beta + \mathbf{q}_j^{\alpha\beta}$.

As for the kinetic term $\epsilon(\boldsymbol{\Pi})$, we first note that the kinetic momentum $\boldsymbol{\Pi}$ commutes with $R_{\hat{\boldsymbol{\tau}}}$ and does not change the $R_{\hat{\boldsymbol{\tau}}}$ eigenvalue (which has one-to-one correspondence with the pair of quantum numbers λ and $\mathbf{Q}_\alpha$). Therefore, the matrix elements of $\epsilon(\boldsymbol{\Pi})$ has to be diagonal in both λ and $\mathbf{Q}_\alpha$. Besides, for each orbital α, we can rewrite the kinetic momentum as $\boldsymbol{\Pi} = \hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha} + \mathbf{k}_0 + \mathbf{Q}_\alpha$, where we have defined a $\mathbf{Q}_\alpha$ dependent operator

$$\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha} = \frac{1}{\sqrt{2}\ell}(a_{\mathbf{Q}_\alpha} + a_{\mathbf{Q}_\alpha}^\dagger, -ia_{\mathbf{Q}_\alpha} + ia_{\mathbf{Q}_\alpha}^\dagger) \ . \tag{50}$$

Note that $\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha}$ only acts on the quantum number n of basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$, which obeys the following rules:

$$a_{\mathbf{Q}_\alpha}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = \sqrt{n}|\lambda, \mathbf{Q}_\alpha, n-1, \alpha\rangle \ , \qquad a_{\mathbf{Q}_\alpha}^\dagger|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = \sqrt{n+1}|\lambda, \mathbf{Q}_\alpha, n+1, \alpha\rangle \ . \tag{51}$$

Therefore, it is easy to find that

$$\begin{aligned}
\langle\lambda', \mathbf{Q}_\beta', m, \beta|\epsilon(\boldsymbol{\Pi})|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle &= \delta_{\lambda'\lambda}\delta_{\mathbf{Q}_\beta', \mathbf{Q}_\alpha}\langle\lambda', \mathbf{Q}_\beta', m, \beta|\epsilon(\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha} + \mathbf{k}_0 + \mathbf{Q}_\alpha)|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle \\
&= \delta_{\lambda'\lambda}\delta_{\mathbf{Q}_\beta', \mathbf{Q}_\alpha}[\epsilon^{\beta\alpha}(\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha} + \mathbf{k}_0 + \mathbf{Q}_\alpha)]_{mn} \ .
\end{aligned} \tag{52}$$

Mathematically, we can define an operator

$$\hat{\boldsymbol{\kappa}} = \frac{1}{\sqrt{2}\ell}(a + a^\dagger, -ia + ia^\dagger) \tag{53}$$

without the $\mathbf{Q}_\alpha$ subindex, where a and $a^\dagger$ are some lowering and raising operators satisfying $[a, a^\dagger] = 1$ (which need not have any relation with $a_{\mathbf{Q}_\alpha}$ and $a_{\mathbf{Q}_\alpha}^\dagger$). It is then easy to see that the matrix element in Eq. (52) is mathematically equal to

$$[\epsilon^{\beta\alpha}(\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha} + \mathbf{k}_0 + \mathbf{Q}_\alpha)]_{mn} = [\epsilon^{\beta\alpha}(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\alpha)]_{mn} = \langle m|\epsilon^{\beta\alpha}(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\alpha)|n\rangle \ , \tag{54}$$

where $|n\rangle$ is the basis of a and $a^\dagger$ defined by $a|n\rangle = \sqrt{n}|n-1\rangle$ and $a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$. Therefore, mathematically one could replace $\hat{\boldsymbol{\kappa}}_{\mathbf{Q}_\alpha}$ by $\hat{\boldsymbol{\kappa}}$ without ambiguity, provided that one remembers that $\hat{\boldsymbol{\kappa}}$ acts on the quantum number n.

From Eqs. (49), (52) and (54), it is easy to see that the matrix elements of both the kinetic term and the lattice potential term of Hamiltonian (34) are diagonal in λ and independent of λ. Therefore, we can divide the entire Hilbert space into subspaces with different (continuous) quantum number λ; the energy spectra of all the λ subspaces are the same. Within the subspace of a fixed λ, the Hamiltonian matrix element from basis $|\lambda, \mathbf{Q}_\beta, n, \beta\rangle$ to $|\lambda, \mathbf{Q}_\alpha', m, \alpha\rangle$ can be written as

$$H_{\mathbf{Q}_\alpha'\mathbf{Q}_\beta, mn}^{\lambda, \alpha\beta} = \left[\epsilon^{\alpha\beta}(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\beta)\right]_{mn}\delta_{\mathbf{Q}_\alpha', \mathbf{Q}_\beta} + \sum_j V_j^{\alpha\beta}\delta_{\mathbf{Q}_\alpha', \mathbf{Q}_\beta + \mathbf{q}_j^{\alpha\beta}}\delta_{mn} \ , \tag{55}$$

where $[\epsilon^{\alpha\beta}(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\beta)]_{mn}$ is defined in Eq. (54). This is nothing but the zero magnetic field momentum space Hamiltonian (33) with the substitution $\mathbf{k} \to \hat{\boldsymbol{\kappa}} + \mathbf{k}_0$ (given also in Eq. (5) of the main text). It is then sufficient to compute the spectrum within just one fixed λ subspace.

We note that different λ sectors have different eigenstate wave functions, although they have identical Hamiltonian matrix elements independent of λ as given by Eq. (55). To see this explicitly, if we take a Landau gauge perpendicular to the $\hat{\boldsymbol{\tau}}$ direction, $\mathbf{A} = B(\mathbf{r} \cdot \hat{\boldsymbol{\tau}})(\hat{\mathbf{z}} \times \hat{\boldsymbol{\tau}})$, we have

$$\langle\mathbf{r}, \alpha|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle = e^{i[\lambda - \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\mathbf{Q}_\alpha)][\mathbf{r}\cdot(\hat{\mathbf{z}}\times\hat{\boldsymbol{\tau}})]/\ell - [\mathbf{r}\cdot\hat{\boldsymbol{\tau}} - \lambda + \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\mathbf{Q}_\alpha)]^2/2\ell^2}h_n\left(\ell^{-1}[\mathbf{r}\cdot\hat{\boldsymbol{\tau}} - \lambda + \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\mathbf{Q}_\alpha)]\right), \tag{56}$$

where $h_n(x)$ is the n-th Hermite polynomial. Therefore, one can see explicitly that the wave functions $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ at different λ have different guiding centers. Assume the subspace Hamiltonian $H_{\mathbf{Q}_\alpha'\mathbf{Q}_\beta, mn}^{\lambda, \alpha\beta}$ in sector λ has an eigenstate $|\lambda, u\rangle = \sum_{\alpha, \mathbf{Q}_\alpha, n} u_{\alpha, \mathbf{Q}_\alpha, n}|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ at energy energy E_u, where the coefficients $u_{\alpha, \mathbf{Q}_\alpha, n}$ are independent of λ. Then the subspace Hamiltonian $H_{\mathbf{Q}_\alpha'\mathbf{Q}_\beta, mn}^{\lambda', \alpha\beta}$ in sector λ' will have an eigenstate $|\lambda', u\rangle = \sum_{\alpha, \mathbf{Q}_\alpha, n} u_{\alpha, \mathbf{Q}_\alpha, n}|\lambda', \mathbf{Q}_\alpha, n, \alpha\rangle$ at the same energy E_u. However, the $\hat{\boldsymbol{\tau}}$ direction guiding center coordinate $R_{\hat{\boldsymbol{\tau}}}$ of the two wave functions $|\lambda, u\rangle$ and $|\lambda', u\rangle$ will differ by $\lambda' - \lambda$, namely, the central position of the two wave functions are different.

We also note that, if there is a disorder potential that breaks the periodicity of the lattice model, e.g., a potential term $\delta\widetilde{V}^{\alpha\beta}e^{i\widetilde{\mathbf{q}}\cdot\mathbf{r}}$ from orbitals β to α with $\widetilde{\mathbf{q}} \neq \mathbf{Q} + \mathbf{p}_\alpha - \mathbf{p}_\beta$ for any reciprocal vector $\mathbf{Q}$, this term will couple different λ sectors, and the Hamiltonian will no longer be diagonal in λ and independent of λ. By Eq. (48) and Eq. (40), such a term $\delta\widetilde{V}^{\alpha\beta}e^{i\widetilde{\mathbf{q}}\cdot\mathbf{r}}$ will couple the sector of coset labeled by λ (defined in Eq. (43)) with the sector of coset of $\lambda - \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}} \times [\widetilde{\mathbf{q}} + \mathbf{p}_\beta - \mathbf{p}_\alpha])$, which does not live in the same coset. In this paper, we shall not consider any disorder potential breaking the periodicity of the lattice model.

In particular, if the kinetic energy $\epsilon(\mathbf{k})$ is a polynomial of $\mathbf{k}$ up to power Δ ($\Delta \in \mathbb{Z}_+$), the matrix element in Eq. (55) has to be zero for $|m - n| > \Delta$. In this case, the Hamiltonian (55) is sparse.

E. Numerical Hofstadter calculations: the cutoffs and the momentum space boundary

In numerical calculations, one needs to take a cutoff in the reciprocal lattice at a boundary enclosing an area of N_Q BZs, and a cutoff in the LL quantum number $n \leq N_L$. The Hamiltonian (55) is then a matrix of size $M N_L N_Q$. We now explain how the cutoffs N_Q and N_L set a momentum space boundary for Hamiltonian (55).

For concreteness, assume the cutoff of the reciprocal lattice encloses a circular momentum space area $N_Q \Omega_{BZ}$ centered at the center momentum $\mathbf{k}_0$, where $\Omega_{BZ} = |\boldsymbol{g}_1 \times \boldsymbol{g}_2| = 4\pi^2/|\boldsymbol{d}_1 \times \boldsymbol{d}_2|$ is the BZ area. This restricts the reciprocal momentum sites within a momentum radius

$$|\mathbf{k}_0 + \mathbf{Q}_\alpha| \leq \sqrt{N_Q \Omega_{BZ}/\pi} \ . \tag{57}$$

In addition, by Eq. (53), with the LL cutoff N_L, for any states we have

$$|\hat{\boldsymbol{\kappa}}|^2 = \langle \hat{\boldsymbol{\kappa}}^2 \rangle = \frac{1}{\ell^2} \langle (2a^\dagger a + 1) \rangle \leq \frac{2N_L + 1}{\ell^2} \approx \frac{2N_L}{\ell^2} \ , \tag{58}$$

where we have defined $|\hat{\boldsymbol{\kappa}}| = \sqrt{\langle \hat{\boldsymbol{\kappa}}^2 \rangle}$ as the norm of the operator $\hat{\boldsymbol{\kappa}}$. Therefore, $\hat{\boldsymbol{\kappa}}$ is restricted within a radius $|\hat{\boldsymbol{\kappa}}| \leq \sqrt{2N_L}/\ell$.

The Hamiltonian (55) can be viewed as a lattice model in the momentum space with "hoppings" V_j between nearby reciprocal sites, and an "on-site potential energy" $\epsilon(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\alpha)$ on each site $\mathbf{Q}_\alpha$, and $\hat{\boldsymbol{\kappa}}$ plays the role of the "position" operator in the momentum space (although its x and y components are noncommuting). The momentum space probability (norm square of amplitude) of the basis wave function $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ is concentrated circularly near a ring of radius $\sqrt{\langle \hat{\boldsymbol{\kappa}}^2 \rangle} \approx \sqrt{2n}/\ell$.

The radius cutoff of $\mathbf{k}_0 + \mathbf{Q}_\alpha$ and the radius cutoff of $\hat{\boldsymbol{\kappa}}$ become equal when

$$\sqrt{2N_L}/\ell = \sqrt{N_Q \Omega_{BZ}/\pi} \ , \qquad \rightarrow \qquad \frac{\varphi}{2\pi} = \frac{B|\boldsymbol{d}_1 \times \boldsymbol{d}_2|}{2\pi} = \frac{2\pi}{\ell^2 \Omega_{BZ}} = \frac{N_Q}{N_L} \ , \tag{59}$$

where $\varphi = B|\boldsymbol{d}_1 \times \boldsymbol{d}_2|$ is the magnetic flux per unit cell, and we have used the Brillouin zone area $\Omega_{BZ} = 4\pi^2/|\boldsymbol{d}_1 \times \boldsymbol{d}_2|$ and $B = 1/\ell^2$. Given N_Q, N_L which we pick in our calculation, this flux $\varphi/2\pi = N_Q/N_L$ then separates the system into a small B regime and a large B regime as follows.

If $\sqrt{2N_L}/\ell < \sqrt{N_Q \Omega_{BZ}/\pi}$, or equivalently $\varphi/2\pi < N_Q/N_L$ (the small B field regime), the momentum space "position" $\hat{\boldsymbol{\kappa}}$ has a hard cutoff $|\hat{\boldsymbol{\kappa}}| \leq \sqrt{2N_L}/\ell$ which serves as the momentum space boundary.

On the contrary, if $\sqrt{2N_L}/\ell > \sqrt{N_Q \Omega_{BZ}/\pi}$, or equivalently $\varphi/2\pi > N_Q/N_L$ (the large B field regime), we have the expectation value of $|\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\alpha| > \sqrt{2N_L}/\ell - \sqrt{N_Q \Omega_{BZ}/\pi}$ for all sites $\mathbf{Q}_\alpha$ if a state has expectation value $|\hat{\boldsymbol{\kappa}}| > \sqrt{N_Q \Omega_{BZ}/\pi}$. In general, the kinetic energy, or "on-site potential energy" in the momentum space $\epsilon(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\alpha)$, is an increasing function of $|\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\alpha|$. If an eigenstate has a large expectation value of $|\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\alpha|$ for any Q_α (because it has expectation value $|\hat{\boldsymbol{\kappa}}| > \sqrt{N_Q \Omega_{BZ}/\pi}$), it will also have a large expectation value of kinetic energy $\epsilon(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}_\alpha)$ for any $\mathbf{Q}_\alpha$, thus its eigenenergy is expected to be large, and cannot be a reliable eigenstate of the low energy Hofstadter bulk bands.

Therefore, from the reasoning the above, we can define a momentum boundary radius given by the cutoffs N_L and N_Q as

$$\kappa_b \approx \min\{\frac{\sqrt{2N_L}}{\ell}, \sqrt{\frac{N_Q \Omega_{BZ}}{\pi}}\} \ , \tag{60}$$

and a trustable low-energy eigenstate in the Hofstadter bulk bands should have its expectation value $|\hat{\boldsymbol{\kappa}}| < \kappa_b$. Any state with expectation value $|\hat{\boldsymbol{\kappa}}| \gtrsim \kappa_b$ are effectively localized on the momentum space boundary at radius κ_b (if we view $\hat{\boldsymbol{\kappa}}$ as a momentum space coordinate), which we will call the *momentum space edge states*. Accordingly, we call the states with $|\hat{\boldsymbol{\kappa}}| < \kappa_b$ the *momentum space bulk states*. The momentum space bulk states give the Hofstadter bulk band spectra we want to calculate.

As we discussed in the main text (below main text Eq. (10)), the momentum space edge states can be identified by a boundary projection operator $P_{\kappa_b, w}$, the matrix elements of which are defined by

$$\langle \lambda', \mathbf{Q}'_\alpha, m, \beta | P_{\kappa_b, w} | \lambda, \mathbf{Q}_\beta, n, \alpha \rangle = \Theta\Big(n - (\kappa_b - w)^2 \ell^2/2\Big) \delta_{\lambda' \lambda} \delta_{mn} \delta_{\alpha\beta} \delta_{\mathbf{Q}'_\alpha, \mathbf{Q}_\beta} \ , \tag{61}$$

where $\Theta(x)$ is the Heaviside unit step function, $\kappa_b \approx \min\{\frac{\sqrt{2N_L}}{\ell}, \sqrt{\frac{N_Q \Omega_{BZ}}{\pi}}\}$ as defined in Eq. (60), and $w > 0$ is a parameter one can vary representing the defining width of the edge states. An eigenstate with large expectation value

$\langle P_{\kappa_b,w} \rangle$ will be mainly concentrated at $|\hat{\boldsymbol{\kappa}}| > \kappa_b - w$ in the momentum space, namely, within distance w inside the boundary radius κ_b, thus are momentum space edge states. These edge eigenstates with large $\langle P_{\kappa_b,w} \rangle > P_c$ above a certain chosen threshold $P_c \in [0,1]$ can then be deleted in the energy spectrum, so that only the bulk spectrum Hofstadter butterfly is kept (see main text Fig. 2(b), which is calculated with cutoffs $N_Q = 37$ and $N_L = 60$). In practice, one may properly adjust w and $\langle P_{\kappa_b,w} \rangle$ for different magnetic fields B and different energy range to reach a cleaner Hofstadter butterfly. Generically, the momentum space edge states in a smaller (larger) Hofstadter gap will be less (more) localized at the momentum boundary, thus requires a larger (smaller) edge width w and a smaller (larger) projection threshold P_c. In calculating main text Fig. 2(b), we have chosen $w = \min\{\ell^{-1}, 1.6\sqrt{\Omega_{BZ}}\}$, and the edge projection threshold $P_c = 0.5$. Here the factor 1.6 in front of $\sqrt{\Omega_{BZ}}$ is numerically tested to be a good choice for eliminating most edge states in the main Hofstadter gaps of the TBG model calculated here. Generically the optimal order 1 factor in front of $\sqrt{\Omega_{BZ}}$ depends on the models calculated (and can even be chosen to be dependent on the magnetic field B and the energy of the eigenstate). Generically, we suggest to choose w to be around the order of the smaller one of ℓ^{-1} and $\sqrt{\Omega_{BZ}}$.

Besides, we note that the Hofstadter butterfly with edge states deleted may have remaining edge state "hairs" near the edges of the Hofstadter gaps (i.e., not a clean Hofstadter butterfly), as shown in Fig. 2(b). This is because the momentum space edge states approaching the Hofstadter gap edges are more and more delocalized from the momentum space boundary, thus some of such edge states cannot satisfy the criteria $\langle P_{\kappa_b,w} \rangle > P_c$ and thus cannot be deleted. By making the momentum boundary radius κ_b larger (which is computationally more expensive) and choose a larger edge width w, one can reduce such edge state "hairs" and improve the clearness of the Hofstadter butterfly.

F. Number of states per Brillouin zone in a fixed λ sector

Here we discuss the number of occupied states per Brillouin zone (i.e., per reciprocal lattice "unit cell") ρ_K of the Hamiltonian $H_{\mathbf{Q}'_\alpha \mathbf{Q}_\beta, mn}^{\lambda, \alpha\beta}$ in a fixed λ sector (coset) given a Fermi energy, which we used in the main text Eq. (8).

Consider a magnetic field B corresponding to flux per unit cell $\varphi = B\Omega = \ell^{-2}\Omega$, where $\Omega = |\boldsymbol{d}_1 \times \boldsymbol{d}_2|$ is the zero-magnetic-field unit cell area. Assume the Fermi energy is ϵ_F, and the number of occupied states (below the Fermi energy ϵ_F) per zero-magnetic-field unit cell area Ω in the real space is ρ.

To find out the number of occupied states in each λ sector, we first count how many λ sectors there are. In an infinite real space and with an infinite reciprocal lattice $\mathbf{Q}$ (i.e., without reciprocal lattice cutoff), the set of λ in Eq. (43) is an infinite set, the size of which cannot be perceived easily. To make the set of λ finite, we assume the system has a large but finite real space area $\Omega_{\text{tot}} = N_\Omega \Omega$ (where N_Ω is the number of zero-magnetic-field unit cells), and take a finite reciprocal lattice number cutoff N_Q (as we did in the numerical calculation described in Sec. II E), but we keep the Landau level number cutoff $N_L \to \infty$. By Eq. (40), for each orbital α, each λ sector consists of the sub-Hilbert spaces with $R_{\hat{\boldsymbol{\tau}}}$ eigenvalues $\lambda - \ell^2 \hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \mathbf{Q}_\alpha)$ with $\mathbf{Q}_\alpha = \mathbf{Q} + \mathbf{p}_\alpha$ (defined in Eq. (29)) running over all $\mathbf{Q}$ within the reciprocal lattice cutoff N_Q. This means each λ sector consists of the sub-Hilbert spaces of N_Q different (discrete) eigenvalues of $R_{\hat{\boldsymbol{\tau}}}$. On the other hand, for each orbital α, the total number of (discrete) $R_{\hat{\boldsymbol{\tau}}}$ eigenvalues in the system is equal to the degeneracy of a single Landau level in the free space, which is

$$N_R = \frac{\Omega_{\text{tot}}}{2\pi\ell^2} \ . \tag{62}$$

This is because in the free space, the Landau level states of a definite LL band can be completely labeled by the eigenvalue of the guiding center along a certain direction (here chosen to be $R_{\hat{\boldsymbol{\tau}}}$). Since each λ sector allows N_Q different eigenvalues of $R_{\hat{\boldsymbol{\tau}}}$, and different λ sectors are orthogonal to each other in the Hilbert space, we conclude the number of λ sectors (cosets) in our method are related by

$$N_\lambda = \frac{N_R}{N_Q} = \frac{\Omega_{\text{tot}}}{2\pi\ell^2 N_Q} \ . \tag{63}$$

Meanwhile, since there are ρ occupied states per zero-magnetic-field unit cell area $\Omega = |\boldsymbol{d}_1 \times \boldsymbol{d}_2|$, the total number of occupied states is given by

$$N_{\text{occ}} = \rho \frac{\Omega_{\text{tot}}}{\Omega} \ . \tag{64}$$

Since the number of $R_{\hat{\boldsymbol{\tau}}}$ eigenvalues is N_R, we can define the number of occupied states in each $R_{\hat{\boldsymbol{\tau}}}$ eigenvalue sector as

$$\rho_K = \frac{N_{\text{occ}}}{N_R} = \rho \frac{2\pi\ell^2}{\Omega} = \frac{2\pi}{\varphi} \rho \ . \tag{65}$$

Note that ρ_K is nothing but the LL filling fraction. On the other hand, the occupied states should evenly belong to each λ sector, since different λ sectors have the same spectrum. Therefore, we conclude the number of states occupied in each λ sector (in the $N_L \to \infty$ and finite N_Q case considered here, i.e., $N_Q/N_L \to 0$) is

$$\mathcal{N}_K^{(\lambda)} = \frac{N_{\text{occ}}}{N_\lambda} = \frac{N_{\text{occ}}}{N_R} N_Q = \frac{2\pi}{\varphi} \rho N_Q = \rho_K N_Q \ , \tag{66}$$

which is independent of λ. Since N_Q is the reciprocal lattice cutoff, or the number of Brillouin zones (BZs) we keep in the reciprocal lattice (note that we assumed $N_L \to \infty$ which does not give a momentum space cutoff), we see that $\rho_K = (2\pi/\varphi)\rho$ can be understood as the number of occupied states per BZ in a fixed λ sector.

At rational flux $\varphi = 2\pi p/q$, according to the Diophantine equation (109), we have $\rho = \nu/q$ where ν is the number of occupied magnetic bands in the magnetic unit cell, and thus we find the number of occupied states per BZ (with λ fixed) is $\rho_K = (2\pi/\varphi)\rho = \nu/p$. Accordingly, the Diophantine equation (109) can be rewritten as

$$t_\nu + s_\nu \frac{2\pi}{\varphi} = \rho_K \ , \tag{67}$$

i.e., main text Eq. (8).

Since the total number of occupied states $\mathcal{N}_K^{(\lambda)} = N_Q \rho_K$ in a λ sector is independent of λ, hereafter we simply denote it as $\mathcal{N}_K$, as appears in the main text Eqs. (9) and (10).

G. Numerical determination of number of occupied states

This subsection discusses how to determine the number of occupied states $\mathcal{N}_K$ in a gap in numerical calculations. We note that $\mathcal{N}_K$ here is counted below the mid-gap energy of a gap, which may disperse as a function of magnetic field B (e.g., the inclined dashed lines in the main text Fig. 2(c)-(d)). Generically, $\mathcal{N}_K$ counted in this way will slightly depend on where the mid-gap energy is chosen, which determines how many in-gap edge states are included in $\mathcal{N}_K$ in addition to the bulk states. However, the relative error in counting $\mathcal{N}_K$ will tend to zero as N_L and N_Q increase, since the ratio between the number of in-gap edge states and the number of bulk band states will tend to zero when the momentum space area increases.

In the main text, we have shown that in the calculation with LL cutoff N_L and reciprocal cutoff N_Q (in a fixed λ sector), there are two regimes:

i) The regime $\varphi/2\pi < N_Q/N_L$, for which the momentum space boundary is at radius $\kappa_b = \sqrt{2N_L}/\ell$, and the momentum space bulk area is $\mathcal{A}_K = 2\pi N_L/\ell^2$, so the number of occupied states $\mathcal{N}_K = \rho_K \mathcal{A}_K/\Omega_{BZ}$ in the fixed λ sector (note that this is different from Eq. (66) where $\varphi/2\pi > N_Q/N_L$) satisfies Eq. (9) in the main text, namely,

$$\mathcal{N}_K = N_L(t_\nu \varphi/2\pi + s_\nu) \ , \tag{68}$$

where t_ν is the Chern number of the gap, and s_ν is another integer which we show in Sec. VI can be understood as a dual Chern number for the momentum space. Accordingly, the in-gap spectral flow rate in a gap is given by

$$\frac{\mathrm{d}\mathcal{N}_K}{\mathrm{d}(\varphi/2\pi)} = N_L t_\nu, \tag{69}$$

which allows us to determine t_ν of a gap in this regime by counting the number of states flowing across the midgap energy per flux number $\varphi/2\pi$, as described in the example given below main text Eq. (9) (shown in main text Fig. 2c). Further, if the number of occupied states $\mathcal{N}_K$ in the gap at some flux φ is known (counted relative to some reference point, which will be discussed below), we could also derive s_ν of the gap from Eq. (68).

ii) The regime $\varphi/2\pi > N_Q/N_L$, for which the momentum space boundary is at radius $\kappa_b = \sqrt{N_Q \Omega_{BZ}/\pi}$, and the momentum space bulk area is $\mathcal{A}_K = N_Q \Omega_{BZ}$, so the number of occupied states $\mathcal{N}_K = \rho_K \mathcal{A}_K/\Omega_{BZ}$ in the fixed λ sector satisfies main text Eq. (10), namely,

$$\mathcal{N}_K = N_Q(t_\nu + 2\pi s_\nu/\varphi) \ . \tag{70}$$

Accordingly, the in-gap spectral flow rate in a gap is given by

$$\frac{\mathrm{d}\mathcal{N}_K}{\mathrm{d}(2\pi/\varphi)} = N_Q s_\nu, \tag{71}$$

which allows us to determine s_ν of a gap in this regime by counting the number of states (without deleting momentum space edge states by the projector method in Sec. II E) flowing across the midgap energy per inverse flux number $2\pi/\varphi$, as described in the example given below main text Eq. (10) (shown in main text Fig. 2d). Further, if the number of occupied states $\mathcal{N}_K$ in the gap at some flux φ is known, we could also derive t_ν of the gap from Eq. (70).

In either regime, finding out $\mathcal{N}_K$ at some φ in the numerical results allows us to derive both t_ν and s_ν of a gap. In numerical calculations, however, the number of occupied states $\mathcal{N}_K$ below a gap should be counted relative to some reference energy level where $\mathcal{N}_K = 0$ is defined. Therefore, we need to find out where the $\mathcal{N}_K = 0$ energy level is defined, which depends on models. This is discussed below.

1. Determination of $\mathcal{N}_K = 0$

A generic method to find out $\mathcal{N}_K = 0$ is the following: first, find a gap which extends over both the small φ regime $\varphi/2\pi < N_Q/N_L$ and the large φ regime $\varphi/2\pi > N_Q/N_L$ (this is the condition for this generic method to work), which we call the reference gap. By extracting out the spectral flow rate in the small and large φ regimes, one can derive t_ν and s_ν of the reference gap from Eqs. (69) and (71), respectively. Then, $\mathcal{N}_K$ of this reference gap at any φ can be determined from Eqs. (68) and (70). Then, the number of occupied states $\mathcal{N}_K$ in any other gap j at a given φ can be determined in the numerical calculation by counting the number of states from the midgap energy of the reference gap to the midgap energy of gap j, plus the number of occupied states in the reference gap which is known. From the reference gap, one could determine which energy level corresponds to $\mathcal{N}_K = 0$.

As an example, in the Hofstadter butterfly of the TBG model in the main text Fig. 2(a) and 2(b) (see Sec. III for details), we can take the $(1, 0)$ gap (labeled in main text Fig. 2(b)) as such a reference gap. This Hofstadter spectrum is calculated by setting $N_Q = 37$ and $N_L = 60$. In the regime $\varphi/2\pi < N_Q/N_L$, the spectral flow rate across the mid-gap energy of gap $(1, 0)$ can be counted along the dashed line in Fig. 2(c) to be $d\mathcal{N}_K/d(\varphi/2\pi) \approx 16/0.25 = 64 = N_L t_\nu$ (16 levels when $\varphi/2\pi$ increases from 0.25 to 0.5), so we find $t_\nu = 1$ being the integer closest to $64/N_L$. Then in the regime $\varphi/2\pi > N_Q/N_L$, there are no levels flowing in the $(1, 0)$ gap except for some horizontal lines, which are spurious Dirac zero modes of the TBG model and should not be counted in $\mathcal{N}_K$ (see Sec. III B for detailed explanation). Therefore, the spectral flow rate in this regime is $d\mathcal{N}_K/d(2\pi/\varphi) = 0 = N_Q s_\nu$, which leads to $s_\nu = 0$. Thus the gap is labeled by quantum numbers $(t_\nu, s_\nu) = (1, 0)$. Then, by Eq. (68), we can find, for instance, $\mathcal{N}_K = N_L(1 \times 0.5 + 0) = 30$ at flux $\varphi/2\pi = 0.5 < N_Q/N_L$. Therefore, the $\mathcal{N}_K = 0$ level can be found by counting 30 levels downwards from the midgap energy of the $(1, 0)$ gap at flux $\varphi/2\pi = 0.5$, which turns out to be approximately the level at zero energy (see the main text Fig. 2(c)).

We further discuss the following two special cases, where the reference point $\mathcal{N}_K = 0$ can be determined more easily:

i) Models with a kinetic energy bounded from below, e.g., a single-orbital model with a Hamiltonian $\widetilde{H}(\mathbf{r}) = \epsilon(-i\nabla) + \sum_j V_j e^{i\mathbf{q}_j \cdot \mathbf{r}}$ with a quadratic kinetic energy $\epsilon(\mathbf{k}) = k^2/2m_0 \geq 0$, where m_0 is the electron effective mass. In this case, one must have Chern number $t_\nu = 0$ below the lowest energy band of the entire spectrum, i.e., when no states are occupied at all. By Eq. (70), one then finds $\mathcal{N}_K = 0$ below the lowest energy band in the limit $\varphi \to \infty$. Since there are no states at lower energies, there should be no spectra flows below the lowest band with respect to φ, so one should have $\mathcal{N}_K = 0$ below the lowest energy band at any flux φ.

ii) Models with a Dirac kinetic energy (which has no lower bound), such as the TBG model in Eq. (73) which has a kinetic term $\epsilon(\mathbf{k}) = v_F \boldsymbol{\sigma}^* \cdot \mathbf{k}$ (where $\boldsymbol{\sigma}^* = (\sigma_x, -\sigma_y)$). In this case, if there is no LL cutoff and reciprocal lattice cutoff, the energy spectrum of the system does not have a lower bound. With a LL cutoff N_L and a reciprocal lattice cutoff N_Q, the Hamiltonian size is $MN_L N_Q$ for M intrinsic orbitals (each Dirac kinetic term $\epsilon(\mathbf{k}) = v_F \boldsymbol{\sigma}^* \cdot \mathbf{k}$ has two intrinsic orbitals). In the $\varphi \to \infty$ limit, the energies of the eigenstates are dominated by the kinetic term $\epsilon(\mathbf{k}) = v_F \boldsymbol{\sigma}^* \cdot \mathbf{k}$ (since $\mathbf{k}$ is replaced by $\hat{\boldsymbol{\kappa}} + \mathbf{k}_0$ and $\hat{\boldsymbol{\kappa}} \propto \ell^{-1} = \sqrt{B}$ which goes to infinity as $\varphi \to \infty$), which should give a (nearly) particle-hole symmetric spectrum because of the particle-hole symmetry of the Dirac kinetic term. Accordingly, all the spectral flows should be (nearly) particle-hole symmetric about the half filling, which fixes $\mathcal{N}_K = 0$ at the half filling point. Therefore, in the limit $\varphi \to \infty$, one has $\mathcal{N}_K = 0$ at the half filling of the Hamiltonian with cutoffs N_L and N_Q, e.g., the filling between the $MN_L N_Q/2$-th level and the $(MN_L N_Q/2) + 1$-th level (energetically sorted). Since the total number of levels $MN_L N_Q$ does not change with respect to φ (for fixed N_L and N_Q), and $\mathcal{N}_K = 0$ is a reference filling independent of φ, we conclude that $\mathcal{N}_K = 0$ is between the $MN_L N_Q/2$-th level and the $(MN_L N_Q/2) + 1$-th level for any φ. In particular, for the TBG model in Eq. (73), the energy spectrum is particle-hole symmetric at all φ, so the half filling point $\mathcal{N}_K = 0$ is at zero energy at any φ. We note that when further counting the number of occupied states $\mathcal{N}_K$ of certain gaps relative to this half filling $\mathcal{N}_K = 0$ point, one needs to exclude the unphysical spurious modes due to LL cutoff N_L as discussed in Sec. III B.

III. THE EXAMPLE OF THE TBG CONTINUUM MODEL

In this section, we explain the Hofstadter spectrum calculation of the one-valley TBG continuum model in Ref. [2], the results of which are given in the main text Fig. 2.

A. Description of the model

The model consists of the Dirac electrons of the same valley of two graphene layers, which are relatively twisted by angle θ. Besides, we only consider one spin, namely, the model is spinless. The model can be written in real space as a 4×4 matrix

$$H_{\text{TBG}} = \begin{pmatrix} -iv_F \boldsymbol{\sigma}^* \cdot \nabla & \sum_{j=1}^3 V_j e^{i\mathbf{q}_j \cdot \mathbf{r}} \\ \sum_{j=1}^3 V_j^\dagger e^{-i\mathbf{q}_j \cdot \mathbf{r}} & -iv_F \boldsymbol{\sigma}^* \cdot \nabla \end{pmatrix} , \tag{72}$$

where the upper (lower) two basis are the A and B sublattices of the upper (lower) monolayer graphene, $\boldsymbol{\sigma}^* = (\sigma_x, -\sigma_y)$ are the Pauli matrices acting on A and B sublattices of the monolayer graphene lattice, the momenta $\mathbf{q}_j$ are given by

$$\mathbf{q}_1 = k_\theta (0, -1)^T , \qquad \mathbf{q}_2 = k_\theta \left(\frac{\sqrt{3}}{2}, \frac{1}{2} \right)^T , \qquad \mathbf{q}_3 = k_\theta \left(-\frac{\sqrt{3}}{2}, \frac{1}{2} \right)^T ,$$

and the interlayer hopping matrices

$$V_1 = w_0 \left(1_2 + \sigma_x\right) , \qquad V_2 = w_0 \left(1_2 - \frac{1}{2}\sigma_x - \frac{\sqrt{3}}{2}\sigma_y \right) , \qquad V_3 = w_0 \left(1_2 - \frac{1}{2}\sigma_x + \frac{\sqrt{3}}{2}\sigma_y \right) ,$$

where 1_2 stands for the 2×2 identity matrix. The parameters are given by $v_F \approx 610\text{meV·nm}$, $w_0 = 110\text{meV}$, and $k_\theta = |\mathbf{q}_j| = (8\pi/3a_0)\sin(\theta/2)$, with the lattice constant $a_0 = 0.246\text{nm}$. In the example shown in main text Fig. 2, we take the twist angle $\theta = 2.2°$.

The reciprocal vectors of the TBG continuum model are given by $\boldsymbol{g}_1 = \mathbf{q}_2 - \mathbf{q}_3$ and $\boldsymbol{g}_2 = \mathbf{q}_3 - \mathbf{q}_1$. By Fourier transforming the zero field Hamiltonian (72) into the momentum space, the Hamiltonian becomes a model in a honeycomb reciprocal lattice, where the orbitals of layer 1 and layer 2 are located at the two different sublattices $\mathbf{Q}_1 \in \mathbf{q}_1 + \boldsymbol{g}_1\mathbb{Z} + \boldsymbol{g}_2\mathbb{Z}$ and $\mathbf{Q}_2 \in -\mathbf{q}_1 + \boldsymbol{g}_1\mathbb{Z} + \boldsymbol{g}_2\mathbb{Z}$ of the honeycomb reciprocal lattice, respectively. Namely, the origin of the momentum in layer $\zeta = 1, 2$ is shifted by $(-1)^{\zeta-1}\mathbf{q}_1$, which is an example of choosing orbital-dependent momentum origins in Eq. (29). Such a shift has the advantage of making the symmetries of the momentum space Hamiltonian more explicit, and thus is adopted in most literatures of TBG. Since the kinetic energy for sublattices $\{\mathbf{Q}_1\}$ and $\{\mathbf{Q}_2\}$ in Eq. (72) are identical, we can use a single notation $\mathbf{Q}$ to denote both reciprocal sublattice sites $\{\mathbf{Q}_1, \mathbf{Q}_2\}$, i.e., the full honeycomb reciprocal lattice sites, and rewrite the Hamiltonian as

$$H_{\mathbf{Q}'\mathbf{Q}}(\mathbf{k}) = v_F \boldsymbol{\sigma}^* \cdot (\mathbf{k} + \mathbf{Q})\delta_{\mathbf{Q}'\mathbf{Q}} + \sum_{j=1}^3 (V_j \delta_{\mathbf{Q}', \mathbf{Q}+\mathbf{q}_j} + h.c.) . \tag{73}$$

One only needs to remember that the two different sublattices of the reciprocal lattice correspond to layers 1 and 2, respectively. The Hamiltonian under magnetic field in our basis is then given by the substitution $\mathbf{k} \to \hat{\boldsymbol{\kappa}} + \mathbf{k}_0$, with $\hat{\boldsymbol{\kappa}} = \frac{1}{\sqrt{2}\ell}(a + a^\dagger, -ia + ia^\dagger)$. In the calculation of main text Fig. 2, we set a twist angle $\theta = 2.2°$, take cutoffs $N_Q = 37$ and $N_L = 60$, and choose the central momentum $\mathbf{k}_0$ at the Γ point of the first TBG BZ (A different choice of $\mathbf{k}_0$ only affects the spectrum at extremely small magnetic fluxes $|\varphi/2\pi| < 1/N_L$, in which regime our calculation reduces to the LL calculation for the $\mathbf{k} \cdot \mathbf{p}$ model expanded at momentum $\mathbf{k}_0$). The same spectrum with edge states present and with them deleted by the edge projection criteria of Eq. (61) in a larger energy range is shown in Fig. 4(a)-(b). The edge states in Fig. 4(b) are deleted following the method described in Sec. II E, where we used edge width $w = \min\{\ell^{-1}, 1.6\sqrt{\Omega_{BZ}}\}$ for the edge projector $P_{\kappa_b, w}$, and the edge projection threshold $P_c = 0.5$. More examples of the TBG Hofstadter spectra calculated with our method can be found in Ref. [12].

As another example, we also calculated the spectrum of the TBG Hamiltonian with a Dirac mass term added (which can arise from hBN substrate alignment in TBG [27, 28]), i.e., a model Hamiltonian

$$H'_{\mathbf{Q}'\mathbf{Q}}(m_D, \mathbf{k}) = [v_F \boldsymbol{\sigma}^* \cdot (\mathbf{k} + \mathbf{Q}) + m_D \sigma_z]\delta_{\mathbf{Q}'\mathbf{Q}} + \sum_{j=1}^3 (V_j \delta_{\mathbf{Q}', \mathbf{Q}+\mathbf{q}_j} + h.c.) . \tag{74}$$

The calculated Hofstadter butterfly with edge states present and deleted for $m_D = 200\text{meV}$ are shown in Fig. 4(c)-(d). We will comment more on this case of nonzero Dirac mass in Sec. III B.
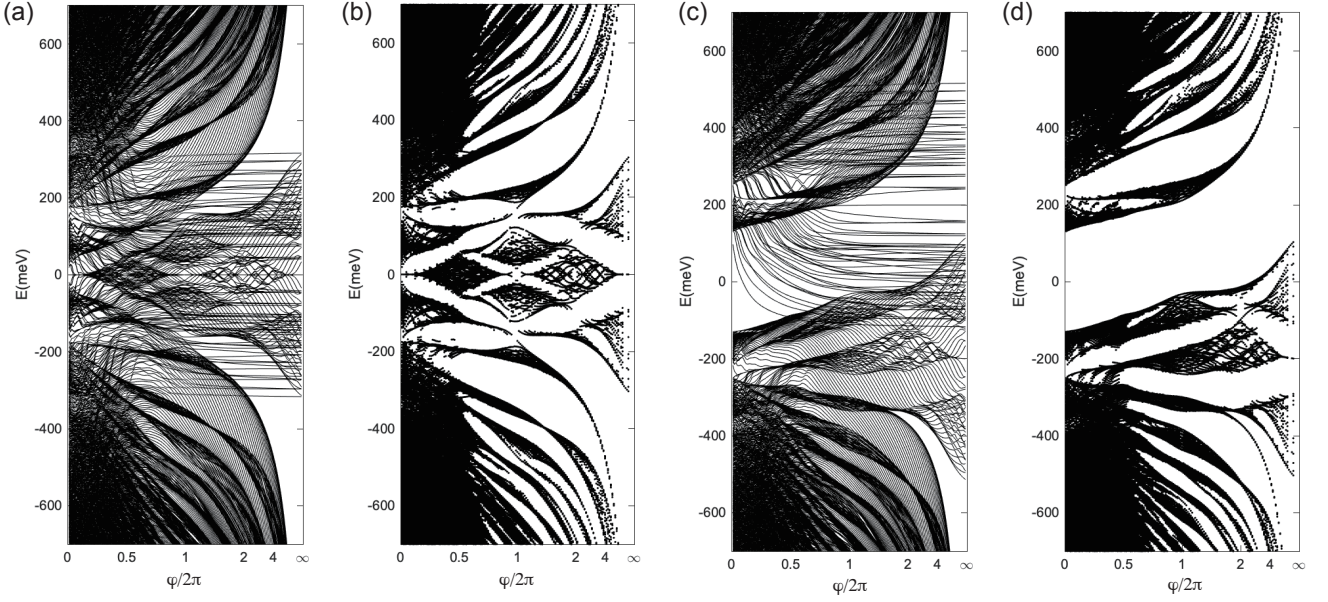
FIG. 4. Hofstadter butterfly and spectral flow of $\theta = 2.2°$ TBG with $N_Q = 37$ and $N_L = 60$ in a larger energy interval than that of main text Fig. 2, where (a)-(b) are calculated for a massless Dirac kinetic term, while (c)-(d) are calculated for a Dirac mass $m_D = 200$meV. The horizontal axis linear coordinate is equal to $\varphi/2\pi$ in the range $[0,1]$, and is equal to $2 - 2\pi/\varphi$ in the range $[1,\infty]$. For convenience, we still label the horizontal axis by the values of φ. The edge states in (b) and (d) are deleted following the method described in Sec. II E, where we used $w = \min\{\ell^{-1}, 1.6\sqrt{\Omega_{BZ}}\}$ for the edge projector $P_{\kappa_b,w}$, and the edge projection threshold $P_c = 0.5$.

B. Spurious zero modes

In the main text, Fig. 2(a) (see also Fig. 4(a)) shows many nondispersive horizontal levels at large fluxes φ, which are roughly distributed in energies from -0.3eV to 0.3eV. In Fig. 4(c) which has a nonzero Dirac mass $m_D = 200$meV, these nondispersive horizontal levels are still present at large φ, but are distributed in an energy range -0.1eV to 0.5eV. When φ is small, these levels become dispersive with respect to φ and tend to higher energies, merging either with the Hofstadter bulk bands or with the dispersive momentum edge states. These nondispersive horizontal levels at large fluxes are understood as spurious zero modes of Dirac fermions due to the LL cutoff N_L, which can be derived as follows.

In the large φ limit, the magnetic length $\ell \to 0$, and the massless Dirac kinetic term of the TBG Hamiltonian (73) at momentum site $\mathbf{Q}$ will tend to

$$v_F \boldsymbol{\sigma}^* \cdot (\hat{\boldsymbol{\kappa}} + \mathbf{k}_0 + \mathbf{Q}) = \frac{\sqrt{2}v_F}{\ell} \begin{pmatrix} 0 & a - p_+\ell/\sqrt{2} \\ a^\dagger - p_-\ell/\sqrt{2} & 0 \end{pmatrix} \to \frac{\sqrt{2}v_F}{\ell} \begin{pmatrix} 0 & a \\ a^\dagger & 0 \end{pmatrix} , \tag{75}$$

where $p_\pm = k_{0x} + Q_x \pm i(k_{0y} + Q_y)$. Therefore, up to error $\mathcal{O}(p_\pm\ell)$, the Dirac kinetic term on each site $\mathbf{Q}$ under LL cutoff N_L has two zero energy modes:

$$\psi_0 = \begin{pmatrix} 0 \\ |0\rangle \end{pmatrix} , \qquad \psi_s = \begin{pmatrix} |N_L\rangle \\ 0 \end{pmatrix} , \tag{76}$$

where we have used the fact that $a|0\rangle = 0$ and $a^\dagger|N_L\rangle = 0$. Note that $a^\dagger|N_L\rangle = 0$ is only true because of the LL cutoff. The first mode ψ_0 is the physical Dirac zero mode, while the second mode ψ_s is an unphysical spurious zero mode due to the LL cutoff N_L. With a reciprocal lattice cutoff of N_Q BZs, we have $2N_Q$ spurious zero modes on the $2N_Q$ honeycomb reciprocal lattice (since there are two Dirac cones in each BZ). These spurious zero modes are located at radius $|\hat{\boldsymbol{\kappa}}| = \sqrt{\langle\hat{\boldsymbol{\kappa}}^2\rangle} \approx \frac{\sqrt{2N_L}}{\ell}$ in the momentum space (main text Fig. 1(b)). For large fluxes $\varphi > 2\pi N_Q/N_L$, the momentum boundary is at radius $\kappa_b \approx \min\{\frac{\sqrt{2N_L}}{\ell}, \sqrt{\frac{N_Q\Omega_{BZ}}{\pi}}\} = \sqrt{\frac{N_Q\Omega_{BZ}}{\pi}} < \frac{\sqrt{2N_L}}{\ell}$, thus these spurious modes at radius $|\hat{\boldsymbol{\kappa}}| \approx \frac{\sqrt{2N_L}}{\ell}$ are outside the momentum space boundary and are not physical states of the Hofstadter butterfly.

Moreover, these spurious modes on different reciprocal sites $\mathbf{Q}$ have zero Dirac kinetic energy, and only hop among nearest reciprocal sites with an amplitude $\approx \psi_s^\dagger V_j \psi_s = w_0 = 110$ meV. Therefore, at large flux φ, they behave as a honeycomb tight-binding model in the reciprocal lattice with hopping amplitude w_0, which has $2N_Q$ energy levels independent of φ distributed between $-3w_0$ and $3w_0$. These levels give the horizontal lines at large φ in the main text Fig. 2(a) (as well as Fig. 4(a)).

When a Dirac mass term $m_D \sigma_z$ is added to the TBG Hamiltonian as given in Eq. (74), one finds the spurious zero mode ψ_s in Eq. (76) at each momentum site $\mathbf{Q}$ no longer has a Dirac zero kinetic energy; instead, it is shifted to energy m_D. Therefore, with the hopping w_0 among nearest momentum sites, one expects the $2N_Q$ spurious zero modes to be distributed in the energy range between $m_D - 3w_0$ and $m_D + 3w_0$. This is exactly the case in Fig. 4(c).

At small φ, one can no longer ignore the $p_\pm$ terms in Eq. (75), thus the spurious modes ψ_s in Eq. 76 is no longer a zero energy mode of the Dirac Hamiltonian in Eq. (75). Therefore, one expect these spurious modes to disperse with respect to φ at small φ, in agreement with the numerical result.

When counting the number of occupied states in a fixed λ sector (in Eqs. (9) and (10) of the main text), these unphysical spurious modes at large magnetic fields should be excluded. Further, since these spurious modes are outside the momentum space boundary (when $\varphi/2\pi > N_Q/N_L$), they will be identified as edge states by the boundary projector $P_{\kappa_b,w}$ in Eq. (61) (their expectation values $\langle P_{\kappa_b,w} \rangle$ are close to 1), and thus will be removed in the edge-state removed spectrum (main text Fig. 2(b), and Fig. 4(b) and (d)).

IV. BASIS COMPLETENESS AND MATRIX ELEMENTS FOR TIGHT-BINDING MODELS

The open momentum space method can also be applied to the numerical calculation of the Hofstadter butterfly of tight-binding models, as we demonstrated in the main text Fig. 3. To understand why the method is also valid for tight-binding models, in this section, we construct the complete and orthonormal basis for tight binding models employed by our method. We prove that under the basis we construct, the tight-binding Hamiltonian under magnetic field (Peierls substitution) is block diagonalized into identical blocks, and each block has matrix elements given by the zero-magnetic-field momentum space Hamiltonian $H^{\alpha\beta}(\mathbf{k})$ with the simple substitution $\mathbf{k} \to \hat{\boldsymbol{\kappa}} + \mathbf{k}_0$, where $\mathbf{k}$ is the quasi-momentum, $\mathbf{k}_0$ is an arbitrary momentum vector, $\hat{\boldsymbol{\kappa}} = \frac{1}{\sqrt{2}\ell}(a + a^\dagger, -ia + ia^\dagger)$, and a and $a^\dagger$ are the Landau level raising and lowering operators.

A. Standard Peierls substitution

We denote the Wannier orbital α in the unit cell labeled by lattice vector $\boldsymbol{D} \in \boldsymbol{d}_1 \mathbb{Z} + \boldsymbol{d}_2 \mathbb{Z}$ in real space as $|\boldsymbol{D}, \alpha\rangle$, and use $\boldsymbol{u}_\alpha$ to denote the position of orbital α in a unit cell. In the continuum space, we have

$$\langle \mathbf{r}, \alpha | \boldsymbol{D}, \alpha \rangle = W_\alpha(\mathbf{r} - \boldsymbol{D} - \boldsymbol{u}_\alpha) , \tag{77}$$

where $W_\alpha(\mathbf{r})$ is the Wannier function of orbital α, and $|\mathbf{r}, \alpha\rangle$ is the underlying continuum space basis at position $\mathbf{r}$ of intrinsic orbital α as defined by Eq. (25). For the discussion of tight-binding models here, without loss of generality, we shall choose the gauge where all $\mathbf{p}_\alpha = 0$ in the definition of $|\mathbf{r}, \alpha\rangle$ in Eq. (25) (recall that $\mathbf{p}_\alpha$ is the momentum space origin of orbital α, which can be chosen freely), namely,

$$|\mathbf{r}, \alpha\rangle = c_\alpha^\dagger(\mathbf{r})|0\rangle . \tag{78}$$

This ensures that the Hamiltonian matrix in the real space basis is invariant under lattice vector translation $T_{\boldsymbol{d}_i}$ (instead of changing by a unitary transformation as shown in Eq. (30) when $\mathbf{p}_\alpha \neq 0$).

We first comment that the derivation of the standard Peierls substitution [21] requires an approximation that the Wannier orbitals $|\boldsymbol{D}, \alpha\rangle$ are infinitely localized, namely, $W_\alpha(\mathbf{r}) = \delta^2(\mathbf{r})$. This is because the Peierls substitution only picks up the gauge phase factor connecting two points of Wannier positions, and is independent of the details (shapes, sizes, etc) of Wannier functions, which can be true only if each Wannier function is infinitely small and thus does not feel the magnetic field inside the orbital itself. This does not require the Wannier charge density to be localized on the site position $\boldsymbol{D}$, instead it can be a delta function localized at any position $\boldsymbol{D} + \boldsymbol{u}_\alpha$ away from the site position $\boldsymbol{D}$.

However, in our paper here, we shall keep the Wannier function $W_\alpha(\mathbf{r})$ in Eq. (77) *generic*, instead of assuming it is a delta function. This makes our proof of the method the most generic, which applies to nonstandard Peierls substitutions discussed in Sec. IV B, too.

The tight-binding model in real space then generically takes the form

$$H = \sum_{j,\alpha,\beta} t_j^{\alpha\beta} T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta}^{\alpha\beta} \ , \tag{79}$$

where $t_j^{\alpha\beta}$ are the hopping amplitudes, and we have defined

$$T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta}^{\alpha\beta} = \sum_{\boldsymbol{D}} e^{i \int_{c_{j,\alpha\beta}} \mathbf{A}(\boldsymbol{D}+\mathbf{r}) \cdot \mathrm{d}\mathbf{r}} |\boldsymbol{D} + \boldsymbol{D}_j, \alpha\rangle\langle \boldsymbol{D}, \beta| \tag{80}$$

as the translation operator from the Wannier orbital $|\boldsymbol{D}, \beta\rangle$ at position $\boldsymbol{D} + \boldsymbol{u}_\beta$ to the Wannier orbital $|\boldsymbol{D} + \boldsymbol{D}_j, \alpha\rangle$ at position $\boldsymbol{D} + \boldsymbol{D}_j + \boldsymbol{u}_\alpha$ under the Peierls substitution of the gauge potential $\mathbf{A}(\mathbf{r})$ in the continuum space, with $\boldsymbol{D}_j \in \boldsymbol{d}_1\mathbb{Z} + \boldsymbol{d}_2\mathbb{Z}$, and $c_{j,\alpha\beta}$ being the straight line segment from $\boldsymbol{u}_\beta$ to $\boldsymbol{D}_j + \boldsymbol{u}_\alpha$.
—— *At zero magnetic field*, we can choose the gauge $\mathbf{A}(\mathbf{r}) = 0$, and the Hamiltonian can be written into the momentum space by Fourier transformation as

$$H^{\alpha\beta}(\mathbf{k}) = \sum_j t_j^{\alpha\beta} e^{-i\mathbf{k}\cdot(\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta)} \ , \tag{81}$$

where $\mathbf{k}$ is the quasi-momentum which takes values in the Brillouin zone, and the basis is the Bloch basis $|\mathbf{k}, \alpha\rangle = \frac{1}{N_\Omega} \sum_{\boldsymbol{D}} e^{i\mathbf{k}\cdot(\boldsymbol{D}+\boldsymbol{u}_\alpha)}|\boldsymbol{D}, \alpha\rangle$, with N_Ω being the number of unit cells in real space.
—— *At nonzero magnetic field*, the gauge potential $\mathbf{A}(\mathbf{r})$ satisfies $\partial_x A_y(\mathbf{r}) - \partial_y A_x(\mathbf{r}) = B$, where B is a uniform magnetic field in the continuum space. We now proceed to define a complete orthonormal basis for Hamiltonian (79) based on the continuum space.

Before starting, we first note that the continuum space has a Hilbert space spanned by the real space basis $|\mathbf{r}, \alpha\rangle$ of all positions $\mathbf{r}$, which is much larger than the Hilbert space of the tight binding model spanned by the Wannier basis $|\boldsymbol{D}, \alpha\rangle$ defined in Eq. (77). In the following, we shall first define a complete basis for the continuum space; then we project the basis into the sub-Hilbert space of the tight-binding model spanned by $|\boldsymbol{D}, \alpha\rangle$ using a projector, and prove that the resulting projected basis forms a complete orthonormal basis for the tight-binding model.

In the continuum space, we can define the kinematic momentum operator $\boldsymbol{\Pi} = -i\nabla - \mathbf{A}(\mathbf{r})$, and the guiding center operator $\mathbf{R} = \mathbf{r} - \frac{\ell^2}{\hbar}\hat{\mathbf{z}} \times \boldsymbol{\Pi}$, as we did in Sec. I. Similar to Sec. I, we define a guiding center $R_{\hat{\boldsymbol{\tau}}}$ along the $\hat{\boldsymbol{\tau}}$ direction, with $\frac{\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_1)}{\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_2)}$ irrational. Moreover, we denote the reciprocal lattice of the tight-binding model as $\mathbf{Q} \in \boldsymbol{g}_1\mathbb{Z} + \boldsymbol{g}_2\mathbb{Z}$. We can then define lowering and raising operators following the same procedure as we did in Sec. II

$$a_{\mathbf{Q}} = \frac{\ell}{\sqrt{2}}[\Pi_x - Q_x - k_{0,x} + i(\Pi_y - Q_y - k_{0,y})] \ , \qquad a_{\mathbf{Q}}^\dagger = \frac{\ell}{\sqrt{2}}[\Pi_x - Q_x - k_{0,x} - i(\Pi_y - Q_y - k_{0,y})] \ , \tag{82}$$

where $\mathbf{k}_0$ is the center momentum which can be chosen freely. Afterwards, we can define a basis $|\lambda, \mathbf{Q}, n, \alpha\rangle$ in the continuum space satisfying

$$R_{\hat{\boldsymbol{\tau}}}|\lambda, \mathbf{Q}, n, \alpha\rangle = [\lambda - \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\mathbf{Q})]|\lambda, \mathbf{Q}, n, \alpha\rangle \ , \qquad a_{\mathbf{Q}}^\dagger a_{\mathbf{Q}}|\lambda, \mathbf{Q}, n, \alpha\rangle = n|\lambda, \mathbf{Q}, n, \alpha\rangle \ , \tag{83}$$

where λ takes values in the quotient set defined in Eq. (43), namely, $\lambda \in \Lambda_{\hat{\boldsymbol{\tau}}} = \mathbb{R}/[\ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_1)\mathbb{Z} + \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_2)\mathbb{Z}]$. The basis definition (83) follows exactly the same derivation of the basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ in Eq. (40) in Sec. II, except that here we have chosen the gauge that the momentum origins of all orbitals α are at $\mathbf{p}_\alpha = \mathbf{0}$ (see Eq. (25) for the definition of $\mathbf{p}_\alpha$, and see Eq. (78) for our gauge choice for tight binding models here), thus $\mathbf{Q}_\alpha = \mathbf{Q}$ for all α (recall the definition of $\mathbf{Q}_\alpha$ in Eq. (II)). As we have proved in Eq. (44), the basis $|\lambda, \mathbf{Q}, n, \alpha\rangle$ in Eq. (83) satisfies

$$\langle\lambda', \mathbf{Q}', n', \beta|\lambda, \mathbf{Q}, n, \alpha\rangle = \delta_{\beta\alpha}\delta_{\lambda\lambda'}\delta_{\mathbf{Q}'\mathbf{Q}}\delta_{n'n} \ , \tag{84}$$

thus forms a complete orthonormal basis for the Hilbert space of the continuum space, where $\lambda \in \Lambda_{\hat{\boldsymbol{\tau}}} = \mathbb{R}/[\ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_1)\mathbb{Z} + \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_2)\mathbb{Z}]$ as defined in Eq. (43), $\mathbf{Q} \in \boldsymbol{g}_1\mathbb{Z} + \boldsymbol{g}_2\mathbb{Z}$ are the reciprocal vectors, and $n \geq 0$ denotes the LL number. We therefore have a completeness condition in the continuum space:

$$\frac{1}{N_Q} \sum_{\lambda,\mathbf{Q},n} |\lambda, \mathbf{Q}, n, \alpha\rangle\langle\lambda, \mathbf{Q}, n, \alpha| = \mathbf{1}_\alpha = \int \mathrm{d}^2\mathbf{r}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha| \ , \tag{85}$$

where N_Q is the number of reciprocal sites $\mathbf{Q}$ (which tends to infinity), and $\mathbf{1}_\alpha$ stands for the identity matrix in the orbital α subspace in the continuum space.

Next, based on the continuum space basis $|\lambda, \mathbf{Q}, n, \alpha\rangle$, we would like to define a complete basis for the Hilbert space of the tight-binding model spanned by Wannier orbitals $|\boldsymbol{D}, \alpha\rangle$. We define the basis for the tight-binding model as

$$\overline{|\lambda, n, \alpha\rangle} = \sum_{\boldsymbol{D}} |\boldsymbol{D}, \alpha\rangle\langle \boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \mathbf{0}, n, \alpha\rangle \ , \tag{86}$$

where $|\boldsymbol{D}, \alpha\rangle$ is the Wannier orbital, $|\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha\rangle$ denotes the position eigenbasis at position $\mathbf{r} = \boldsymbol{D} + \boldsymbol{u}_\alpha$ (the center of the Wannier orbital), and $\lambda \in \Lambda_{\hat{\boldsymbol{\tau}}} = \mathbb{R}/[\ell^2\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_1)\mathbb{Z} + \ell^2\hat{\boldsymbol{\tau}} \cdot (\hat{\mathbf{z}} \times \boldsymbol{g}_2)\mathbb{Z}]$ as defined in Eq. (43). One may wonder why only the continuum space states $|\lambda, \mathbf{0}, n, \alpha\rangle$ at $\mathbf{Q} = \mathbf{0}$ are used in defining the basis (86). In fact, by Eq. (48), we have the following identity for any lattice vector $\boldsymbol{D}$ and reciprocal vector $\mathbf{Q}$ (in the equation below, $\mathbf{r}$ is understood as a number denoting the continuous space coordinates)

$$\langle \boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \mathbf{Q}, n, \alpha\rangle = \int \mathrm{d}^2\mathbf{r}\,\delta^2(\mathbf{r} - \boldsymbol{D} - \boldsymbol{u}_\alpha)\langle\mathbf{r}, \alpha|e^{i\mathbf{Q}\cdot\mathbf{r}}|\lambda, \mathbf{0}, n, \alpha\rangle = e^{i\mathbf{Q}\cdot(\boldsymbol{D}+\boldsymbol{u}_\alpha)}\langle\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \mathbf{0}, n, \alpha\rangle$$
$$= e^{i\mathbf{Q}\cdot\boldsymbol{u}_\alpha}\langle\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \mathbf{0}, n, \alpha\rangle \ . \tag{87}$$

Thus, one can equivalently rewrite the basis definition as

$$\overline{|\lambda, n, \alpha\rangle} = \frac{1}{\sqrt{N_Q}} \sum_{\mathbf{Q}, \boldsymbol{D}} e^{-i\mathbf{Q}\cdot\boldsymbol{u}_\alpha} |\boldsymbol{D}, \alpha\rangle\langle \boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \mathbf{Q}, n, \alpha\rangle \ , \tag{88}$$

which is expressed using the continuum space basis of all reciprocal sites $\mathbf{Q}$.

By the orthonormal relation (84) and the basis expression in Eq. (86), it is easy to see the subset of basis $\overline{|\lambda, n, \alpha\rangle}$ is orthonormal:

$$\overline{\langle\lambda', n', \beta|}\overline{|\lambda, n, \alpha\rangle} = \delta_{\beta\alpha} \sum_{\boldsymbol{D}} \langle\lambda', \mathbf{0}, n', \alpha|\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha\rangle\langle\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \mathbf{0}, n, \alpha\rangle$$
$$= \delta_{\beta\alpha}\langle\lambda', \mathbf{0}, n', \alpha| \sum_{\mathbf{Q}} e^{i\mathbf{Q}\cdot(\mathbf{r}-\boldsymbol{u}_\alpha)}|\lambda, \mathbf{0}, n, \alpha\rangle = \delta_{\beta\alpha} \sum_{\mathbf{Q}} e^{-i\mathbf{Q}\cdot\boldsymbol{u}_\alpha}\langle\lambda', \mathbf{0}, n', \alpha|\lambda, \mathbf{Q}, n, \alpha\rangle = \delta_{\lambda\lambda'}\delta_{n'n}\delta_{\beta\alpha} \ , \tag{89}$$

where in the 2nd line $\mathbf{r}$ is understood as the position operator (not number), and we have used Eq. (48) which implies $|\lambda, \mathbf{Q}, n, \alpha\rangle = e^{i\mathbf{Q}\cdot\mathbf{r}}|\lambda, \mathbf{0}, n, \alpha\rangle$. Besides, from the 1st line to the 2nd line we have used the following identity for each orbital α:

$$\sum_{\boldsymbol{D}} |\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha\rangle\langle\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha| = \int \mathrm{d}^2\mathbf{r}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha| \sum_{\boldsymbol{D}} \delta^2(\mathbf{r} - \boldsymbol{D} - \boldsymbol{u}_\alpha)$$
$$= \int \mathrm{d}^2\mathbf{r}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha| \sum_{\mathbf{Q}} e^{i\mathbf{Q}\cdot(\mathbf{r}-\boldsymbol{u}_\alpha)} = \sum_{\mathbf{Q}} e^{i\mathbf{Q}\cdot(\mathbf{r}-\boldsymbol{u}_\alpha)} \ , \tag{90}$$

where $\mathbf{r}$ on the right-most hand side of the 2nd line is understood as the position operator (instead of a number).

Furthermore, by Eq. (87), we can prove the completeness of the basis (86) for the tight-binding model as follows:

$$\sum_{\lambda, n, \alpha} \overline{|\lambda, n, \alpha\rangle}\overline{\langle\lambda, n, \alpha|} = \sum_{\lambda, n, \alpha} \sum_{\boldsymbol{D}, \boldsymbol{D}'} |\boldsymbol{D}, \alpha\rangle\langle\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \mathbf{0}, n, \alpha\rangle\langle\lambda, \mathbf{0}, n, \alpha|\boldsymbol{D}' + \boldsymbol{u}_\alpha, \alpha\rangle\langle\boldsymbol{D}', \alpha|$$
$$= \frac{1}{N_Q} \sum_{\lambda, \mathbf{Q}, n, \alpha} \sum_{\boldsymbol{D}, \boldsymbol{D}'} |\boldsymbol{D}, \alpha\rangle e^{-i\mathbf{Q}\cdot\boldsymbol{u}_\alpha}\langle\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha|\lambda, \mathbf{Q}, n, \alpha\rangle\langle\lambda, \mathbf{Q}, n, \alpha|\boldsymbol{D}' + \boldsymbol{u}_\alpha, \alpha\rangle e^{i\mathbf{Q}\cdot\boldsymbol{u}_\alpha}\langle\boldsymbol{D}', \alpha|$$
$$= \sum_{\alpha} \sum_{\boldsymbol{D}, \boldsymbol{D}'} |\boldsymbol{D}, \alpha\rangle\langle\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha| \left(\frac{1}{N_Q} \sum_{\lambda, \mathbf{Q}, n} |\lambda, \mathbf{Q}, n, \alpha\rangle\langle\lambda, \mathbf{Q}, n, \alpha|\right) |\boldsymbol{D}' + \boldsymbol{u}_\alpha, \alpha\rangle\langle\boldsymbol{D}', \alpha|$$
$$= \sum_{\boldsymbol{D}, \boldsymbol{D}', \alpha} |\boldsymbol{D}, \alpha\rangle\langle\boldsymbol{D} + \boldsymbol{u}_\alpha, \alpha| \left(\int \mathrm{d}^2\mathbf{r}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha|\right) |\boldsymbol{D}' + \boldsymbol{u}_\alpha, \alpha\rangle\langle\boldsymbol{D}', \alpha| = \sum_{\boldsymbol{D}, \alpha} |\boldsymbol{D}, \alpha\rangle\langle\boldsymbol{D}, \alpha| \ . \tag{91}$$

This proves that the basis $\overline{|\lambda, n, \alpha\rangle}$ forms a complete orthonormal basis for the Hilbert space of the tight-binding model spanned by Wannier orbitals $|\boldsymbol{D}, \alpha\rangle$.

Now we discuss the translation operator in Eq. (80) under magnetic field. First, we prove the following identity of the position basis $|\mathbf{r}, \alpha\rangle$ in the continuum space. Assume c_{f0} is a path (not necessarily straight) from position $\mathbf{r}_0$ to

position $\mathbf{r}_f$ in the continuum space. By path partitioning c_{f0} into N small segments $\delta\mathbf{r}_j = \mathbf{r}_j - \mathbf{r}_{j-1}$ ($1 \leq j \leq N$), and note that $e^{-i\mathbf{\Pi}\cdot\delta\mathbf{r}} = e^{i\mathbf{A}(\mathbf{r})\cdot\delta\mathbf{r}}e^{-\delta\mathbf{r}\cdot\nabla}$ in the $\delta\mathbf{r} \to 0$ limit, we can prove that

$$\mathcal{P}e^{-i\int_{c_{f0}}\mathbf{\Pi}\cdot d\mathbf{r}}|\mathbf{r}_0,\alpha\rangle = \lim_{N\to\infty}\prod_{j=1}^{N}e^{-i\mathbf{\Pi}\cdot\delta\mathbf{r}_j}|\mathbf{r}_0,\alpha\rangle$$

$$= \lim_{N\to\infty}\int d\mathbf{r}_N|\mathbf{r}_N,\alpha\rangle\langle\mathbf{r}_N,\alpha|e^{-i\mathbf{\Pi}\cdot\delta\mathbf{r}_N}\int d\mathbf{r}_{N-1}|\mathbf{r}_{N-1},\alpha\rangle\langle\mathbf{r}_{N-1},\alpha|e^{-i\mathbf{\Pi}\cdot\delta\mathbf{r}_{N-1}}\cdots\int d\mathbf{r}_1|\mathbf{r}_1,\alpha\rangle\langle\mathbf{r}_1,\alpha|e^{-i\mathbf{\Pi}\cdot\delta\mathbf{r}_1}|\mathbf{r}_0,\alpha\rangle$$

$$= \lim_{N\to\infty}\prod_{j=1}^{N}e^{i\mathbf{A}(\mathbf{r}_0+\sum_{i\leq j}\delta\mathbf{r}_i)\cdot\delta\mathbf{r}_j}|\mathbf{r}_0+\sum_{j=1}^{N}\delta\mathbf{r}_j,\alpha\rangle$$

$$= e^{i\int_{c_{f0}}\mathbf{A}(\mathbf{r})\cdot d\mathbf{r}}|\mathbf{r}_f,\alpha\rangle\ ,$$

$$(92)$$

where $\mathcal{P}$ stands for path ordering, and we have used the fact that $e^{-\delta\mathbf{r}\cdot\nabla}|\mathbf{r},\alpha\rangle = |\mathbf{r}+\delta\mathbf{r},\alpha\rangle$. Therefore, note that the definition of the basis $\overline{|\lambda,n,\alpha\rangle}$ in Eq. (86) contains the position eigenstate $|\boldsymbol{D}+\boldsymbol{u}_\alpha,\alpha\rangle$ at position $\mathbf{r} = \boldsymbol{D}+\boldsymbol{u}_\alpha$, we can rewrite the action of the translation operator in Eq. (80) on the basis $\overline{|\lambda,n,\alpha\rangle}$ as

$$T_{\boldsymbol{D}_j+\boldsymbol{u}_\alpha-\boldsymbol{u}_\beta}^{\alpha\beta}\overline{|\lambda,n,\beta'\rangle} = \delta_{\beta\beta'}\sum_{\boldsymbol{D}}e^{i\int_{c_{\alpha\beta}}\mathbf{A}(\boldsymbol{D}+\mathbf{r})\cdot d\mathbf{r}}|\boldsymbol{D}+\boldsymbol{D}_j,\alpha\rangle\langle\boldsymbol{D}+\boldsymbol{u}_\beta,\beta|\lambda,\boldsymbol{0},n,\beta\rangle$$

$$= \delta_{\beta\beta'}\sum_{\boldsymbol{D}}|\boldsymbol{D}+\boldsymbol{D}_j,\alpha\rangle\langle\boldsymbol{D}+\boldsymbol{D}_j+\boldsymbol{u}_\alpha,\beta|\mathcal{P}e^{-i\int_{c_{j,\alpha\beta}}\mathbf{\Pi}\cdot d\mathbf{r}}|\lambda,\boldsymbol{0},n,\beta\rangle\ ,$$

$$(93)$$

where notations of the form $|\boldsymbol{D}+\boldsymbol{u}_\alpha,\beta\rangle$ denotes the position basis of orbital β in the continuum space at position $\boldsymbol{D}+\boldsymbol{u}_\alpha$ (note that β need not be equal to α). In contrast, $|\boldsymbol{D}+\boldsymbol{D}_j,\alpha\rangle$ denotes the Wannier orbital basis of the unit cell at $\boldsymbol{D}+\boldsymbol{D}_j$, as we have defined. We note that in Eq. (93), from the 1st line to the 2nd line we have used the fact that the position basis in the continuum space satisfies $\langle\boldsymbol{D}+\boldsymbol{u}_\beta,\beta| = \langle\boldsymbol{D}+\boldsymbol{D}_j+\boldsymbol{u}_\alpha,\beta|\mathcal{P}e^{-i\int_{c_{j,\alpha\beta}}\mathbf{\Pi}\cdot d\mathbf{r}}$, which we have proved in Eq. (92).

We then define lowering and raising operators $a = a_{\boldsymbol{0}} = \frac{\ell}{\sqrt{2}}[\Pi_x - k_{0,x} + i(\Pi_y - k_{0,y})]$ and $a = a_{\boldsymbol{0}}^\dagger$, where $a_{\boldsymbol{0}}, a_{\boldsymbol{0}}^\dagger$ are the lowering and raising operators in Eq. (82) at reciprocal site $\mathbf{Q} = \boldsymbol{0}$. By further defining $\hat{\boldsymbol{\kappa}} = \frac{1}{\sqrt{2}\ell}(a+a^\dagger, -ia+ia^\dagger)$, we can rewrite the kinematic momentum as $\mathbf{\Pi} = \hat{\boldsymbol{\kappa}}+\mathbf{k}_0$, where $\mathbf{k}_0$ is the center momentum in Eq. (82). We can then further simplify Eq. (93) as

$$T_{\boldsymbol{D}_j+\boldsymbol{u}_\alpha-\boldsymbol{u}_\beta}^{\alpha\beta}\overline{|\lambda,n,\beta'\rangle} = \delta_{\beta\beta'}\sum_{\boldsymbol{D}}|\boldsymbol{D}+\boldsymbol{D}_j,\alpha\rangle\langle\boldsymbol{D}+\boldsymbol{D}_j+\boldsymbol{u}_\alpha,\beta|\mathcal{P}e^{-i\int_{c_{j,\alpha\beta}}(\hat{\boldsymbol{\kappa}}+\mathbf{k}_0)\cdot d\mathbf{r}}|\lambda,\boldsymbol{0},n,\beta\rangle\ ,$$

$$= \delta_{\beta\beta'}\sum_{\boldsymbol{D}}|\boldsymbol{D}+\boldsymbol{D}_j,\alpha\rangle\langle\boldsymbol{D}+\boldsymbol{D}_j+\boldsymbol{u}_\alpha,\alpha|\mathcal{P}e^{-i\int_{c_{j,\alpha\beta}}(\hat{\boldsymbol{\kappa}}+\mathbf{k}_0)\cdot d\mathbf{r}}|\lambda,\boldsymbol{0},n,\alpha\rangle\ ,$$

$$= \delta_{\beta\beta'}\sum_{\boldsymbol{D}}|\boldsymbol{D},\alpha\rangle\langle\boldsymbol{D}+\boldsymbol{u}_\alpha,\alpha|\mathcal{P}e^{-i\int_{c_{j,\alpha\beta}}(\hat{\boldsymbol{\kappa}}+\mathbf{k}_0)\cdot d\mathbf{r}}|\lambda,\boldsymbol{0},n,\alpha\rangle$$

$$= \delta_{\beta\beta'}\mathcal{P}e^{-i\int_{c_{j,\alpha\beta}}(\hat{\boldsymbol{\kappa}}+\mathbf{k}_0)\cdot d\mathbf{r}}\overline{|\lambda,n,\alpha\rangle}\ ,$$

$$(94)$$

where $\mathcal{P}$ stands for path ordering. From the 1st line to the 2nd line of Eq. (94), we have used the fact that the matrix elements of the operator $\mathcal{P}e^{-i\int_{c_{j,\alpha\beta}}(\hat{\boldsymbol{\kappa}}+\mathbf{k}_0)\cdot d\mathbf{r}}$ (acting from the right on the state $\langle\boldsymbol{D}+\boldsymbol{D}_j+\boldsymbol{u}_\alpha,\beta|$) only depend on the position $\mathbf{r} = \boldsymbol{D}+\boldsymbol{D}_j+\boldsymbol{u}_\alpha$ of the state (since $\hat{\boldsymbol{\kappa}}$ is defined in terms of $\mathbf{\Pi}$ which displaces $\mathbf{r}$), and are independent of the orbital index β. So we can simply change the orbital index β in the 1st line into α in the 2nd line. In the derivation from the 3rd line to the 4-th line, we have also used the fact that the matrix elements of operator $\hat{\boldsymbol{\kappa}}$ only depends on the LL quantum number n, so in the last line of Eq. (94) one should understand the operator $\hat{\boldsymbol{\kappa}}$ as solely acting on the quantum number n of basis $\overline{|\lambda,n,\alpha\rangle}$. More concretely, the LL lowering and raising operators in $\hat{\boldsymbol{\kappa}}$ act as $a\overline{|\lambda,n,\alpha\rangle} = \sqrt{n}\overline{|\lambda,n-1,\alpha\rangle}$ and $a^\dagger\overline{|\lambda,n,\alpha\rangle} = \sqrt{n+1}\overline{|\lambda,n+1,\alpha\rangle}$.

For standard Peierls substitution, the path $c_{j,\alpha\beta}$ in Eq. (94) is a straight line segment (see the paragraph below Eq. (80)), so the integral on the exponent of $\mathcal{P}e^{-i\int_{c_{j,\alpha\beta}}(\hat{\boldsymbol{\kappa}}+\mathbf{k}_0)\cdot d\mathbf{r}}$ is along the straight line segment from $\boldsymbol{D}+\boldsymbol{u}_\beta$ to $\boldsymbol{D}+\boldsymbol{D}_j+\boldsymbol{u}_\alpha$. Since the matrix elements of $\hat{\boldsymbol{\kappa}}$ only depend on n when acting on $\overline{|\lambda,n,\alpha\rangle}$, the matrix $\hat{\boldsymbol{\kappa}}+\mathbf{k}_0$ should be a constant along the path of integration, and thus we can further simplify Eq. (94) into

$$T_{\boldsymbol{D}_j+\boldsymbol{u}_\alpha-\boldsymbol{u}_\beta}^{\alpha\beta}\overline{|\lambda,n,\beta'\rangle} = \delta_{\beta\beta'}e^{-i(\hat{\boldsymbol{\kappa}}+\mathbf{k}_0)\cdot(\boldsymbol{D}_j+\boldsymbol{u}_\alpha-\boldsymbol{u}_\beta)}\overline{|\lambda,n,\alpha\rangle}\ .$$
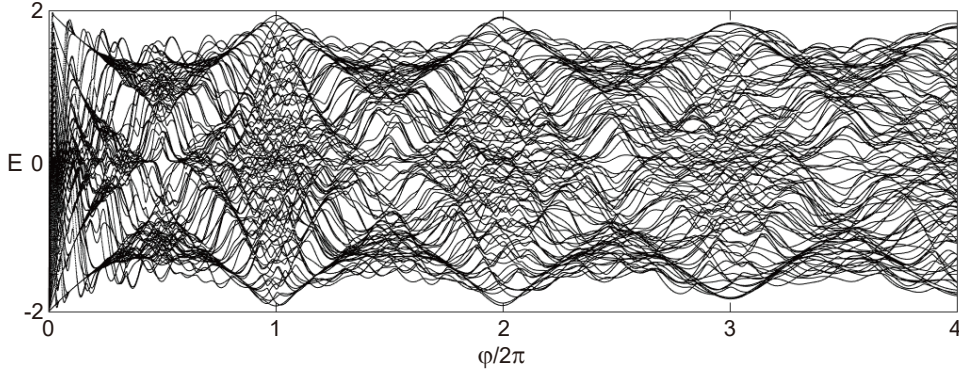
$$(95)$$

FIG. 5. The first 4 periods of Hofstadter butterfly of the tight-binding model $H(\mathbf{k}) = -\cos k_x - \cos k_y$ in the square lattice calculated using our method of substitution $\mathbf{k} \to \hat{\boldsymbol{\kappa}} + \mathbf{k}_0$, where we take a LL cutoff $N_L = 100$. The cutoff N_L breaks the periodicity of the Hofstadter spectrum, and the larger the magnetic field B is, the less clear the Hofstadter butterfly is. For sufficiently large LL cutoffs (e.g., $N_L \sim 500$), the periodicity of Hofstadter spectrum will be well-preserved in the first few periods.

It is evident from Eq. (95) that the matrix elements of $T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta}$ is diagonal in quantum number λ and independent of λ. Therefore, we can rewrite the translation operator in a fixed λ subspace from basis $\overline{|\lambda, n, \beta\rangle}$ to basis $\overline{|\lambda, n', \alpha\rangle}$ ($n, n' \in \mathbb{Z}$) are nonnegative integers) as

$$T^{\lambda, \alpha\beta}_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta} = e^{-i(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0) \cdot (\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta)} \ , \tag{96}$$

where $\hat{\boldsymbol{\kappa}}$ acting on the LL quantum number n (recall that $\boldsymbol{\Pi} = \hat{\boldsymbol{\kappa}} + \mathbf{k}_0$). Accordingly, the tight-binding Hamiltonian in magnetic field in a fixed λ subspace from basis $\overline{|\lambda, n, \beta\rangle}$ to basis $\overline{|\lambda, n', \alpha\rangle}$ ($n, n' \in \mathbb{Z}$) takes the form

$$H^{\lambda, \alpha\beta} = \sum_j t_j^{\alpha\beta} T^{\lambda, \alpha\beta}_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta} = \sum_j t_j^{\alpha\beta} e^{-i(\hat{\boldsymbol{\kappa}} + \mathbf{k}_0) \cdot (\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta)} \ , \tag{97}$$

which is exactly the zero field momentum space Hamiltonian (81) with the substitution $\mathbf{k} \to \hat{\boldsymbol{\kappa}} + \mathbf{k}_0$.

As an example of the above method, we numerically calculate the Hofstadter butterfly of the simplest square lattice model with one orbital per site and nearest hopping amplitude $-1/2$, which has a momentum space Hamiltonian $H(\mathbf{k}) = -\cos k_x - \cos k_y$. The Hamiltonian matrix with LL cutoff N_L is constructed by replacing k_x by the Hermitian matrix $k_{0,x} \mathbf{1}_{N_L} + (a + a^\dagger)/(\sqrt{2}\ell)$ and similarly for k_y, where a is a matrix with matrix elements $[a]_{mn} = \langle m|a|n\rangle = \sqrt{n}\delta_{m,n-1}$ (m, n are integers from 0 to N_L). The cosine of a Hermitian matrix W can be calculated by first diagonalizing the matrix into $W = U^\dagger D_W U$ where D_W is diagonal and U is unitary. Then $\cos W$ can be calculated efficiently using the identity $\cos W = U^\dagger(\cos D_W)U$ and the fact that $\cos D_W$ is simply the cosine of each element of the diagonal matrix D_W.

The spectra with cutoffs $N_L = 100$ and $N_L = 500$ are shown in the main text Fig. 3. Fig. 3 shows that the higher cutoff N_L is, the clearer the Hofstadter butterfly is. Furthermore, the cutoff N_L breaks the periodicity of the Hofstadter spectrum with respect to magnetic flux per unit cell φ. Fig. 5 shows the first 4 periods of the numerical Hofstadter butterfly with $N_L = 100$. In particular, the larger φ is, the less clear the Hofstadter butterfly is. This is because in Eq. (97) the operator $\hat{\boldsymbol{\kappa}} \propto \ell^{-1}(a, a^\dagger)$; for a fixed cutoff N_L, the error in the LL operators $(a, a^\dagger)$ is fixed, so the error in operator $\hat{\boldsymbol{\kappa}}$ is larger for larger φ (which corresponds to smaller ℓ).

Lastly, we note that throughout our derivation, we have not used the details of the Wannier function $W_\alpha(\mathbf{r})$ in Eq. (77). Therefore, our derivation holds generically as long as the Peierls substitution is valid, independent of the shape of the Wannier orbitals.

B. Nonstandard Peierls substitution

In some models the Wannier orbitals cannot be approximated as localized at one position (but dominantly localizes at several positions respecting the symmetries of the lattice), so the standard Peierls approximations are no longer valid. However, in certain models, nonstandard Peierls substitution can be derived under certain approximations

[12]. Accordingly, the Peierls substitution might not be along the straight line segment path $c_{\alpha\beta}$, but may be along a nonstraight path or the sum of multiple nonstraight paths from $\boldsymbol{D} + \boldsymbol{u}_\beta$ to $\boldsymbol{D} + \boldsymbol{D}_j + \boldsymbol{u}_\alpha$. For instance, in the 4-band tight-binding model for TBG in Ref. [29], the Wannier orbital is centered at AB and BA stackings of TBG but extends to the 3 closest AA stacking centers, and it is shown to have an approximate nonstandard Peierls substitution given by the summation of contributions of multiple broken-line paths [12].

For such a tight-binding model with nonstandard Peierls substitution, the tight-binding Hamiltonian is still of the form of Eq. (79), except that the translation operator in Eq. (79) now reads

$$T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta}^{\alpha\beta} = \sum_{\boldsymbol{D},\mu} e^{i\int_{c_{j,\alpha\beta}^\mu} \mathbf{A}(\boldsymbol{D}+\mathbf{r})\cdot\mathrm{d}\mathbf{r}} |\boldsymbol{D} + \boldsymbol{D}_j, \alpha\rangle\langle\boldsymbol{D}, \beta| \ , \tag{98}$$

where the phase factor involves the summation of multiple different paths (e.g., broken-lines) $c_{j,\alpha\beta}^\mu$ (labeled by index $\mu = 1, 2, \cdots$) from position $\boldsymbol{u}_\beta$ to $\boldsymbol{D}_j + \boldsymbol{u}_\alpha$. For instance, in the 4-band tight-binding model for TBG studied in [12], the Peierls substitution between the nearest neighbors is given by the summation of the gauge phase factor of 2 different paths (i.e., $\mu = 1, 2$) from one AB site to another AB site via the nearest 2 AA sites.

Our method can still apply to such tight-binding models with nonstandard Peierls substitutions. To see this, we first note that throughout our derivations in Sec. IV A for the standard Peierls substitution case, we do not require at all the Wannier orbitals in Eq. (77) to be localized. Therefore, in the nonstandard Peierls substitution case here, we can still define the eigenbasis by Eq. (86), which still satisfies the orthogonality in Eq. (89) and the completeness in Eq. (91).

The action of the translation operator is still given by Eq. (94), except that one need to sum over all the paths $c_{j,\alpha\beta}^\mu$ (i.e., replace $c_{j,\alpha\beta}$ in Eq. (94) by $c_{j,\alpha\beta}^\mu$ and sum over the path index μ). However, since $c_{j,\alpha\beta}^\mu$ are no longer straight paths, Eq. (94) cannot be further reduced to Eq. (95). Therefore, the translation operator in a fixed λ sector from basis $\overline{|\lambda, n, \beta\rangle}$ to $\overline{|\lambda, n', \alpha\rangle}$ can be expressed by Eq. (94) as

$$T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta}^{\lambda,\alpha\beta} = \sum_\mu \mathcal{P} e^{-i\int_{c_{\alpha\beta}^\mu}(\hat{\boldsymbol{\kappa}}+\mathbf{k}_0)\cdot\mathrm{d}\mathbf{r}} \ , \tag{99}$$

where $\hat{\boldsymbol{\kappa}} = \frac{1}{\sqrt{2}\ell}(a+a^\dagger, -ia+ia^\dagger)$ acts on the LL quantum number n, and $\mathcal{P}$ stands for path ordering. The Hamiltonian in a fixed λ subspace then reads $H^{\lambda,\alpha\beta} = \sum_j t_j^{\alpha\beta} T_{\boldsymbol{D}_j + \boldsymbol{u}_\alpha - \boldsymbol{u}_\beta}^{\lambda,\alpha\beta}$, the matrix elements of which are independent of λ.

V. REVIEW OF THE DIOPHANTINE EQUATION

In this section, we briefly review the proof of the Diophantine equation.

For a lattice model with magnetic flux $\varphi = 2\pi\frac{p}{q}$ per unit cell, where p and q are two coprime numbers, the energy spectrum forms a set of Hofstadter bands. In particular, each Hofstadter gap is characterized by two integers t_ν and s_ν, which satisfy the Diophantine equation

$$t_\nu p + s_\nu q = \nu \ , \tag{100}$$

where ν is an integer. In particular, t_ν is the Chern number of the Hofstadter gap.

Here we briefly review how the Diophantine equation is proved following Ref. [15] (See also Ref. [30] Chapter 5.3 for a different proof). In fact, one can prove an equivalent statement, that each Hofstadter band satisfies a Diophantine equation

$$\sigma p + mq = 1 \ , \tag{101}$$

where σ is the Chern number of the Hofstadter band, and m is another integer characterizing the band.

The proof is as follows. First, we note that a lattice Hamiltonian H (either a continuum model in Eq. (27) or a tight-binding model as in Eq. (79) we considered) in a uniform magnetic field $\mathbf{B} = B\hat{\mathbf{z}}$ still has translation symmetries along the Bravais lattice vectors $\boldsymbol{d}_1$ and $\boldsymbol{d}_2$. However, the translation symmetry operators are not the simple translation operators

$$T_{\boldsymbol{d}_j} = e^{-i\mathbf{\Pi}\cdot\boldsymbol{d}_j} \qquad (j = 1, 2) \tag{102}$$

in magnetic field in the continuum space, since the operator $T_{\boldsymbol{d}_j}$ do not commute with Hamiltonian H as one can easily verify (since H contains operator $\mathbf{\Pi}$, and $[\Pi_x, \Pi_y] \neq 0$). Instead, the (magnetic) translation symmetry operators which commute with the Hamiltonian H are given by (for simplicity here we choose all the $\mathbf{p}_\alpha = \mathbf{0}$ in Eq. (29)):

$$\widetilde{T}_{\boldsymbol{d}_j} = T_{\boldsymbol{d}_j} e^{-i(\ell^{-2}\hat{\mathbf{z}}\times\mathbf{r})\cdot\boldsymbol{d}_j} = e^{-i(\mathbf{\Pi}+\ell^{-2}\hat{\mathbf{z}}\times\mathbf{r})\cdot\boldsymbol{d}_j} = e^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}_j} \ , \qquad [\widetilde{T}_{\boldsymbol{d}_j}, H] = 0 \ , \qquad (j = 1, 2) \tag{103}$$
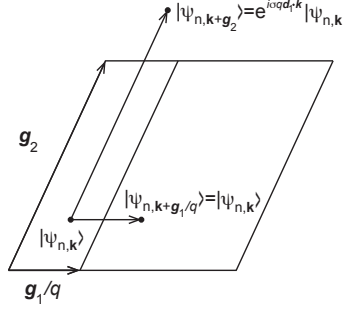
FIG. 6. Illustration of the gauge choices in Eq. (106) for proving the Diophantine equation, where the magnetic flux is $\varphi/2\pi = p/q$.

where $\mathbf{R}$ is the guiding center operator. Namely, the translation operator $T_{\boldsymbol{d}_j}$ and the translation symmetry operator $\widetilde{T}_{\boldsymbol{d}_j}$ differ by a unitary transformation $e^{-i(\ell^{-2}\hat{\mathbf{z}}\times\mathbf{r})\cdot\boldsymbol{d}_j}$. At zero magnetic field, $T_{\boldsymbol{d}_j}$ and $\widetilde{T}_{\boldsymbol{d}_j}$ become the same.

At rational flux $\varphi = 2\pi\frac{p}{q}$ per unit cell, it is straightforward to show that the translation symmetry operators satisfy the commutation relation

$$\widetilde{T}_{\boldsymbol{d}_1}\widetilde{T}_{\boldsymbol{d}_2} = e^{-i2\pi p/q}\widetilde{T}_{\boldsymbol{d}_2}\widetilde{T}_{\boldsymbol{d}_1} \ . \tag{104}$$

Note that in contrast we have $T_{\boldsymbol{d}_1}T_{\boldsymbol{d}_2} = e^{i2\pi p/q}T_{\boldsymbol{d}_2}T_{\boldsymbol{d}_1}$. One can therefore define two magnetic translation symmetry operators $\widetilde{T}_{q\boldsymbol{d}_1} = (\widetilde{T}_{\boldsymbol{d}_1})^q$ and $\widetilde{T}_{\boldsymbol{d}_2}$ which commute with each other, i.e., $[\widetilde{T}_{q\boldsymbol{d}_1}, \widetilde{T}_{\boldsymbol{d}_2}] = 0$. Since they also commute with the Hamiltonian H, we can define the Bloch wave function eigenstates $|\psi_{n,\mathbf{k}}\rangle$ of a Hofstadter band n of H, with quasimomentum $\mathbf{k}$ defined by

$$\widetilde{T}_{q\boldsymbol{d}_1}|\psi_{n,\mathbf{k}}\rangle = e^{iq\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}}\rangle \ , \qquad \widetilde{T}_{\boldsymbol{d}_2}|\psi_{n,\mathbf{k}}\rangle = e^{i\boldsymbol{d}_2\cdot\mathbf{k}}|\psi_{n,\mathbf{k}}\rangle \ . \tag{105}$$

The magnetic BZ is then a parallelogram spanned by momentum vectors $\boldsymbol{g}_1/q$ and $\boldsymbol{g}_2$, where $\boldsymbol{g}_i$ satisfies $\boldsymbol{g}_i\cdot\boldsymbol{d}_j = 2\pi\delta_{ij}$ $(i,j = 1,2)$. For simplicity, we shall assume all the Hofstadter bands are nondegenerate in one magnetic BZ (namely, at each momentum $\mathbf{k}$ in the magnetic BZ, the energy eigenvalues are nondegenerate), which is generically true when there is no other symmetries (which may protect degeneracies) except for the translation symmetries. When the Hofstadter band n has a Chern number σ, one can choose the Bloch wave function $|\psi_{n,\mathbf{k}}\rangle$ as a continuous function of $\mathbf{k}$ satisfying

$$|\psi_{n,\mathbf{k}+\boldsymbol{g}_1/q}\rangle = |\psi_{n,\mathbf{k}}\rangle \ , \qquad |\psi_{n,\mathbf{k}+\boldsymbol{g}_2}\rangle = e^{i\sigma q\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}}\rangle \ , \tag{106}$$

as illustrated in Fig. 6. Here we do not restrict $\mathbf{k}$ within the magnetic BZ. One can easily verify that the above choice gives a Berry phase $\sigma q\boldsymbol{d}_1\cdot(\boldsymbol{g}_1/q) = 2\pi\sigma$ circulating the boundary of the magnetic BZ (starting from momentum $(\mathbf{0},\mathbf{0})$ to $(\mathbf{0},\boldsymbol{g}_2)$ to $(\boldsymbol{g}_1/q,\boldsymbol{g}_2)$ to $(\boldsymbol{g}_1/q,\mathbf{0})$ and then back to $(\mathbf{0},\mathbf{0})$), which is required by the Chern number σ.

By Eq. (104) and the definition of quasimomentum $\mathbf{k}$ in Eq. (105), the state $\widetilde{T}_{\boldsymbol{d}_1}|\psi_{n,\mathbf{k}}\rangle$ should be equal to the Bloch state $|\psi_{n,\mathbf{k}+p\boldsymbol{g}_2/q}\rangle$ up to a phase factor, and should be degenerate with the state $|\psi_{n,\mathbf{k}}\rangle$. In consistency with Eq. (106), we can in general choose the phase such that

$$\widetilde{T}_{\boldsymbol{d}_1}|\psi_{n,\mathbf{k}}\rangle = e^{imq\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}+p\boldsymbol{g}_2/q}\rangle \ , \tag{107}$$

where m is some coefficient to be determined. First, one can prove that $m \in \mathbb{Z}$ is an integer: this is because the first equation in (106) requires $e^{imq\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}+p\boldsymbol{g}_2/q}\rangle = \widetilde{T}_{\boldsymbol{d}_1}|\psi_{n,\mathbf{k}}\rangle = \widetilde{T}_{\boldsymbol{d}_1}|\psi_{n,\mathbf{k}+\boldsymbol{g}_1/q}\rangle = e^{imq\boldsymbol{d}_1\cdot(\mathbf{k}+\boldsymbol{g}_1/q)}|\psi_{n,\mathbf{k}+\boldsymbol{g}_1/q+p\boldsymbol{g}_2/q}\rangle = e^{imq\boldsymbol{d}_1\cdot(\mathbf{k}+\boldsymbol{g}_1/q)}|\psi_{n,\mathbf{k}+p\boldsymbol{g}_2/q}\rangle$, namely, $e^{imq\boldsymbol{d}_1\cdot(\boldsymbol{g}_1/q)} = e^{i2\pi m} = 1$. We can then apply $\widetilde{T}_{\boldsymbol{d}_1}$ by q times to obtain

$$\widetilde{T}_{q\boldsymbol{d}_1}|\psi_{n,\mathbf{k}}\rangle = e^{imq^2\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}+p\boldsymbol{g}_2}\rangle = e^{i(\sigma p+mq)q\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}}\rangle = e^{iq\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}}\rangle \ . \tag{108}$$

Since $\mathbf{k}$ can take any value, we conclude that the Diophantine equation 101 for the band has to hold.

It is then straightforward to see the Diophantine equation 100 holds in a Hofstadter gap: ν is simply the number of Hofstadter bands below the gap counted from some reference point of filling (see Sec. II G for discussion of the reference point of filling). The Hofstadter band between the ν-th gap and the $(\nu-1)$-th gap then has Chern number $\sigma = t_\nu - t_{\nu-1}$, and the other quantum number $m = s_\nu - s_{\nu-1}$.

Since ν has the physical meaning of number of occupied magnetic bands, and the magnetic unit cell has an area q times of the original unit cell, we can define the number of electrons per original unit cell as $\rho = \nu/q$. One can then divide the Diophantine equation by q to obtain our main text Eq. (7):

$$t_\nu \frac{\varphi}{2\pi} + s_\nu = \rho \ , \tag{109}$$

which holds even when the flux quanta $\varphi/2\pi$ is irrational. The derivative of this equation with respect to φ gives the Streda formula $2\pi \frac{dn_r}{d\varphi} = t_r$ [20].

VI. QUANTIZED LORENTZ SUSCEPTIBILITY FROM s_ν

In this section, we show that the quantum number s_ν, which can be viewed as a dual Chern number in the momentum space, yields a quantized Lorentz susceptibility $\gamma_{xy} = eBs_\nu$. This is analogous to the Hall conductance given by the Chern number t_ν.

A. Review of the quantized Hall conductance

We first briefly review how the Chern number t_ν leads to a quantized Hall conductance $\sigma_{xy} = t_\nu \frac{e^2}{h}$ [10]. We assume the flux per unit cell is $\varphi = 2\pi p/q$, so that we can define the magnetic BZ quasimomentum (k_x, k_y) by Eq. (105). Using the Kubo formula, the Hall conductance under a uniform electric field E_y can be written as

$$\sigma_{xy} = i\frac{\partial}{\partial \omega} \int_{-\infty}^{\infty} d\omega' \langle G_{\omega+\omega'} \hat{j}_x G_{\omega'} \hat{j}_y \rangle \Big|_{\omega \to 0} \tag{110}$$

where $\hat{\boldsymbol{j}}$ is the *spatially uniform* current operator which is conjugate to a spatially uniform gauge field $\mathbf{A}$., while $G_\omega = (\omega - H)^{-1}$ is the Green's function in the magnetic field B at energy ω, and H is the (first quantized) full single-particle Hamiltonian H under magnetic field in the real space (under the position basis $|\mathbf{r}, \alpha\rangle$). For example, for continuum models H is as defined in Eq. (34), while for tight-binding models H is as given in Eq. (79) embedded into the continuum space by the infinitely localized Wannier orbital assumption (77). The coefficient σ_{xy} corresponds to a Hall response $j_x = -\sigma_{xy} dA_y/dt = \sigma_{xy} E_y$, where $E_y = -dA_y/dt$ is the uniform electric field in the y direction.

We denote the Bloch eigenstates of H in the n-th band as $|\psi_{n,\mathbf{k}}\rangle$ satisfying

$$H|\psi_{n,\mathbf{k}}\rangle = \epsilon_{n,\mathbf{k}}|\psi_{n,\mathbf{k}}\rangle \ , \tag{111}$$

where the magnetic BZ quasimomentum $\mathbf{k}$ is defined by Eq. (105) using the magnetic translation symmetry operators $\widetilde{T}_{q\boldsymbol{d}_1} = e^{-iq\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}_1}$ and $\widetilde{T}_{\boldsymbol{d}_2} = e^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}_2}$, and $\epsilon_{n,\mathbf{k}}$ is the energy of Bloch state $|\psi_{n,\mathbf{k}}\rangle$. Since H is diagonal in the magnetic BZ quasimomentum $\mathbf{k}$, the Green's function (in the magnetic field B) is also diagonal in $\mathbf{k}$, and can be written as

$$G_\omega = \sum_{n,\mathbf{k}} \frac{|\psi_{n,\mathbf{k}}\rangle\langle\psi_{n,\mathbf{k}}|}{\omega - \epsilon_{n,\mathbf{k}}} \ . \tag{112}$$

We comment that under the continuum space basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ in Eq. (40) employed by our Hofstadter method, the Bloch state $|\psi_{n,\mathbf{k}}\rangle$ does not live in a definite λ sector. This is can be seen by noting that $|\psi_{n,\mathbf{k}}\rangle$ is the eigenstate of $\widetilde{T}_{q\boldsymbol{d}_1}$ and $\widetilde{T}_{\boldsymbol{d}_2}$, while we have $\widetilde{T}_{q\boldsymbol{d}_1} R_{\hat{\boldsymbol{\tau}}} \widetilde{T}_{q\boldsymbol{d}_1}^\dagger = R_{\hat{\boldsymbol{\tau}}} + q\hat{\boldsymbol{\tau}}\cdot\boldsymbol{d}_1 = R_{\hat{\boldsymbol{\tau}}} - p\ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{z}\times\boldsymbol{g}_2)$ and $\widetilde{T}_{\boldsymbol{d}_2} R_{\hat{\boldsymbol{\tau}}} \widetilde{T}_{\boldsymbol{d}_2}^\dagger = R_{\hat{\boldsymbol{\tau}}} + \hat{\boldsymbol{\tau}}\cdot\boldsymbol{d}_2 = R_{\hat{\boldsymbol{\tau}}} + \frac{p}{q}\ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{z}\times\boldsymbol{g}_1)$. Since the quantum number $\lambda \in \Lambda_{\hat{\boldsymbol{\tau}}} = \mathbb{R}/[\ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_1)\mathbb{Z} + \ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{\mathbf{z}}\times\boldsymbol{g}_2)\mathbb{Z}]$ as defined in Eq. (43), we find that the action of $\widetilde{T}_{q\boldsymbol{d}_1}$ preserves the coset of λ, while $\widetilde{T}_{\boldsymbol{d}_2}$ maps the sector of coset λ to the sector of coset $\lambda + \frac{p}{q}\ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{z}\times\boldsymbol{g}_1)$. Therefore, each Bloch state $|\psi_{n,\mathbf{k}}\rangle$ must be the superposition of the degenerate eigenstates in multiple of q sectors labeled by quantum numbers $\lambda + m\frac{p}{q}\ell^2\hat{\boldsymbol{\tau}}\cdot(\hat{z}\times\boldsymbol{g}_1)$ (λ belongs to a certain set, and $m = 0, \cdots, q-1$). However, the wavefunction coefficients under the basis $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ will be complicated, since the basis state $|\lambda, \mathbf{Q}_\alpha, n, \alpha\rangle$ with definite quantum numbers $\lambda, \mathbf{Q}_\alpha$, n is not translationally invariant.

For any state preserving the translation symmetry, the expectation value of the current operator $\hat{\boldsymbol{j}}$ satisfies (recall we have set charge $e = 1$)

$$\langle \hat{\boldsymbol{j}} \rangle = \langle \frac{d\mathbf{r}}{dt} \rangle = -i\langle [\mathbf{r}, H] \rangle \ . \tag{113}$$

The uniform current operator $\hat{\boldsymbol{j}}$ preserves the translation symmetry and is thus diagonal in magnetic BZ quasimomentum $\mathbf{k}$. In contrast, the position operator $\mathbf{r}$ does not commute with translation symmetry operators $\widetilde{T}_{q\boldsymbol{d}_1}$ and $\widetilde{T}_{\boldsymbol{d}_2}$, thus $\mathbf{r}$ is not fully diagonal in magnetic BZ quasimomentum $\mathbf{k}$, and accordingly $[\mathbf{r}, H]$ is not fully diagonal in $\mathbf{k}$ (see Ref. [31]). Therefore, Eq. (113) shows that the current operator $\hat{\boldsymbol{j}}$ is given by the part of $-i[\mathbf{r}, H]$ that is diagonal in magnetic BZ quasimomentum $\mathbf{k}$, namely,

$$\hat{\boldsymbol{j}} = \sum_{m,n,\mathbf{k}} \hat{\boldsymbol{j}}_{m\mathbf{k},n\mathbf{k}} |\psi_{m,\mathbf{k}}\rangle\langle\psi_{n,\mathbf{k}}| , \qquad \hat{\boldsymbol{j}}_{m\mathbf{k},n\mathbf{k}} = -i\langle\psi_{m,\mathbf{k}}|[\mathbf{r}, H]|\psi_{n,\mathbf{k}}\rangle . \tag{114}$$

For a given magnetic BZ quasimomentum $\mathbf{k}$ (within the magnetic BZ spanned by vector $\boldsymbol{g}_1/q$ and $\boldsymbol{g}_2$), we can define a reduced magnetic BZ momentum space Hamiltonian $\hat{H}(\mathbf{k})$ (we label it by a hat, since it is distinguished from the full Hamiltonian H in field in Eq. (34) for continuum models or in Eq. (79) for tight-binding models):

$$\hat{H}(\mathbf{k}) = P_{\mathbf{0}}(e^{-i\mathbf{k}\cdot\mathbf{r}} H e^{i\mathbf{k}\cdot\mathbf{r}})P_{\mathbf{0}} , \qquad P_{\mathbf{0}} = \frac{q}{N_\Omega} \sum_{j_1,j_2\in\mathbb{Z}} \widetilde{T}_{qj_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2} , \tag{115}$$

where $\mathbf{k}$ is understood as a parameter, $\mathbf{r}$ is the position operator (understood as $\sum_\alpha \int \mathbf{r}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha|\mathrm{d}^2\mathbf{r}$ when expanded in position basis, with α being the orbital), $P_{\mathbf{0}}$ is the projection operator into the magnetic translationally invariant subspace, $\widetilde{T}_{\boldsymbol{d}} = e^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}$ is the translation operator of displacement $\boldsymbol{d}$ by the guiding center operator $\mathbf{R}$ as defined in Eq. (103), and N_Ω is the number of zero-magnetic-field unit cells (there are N_Ω/q magnetic unit cells at magnetic flux per unit cell $\varphi = 2\pi p/q$). Note that any state with a nonzero momentum $\mathbf{k} \neq 0$ in the magnetic BZ, for which the translation operator $\widetilde{T}_{qj_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2}$ has an eigenvalue $e^{i(qj_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2)\cdot\mathbf{k}}$, will be annihilated by $P_{\mathbf{0}}$ (since $\sum_{j_1,j_2} e^{i(qj_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2)\cdot\mathbf{k}} = \frac{N_\Omega}{q}\delta_{\mathbf{k},\mathbf{0}}$). Thus, unlike the full Hamiltonian H in the real space which has a large Hilbert space dimension equal to the number of bands times the number of $\mathbf{k}$ in the magnetic BZ, the reduced Hamiltonian $\hat{H}(\mathbf{k})$ is defined at a fixed $\mathbf{k}$ and has a smaller Hilbert dimension equal to the number of bands, which can be understood as the full Hamiltonian H projected into the sub-Hilbert space at magnetic BZ momentum $\mathbf{k}$. An explicit example of $\hat{H}(\mathbf{k})$ for a tight-binding model is given in Eqs. (158) and (159), where one can see $\hat{H}(\mathbf{k})$ is simply the magnetic BZ momentum space Hamiltonian under the Fourier transformed basis of the Wannier orbitals. The Hamiltonian $\hat{H}(\mathbf{k})$ has eigenstates

$$|u_{n,\mathbf{k}}\rangle = \sum_\alpha \int \mathrm{d}^2\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha|\psi_{n,\mathbf{k}}\rangle , \qquad \hat{H}(\mathbf{k})|u_{n,\mathbf{k}}\rangle = \epsilon_{n,\mathbf{k}}|u_{n,\mathbf{k}}\rangle , \tag{116}$$

where n runs over all the bands in the magnetic BZ. Note that the state $|u_{n,\mathbf{k}}\rangle$ has is magnetically translationally invariant, namely, $\widetilde{T}_{q\boldsymbol{d}_1}|u_{n,\mathbf{k}}\rangle = \widetilde{T}_{\boldsymbol{d}_2}|u_{n,\mathbf{k}}\rangle = |u_{n,\mathbf{k}}\rangle$. Therefore, $P_{\mathbf{0}}|u_{n,\mathbf{k}}\rangle = |u_{n,\mathbf{k}}\rangle$, and $|u_{n,\mathbf{k}}\rangle$ can be understood as the periodic part of the Bloch wave function $|\psi_{n,\mathbf{k}}\rangle$. In particular, from the definition of $\hat{H}(\mathbf{k})$ in Eq. (115) and the definition of current operator $\hat{\boldsymbol{j}}$ in Eq. (114), we have (calculated by inserting the position basis)

$$\begin{aligned}
\langle\psi_{m,\mathbf{k}}|\hat{\boldsymbol{j}}|\psi_{n,\mathbf{k}}\rangle &= -i\sum_{\alpha,\alpha'}\int \mathrm{d}^2\mathbf{r}d^2\mathbf{r}'\langle\psi_{m,\mathbf{k}}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha|(\mathbf{r}H - H\mathbf{r}')|\mathbf{r}', \alpha'\rangle\langle\mathbf{r}', \alpha'|\psi_{n,\mathbf{k}}\rangle \\
&= -i\sum_{\alpha,\alpha'}\int \mathrm{d}^2\mathbf{r}d^2\mathbf{r}'\langle u_{m,\mathbf{k}}|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha|e^{-i\mathbf{k}\cdot\mathbf{r}}(\mathbf{r}H - H\mathbf{r}')e^{i\mathbf{k}\cdot\mathbf{r}'}|\mathbf{r}', \alpha'\rangle\langle\mathbf{r}', \alpha'|u_{n,\mathbf{k}}\rangle \\
&= \langle u_{m,\mathbf{k}}|\frac{\partial}{\partial\mathbf{k}}\left(P_{\mathbf{0}}\int \mathrm{d}^2\mathbf{r}d^2\mathbf{r}'|\mathbf{r}, \alpha\rangle\langle\mathbf{r}, \alpha|e^{-i\mathbf{k}\cdot\mathbf{r}}H e^{i\mathbf{k}\cdot\mathbf{r}'}|\mathbf{r}', \alpha'\rangle\langle\mathbf{r}', \alpha'|P_{\mathbf{0}}\right)|u_{n,\mathbf{k}}\rangle \\
&= \langle u_{m,\mathbf{k}}|\frac{\partial\hat{H}(\mathbf{k})}{\partial\mathbf{k}}|u_{n,\mathbf{k}}\rangle ,
\end{aligned} \tag{117}$$

where we have used the first equation in (116) and the fact that $P_{\mathbf{0}}|u_{n,\mathbf{k}}\rangle = |u_{n,\mathbf{k}}\rangle$. Therefore, by carrying out the ω' integral in Eq. (110), we arrive at

$$\begin{aligned}
\sigma_{xy} &= i\frac{e^2}{\hbar}\sum_{n,m}\int_{\text{MBZ}}\frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\frac{n_F(\epsilon_{n,\mathbf{k}}) - n_F(\epsilon_{m,\mathbf{k}})}{(\epsilon_{m,\mathbf{k}} - \epsilon_{n,\mathbf{k}})^2}\langle\psi_{n,\mathbf{k}}|\hat{j}_x|\psi_{m,\mathbf{k}}\rangle\langle\psi_{m,\mathbf{k}}|\hat{j}_y|\psi_{n,\mathbf{k}}\rangle \\
&= i\frac{e^2}{\hbar}\sum_{n,m}\int_{\text{MBZ}}\frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\frac{n_F(\epsilon_{n,\mathbf{k}}) - n_F(\epsilon_{m,\mathbf{k}})}{(\epsilon_{m,\mathbf{k}} - \epsilon_{n,\mathbf{k}})^2}\langle u_{n,\mathbf{k}}|\frac{\partial\hat{H}(\mathbf{k})}{\partial k_x}|u_{m,\mathbf{k}}\rangle\langle u_{m,\mathbf{k}}|\frac{\partial\hat{H}(\mathbf{k})}{\partial k_y}|u_{n,\mathbf{k}}\rangle \\
&= i\frac{e^2}{\hbar}\sum_{n\in\text{occ},m\notin\text{occ}}\int_{\text{MBZ}}\frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\frac{\langle u_{n,\mathbf{k}}|\frac{\partial\hat{H}(\mathbf{k})}{\partial k_x}|u_{m,\mathbf{k}}\rangle\langle u_{m,\mathbf{k}}|\frac{\partial\hat{H}(\mathbf{k})}{\partial k_y}|u_{n,\mathbf{k}}\rangle - \langle u_{n,\mathbf{k}}|\frac{\partial\hat{H}(\mathbf{k})}{\partial k_y}|u_{m,\mathbf{k}}\rangle\langle u_{m,\mathbf{k}}|\frac{\partial\hat{H}(\mathbf{k})}{\partial k_x}|u_{n,\mathbf{k}}\rangle}{(\epsilon_{m,\mathbf{k}} - \epsilon_{n,\mathbf{k}})^2} ,
\end{aligned} \tag{118}$$

where the integral of $\mathbf{k}$ runs over one magnetic BZ (MBZ), $n_F(\epsilon) = \frac{1-\mathrm{sgn}(\epsilon)}{2}$ is the zero temperature Fermi-Dirac distribution function, $n \in$ occ means n runs over all the occupied bands, and $m \notin$ occ means m runs over all the empty bands. Since we are in a Hofstadter gap, $n_F(\epsilon_{n,\mathbf{k}})$ only depends on the band index n. Then, using the fact that

$$
\langle u_{m,\mathbf{k}}|\frac{\partial \hat{H}(\mathbf{k})}{\partial k_i}|u_{n,\mathbf{k}}\rangle = \partial_{k_i}(\langle u_{m,\mathbf{k}}|\hat{H}(\mathbf{k})|u_{n,\mathbf{k}}\rangle) - \langle \partial_{k_i}u_{m,\mathbf{k}}|\hat{H}(\mathbf{k})|u_{n,\mathbf{k}}\rangle - \langle u_{m,\mathbf{k}}|\hat{H}(\mathbf{k})|\partial_{k_i}u_{n,\mathbf{k}}\rangle
$$
$$
= 0 - \epsilon_{n,\mathbf{k}}\langle \partial_{k_i}u_{m,\mathbf{k}}|u_{n,\mathbf{k}}\rangle - \epsilon_{m,\mathbf{k}}\langle u_{m,\mathbf{k}}|\partial_{k_i}u_{n,\mathbf{k}}\rangle = (\epsilon_{n,\mathbf{k}} - \epsilon_{m,\mathbf{k}})\langle u_{m,\mathbf{k}}|\partial_{k_i}u_{n,\mathbf{k}}\rangle = (\epsilon_{m,\mathbf{k}} - \epsilon_{n,\mathbf{k}})\langle \partial_{k_i}u_{m,\mathbf{k}}|u_{n,\mathbf{k}}\rangle ~,
$$
(119)

we can rewrite the Hall conductance as the TKNN formula [10]

$$
\sigma_{xy} = i\frac{e^2}{\hbar}\sum_{n\in\mathrm{occ}}\sum_{m\notin\mathrm{occ}}\int_{\mathrm{MBZ}}\frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\left(\langle \partial_{k_x}u_{n,\mathbf{k}}|u_{m,\mathbf{k}}\rangle\langle u_{m,\mathbf{k}}|\partial_{k_y}u_{n,\mathbf{k}}\rangle - \langle \partial_{k_y}u_{n,\mathbf{k}}|u_{m,\mathbf{k}}\rangle\langle u_{m,\mathbf{k}}|\partial_{k_x}u_{n,\mathbf{k}}\rangle\right)
$$
$$
= i\frac{e^2}{\hbar}\sum_{n\in\mathrm{occ}}\sum_{m}\int_{\mathrm{MBZ}}\frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\left(\langle \partial_{k_x}u_{n,\mathbf{k}}|u_{m,\mathbf{k}}\rangle\langle u_{m,\mathbf{k}}|\partial_{k_y}u_{n,\mathbf{k}}\rangle - \langle \partial_{k_y}u_{n,\mathbf{k}}|u_{m,\mathbf{k}}\rangle\langle u_{m,\mathbf{k}}|\partial_{k_x}u_{n,\mathbf{k}}\rangle\right)
$$
$$
= i\frac{e^2}{\hbar}\sum_{n\in\mathrm{occ}}\int_{\mathrm{MBZ}}\frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\left(\langle \partial_{k_x}u_{n,\mathbf{k}}|\partial_{k_y}u_{n,\mathbf{k}}\rangle - \langle \partial_{k_y}u_{n,\mathbf{k}}|\partial_{k_x}u_{n,\mathbf{k}}\rangle\right)
$$
$$
= \frac{e^2}{\hbar}\sum_{n\in\mathrm{occ}}\int_{\mathrm{MBZ}}\frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\mathcal{F}^n_{xy}(\mathbf{k}) = t_\nu\frac{e^2}{h} ~,
$$
(120)

where we have used the fact that t_ν in the Streda formula (Eq. (109)) is the total Chern number of occupied bands, $\sum_m |u_{m,\mathbf{k}}\rangle\langle u_{m,\mathbf{k}}|$ is the identity matrix for the hatted Hamiltonian $\hat{H}(\mathbf{k})$ at magnetic BZ quasimomentum $\mathbf{k}$, and $\mathcal{F}^n_{xy}(\mathbf{k})$ is the Berry curvature of the n-th band. The corresponding U(1) Berry gauge field is

$$
\mathcal{A}^n(\mathbf{k}) = i\langle u_{n,\mathbf{k}}|\partial_{\mathbf{k}}u_{n,\mathbf{k}}\rangle ~, \qquad \mathcal{F}^n_{xy}(\mathbf{k}) = \partial_x\mathcal{A}^n_y(\mathbf{k}) - \partial_y\mathcal{A}^n_x(\mathbf{k}) ~.
$$
(121)

B. Quantized Lorentz susceptibility from the Kubo formula

We now try to find a quantity that can be viewed as the momentum space dual of the Hall conductance. The Hall conductance gives a current density $j_x = \partial H/\partial k_x$ in response to a uniform electric field $E_y = \mathrm{d}k_y/\mathrm{d}t$ where $\mathbf{k}$ is the canonical momentum. Therefore, a natural dual response can be defined by exchanging the roles of the real space position $\mathbf{r}$ and the momentum $\mathbf{k}$. Such a position-momentum dual response (dual to the Hall conductance response) corresponds to a force density (force per unit cell) $F_x = -\partial H/\partial x$ acting on the lattice in response to a position pumping, i.e., velocity $v_y = \mathrm{d}y/\mathrm{d}t$ of the system, which is nothing but the Lorentz force. Therefore, we name the coefficient γ_{xy} in the response

$$
F_x = \gamma_{xy}v_y
$$
(122)

as the Lorentz susceptibility.

We first identify the spatially uniform force operator $\hat{\boldsymbol{F}}$ representing the Lorentz force felt by the electrons as the lattice moves rigidly. Recall that an electron in a magnetic field is centered at the guiding center $\mathbf{R}$. In a translationally invariant state, if the guiding centers $\mathbf{R}$ of the electrons drift at velocity $\langle\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t}\rangle$, the electrons will feel a Lorentz force

$$
\langle\hat{\boldsymbol{F}}\rangle = B\hat{\mathbf{z}}\times\langle\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t}\rangle = -iB\hat{\mathbf{z}}\times\langle[\mathbf{R},H]\rangle = -i\langle[\mathbf{\Pi},H]\rangle - iB\hat{\mathbf{z}}\times\langle[\mathbf{r},H]\rangle ~,
$$
(123)

where in the last equality we have used the definition of guiding center $\mathbf{R} = \mathbf{r} - \ell^2\hat{\mathbf{z}}\times\mathbf{\Pi}$. For adiabatic processes, the electrons are in equilibrium, so the Lorentz force $\langle\hat{\boldsymbol{F}}\rangle$ the electrons felt has to be balanced by a force $-\langle\hat{\boldsymbol{F}}\rangle$ the lattice exerts on the electrons. According to Newton's third law, the electrons will then exert a force $\langle\hat{\boldsymbol{F}}\rangle$ on the lattice. Therefore, experimentally one could measure the force felt by the lattice to find the Lorentz force $\langle\hat{\boldsymbol{F}}\rangle$ felt by the electrons.

Since the force operator $\hat{\boldsymbol{F}}$ is spatially uniform and preserves the translation symmetry, it is diagonal in magnetic BZ quasimomentum $\mathbf{k}$. On the right hand side of Eq. (123), the operator $[\mathbf{\Pi}, H]$ is readily diagonal in magnetic BZ quasi-momentum $\mathbf{k}$, since $\mathbf{\Pi}$ and hence H both commute with the translation symmetry operators $\widetilde{T}_{q\boldsymbol{d}_1} = e^{-iq\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}_1}$

and $\widetilde{T}_{d_2} = e^{-i\ell^{-2}(\hat{\mathbf{z}} \times \mathbf{R}) \cdot d_2}$. The other operator $[\mathbf{r}, H]$ is not diagonal in $\mathbf{k}$ (Ref. [31]), but its diagonal part is simply the current operator $\hat{\boldsymbol{j}}$ as we discussed in Sec. VI A. Therefore, we conclude the force operator is given by

$$\hat{\boldsymbol{F}} = -i[\boldsymbol{\Pi}, H] + B\hat{\mathbf{z}} \times \hat{\boldsymbol{j}} , \tag{124}$$

which is diagonal in magnetic BZ quasimomentum $\mathbf{k}$. Note that here H is the full Hamiltonian defined in Eq. (34) for continuum models or in Eq. (79) for tight-binding models (embedded in the continuum space).

In particular, if there is no periodic lattice potential for the electrons at all, H will be solely a function of $\boldsymbol{\Pi}$ (i.e., the kinetic term of electrons in the free space), and one will have $\langle \hat{\boldsymbol{F}} \rangle = -iB\hat{\mathbf{z}} \times \langle [\mathbf{R}, H] \rangle \equiv \mathbf{0}$ since $[\mathbf{R}, \boldsymbol{\Pi}] = 0$. Physically, this is because without a lattice potential, the electrons cannot feel a force in respond to the movement of the lattice and thus cannot exert a force on the lattice.

Similar to σ_{xy} which is given by the Kubo formula of the spatially uniform current operator $\hat{\boldsymbol{j}}$, we can define the Lorentz susceptibility in Eq. (122) by the Kubo formula of the spatially uniform force operator $\hat{\boldsymbol{F}}$ as

$$\gamma_{xy} = -i\frac{\partial}{\partial \omega} \int_{-\infty}^{\infty} \mathrm{d}\omega' \langle G_{\omega+\omega'} \hat{F}_x G_{\omega'} \hat{F}_y \rangle , \tag{125}$$

where G_ω is the Green's function defined in Eq. (112).

In the below, we first calculate the quantized value of the Lorentz susceptibility in Sec. VI B 1, and then derive a dual formula for the Lorentz susceptibility analogous to the TKNN formula for Hall conductance in Sec. VI B 2.

1. Calculation of the Lorentz susceptibility from the Diophantine equation

By the expression of $\hat{\boldsymbol{F}}$ in Eq. (124), we can rewrite the Lorentz susceptibility defined in Eq. (125) as

$$
\begin{aligned}
\gamma_{xy} = & -ie\Omega \sum_{n \in \mathrm{occ}, m \notin \mathrm{occ}} \int_{\mathrm{MBZ}} \frac{\mathrm{d}^2\mathbf{k}}{4\pi^2} \frac{1}{(\epsilon_{m,\mathbf{k}} - \epsilon_{n,\mathbf{k}})^2} \left[\langle \psi_{n,\mathbf{k}} | \hat{F}_x | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | \hat{F}_y | \psi_{n,\mathbf{k}} \rangle - (x \leftrightarrow y) \right] \\
= & -ie\Omega \sum_{n \in \mathrm{occ}, m \notin \mathrm{occ}} \frac{1}{(\epsilon_{m,\mathbf{k}} - \epsilon_{n,\mathbf{k}})^2} \int_{\mathrm{MBZ}} \frac{\mathrm{d}^2\mathbf{k}}{4\pi^2} \left[-B^2 \left(\langle \psi_{n,\mathbf{k}} | \hat{j}_y | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | \hat{j}_x | \psi_{n,\mathbf{k}} \rangle - (x \leftrightarrow y) \right) \right. \\
& - iB \left(\langle \psi_{n,\mathbf{k}} | [\Pi_x, H] | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | \hat{j}_x | \psi_{n,\mathbf{k}} \rangle - \langle \psi_{n,\mathbf{k}} | \hat{j}_x | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | [\Pi_x, H] | \psi_{n,\mathbf{k}} \rangle \right. \\
& \left. - \langle \psi_{n,\mathbf{k}} | [\Pi_y, H] | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | \hat{j}_y | \psi_{n,\mathbf{k}} \rangle + \langle \psi_{n,\mathbf{k}} | \hat{j}_y | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | [\Pi_y, H] | \psi_{n,\mathbf{k}} \rangle \right) \\
& \left. - \langle \psi_{n,\mathbf{k}} | [\Pi_x, H] | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | [\Pi_y, H] | \psi_{n,\mathbf{k}} \rangle + \langle \psi_{n,\mathbf{k}} | [\Pi_x, H] | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | [\Pi_y, H] | \psi_{n,\mathbf{k}} \rangle \right] ,
\end{aligned}
\tag{126}
$$

where the integral of $\mathbf{k}$ runs over one magnetic BZ (MBZ). Note that so far everything is expressed in the eigenbasis $|\psi_{m,\mathbf{k}}\rangle$ of the full Hamiltonian H. The first two terms of correlations of $\hat{j}_x$ and $\hat{j}_y$ in Eq. (126) simply yield the Berry curvature, which can be calculated by transforming from the eigenbasis $|\psi_{m,\mathbf{k}}\rangle$ of the full Hamiltonian H into the reduced basis $|u_{m,\mathbf{k}}\rangle$ of the reduced magnetic BZ momentum space Hamiltonian $\hat{H}(\mathbf{k})$ as we have shown in Sec. VI A. We now examine the remaining terms of Eq. (126) (which can be calculated simply in the eigenbasis $|\psi_{m,\mathbf{k}}\rangle$ of the full Hamiltonian H). Since $[\boldsymbol{\Pi}, H]$ is diagonal in $\mathbf{k}$, and $\hat{\boldsymbol{j}}$ is the part of operator $-i[\mathbf{r}, H]$ that is diagonal in $\mathbf{k}$, we can rewrite the terms of correlations of $[\Pi_x, H]$ and $\hat{j}_x$ as

$$
\begin{aligned}
& \sum_{n \in \mathrm{occ}, m \notin \mathrm{occ}} \frac{\langle \psi_{n,\mathbf{k}} | [\Pi_x, H] | \psi_{m,\mathbf{k}} \rangle \langle \psi_{m,\mathbf{k}} | \hat{j}_x | \psi_{n,\mathbf{k}} \rangle}{(\epsilon_{m,\mathbf{k}} - \epsilon_{n,\mathbf{k}})^2} - \left([\Pi_x, H] \leftrightarrow \hat{j}_x \right) \\
= & \sum_{n \in \mathrm{occ}} \sum_{m \notin \mathrm{occ}, \mathbf{k}'} \frac{\langle \psi_{n,\mathbf{k}} | [\Pi_x, H] | \psi_{m,\mathbf{k}'} \rangle \langle \psi_{m,\mathbf{k}'} | (-i[x, H]) | \psi_{n,\mathbf{k}} \rangle}{(\epsilon_{m,\mathbf{k}'} - \epsilon_{n,\mathbf{k}})^2} - \left([\Pi_x, H] \leftrightarrow -i[x, H] \right) \\
= & \sum_{n \in \mathrm{occ}} \sum_{m \notin \mathrm{occ}, \mathbf{k}'} \frac{i(\epsilon_{m,\mathbf{k}'} - \epsilon_{n,\mathbf{k}})^2 \langle \psi_{n,\mathbf{k}} | \Pi_x | \psi_{m,\mathbf{k}'} \rangle \langle \psi_{m,\mathbf{k}'} | x | \psi_{n,\mathbf{k}} \rangle}{(\epsilon_{m,\mathbf{k}'} - \epsilon_{n,\mathbf{k}})^2} - \left(\Pi_x \leftrightarrow x \right) \\
= & \sum_{n \in \mathrm{occ}} \sum_{m,\mathbf{k}'} i \langle \psi_{n,\mathbf{k}} | \Pi_x | \psi_{m,\mathbf{k}'} \rangle \langle \psi_{m,\mathbf{k}'} | x | \psi_{n,\mathbf{k}} \rangle - \left(\Pi_x \leftrightarrow x \right) \\
= & \sum_{n \in \mathrm{occ}} i \langle \psi_{n,\mathbf{k}} | [\Pi_x, x] | \psi_{n,\mathbf{k}} \rangle ,
\end{aligned}
\tag{127}
$$

where we have used the fact that $\sum_{m,\mathbf{k}'} |\psi_{m,\mathbf{k}'}\rangle\langle\psi_{m,\mathbf{k}'}|$ is identity for the entire Hilbert space. A similar equality holds for the terms of correlations of $[\Pi_y, H]$ and $\hat{j}_y$. Lastly, by the same derivation technique, we have

$$\sum_{n\in\text{occ},m\notin\text{occ}} \frac{\langle\psi_{n,\mathbf{k}}|[\Pi_x,H]|\psi_{m,\mathbf{k}}\rangle\langle\psi_{m,\mathbf{k}}|[\Pi_y,H]|\psi_{n,\mathbf{k}}\rangle}{(\epsilon_{m,\mathbf{k}}-\epsilon_{n,\mathbf{k}})^2} = -\sum_{n\in\text{occ}} \langle\psi_{n,\mathbf{k}}|[\Pi_x,\Pi_y]|\psi_{n,\mathbf{k}}\rangle \ . \tag{128}$$

Therefore, we can rewrite Eq. (126) as

$$\gamma_{xy} = e\Omega \sum_{n\in\text{occ}} \int_{\text{MBZ}} \frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\Big(-B^2\mathcal{F}_{xy}^n(\mathbf{k}) - iB\langle\psi_{n,\mathbf{k}}|[\Pi_x,x]|\psi_{n,\mathbf{k}}\rangle + iB\langle\psi_{n,\mathbf{k}}|[\Pi_y,y]|\psi_{n,\mathbf{k}}\rangle - i\langle\psi_{n,\mathbf{k}}|[\Pi_x,\Pi_y]|\psi_{n,\mathbf{k}}\rangle \Big)$$

$$= e\Omega \sum_{n\in\text{occ}} \int_{\text{MBZ}} \frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\Big(-B^2\mathcal{F}_{xy}^n(\mathbf{k}) + B \Big) = eB\left(\rho - t_\nu \frac{\varphi}{2\pi}\right) = eBs_\nu \ , \tag{129}$$

where we have used commutation relations $[\Pi_x, x] = [\Pi_y, y] = -i$, $[\Pi_x, \Pi_y] = i\ell^{-2} = iB$, and the Diophantine equation rewritten in the form of Eq. (109). As we discussed in both the main text and the supplementary Sec. V, $\rho = \nu/q$ is the number of electrons per zero-magnetic-field unit cell. This proves that the Lorentz susceptibility is quantized in terms of s_ν, and corresponds to a Lorentz force per original (zero field) unit cell

$$F_x = \gamma_{xy}v_y = eBs_\nu v_y \ . \tag{130}$$

2. A formula for Lorentz susceptibility similar to the TKNN formula for Hall conductance

In Eq. (117), we know the matrix elements of the spatially uniform current operator $\hat{\boldsymbol{j}}$ can be calculated by $\frac{\partial\hat{H}(\mathbf{k})}{\partial\mathbf{k}}$, where $\hat{H}(\mathbf{k})$ is the reduced Hamiltonian defined in Eq. (115). Here for the force operator $\hat{\boldsymbol{F}}$, we can find a similar expression as follows. First, we define another reduced Hamiltonian

$$\hat{\tilde{H}}(\boldsymbol{d}) = P_{\mathbf{0}}(\widetilde{T}_{-\boldsymbol{d}}H\widetilde{T}_{\boldsymbol{d}})P_{\mathbf{0}} = P_{\mathbf{0}}(e^{i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}He^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}})P_{\mathbf{0}} \ , \qquad P_{\mathbf{0}} = \frac{q}{N_\Omega} \sum_{j_1,j_2\in\mathbb{Z}} \widetilde{T}_{qj_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2} \ , \tag{131}$$

where $\widetilde{T}_{\boldsymbol{d}} = \widetilde{T}_{-\boldsymbol{d}}^\dagger = e^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}$ (similar to $e^{i\mathbf{k}\cdot\mathbf{r}}$) is the translation operator generated by guiding center $\mathbf{R}$, and $P_{\mathbf{0}}$ is still the projection operator into the magnetic translationally invariant subspace as defined in Eq. (115). Using the Bloch eigenstates $|\psi_{n,\mathbf{k}}\rangle$ of the full Hamiltonian H, we can define a set of states

$$|w_{n,\boldsymbol{d}}\rangle = \widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \ , \tag{132}$$

where $|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle$ is the Bloch wave function at momentum $\mathbf{k} = \ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}$. More discussions and an explicit example of the Hamiltonian $\hat{\tilde{H}}(\boldsymbol{d})$ and wavefunction $|w_{n,\boldsymbol{d}}\rangle$ are given in Sec. VI C. Making use of the fact that $\widetilde{T}_{\boldsymbol{d}}\widetilde{T}_{\boldsymbol{d}'} = e^{-i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}\times\boldsymbol{d}')}\widetilde{T}_{\boldsymbol{d}'}\widetilde{T}_{\boldsymbol{d}} = e^{-i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}\times\boldsymbol{d}')/2}\widetilde{T}_{\boldsymbol{d}+\boldsymbol{d}'}$ (since $[R_x, R_y] = -i\ell^2$), we have

$$\begin{aligned}
\widetilde{T}_{q\boldsymbol{d}_1}|w_{n,\boldsymbol{d}}\rangle &= \widetilde{T}_{q\boldsymbol{d}_1}\widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{i\ell^{-2}\hat{\mathbf{z}}\cdot(q\boldsymbol{d}_1\times\boldsymbol{d})}\widetilde{T}_{-\boldsymbol{d}}\widetilde{T}_{q\boldsymbol{d}_1}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \\
&= e^{-iq\boldsymbol{d}_1\cdot(\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d})}\widetilde{T}_{-\boldsymbol{d}}e^{iq\boldsymbol{d}_1\cdot(\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d})}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = |w_{n,\boldsymbol{d}}\rangle \ , \\
\widetilde{T}_{\boldsymbol{d}_2}|w_{n,\boldsymbol{d}}\rangle &= \widetilde{T}_{\boldsymbol{d}_2}\widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_2\times\boldsymbol{d})}\widetilde{T}_{-\boldsymbol{d}}\widetilde{T}_{\boldsymbol{d}_2}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \\
&= e^{-i\boldsymbol{d}_2\cdot(\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d})}\widetilde{T}_{-\boldsymbol{d}}e^{i\boldsymbol{d}_2\cdot(\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d})}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = |w_{n,\boldsymbol{d}}\rangle \ ,
\end{aligned} \tag{133}$$

where we have used the definition of Bloch momentum in Eq. (105). Therefore, we find the states $|w_{n,\boldsymbol{d}}\rangle$ are translationally invariant, namely,

$$P_{\mathbf{0}}|w_{n,\boldsymbol{d}}\rangle = |w_{n,\boldsymbol{d}}\rangle \ . \tag{134}$$

It is then easy to see that the eigenstates of the reduced Hamiltonian $\hat{\tilde{H}}(\boldsymbol{d})$ are given by $|w_{n,\boldsymbol{d}}\rangle$, namely,

$$\begin{aligned}
\hat{\tilde{H}}(\boldsymbol{d})|w_{n,\boldsymbol{d}}\rangle &= P_{\mathbf{0}}\widetilde{T}_{-\boldsymbol{d}}H\widetilde{T}_{\boldsymbol{d}}|w_{n,\boldsymbol{d}}\rangle = P_{\mathbf{0}}\widetilde{T}_{-\boldsymbol{d}}H|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \\
&= P_{\mathbf{0}}\widetilde{T}_{-\boldsymbol{d}}\epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = \epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|w_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \ ,
\end{aligned} \tag{135}$$

where $\epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}$ are the eigenenergies at momentum $\mathbf{k} = \ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}$. In particular, we can further prove that the force operator satisfies (recall Eq. (123))

$$
\begin{aligned}
\langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\hat{\boldsymbol{F}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle &= -i\langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|B[\hat{\mathbf{z}}\times\mathbf{R},H]|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \\
&= -i\langle w_{n,\boldsymbol{d}}|\ell^{-2}e^{i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}[\hat{\mathbf{z}}\times\mathbf{R},H]e^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}|w_{n,\boldsymbol{d}}\rangle \\
&= -\langle w_{n,\boldsymbol{d}}|\frac{\partial}{\partial\boldsymbol{d}}\left(P_{\mathbf{0}}e^{i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}He^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}P_0\right)|w_{n,\boldsymbol{d}}\rangle \\
&= -\langle w_{n,\boldsymbol{d}}|\frac{\partial\hat{\tilde{H}}(\boldsymbol{d})}{\partial\boldsymbol{d}}|w_{n,\boldsymbol{d}}\rangle.
\end{aligned}
\tag{136}
$$

Eq. (136) is then the dual analog to Eq. (117) (see Sec. VI C for an explicit example).

We then consider the periodicity of the wave function $|w_{n,\boldsymbol{d}}\rangle$ in the space of parameter $\boldsymbol{d}$. We first note that by the definition of the projection operator $P_{\mathbf{0}}$ in Eq. (131), one has

$$
P_{\mathbf{0}}\widetilde{T}_{\boldsymbol{d}_2/p} = \frac{q}{N_\Omega}\widetilde{T}_{\boldsymbol{d}_2/p}\sum_{j_1,j_2\in\mathbb{Z}}e^{-i\ell^{-2}\hat{\mathbf{z}}\cdot[(qj_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2)\times(\boldsymbol{d}_2/p)]}\widetilde{T}_{qj_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2} = \widetilde{T}_{\boldsymbol{d}_2/p}P_{\mathbf{0}}\ ,
\tag{137}
$$

where we have used the fact that $\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_1\times\boldsymbol{d}_2) = 2\pi p/q$. Therefore, we find

$$
\begin{aligned}
\hat{\tilde{H}}(\boldsymbol{d}+\boldsymbol{d}_2/p) &= P_{\mathbf{0}}(\widetilde{T}_{-\boldsymbol{d}-\boldsymbol{d}_2/p}H\widetilde{T}_{\boldsymbol{d}+\boldsymbol{d}_2/p})P_{\mathbf{0}} = P_{\mathbf{0}}\widetilde{T}_{-\boldsymbol{d}_2/p}\widetilde{T}_{-\boldsymbol{d}}H\widetilde{T}_{\boldsymbol{d}}\widetilde{T}_{\boldsymbol{d}_2/p}P_{\mathbf{0}} = \widetilde{T}_{-\boldsymbol{d}_2/p}(P_{\mathbf{0}}\widetilde{T}_{-\boldsymbol{d}}H\widetilde{T}_{\boldsymbol{d}}P_{\mathbf{0}})\widetilde{T}_{\boldsymbol{d}_2/p} \\
&= \widetilde{T}_{-\boldsymbol{d}_2/p}\hat{\tilde{H}}(\boldsymbol{d})\widetilde{T}_{\boldsymbol{d}_2/p}\ ,
\end{aligned}
\tag{138}
$$

namely, the reduced Hamiltonian $\hat{\tilde{H}}(\boldsymbol{d})$ and $\hat{\tilde{H}}(\boldsymbol{d}+\boldsymbol{d}_2/p)$ differ by a $\boldsymbol{d}$-independent unitary transformation $\widetilde{T}_{\boldsymbol{d}_2/p}$. Moreover, the lattice translation symmetry tells us that

$$
\hat{\tilde{H}}(\boldsymbol{d}+\boldsymbol{d}_1) = P_{\mathbf{0}}(\widetilde{T}_{-\boldsymbol{d}-\boldsymbol{d}_1}H\widetilde{T}_{\boldsymbol{d}+\boldsymbol{d}_1})P_{\mathbf{0}} = P_{\mathbf{0}}\widetilde{T}_{-\boldsymbol{d}}\widetilde{T}_{-\boldsymbol{d}_1}H\widetilde{T}_{\boldsymbol{d}_1}\widetilde{T}_{\boldsymbol{d}}P_{\mathbf{0}} = P_{\mathbf{0}}\widetilde{T}_{-\boldsymbol{d}}H\widetilde{T}_{\boldsymbol{d}}P_{\mathbf{0}} = \hat{\tilde{H}}(\boldsymbol{d}),
\tag{139}
$$

where we have used the fact that the full Hamiltonian satisfies $[\widetilde{T}_{\boldsymbol{d}_1},H] = 0$ (translational symmetry), as given in Eq. (103). Therefore, we conclude that the eigenstates of $\hat{\tilde{H}}(\boldsymbol{d})$ satisfies

$$
|w_{n,\boldsymbol{d}+\boldsymbol{d}_2/p}\rangle = e^{i\varphi_2(\boldsymbol{d})}\widetilde{T}_{-\boldsymbol{d}_2/p}|w_{n,\boldsymbol{d}}\rangle\ ,\qquad |w_{n,\boldsymbol{d}+\boldsymbol{d}_1}\rangle = e^{i\varphi_1(\boldsymbol{d})}|w_{n,\boldsymbol{d}}\rangle\ ,
\tag{140}
$$

where $\varphi_i(\boldsymbol{d})$ are some $\boldsymbol{d}$-dependent phases. This effectively defines a dual "Brillouin zone" Ω_M in the $\boldsymbol{d}$ space with periods $\boldsymbol{d}_1$ and $\boldsymbol{d}_2/p$ (which is a torus). The unitary transformation $\widetilde{T}_{-\boldsymbol{d}_2/p}$ serves as an embedding matrix for mapping state $|w_{n,\boldsymbol{d}}\rangle$ to $|w_{n,\boldsymbol{d}+\boldsymbol{d}_2/p}\rangle$. Note that Ω_M is simple $1/p$ fraction of the zero magnetic field unit cell. Furthermore, this allows us to define a U(1) Berry gauge field and its field strength for band n on the dual "Brillouin zone" Ω_m (which is a closed manifold):

$$
\widetilde{\boldsymbol{a}}^n(\boldsymbol{d}) = -i\langle w_{n,\boldsymbol{d}}|\partial_{\boldsymbol{d}}w_{n,\boldsymbol{d}}\rangle\ ,\qquad \widetilde{f}_{xy}^n(\boldsymbol{d}) = \partial_x\widetilde{a}_y^n(\boldsymbol{d}) - \partial_y\widetilde{a}_x^n(\boldsymbol{d})\ .
\tag{141}
$$

In the following, we will show that the Lorentz susceptibility is given by a formula similar to the TKNN formula [10], but with the usual Berry gauge field in Eq. (121) replaced by the new gauge field we defined in Eq. (141).

We first note that the Lorentz susceptibility in Eq. (126) can be re-expressed in the $\boldsymbol{d}$ space (by variable substitution $\mathbf{k} = \ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}$) as

$$
\begin{aligned}
\gamma_{xy} &= -ie\Omega\sum_{n\in\text{occ},m\notin\text{occ}}\int_{\text{MBZ}}\frac{\mathrm{d}^2\mathbf{k}}{4\pi^2}\frac{1}{(\epsilon_{m,\mathbf{k}}-\epsilon_{n,\mathbf{k}})^2}\left[\langle\psi_{n,\mathbf{k}}|\hat{F}_x|\psi_{m,\mathbf{k}}\rangle\langle\psi_{m,\mathbf{k}}|\hat{F}_y|\psi_{n,\mathbf{k}}\rangle - (x\leftrightarrow y)\right] \\
&= -i\frac{eBp}{2\pi q}\sum_{n\in\text{occ},m\notin\text{occ}}\int_{\boldsymbol{d}\in\mathcal{M}_{\{q\boldsymbol{d}_1/p,\boldsymbol{d}_2/p\}}}\mathrm{d}^2\boldsymbol{d}\frac{\left[\langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\hat{F}_x|\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle\langle\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\hat{F}_y|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle - (x\leftrightarrow y)\right]}{(\epsilon_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}} - \epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}})^2},
\end{aligned}
\tag{142}
$$

where Ω is the zero magnetic field unit cell area, MBZ stands for the magnetic BZ spanned by $\boldsymbol{g}_1/q$ and $\boldsymbol{g}_2$, and $\mathcal{M}_{\{q\boldsymbol{d}_1/p,\boldsymbol{d}_2/p\}}$ denotes a torus with periods $q\boldsymbol{d}_1/p$ and $\boldsymbol{d}_2/p$, which is nothing but the MBZ rotated by $\pi/2$ and size scaled by a factor ℓ^2 (this can be seen by noting that $\boldsymbol{d}_2 = \frac{\Omega}{2\pi}\hat{\mathbf{z}}\times\boldsymbol{g}_1$, and $\boldsymbol{d}_1 = -\frac{\Omega}{2\pi}\hat{\mathbf{z}}\times\boldsymbol{g}_2$, and $B\Omega = \ell^{-2}\Omega = 2\pi p/q$).

Then, by Eq. (107) we know that $\widetilde{T}_{\boldsymbol{d}_1}|\psi_{n,\mathbf{k}}\rangle$ is the same as $|\psi_{n,\mathbf{k}-p\boldsymbol{g}_2/q}\rangle$ up to a phase factor. Since p and q are coprime, there exists an integer j so that $jp \equiv -1(\mathrm{mod}\, q)$, and thus $\widetilde{T}_{j\boldsymbol{d}_1}|\psi_{n,\mathbf{k}}\rangle = e^{i\alpha_{\mathbf{k}}}|\psi_{n,\mathbf{k}+jp\boldsymbol{g}_2/q}\rangle = e^{i\alpha'_{\mathbf{k}}}|\psi_{n,\mathbf{k}-\boldsymbol{g}_2/q}\rangle = e^{i\alpha'_{\mathbf{k}}}|\psi_{n,\mathbf{k}+\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}_1/p}\rangle$, where $\alpha_{\mathbf{k}}$ and $\alpha'_{\mathbf{k}}$ are some phase factors. Making use of the fact that $[\hat{\boldsymbol{F}}, \widetilde{T}_{\boldsymbol{d}_1}] = 0$ (translationally invariant), we then have

$$\langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\hat{\boldsymbol{F}}|\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = \langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\widetilde{T}_{-j\boldsymbol{d}_1}\hat{\boldsymbol{F}}\widetilde{T}_{j\boldsymbol{d}_1}|\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = \langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times(\boldsymbol{d}+\boldsymbol{d}_1/p)}|\hat{\boldsymbol{F}}|\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times(\boldsymbol{d}+\boldsymbol{d}_1/p)}\rangle\,, \quad (143)$$

namely, the quantity $\langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\hat{\boldsymbol{F}}|\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle$ invariant under displacement $\boldsymbol{d} \to \boldsymbol{d} + \boldsymbol{d}_1/p$. Therefore, the integral in region $\mathcal{M}_{\{q\boldsymbol{d}_1/p,\boldsymbol{d}_2/p\}}$ in Eq. (142) is equal to the same integral in a region $\mathcal{M}_{\{\boldsymbol{d}_1,\boldsymbol{d}_2/p\}}$ times a global factor q/p. Note that region $\mathcal{M}_{\{\boldsymbol{d}_1,\boldsymbol{d}_2/p\}}$ is nothing but the torus dual "Brillouin zone" Ω_M on which we defined the U(1) Berry gauge field in Eq. (141). We can then rewrite the Lorentz susceptibility in Eq. (142) as

$$\gamma_{xy} = -i\frac{eB}{2\pi} \sum_{n\in\mathrm{occ},m\notin\mathrm{occ}} \int_{\boldsymbol{d}\in\Omega_M} \mathrm{d}^2\boldsymbol{d}\frac{\left[\langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\hat{F}_x|\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle\langle\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\hat{F}_y|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle - (x\leftrightarrow y)\right]}{(\epsilon_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}} - \epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}})^2}$$

$$= -i\frac{eB}{2\pi} \sum_{n\in\mathrm{occ},m\notin\mathrm{occ}} \int_{\boldsymbol{d}\in\Omega_M} \mathrm{d}^2\boldsymbol{d}\frac{\left[\langle\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\frac{\partial\tilde{\tilde{H}}(\boldsymbol{d})}{\partial d_x}|\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle\langle\psi_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\frac{\partial\tilde{\tilde{H}}(\boldsymbol{d})}{\partial d_y}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle - (x\leftrightarrow y)\right]}{(\epsilon_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}} - \epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}})^2}\,, \quad (144)$$

where we have used Eq. (136). Then, by noting that

$$\langle w_{m,\boldsymbol{d}}|\frac{\partial\tilde{\tilde{H}}(\boldsymbol{d})}{\partial d_i}|w_{n,\boldsymbol{d}}\rangle = \partial_{d_i}(\langle w_{m,\boldsymbol{d}}|\tilde{\tilde{H}}(\boldsymbol{d})|w_{n,\boldsymbol{d}}\rangle) - \langle\partial_{d_i}w_{m,\boldsymbol{d}}|\tilde{\tilde{H}}(\boldsymbol{d})|w_{n,\boldsymbol{d}}\rangle - \langle w_{m,\boldsymbol{d}}|\tilde{\tilde{H}}(\boldsymbol{d})|\partial_{d_i}w_{n,\boldsymbol{d}}\rangle$$

$$= 0 - \epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\langle\partial_{d_i}w_{m,\boldsymbol{d}}|w_{n,\boldsymbol{d}}\rangle - \epsilon_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\langle w_{m,\boldsymbol{d}}|\partial_{d_i}w_{n,\boldsymbol{d}}\rangle$$

$$= (\epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}} - \epsilon_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}})\langle w_{m,\boldsymbol{d}}|\partial_{d_i}w_{n,\boldsymbol{d}}\rangle = (\epsilon_{m,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}} - \epsilon_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}})\langle\partial_{k_i}w_{m,\boldsymbol{d}}|w_{n,\boldsymbol{d}}\rangle\,, \quad (145)$$

we can then derive the following:

$$\gamma_{xy} = -i\frac{eB}{2\pi} \sum_{n\in\mathrm{occ},m\notin\mathrm{occ}} \int_{\boldsymbol{d}\in\Omega_M} \mathrm{d}^2\boldsymbol{d}\left(\langle\partial_{d_x}w_{n,\boldsymbol{d}}|w_{m,\boldsymbol{d}}\rangle\langle w_{m,\boldsymbol{d}}|\partial_{d_y}w_{n,\boldsymbol{d}}\rangle - \langle\partial_{d_y}w_{n,\boldsymbol{d}}|w_{m,\boldsymbol{d}}\rangle\langle w_{m,\boldsymbol{d}}|\partial_{k_x}w_{n,\boldsymbol{d}}\rangle\right)$$

$$= -i\frac{eB}{2\pi} \sum_{n\in\mathrm{occ},m} \int_{\boldsymbol{d}\in\Omega_M} \mathrm{d}^2\boldsymbol{d}\left(\langle\partial_{d_x}w_{n,\boldsymbol{d}}|w_{m,\boldsymbol{d}}\rangle\langle w_{m,\boldsymbol{d}}|\partial_{d_y}w_{n,\boldsymbol{d}}\rangle - \langle\partial_{d_y}w_{n,\boldsymbol{d}}|w_{m,\boldsymbol{d}}\rangle\langle w_{m,\boldsymbol{d}}|\partial_{d_x}w_{n,\boldsymbol{d}}\rangle\right)$$

$$= -i\frac{eB}{2\pi} \sum_{n\in\mathrm{occ}} \int_{\boldsymbol{d}\in\Omega_M} \mathrm{d}^2\boldsymbol{d}\left(\langle\partial_{d_x}w_{n,\boldsymbol{d}}|\partial_{d_y}w_{n,\boldsymbol{d}}\rangle - \langle\partial_{d_y}w_{n,\boldsymbol{d}}|\partial_{d_x}w_{n,\boldsymbol{d}}\rangle\right)$$

$$= eB \sum_{n\in\mathrm{occ}} \int_{\boldsymbol{d}\in\Omega_M} \frac{\mathrm{d}^2\boldsymbol{d}}{2\pi}\tilde{f}^n_{xy}(\boldsymbol{d})\,, \quad (146)$$

where $\tilde{f}^n_{xy}(\boldsymbol{d})$ is the new Berry curvature in the $\boldsymbol{d}$ space we defined in Eq. (141). This is then a dual analogy to the TKNN formula. In particular, our results for γ_{xy} in Sec. VI B 1 implies

$$\sum_{n\in\mathrm{occ}} \int_{\boldsymbol{d}\in\Omega_M} \frac{\mathrm{d}^2\boldsymbol{d}}{2\pi}\tilde{f}^n_{xy}(\boldsymbol{d}) = s_\nu\,. \quad (147)$$

Therefore, s_ν can be viewed as a dual Chern number, where the integral is within a torus "dual magnetic Brillouin zone" Ω_M with periods $\boldsymbol{d}_1$ and $\boldsymbol{d}_2/p$.

— Eq. (147) can also be proved by the gauge choices of the Bloch wave functions in Eqs. (106) and (107) (which we employed to prove the Diophantine equation), namely, one can fix $|\psi_{n,\mathbf{k}+\boldsymbol{g}_1/q}\rangle = |\psi_{n,\mathbf{k}}\rangle$, $|\psi_{n,\mathbf{k}+\boldsymbol{g}_2}\rangle = e^{-i\sigma q\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}}\rangle$ and $\widetilde{T}_{\boldsymbol{d}_1}|\psi_{n,\mathbf{k}}\rangle = e^{imq\boldsymbol{d}_1\cdot\mathbf{k}}|\psi_{n,\mathbf{k}+p\boldsymbol{g}_2/q}\rangle$ for a Hofstadter band n between the $(\nu-1)$-th gap and the ν-th gap, where $\sigma = t_\nu - t_{\nu-1}$ and $m = s_\nu - s_{\nu-1}$. Then by Eq. (132), we find

$$|w_{n,\boldsymbol{d}+\boldsymbol{d}_2/p}\rangle = \widetilde{T}_{-\boldsymbol{d}-\boldsymbol{d}_2/p}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times(\boldsymbol{d}+\boldsymbol{d}_2/p)}\rangle = e^{i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_2\times\boldsymbol{d})/2p}\widetilde{T}_{-\boldsymbol{d}_2/p}\widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}-\boldsymbol{g}_1/q}\rangle$$

$$= e^{i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_2\times\boldsymbol{d})/2p}\widetilde{T}_{-\boldsymbol{d}_2/p}\widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_2\times\boldsymbol{d})/2p}\widetilde{T}_{-\boldsymbol{d}_2/p}|w_{n,\boldsymbol{d}}\rangle\,,$$

$$|w_{n,\boldsymbol{d}+\boldsymbol{d}_1}\rangle = \widetilde{T}_{-\boldsymbol{d}-\boldsymbol{d}_1}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times(\boldsymbol{d}+\boldsymbol{d}_1)}\rangle = e^{-i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_1\times\boldsymbol{d})/2}\widetilde{T}_{-\boldsymbol{d}}\widetilde{T}_{-\boldsymbol{d}_1}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}+p\boldsymbol{g}_2/q}\rangle$$

$$= e^{-i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_1\times\boldsymbol{d})/2-imq\boldsymbol{d}_1\cdot(\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d})}\widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{-i\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_1\times\boldsymbol{d})/2+imq\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_1\times\boldsymbol{d})}|w_{n,\boldsymbol{d}}\rangle\,. \quad (148)$$

Therefore, for band n between the $(\nu-1)$-th gap and the ν-th gap, we find the dual Berry gauge field in Eq. (141) satisfies

$$
\begin{aligned}
\int_{\boldsymbol{d}\in\Omega_M}\frac{\mathrm{d}^2\boldsymbol{d}}{2\pi}\widetilde{f}^n_{xy}(\boldsymbol{d}) &= \frac{1}{2\pi}\oint_{\partial\Omega_M}\widetilde{\boldsymbol{a}}^n(\boldsymbol{d})\cdot\mathrm{d}\boldsymbol{d} \\
&= \frac{1}{2\pi}[mq\ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_1\times\boldsymbol{d}_2/p) - \ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_1\times\boldsymbol{d}_2/p)/2 + \ell^{-2}\hat{\mathbf{z}}\cdot(\boldsymbol{d}_1\times\boldsymbol{d}_2)/2p] = \frac{mq}{2\pi p}\ell^{-2}\Omega \\
&= m = s_\nu - s_{\nu-1}\ .
\end{aligned}
\tag{149}
$$

Summing over all occupied bands then proves Eq. (147).

C. Relation between the dual wave functions, and a model example

This subsection is devoted for a better understanding of the relation between the two wave functions $|u_{n,\mathbf{k}}\rangle$ defined in Eq. (116) and $|w_{n,\boldsymbol{d}}\rangle$ defined in Eq. (132). To do this, we examine their form in the real space basis $|\mathbf{r}\rangle$. If we denote the Bloch wave function of momentum $\mathbf{k}$ as $|\psi_{n,\mathbf{k}}\rangle$, by the definition of the periodic wavefunction $|u_{n,\mathbf{k}}\rangle$ in Eq. (116), we have

$$
\psi_{n,\mathbf{k}}(\mathbf{r}) = \langle\mathbf{r}|\psi_{n,\mathbf{k}}\rangle\ , \qquad u_{n,\mathbf{k}}(\mathbf{r}) = \langle\mathbf{r}|u_{n,\mathbf{k}}\rangle = e^{-i\mathbf{k}\cdot\mathbf{r}}\psi_{n,\mathbf{k}}(\mathbf{r})\ .
\tag{150}
$$

To find the real space wave function of $|w_{n,\boldsymbol{d}}\rangle = \widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle$ as defined in Eq. (132), we first note that

$$
\begin{aligned}
\widetilde{T}_{\boldsymbol{d}}|\mathbf{r}\rangle &= e^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}|\mathbf{r}\rangle = e^{-i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{r})\cdot\boldsymbol{d}-i\mathbf{\Pi}\cdot\boldsymbol{d}}|\mathbf{r}\rangle = e^{i\int_{\mathbf{r}}^{\mathbf{r}+\boldsymbol{d}}[\mathbf{A}(\mathbf{r}')-\ell^{-2}\hat{\mathbf{z}}\times\mathbf{r}']\cdot\mathrm{d}\mathbf{r}'}|\mathbf{r}+\boldsymbol{d}\rangle \\
&= e^{i\ell^{-2}(\hat{\mathbf{z}}\times\boldsymbol{d})\cdot(\mathbf{r}+\boldsymbol{d})}e^{i\int_{\mathbf{r}}^{\mathbf{r}+\boldsymbol{d}}\mathbf{A}(\mathbf{r}')\cdot\mathrm{d}\mathbf{r}'}|\mathbf{r}+\boldsymbol{d}\rangle\ ,
\end{aligned}
\tag{151}
$$

which can be seen by following a derivation similar to Eq. (92), where the integral on the exponent is along the straight line segment from $\mathbf{r}$ to $\mathbf{r}+\boldsymbol{d}$. Therefore, we find

$$
\begin{aligned}
w_{n,\boldsymbol{d}}(\mathbf{r}) &= \langle\mathbf{r}|w_{n,\boldsymbol{d}}\rangle = \langle\mathbf{r}|\widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{-i\ell^{-2}(\hat{\mathbf{z}}\times\boldsymbol{d})\cdot(\mathbf{r}+\boldsymbol{d})}e^{-i\int_{\mathbf{r}}^{\mathbf{r}+\boldsymbol{d}}\mathbf{A}(\mathbf{r}')\cdot\mathrm{d}\mathbf{r}'}\langle\mathbf{r}+\boldsymbol{d}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \\
&= e^{-i\int_{\mathbf{r}}^{\mathbf{r}+\boldsymbol{d}}\mathbf{A}(\mathbf{r}')\cdot\mathrm{d}\mathbf{r}'}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}(\mathbf{r}+\boldsymbol{d})\ .
\end{aligned}
\tag{152}
$$

This is the explicit form of the wave function $w_{n,\boldsymbol{d}}(\mathbf{r})$.

Without going into the real space position basis, we can also derive that (here $\mathbf{r}$ below denotes the position operator)

$$
|w_{n,\boldsymbol{d}}\rangle = \widetilde{T}_{-\boldsymbol{d}}|\psi_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{i\ell^{-2}(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d}}e^{i\ell^{-2}(\hat{\mathbf{z}}\times\boldsymbol{d})\cdot\mathbf{r}}|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{i\mathbf{\Pi}\cdot\boldsymbol{d}}|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle\ ,
\tag{153}
$$

where we have used the fact that $[(\hat{\mathbf{z}}\times\mathbf{R})\cdot\boldsymbol{d},(\hat{\mathbf{z}}\times\mathbf{r})\cdot\boldsymbol{d}]=0$, and $\mathbf{R}-\mathbf{r}=-\ell^2\hat{\mathbf{z}}\times\mathbf{\Pi}$. By this relation, the dual Berry curvature $\widetilde{f}^n_{xy}(\boldsymbol{d})$ defined in Eq. (141) can be explicitly related to the Berry curvature $\mathcal{F}^n_{xy}(\mathbf{k})$ in Eq. (121). We have

$$
|\partial_{d_i}w_{n,\boldsymbol{d}}\rangle = (\partial_{d_i}e^{i\mathbf{\Pi}\cdot\boldsymbol{d}})|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle + e^{i\mathbf{\Pi}\cdot\boldsymbol{d}}|\partial_{d_i}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle = e^{i\mathbf{\Pi}\cdot\boldsymbol{d}}[i(\Pi_i+\epsilon_{ij}d_j/2)|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle + |\partial_{d_i}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle]\ ,
\tag{154}
$$

where ϵ_{ji} $(i,j=x,y)$ is the Levi-Civita tensor $(\epsilon_{xy}=-\epsilon_{yx}=1)$. Therefore, we find the dual Berry curvature

$$
\begin{aligned}
\widetilde{f}^n_{xy}(\boldsymbol{d}) &= -i\langle\partial_{d_x}w_{n,\boldsymbol{d}}|\partial_{d_y}w_{n,\boldsymbol{d}}\rangle + i\langle\partial_{d_y}w_{n,\boldsymbol{d}}|\partial_{d_x}w_{n,\boldsymbol{d}}\rangle \\
&= -i\langle u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|[\Pi_x,\Pi_y]|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle - i\langle\partial_{d_x}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\partial_{d_y}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle + i\langle\partial_{d_y}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\partial_{d_x}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \\
&\quad - \langle\partial_{d_y}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\Pi_x|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle - \langle u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\Pi_x|\partial_{d_y}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \\
&\quad + \langle\partial_{d_x}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\Pi_y|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle + \langle u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\Pi_y|\partial_{d_x}u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle \\
&= \ell^{-2} - \ell^{-4}\mathcal{F}^n_{xy}(\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}) + \partial_{d_x}(\langle u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\Pi_y|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle) - \partial_{d_y}(\langle u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}|\Pi_x|u_{n,\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}}\rangle) \\
&= B - B^2\mathcal{F}^n_{xy}(\ell^{-2}\hat{\mathbf{z}}\times\boldsymbol{d}) + \partial_{d_x}(\langle w_{n,\boldsymbol{d}}|\Pi_y|w_{n,\boldsymbol{d}}\rangle) - \partial_{d_y}(\langle w_{n,\boldsymbol{d}}|\Pi_x|w_{n,\boldsymbol{d}}\rangle)\ .
\end{aligned}
\tag{155}
$$

This gives the relationship between the dual Berry curvature $\widetilde{f}^n_{xy}(\boldsymbol{d})$ and the Berry curvature $\mathcal{F}^n_{xy}(\mathbf{k})$. In particular, we note that the two terms $\partial_{d_x}(\langle w_{n,\boldsymbol{d}}|\Pi_y|w_{n,\boldsymbol{d}}\rangle) - \partial_{d_y}(\langle w_{n,\boldsymbol{d}}|\Pi_x|w_{n,\boldsymbol{d}}\rangle)$ in the last line of Eq. (155) are total derivatives which do not contribute to the dual Chern number s_ν, since $\langle w_{n,\boldsymbol{d}}|\Pi_i|w_{n,\boldsymbol{d}}\rangle$ are well-defined periodic functions (physical quantities) in the dual Brillouin zone Ω_M spanned by $\boldsymbol{d}_1$ and $\boldsymbol{d}_2/p$.

1. An example

For simplicity, we consider a single-orbital tight-binding model on a square lattice with a magnetic flux per unit cell $\varphi = 2\pi/3$. Assume the lattice vectors are $\boldsymbol{d}_1 = (1,0)$ and $\boldsymbol{d}_2 = (0,1)$, and the Wannier orbitals $|\boldsymbol{D}_{n_1,n_2}\rangle$ are delta functions at lattice sites $\boldsymbol{D}_{n_1,n_2} = (n_1, n_2) \in \boldsymbol{d}_1\mathbb{Z} + \boldsymbol{d}_2\mathbb{Z}$ in the continuous space. At zero magnetic field, we assume all the nearest bonds have hopping $-t$ (with t being a real number), while all the longer range hoppings are zero. We then add a uniform magnetic field $B = 2\pi p/q = 4\pi/3$ (so that the flux $\varphi = B|\boldsymbol{d}_1 \times \boldsymbol{d}_2| = 4\pi/3$), and we adopt the Landau gauge $\mathbf{A}(\mathbf{r}) = (0, Bx)$. This correspond to $p = 2$ and $q = 3$. Accordingly, the magnetic length ℓ satisfies $\ell^{-2} = 4\pi/3$.

The tight-binding Hamiltonian in magnetic field B can then be written as

$$H = -t \sum_{n_1, n_2 \in \mathbb{Z}} \left(|\boldsymbol{D}_{n_1+1,n_2}\rangle\langle\boldsymbol{D}_{n_1,n_2}| + e^{i4\pi n_1/3}|\boldsymbol{D}_{n_1,n_2+1}\rangle\langle\boldsymbol{D}_{n_1,n_2}| + h.c. \right) . \tag{156}$$

Therefore, we have

$$e^{-i\mathbf{k}\cdot\mathbf{r}} H e^{i\mathbf{k}\cdot\mathbf{r}} = -t \sum_{n_1, n_2 \in \mathbb{Z}} \left(e^{-ik_x}|\boldsymbol{D}_{n_1+1,n_2}\rangle\langle\boldsymbol{D}_{n_1,n_2}| + e^{-ik_y+i4\pi n_1/3}|\boldsymbol{D}_{n_1,n_2+1}\rangle\langle\boldsymbol{D}_{n_1,n_2}| + h.c. \right) , \tag{157}$$

where $\mathbf{r}$ is the position operator (instead of a parameter), while $\mathbf{k} = (k_x, k_y)$ is the momentum parameter. By Eq. (151), the projector P_0 in Eq. (131) in the Hilbert space of Wannier basis $|\boldsymbol{D}_{n_1,n_2}\rangle$ can be rewritten as

$$P_0 = \frac{3}{N_\Omega} \sum_{j_1, j_2 \in \mathbb{Z}} \widetilde{T}_{3j_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2} = \sum_{j=1}^{3} |j\rangle\langle j| , \qquad |j\rangle = \sqrt{\frac{3}{N_\Omega}} \sum_{n_1, n_2 \in \mathbb{Z}} |\boldsymbol{D}_{3n_1+j-1,n_2}\rangle , \quad (j = 1, 2, 3). \tag{158}$$

The three basis $|j\rangle$ $(j = 1, 2, 3)$ correspond exactly to the three orbitals in a magnetic unit cell. Therefore, under the three basis $|j\rangle$, we find the reduced momentum space Hamiltonian $\hat{H}(\mathbf{k})$ is given by

$$\hat{H}(\mathbf{k}) = P_0 e^{-i\mathbf{k}\cdot\mathbf{r}} H e^{i\mathbf{k}\cdot\mathbf{r}} P_0 = -t \begin{pmatrix} 2\cos(k_y) & e^{ik_x} & e^{-ik_x} \\ e^{-ik_x} & 2\cos(k_y + 2\pi/3) & e^{ik_x} \\ e^{ik_x} & e^{-ik_x} & 2\cos(k_y - 2\pi/3) \end{pmatrix} . \tag{159}$$

This is exactly the momentum space Hamiltonian we are familiar with.

We now consider the other reduced Hamiltonian $\hat{\widetilde{H}}(\boldsymbol{d})$. By Eq. (151), for $\boldsymbol{d} = (d_x, d_y)$ we have

$$\widetilde{T}_{-\boldsymbol{d}} H \widetilde{T}_{\boldsymbol{d}} = -t \sum_{n_1, n_2 \in \mathbb{Z}} \left(|\boldsymbol{D}_{n_1+1,n_2} - \boldsymbol{d}\rangle\langle\boldsymbol{D}_{n_1,n_2} - \boldsymbol{d}| + e^{-i4\pi d_x/3+i4\pi n_1/3}|\boldsymbol{D}_{n_1,n_2+1} - \boldsymbol{d}\rangle\langle\boldsymbol{D}_{n_1,n_2} - \boldsymbol{d}| + h.c. \right), \tag{160}$$

where we have used $\mathbf{A}(\mathbf{r}) = (0, Bx)$ in the continuous space. We note that the Hamiltonian is transformed into a shifted Wannier basis $|\boldsymbol{D}_{n_1+1,n_2} - \boldsymbol{d}\rangle$. Note that by Eq. (151), the lattice translation operators act as

$$\widetilde{T}_{3\boldsymbol{d}_1}|\boldsymbol{D}_{n_1,n_2} - \boldsymbol{d}\rangle = e^{-i4\pi d_y}|\boldsymbol{D}_{n_1+3,n_2} - \boldsymbol{d}\rangle , \qquad \widetilde{T}_{\boldsymbol{d}_2}|\boldsymbol{D}_{n_1,n_2} - \boldsymbol{d}\rangle = |\boldsymbol{D}_{n_1,n_2+1} - \boldsymbol{d}\rangle . \tag{161}$$

Therefore, the projector P_0 in the Hilbert space of Wannier basis $|\boldsymbol{D}_{n_1+1,n_2} - \boldsymbol{d}\rangle$ is

$$P_0 = \frac{3}{N_\Omega} \sum_{j_1, j_2 \in \mathbb{Z}} \widetilde{T}_{3j_1\boldsymbol{d}_1+j_2\boldsymbol{d}_2} = \sum_{j=1}^{3} |j, \boldsymbol{d}\rangle\langle j, \boldsymbol{d}| , \qquad |j, \boldsymbol{d}\rangle = \sqrt{\frac{3}{N_\Omega}} \sum_{n_1, n_2 \in \mathbb{Z}} e^{-i4\pi(3n_1+j)d_y/3}|\boldsymbol{D}_{3n_1+j-1,n_2} - \boldsymbol{d}\rangle, \tag{162}$$

where $j = 1, 2, 3$. Under the basis $|j, \boldsymbol{d}\rangle$, we find

$$\hat{\widetilde{H}}(\boldsymbol{d}) = P_0 \widetilde{T}_{-\boldsymbol{d}} H \widetilde{T}_{\boldsymbol{d}} P_0 = -t \begin{pmatrix} 2\cos(4\pi d_x/3) & e^{-i4\pi d_y/3} & e^{i4\pi d_y/3} \\ e^{i4\pi d_y/3} & 2\cos(4\pi d_x/3 + 2\pi/3) & e^{-i4\pi d_y/3} \\ e^{-i4\pi d_y/3} & e^{i4\pi d_y/3} & 2\cos(4\pi d_x/3 - 2\pi/3) \end{pmatrix} . \tag{163}$$

Therefore, under the Landau gauge, we find $\hat{\widetilde{H}}(\boldsymbol{d})$ takes a form analogous to $\hat{H}(\mathbf{k})$ with $\mathbf{k} = (4\pi/3)\hat{\mathbf{z}} \times \boldsymbol{d}$, except that the basis $|j, \boldsymbol{d}\rangle$ is $\boldsymbol{d}$ dependent. Therefore, if the eigenstates of $\hat{H}(\mathbf{k})$ are $|u_{n,\mathbf{k}}\rangle = \sum_{j=1}^{3} u_{n,j}(\mathbf{k})|j\rangle$, the eigenstates of $\hat{\widetilde{H}}(\boldsymbol{d})$ will be $|w_{n,\boldsymbol{d}}\rangle = \sum_{j=1}^{3} u_{n,j}(\frac{4\pi}{3}\hat{\mathbf{z}} \times \boldsymbol{d})|j, \boldsymbol{d}\rangle$, in agreement with Eq. (152).

The Chern numbers of the 3 bands of $\hat{H}(\mathbf{k})$ can be calculated to be $\sigma_n = \{-1, 2, -1\}$ (from the lowest band to the highest band). According to Eq. (155), we find the dual Chern numbers of the 3 bands of $\hat{\widetilde{H}}(\boldsymbol{d})$ are given by $m_n = \frac{1}{3} - \frac{2}{3}\sigma_n = \{1, -1, 1\}$ (from the lowest band to the highest band).
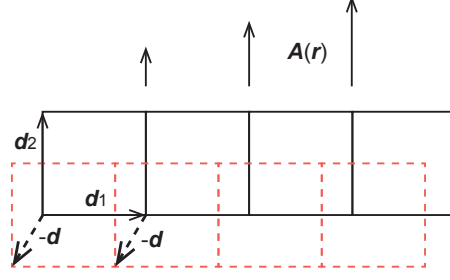
FIG. 7. An example of tight binding model on a square lattice in a magnetic field, where Landau gauge is chosen.

D. Physical understanding of the quantized Lorentz susceptibility

The quantized Lorentz susceptibility $\gamma_{xy} = eBs_\nu$ can be easily understood from the following physical picture. Consider a lattice model in the x-y plane moving with a velocity $\mathbf{v} = v_y\hat{\mathbf{y}}$ in the $\hat{\mathbf{z}}$ direction magnetic field B, and the Fermi level is in a Hofstadter gap ν. In the rest frame of the lattice system, the magnetic field B will move with a velocity $-\mathbf{v} = -v_y\hat{\mathbf{y}}$, which produces an electric field $\mathbf{E} = v_yB\hat{\mathbf{x}}$. This electric field then produces a Hall current density $\tilde{j}_y = -\sigma_{xy}v_yB$ in the rest frame of the lattice. If we go back to the laboratory frame where the system moves with velocity $\mathbf{v} = v_y\hat{\mathbf{y}}$, we would find a total current density

$$ j_y = \frac{e\rho v_y}{\Omega} + \tilde{j}_y = \left(\frac{e\rho}{\Omega} - \sigma_{xy}B\right)v_y \ , $$

where ρ is the number of occupied electron per unit cell, and Ω is the unit cell area. Therefore, the Lorentz force per unit cell is given by

$$ F_x = j_y\Omega B = (e\rho - \sigma_{xy}\Omega B)Bv_y = \left(e\rho - t_\nu\frac{e^2}{h}\Omega B\right)Bv_y = eB\left(\rho - t_\nu\frac{\varphi}{2\pi}\right)v_y = eBs_\nu v_y \ , \tag{164} $$

where we have used the Diophantine equation. Therefore, we find the Lorentz susceptibility is $\gamma_{xy} = eBs_\nu$.