

Describing Landau Level Mixing in Fractional Quantum Hall States with Deep Learning

Yubing Qian^{1,2,*}, Tongzhou Zhao^{3,*†}, Jianxiao Zhang⁴, Tao Xiang³, Xiang Li^{2,‡}, and Ji Chen^{1,5,§}¹*School of Physics, Peking University, Beijing 100871, People's Republic of China*²*ByteDance Research China, Fangheng Fashion Center,**No. 27, North 3rd Ring West Road, Haidian District, Beijing 100098, People's Republic of China*³*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*⁴*Department of Physics, 104 Davey Lab, Pennsylvania State University, University Park, Pennsylvania 16802, USA*⁵*Interdisciplinary Institute of Light-Element Quantum Materials, Frontiers Science Center for Nano-Optoelectronics, Peking University, Beijing 100871, People's Republic of China*

(Received 20 December 2024; accepted 8 April 2025; published 30 April 2025)

Strong correlation brings a rich array of emergent phenomena, as well as a daunting challenge to theoretical physics study. In condensed matter physics, the fractional quantum Hall effect is a prominent example of strong correlation, with Landau level mixing being one of the most challenging aspects to address using traditional computational methods. Deep learning real-space neural network wave function methods have emerged as promising architectures to describe electron correlations in molecules and materials, but their power has not been fully tested for exotic quantum states. In this work, we employ real-space neural network wave function techniques to investigate fractional quantum Hall systems. On both $1/3$ and $2/5$ filling systems, we achieve energies consistently lower than exact diagonalization results which only consider the lowest Landau level. We also demonstrate that the real-space neural network wave function can naturally capture the extent of Landau level mixing up to a very high level, overcoming the limitations of traditional methods. Our work underscores the potential of neural networks for future studies of strongly correlated systems and opens new avenues for exploring the rich physics of the fractional quantum Hall effect.

DOI: [10.1103/PhysRevLett.134.176503](https://doi.org/10.1103/PhysRevLett.134.176503)

The fractional quantum Hall (FQH) effect is one of the most notable examples in condensed matter physics, highlighting the fascinating emergent phenomena driven by strong correlation effects and non-trivial topology [1,2]. In FQH systems, the kinetic energy is quenched by the strong magnetic field, and the Coulomb interaction dominates the physics. Consequently, the wave function of a FQH system cannot be adiabatically connected to a simple, noninteracting state described by a single Slater determinant. This complexity, combined with their intriguing topological properties, makes it challenging to represent these ground states efficiently using conventional methods.

Early studies usually assume infinitely strong magnetic fields, confining all electrons to the lowest Landau level (LLL). However, in typical experiments, the Coulomb interaction strength is comparable to the cyclotron energy [3], and the Landau level mixing (LLM) can induce new physics. Recent experiments have linked LLM to phenomena such as Wigner-crystal phase transitions [4,5], spin transitions [6], and novel non-Abelian FQH states [7,8]. Nevertheless, traditional numerical approaches have

limitations in addressing LLM. Exact diagonalization (ED) typically explicitly handles only the lowest Landau level (LLL) due to the large Hilbert space dimension [9], with very few studies extending it to two Landau levels [10]. The density matrix renormalization group (DMRG) method can include up to five Landau levels but struggles with states near the critical point with high entanglement [11,12]. The fixed-phase diffusion Monte Carlo method (fp-DMC) can naturally handle LLM, but is limited by the accuracy of its phase approximation and does not provide an explicit wave function [13–15]. The recent extension of density functional theory to FQH problems has demonstrated promising results in handling LLM, but it requires explicit knowledge of composite fermions to construct the exchange-correlation functional [16–18].

In recent years, deep learning methods have emerged as promising alternatives for studying strongly correlated systems [19,20]. These methods have achieved remarkable success in representing the wave functions of various strongly correlated systems, including lattice models [21,22], molecules [23–27], solids [28–30], electron gases [30–33], and Moiré systems [34,35]. In these architectures, wave functions are presented in real space with neural networks and, when applied to FQH systems, can naturally include contributions from higher Landau levels within the ansatz. This capability makes deep learning based on

*These authors contributed equally to this work.

†Contact author: tzzhao_2022@iphy.ac.cn‡Contact author: lixiang.62770689@bytedance.com§Contact author: ji.chen@pku.edu.cn

real-space neural networks a promising tool to break through the limitations of traditional methods and gain deeper insights into the ground states of FQH systems with LLM.

In this study, we employ real-space neural network methods to investigate FQH systems in spherical geometry [36]. In the literature, disk [2] and torus [37] geometries are also employed to study FQH effects, compared to which the spherical geometry has several advantages, including the edgeless structure, the well-defined filling factor, and the simplicity in mathematical treatment. As shown in Fig. 1(a), N spin-polarized electrons are confined on a spherical surface, with a magnetic monopole placed at the center. The monopole creates a total flux of $2Q\phi_0$ through the surface, where ϕ_0 is the flux quantum and $2Q$ is an integer. The radius of the sphere R is given by $\sqrt{Q}\ell$, where $\ell = \sqrt{\hbar c/eB}$ is the magnetic length, and B is the strength of the uniform magnetic field on the sphere. The flux-particle relationship on the sphere is determined by $2Q = N/\nu - S$, where ν is the filling factor and S is the shift, which characterizes the topological properties of the corresponding FQH state [38].

The Hamiltonian on the spherical geometry is formulated as

$$\hat{H} = \sum_i \frac{\ell^2 \hbar \omega_c}{2R^2} |\hat{\Lambda}_i|^2 + \frac{e^2}{\epsilon} \sum_{i < j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (1)$$

where $\omega_c = eB/mc$ is the cyclotron frequency, m is the band mass of the electrons, ϵ is the dielectric constant of the material, and \mathbf{r}_i is the coordinate of the i th electron. $\hat{\Lambda}_i$ is proportional to the canonical momentum tangential to the surface:

$$|\hat{\Lambda}_i|^2 = -\frac{1}{\sin\theta_i} \frac{\partial}{\partial\theta_i} \sin\theta_i \frac{\partial}{\partial\theta_i} + \left(Q \cot\theta_i + \frac{i}{\sin\theta_i} \frac{\partial}{\partial\phi_i} \right)^2. \quad (2)$$

We define $\kappa = (e^2/\epsilon\ell)/(\hbar\omega_c)$ as a measure of the interaction strength, which also effectively tunes the LLM. All energies shown are in units of $\hbar\omega_c\kappa$.

Crafting a valid and expressive neural network wave function ansatz is essential for obtaining fractional quantum Hall states with deep learning. To this end, we introduce DeepHall, a novel and flexible framework designed to model strongly correlated electrons in FQH systems. The trial wave function ψ_T^{DeepHall} is constructed as follows:

$$\psi_T^{\text{DeepHall}}(\mathbf{r}) = e^{\mathcal{J}(\mathbf{r})} \det[\varphi_i(\mathbf{r}_j; \{\mathbf{r}_{\neq j}\})], \quad (3)$$

where \mathbf{r} denotes the coordinates of all electrons, and the notation $\{\mathbf{r}_{\neq j}\}$ indicates that the permutation of other electrons does not change the output [Fig. 1(b)]. The Jastrow factor

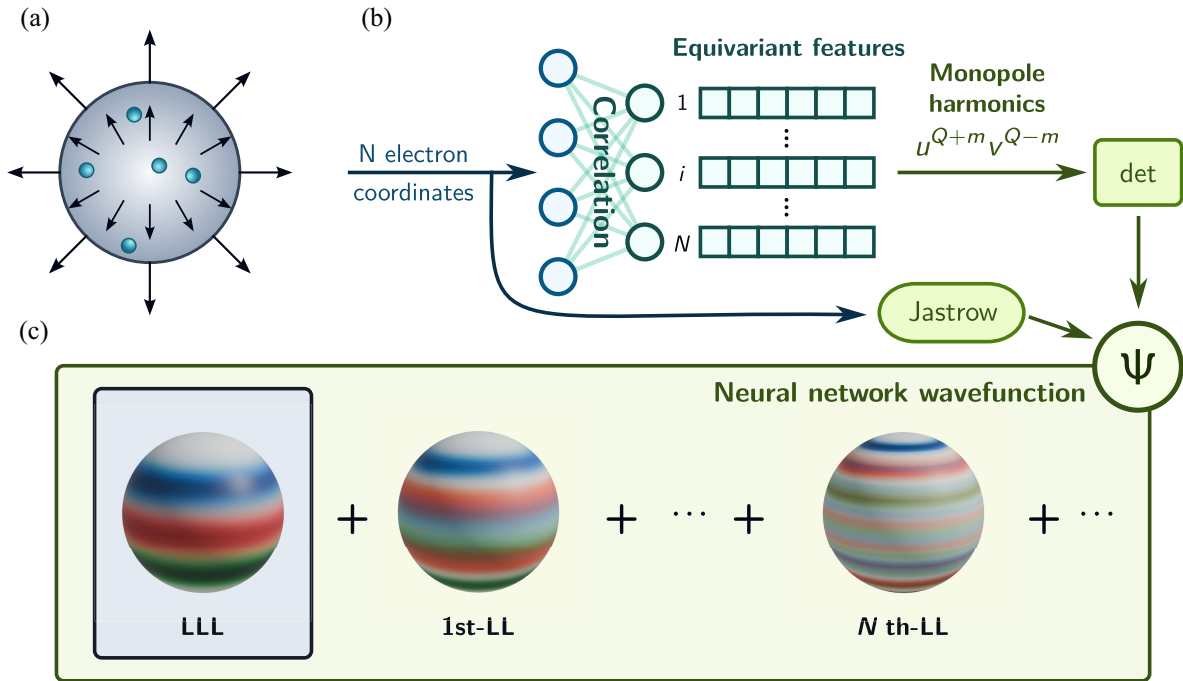


FIG. 1. (a) Illustration of the system with spherical geometry. Electrons are confined to the spherical surface, with a magnetic monopole at the center. (b) DeepHall framework. The neural network takes electron coordinates as inputs and outputs permutation-equivariant features that encode correlations of all electrons. These features are then multiplied by monopole harmonics to form multielectron orbitals. The fermionic wave function is constructed by the determinant of the orbital matrix multiplied by a Jastrow factor. (c) A schematic illustration of Landau level mixing. Real-space neural network wave function goes beyond the lowest Landau level, and describes a mixture of an infinite number of Landau levels. Only 3 orbitals are drawn per Landau level, colored as blue, red, and green.

$$\mathcal{J}(\mathbf{r}) = -\frac{1}{4} \sum_{i < j} \frac{a^2}{a + |\mathbf{r}_i - \mathbf{r}_j|} \quad (4)$$

is included to satisfy the electron-electron Coulomb cusp condition [39], where a is a free parameter. And the multielectron orbitals $\varphi_i(\mathbf{r}_j; \{\mathbf{r}_{\neq j}\})$ on the sphere are given by

$$\varphi_i(\mathbf{r}_j; \{\mathbf{r}_{\neq j}\}) = \sum_{k,m} w_{ikm} f_k(\mathbf{r}_j; \{\mathbf{r}_{\neq j}\}) u_j^{Q+m} v_j^{Q-m}, \quad (5)$$

where w_{ikm} are complex parameters and $u_j^{Q+m} v_j^{Q-m}$ are monopole harmonics [40] on the LLL, with $m = -|Q|, -|Q| + 1, \dots, |Q|$. The spinor coordinates of the j th electron are given by $u_j = \cos(\theta_j/2)e^{i\phi_j/2}$ and $v_j = \sin(\theta_j/2)e^{-i\phi_j/2}$. The use of monopole harmonics largely avoids the divergence problem in local energy at the poles of the sphere originating from Dirac strings, and the physical information such as the correlations and topological properties is encoded in the neural network $f_k(\mathbf{r}_j; \{\mathbf{r}_{\neq j}\})$. In the DeepHall framework, the neural network part can adopt any architecture originally designed for molecular systems. Here, we use Psiformer [27], an effective neural network in modeling strongly correlated electrons of molecules. We note that the neural network part and the other parts are explicitly correlated in the DeepHall framework, which is essential for accurately modeling FQH systems. In Fig. S2, we present tests with a slightly different architecture among various other architectures designed based on Psiformer, where the results are substantially worse [41].

With the ansatz constructed, we then employ the variational Monte Carlo process to optimize the wave function. The loss function during training is chosen as the energy E_v evaluated using Monte Carlo integration.

The Kronecker-factored approximate curvature method [42] is used to optimize the parameters. With the expressiveness of the neural network, it is possible to describe the quantum Hall states using a single determinant constructed with multielectron orbitals $\varphi_i(\mathbf{r}_j; \{\mathbf{r}_{\neq j}\})$. Moreover, the DeepHall wave function can naturally include the effects of LLM [Fig. 1(c)], surpassing the calculations that only include the lowest Landau level.

We first apply DeepHall to investigate the most robust FQH effect at $\nu = 1/3$, for which the Laughlin wave function provides a good description of the ground state [2]. Figure 2(a) shows a simulation of $N = 8$ electrons with flux $2Q = 21$ and $\kappa = 1$. The energy obtained from the DeepHall falls below the ED result, which is restricted to the LLL, and aligns with the fp-DMC results. We further apply our approach to study the more complex $\nu = 2/5$ system. As demonstrated in Fig. 2(b), DeepHall achieves a significantly lower energy than the LLL ED result and remains consistent with the fp-DMC energy. However, both DeepHall and fp-DMC energies are still higher than the ED results when two Landau levels are included, suggesting the need for further improvements. These results demonstrate that DeepHall effectively captures electron correlations and can be applied to investigate FQH systems at various fillings.

To further validate the robustness of our approach, we examine the energy as a function of the number of electrons for the $\nu = 1/3$ state. As shown in Fig. 2(c), the DeepHall results for systems ranging from 6 to 12 electrons align with the fp-DMC results and are consistently lower than the ED results. This demonstrates DeepHall's ability to accurately capture ground state energies across different system sizes.

The above results are obtained with $\kappa = 1$, where LLM is mild. To examine the ability of DeepHall to learn FQH states for different levels of LLM, we study the $\nu = 1/3$ system with $N = 6$ electrons across different κ . A direct

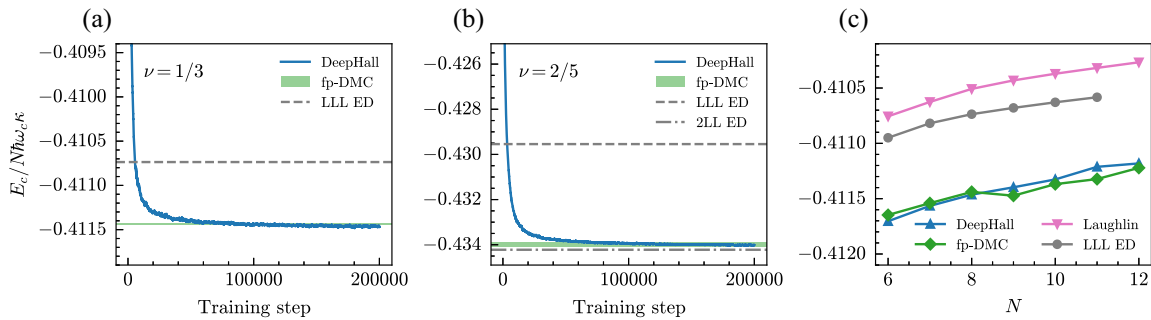


FIG. 2. (a) The training curve of a $\nu = 1/3$ system with $N = 8$, $2Q = 21$, and $\kappa = 1$. A moving window average over 500 steps is applied, and outliers are removed. The green shaded area represents the fp-DMC energy result along with its corresponding error range. The dashed line indicates the energy result obtained from ED. (b) Training curve for a $\nu = 2/5$ system with $N = 8$, $2Q = 16$, and $\kappa = 1$. In addition to the LLL ED result, the ED energy considering two Landau levels is included, represented by a dash-dotted line. (c) Laughlin (pink), ED (gray), DeepHall (blue), and fp-DMC (green) energy results with different number of electrons for $\nu = 1/3$ filling. The energy error bars are smaller than the size of the markers. The energy (E_c) includes the contribution from background charge [43]. In addition, it is shifted by $N\omega_c/2$ and a density correction to reduce the system-size dependence is applied [43,44]. The energy unit is $\hbar\omega_c\kappa$.

comparison with ED at $\kappa = 0$ is impossible because $\kappa = 0$ removes interactions, making the ratio E_c/κ indeterminate. However, for $\kappa > 0$, DeepHall and fp-DMC yield energy results in good agreement. As κ increases, ED calculations restricted to the lowest Landau level progressively overestimate the energy, while DeepHall and fp-DMC achieve substantially lower energies. Notably, as LLM becomes more pronounced, DeepHall demonstrates advantages over fp-DMC in the strong LLM regime, achieving lower variational energies. This highlights the potential of our approach for studying FQH systems under significant LLM.

To further reveal the LLM effects, we monitor the overlap S between the DeepHall wave function and the Laughlin wave function, which measures the similarity between them. The overlap is defined as

$$S = \frac{\langle \psi^{\text{DeepHall}} | \psi^{\text{Laughlin}} \rangle}{(\langle \psi^{\text{DeepHall}} | \psi^{\text{DeepHall}} \rangle \langle \psi^{\text{Laughlin}} | \psi^{\text{Laughlin}} \rangle)^{1/2}}. \quad (6)$$

In Fig. 3(b), we plot the modulus of the overlap ($|S|$). When κ is small, a large $|S|$ is obtained, since the Laughlin wave function is an excellent approximation of the ground state wave function in the absence of LLM. As LLM grows stronger, the modulus of overlap $|S|$ decreases, reasonably reflecting the influence of LLM.

To gain deeper insights about the contributions from higher Landau levels, we examine the number of electrons on the LLL, denoted as N_{LLL} . Our results in Fig. 3(b) show that N_{LLL} follows the same trend as $|S|$, being large at small κ and decreasing with larger κ . Although approximately 96% of the electrons remain on the LLL even at $\kappa = 10$, the contribution from the remaining 4% of electrons in higher Landau levels is not negligible, as evidenced by the energy results. This highlights the significant role of higher Landau levels in the presence of LLM.

With DeepHall we can also model the behavior of the pair correlation function (PCF) $g(\theta)$, which describes the probability of finding two electrons with an angular

separation of θ and reveals the correlations between electrons. The $\nu = 1/3$ FQH state is known to resemble a liquid [45], as illustrated by the PCF of the Laughlin wave function in Fig. 3(c). The PCF from the DeepHall wave function closely matches that of the Laughlin wave function when $\kappa = 0.5$. As LLM increases, the PCF becomes more structured. Eventually, at $\kappa = 10$, the PCF no longer exhibits a decaying trend and displays a prominent peak at $\theta = \pi$, indicating a phase transition away from the liquid phase [46–48]. However, the current system size in our approach is not sufficient to definitively establish the phase boundary. This is also evident in the subsequent energy gap calculations, where the charge gap persists up to $\kappa = 10$, as shown in Fig. 4(c). Nevertheless, the transition trend observed in our calculations aligns with the classical limit $\kappa \rightarrow \infty$, where the system is known to form a classical Wigner crystal with electrons arranged in a triangular lattice [49,50].

Besides the ground state, the corresponding quasiparticle and quasihole excitations in FQH systems also exhibit interesting physics. These excitations carry fractional electric charges, a remarkable departure from the usual integer charges, and they exhibit anyonic braiding statistics, which are neither bosonic nor fermionic. On the spherical geometry, the quasiparticle and quasihole states are actually *ground states* for different flux values $2Q^{\text{qp/qh}} = 3(N-1) \mp 1$. To reveal the charge density for these excitations, we need to break the degeneracy of different excitations by adding a penalty term $\beta(\hat{L}_z - m_t)^2$ to the loss function during the network training, where L_z is the total angular momentum in the z -direction, m_t is the target L_z , and β is a tunable hyperparameter. The training is first converged with $\beta = 0$, and then we turn on β to select the state whose $L_z = m_t$. In Figs. 4(a) and 4(b), we plot the charge density for quasiparticle and quasihole excitations from DeepHall for the $\nu = 1/3$ system with $\kappa = 1$, scaled by the square of sphere radius R . Notably, since the Laughlin wave function captures short-range physics [36], the excitations are more localized compared to the ED and

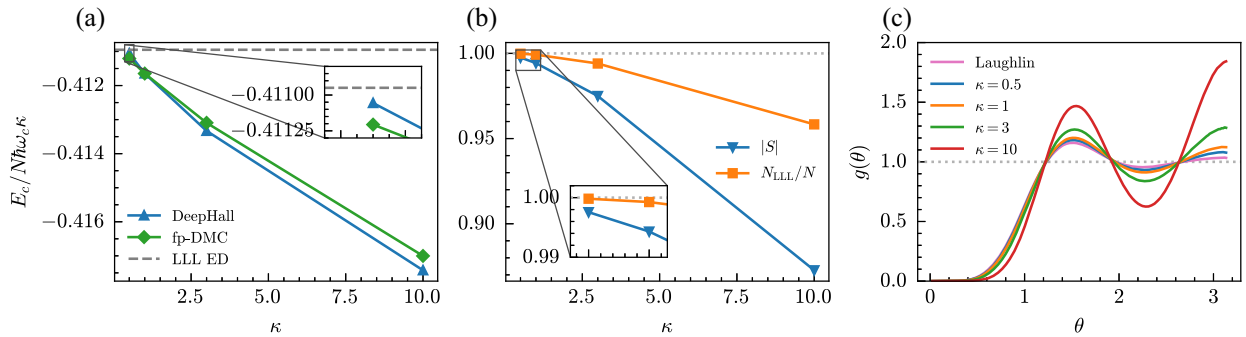


FIG. 3. (a) Energy per electron from DeepHall (blue) and fp-DMC (green) as a function of the Landau level mixing parameter κ . The error bars are smaller than the size of the markers. ED results, computed within the lowest Landau level, remain constant across κ and are shown as a horizontal dashed line. (b) Overlap modulus of the DeepHall wave function with the Laughlin wave function (blue) and the ratio of electrons on the lowest Landau level (orange) as a function of κ . The error bars are smaller than the size of the markers. (c) Pair correlation function $g(\theta)$ for different values of κ . All simulations are performed with $N = 6$ electrons, flux $2Q = 15$.

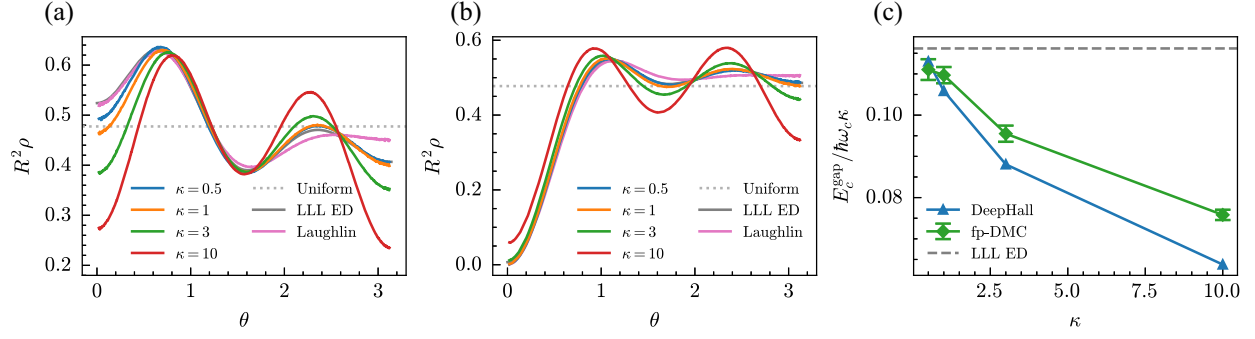


FIG. 4. (a) Density of quasiparticle states derived from ED, the Laughlin wave function, and DeepHall with various κ . The quasiparticle is localized at the north pole ($\theta = 0$). The density is scaled by the square of the radius R of the spherical geometry. (b) Similar to (a), but for the density of quasihole states. (c) Corrected transport gap from DeepHall (blue) and fp-DMC (green) as a function of the Landau level mixing parameter κ . The error bars for DeepHall are smaller than the marker size. Calculations are performed for $N = 6$ electrons, with fluxes $2Q = 14$ (quasiparticle), 15 (ground state), and 16 (quasihole).

the DeepHall results. For small κ , the DeepHall results agree well with the ED results, further demonstrating that our DeepHall wave function accurately captures the physics of the FQH system. As κ increases, the density fluctuations are enhanced compared to the ED results, consistent with the behavior of the PCF shown in Fig. 3(c). At a large $\kappa = 10$, the quasihole excitation density near the center is no longer close to zero, indicating a phase transition at high κ .

Having demonstrated the accuracy of DeepHall wave functions in describing excitations, we can now study the transport gap of FQH systems and the effects of LLM on it, which are important topics that pose challenges to the theoretical community [51–53]. The transport gap corresponds to the energy cost of moving a quasiparticle far from the system, leaving a quasihole behind. This gap can be measured in finite-temperature transport experiments [54,55]. It also reflects the composite fermion mass and can be deduced by the composite fermion Chern-Simons field theory [56–59]. The transport gap is defined as $E^{\text{gap}} = E_v^{\text{qp}} + E_v^{\text{qh}} - 2E_v$, where E_v^{qp} and E_v^{qh} are quasiparticle and quasihole energies, respectively. Similar to Fig. 2, we denote the corrected gap as E_c^{gap} . As shown in Fig. 4(c), the transport gap from DeepHall and fp-DMC align and match ED results for small κ . As κ increases, the transport gap decreases progressively. In particular, DeepHall predicts a smaller gap than fp-DMC in the strong LLM regime. Furthermore, Fig. S5 [41] demonstrates that DeepHall yields lower quasiparticle and quasihole excitation energies than fp-DMC in this regime, indicating its superior accuracy and the tendency of fp-DMC to overestimate the transport gap. Notably, even at $\kappa = 10$, the transport gap does not fully close, confirming that the system remains in a gapped phase despite strong LLM. These findings are consistent with previous studies [53,60,61].

In conclusion, we have demonstrated the effectiveness of deep learning methods in investigating FQH systems,

capturing contributions from higher Landau levels. We employ the spherical geometry and examine the bulk properties and quasiparticle excitations of the $\nu = 1/3$ and $\nu = 2/5$ states, where we show that real-space neural networks can learn more accurate wave functions than those obtained by exact diagonalization based on the lowest Landau levels. In a concurrent study, a similar approach to this work is devised to study FQH on disk geometry, where similar conclusions were drawn, further highlighting the strengths of neural network methods [62]. Together with this work, these successes highlight the potential of deep learning for future explorations, including investigations into additional filling factors and other topological properties. Our results also show hints of the phase transition, but further investigations are desired to elucidate the issue. In the future, deep learning frameworks can be expanded to other FQH geometries and tailored to study states with specific symmetries or topologies, as well as real materials with nonflat bands where exotic states such as the fractional Chern insulators are observed.

Acknowledgments—The authors would like to thank Yinghai Wu for insightful discussions and for sharing the 2LL ED data for the $2/5$ system. We are also grateful to Nicholas Regnault, Songyang Pu, Jiequn Han, and Xi Dai for their valuable discussions. Additionally, we acknowledge the ByteDance Research Group for their inspiration and encouragement. J. C. acknowledges supports from the National Key R&D Program of China under Grant No. 2021YFA1400500, the National Natural Science Foundation of China under Grant No. 12334003, and the Beijing Municipal Natural Science Foundation under Grant No. JQ22001. T. X. acknowledges supports by the NSFC Grant No. 12488201. T. Z. acknowledges support by the China Postdoctoral Science Foundation Grant No. 2023M743742.

Data availability—The data that support the findings of this Letter are openly available [63,64].

- [1] H. L. Stormer, D. C. Tsui, and A. C. Gossard, *Rev. Mod. Phys.* **71**, S298 (1999).
- [2] R. B. Laughlin, *Phys. Rev. Lett.* **50**, 1395 (1983).
- [3] I. Sodemann and A. H. MacDonald, *Phys. Rev. B* **87**, 245425 (2013).
- [4] V. J. Goldman, M. Santos, M. Shayegan, and J. E. Cunningham, *Phys. Rev. Lett.* **65**, 2189 (1990).
- [5] N. Thiebaut, N. Regnault, and M. O. Goerbig, *Phys. Rev. B* **92**, 245401 (2015).
- [6] J. P. Eisenstein, H. L. Stormer, L. Pfeiffer, and K. W. West, *Phys. Rev. Lett.* **62**, 1540 (1989).
- [7] D. R. Luhman, W. Pan, D. C. Tsui, L. N. Pfeiffer, K. W. Baldwin, and K. W. West, *Phys. Rev. Lett.* **101**, 266804 (2008).
- [8] Y.-L. Wu, B. Estienne, N. Regnault, and B. A. Bernevig, *Phys. Rev. Lett.* **113**, 116801 (2014).
- [9] N. Regnault, J. Maciejko, S. A. Kivelson, and S. L. Sondhi, *Phys. Rev. B* **96**, 035150 (2017).
- [10] D. Yoshioka, *J. Phys. Soc. Jpn.* **53**, 3740 (1984).
- [11] A. E. Feiguin, E. Rezayi, C. Nayak, and S. Das Sarma, *Phys. Rev. Lett.* **100**, 166803 (2008).
- [12] M. P. Zaletel, R. S. K. Mong, F. Pollmann, and E. H. Rezayi, *Phys. Rev. B* **91**, 045115 (2015).
- [13] G. Ortiz, D. M. Ceperley, and R. M. Martin, *Phys. Rev. Lett.* **71**, 2777 (1993).
- [14] J. Zhao, Y. Zhang, and J. K. Jain, *Phys. Rev. Lett.* **121**, 116802 (2018).
- [15] T. Zhao, A. C. Balram, and J. K. Jain, *Phys. Rev. Lett.* **130**, 186302 (2023).
- [16] Y.-H. Zhang and J.-R. Shi, *Chin. Phys. Lett.* **32**, 037101 (2015).
- [17] Y. Hu and J. K. Jain, *Phys. Rev. Lett.* **123**, 176802 (2019).
- [18] Y. Hu, Y. Ge, J.-X. Zhang, and J. K. Jain, *Phys. Rev. B* **104**, 035122 (2021).
- [19] J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W. M. C. Foulkes, D. Pfau, G. Carleo, and F. Noé, *Nat. Rev. Chem.* **7**, 692 (2023).
- [20] Y. Qian, X. Li, Z. Li, W. Ren, and J. Chen, *arXiv*: 2407.00707.
- [21] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [22] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, *Phys. Rev. Lett.* **122**, 250503 (2019).
- [23] J. Han, L. Zhang, and W. E, *J. Comput. Phys.* **399**, 108929 (2019).
- [24] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Phys. Rev. Res.* **2**, 033429 (2020).
- [25] J. Hermann, Z. Schätzle, and F. Noé, *Nat. Chem.* **12**, 891 (2020).
- [26] K. Choo, A. Mezzacapo, and G. Carleo, *Nat. Commun.* **11**, 2368 (2020).
- [27] I. von Glehn, J. S. Spencer, and D. Pfau, in *The Eleventh International Conference on Learning Representations, ICLR 2023* (OpenReview.net, kigali, Rwanda, 2023).
- [28] N. Yoshioka, W. Mizukami, and F. Nori, *Commun. Phys.* **4**, 1 (2021).
- [29] X. Li, Y. Qian, and J. Chen, *Phys. Rev. Lett.* **132**, 176401 (2024).
- [30] X. Li, Z. Li, and J. Chen, *Nat. Commun.* **13**, 7895 (2022).
- [31] M. Wilson, S. Moroni, M. Holzmann, N. Gao, F. Wudarski, T. Vegge, and A. Bhowmik, *Phys. Rev. B* **107**, 235139 (2023).
- [32] G. Cassella, H. Sutterud, S. Azadi, N. D. Drummond, D. Pfau, J. S. Spencer, and W. M. C. Foulkes, *Phys. Rev. Lett.* **130**, 036401 (2023).
- [33] J. Kim, G. Pescia, B. Fore, J. Nys, G. Carleo, S. Gandolfi, M. Hjorth-Jensen, and A. Lovato, *Commun. Phys.* **7**, 148 (2024).
- [34] X. Li, Y. Qian, W. Ren, Y. Xu, and J. Chen, *arXiv*:2406.11134.
- [35] D. Luo, D. D. Dai, and L. Fu, *arXiv*:2406.17645.
- [36] F. D. M. Haldane, *Phys. Rev. Lett.* **51**, 605 (1983).
- [37] D. Yoshioka, B. I. Halperin, and P. A. Lee, *Phys. Rev. Lett.* **50**, 1219 (1983).
- [38] X. G. Wen and A. Zee, *Phys. Rev. Lett.* **69**, 953 (1992).
- [39] T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- [40] T. T. Wu and C. N. Yang, *Nucl. Phys.* **B107**, 365 (1976).
- [41] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.134.176503> for training hyperparameters, tabular data, and additional discussions.
- [42] J. Martens and R. Grosse, in *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, 2015), pp. 2408–2417.
- [43] J. K. Jain, *Composite Fermions*, 1st ed. (Cambridge University Press, Cambridge; New York, 2007).
- [44] R. Morf, N. d’Ambrumenil, and B. I. Halperin, *Phys. Rev. B* **34**, 3037 (1986).
- [45] R. K. Kamilla, J. K. Jain, and S. M. Girvin, *Phys. Rev. B* **56**, 12411 (1997).
- [46] C. Yannouleas and U. Landman, *Phys. Rev. B* **68**, 035326 (2003).
- [47] C. Yannouleas and U. Landman, *Phys. Rev. B* **66**, 115315 (2002).
- [48] N. Shibata and D. Yoshioka, *Phys. Rev. Lett.* **86**, 5755 (2001).
- [49] J. J. Thomson, *Phil. Mag.* **7**, 237 (1904).
- [50] E. Wigner, *Phys. Rev.* **46**, 1002 (1934).
- [51] F. D. M. Haldane and E. H. Rezayi, *Phys. Rev. Lett.* **54**, 237 (1985).
- [52] R. H. Morf, N. d’Ambrumenil, and S. Das Sarma, *Phys. Rev. B* **66**, 075408 (2002).
- [53] T. Zhao, K. Kudo, W. N. Faugno, A. C. Balram, and J. K. Jain, *Phys. Rev. B* **105**, 205147 (2022).
- [54] G. S. Boebinger, A. M. Chang, H. L. Stormer, and D. C. Tsui, *Phys. Rev. Lett.* **55**, 1606 (1985).
- [55] R. R. Du, H. L. Stormer, D. C. Tsui, L. N. Pfeiffer, and K. W. West, *Phys. Rev. Lett.* **70**, 2944 (1993).
- [56] B. I. Halperin, P. A. Lee, and N. Read, *Phys. Rev. B* **47**, 7312 (1993).
- [57] G. Murthy and R. Shankar, *Rev. Mod. Phys.* **75**, 1101 (2003).
- [58] Y. Yu, Z.-B. Su, and X. Dai, *Phys. Rev. B* **57**, 9897 (1998).
- [59] A. Praz, *Phys. Rev. B* **75**, 205342 (2007).
- [60] V. Melik-Alaverdian and N. E. Bonesteel, *Phys. Rev. B* **52**, R17032 (1995).
- [61] G. Murthy and R. Shankar, *Phys. Rev. B* **65**, 245309 (2002).
- [62] Y. Teng, D. D. Dai, and L. Fu, *arXiv*:2412.00618.

- [63] Y. Qian, T. Zhao, J. Zhang, T. Xiang, X. Li, and J. Chen, [10.5281/zenodo.14591897](https://doi.org/10.5281/zenodo.14591897) (2025).
 [64] Y. Qian, DeepHall (v0.1.0), Zenodo (2025), [10.5281/zenodo.14849253](https://doi.org/10.5281/zenodo.14849253).
 [65] DiagHam, <https://nick-ux.org/diagham>.
 [66] M. Fishman, S. R. White, and E. M. Stoudenmire, *SciPost Phys. Codebases*, **4** (2022).
 [67] V. Melik-Alaverdian, N. E. Bonesteel, and G. Ortiz, *Phys. Rev. Lett.* **79**, 5286 (1997).

End Matter

Insights on monopole harmonics—Here we provide some insights on how the monopole harmonics $u_j^{Q+m} v_j^{Q-m}$ handle the complex phases arising from the Dirac string. Let us first consider the case without the Dirac string, where divergences in local energy near the poles also exist. Consider a one-electron wave function $\psi \propto \theta^n e^{im\phi}$ near the north pole. To avoid divergence in local energy, the condition $m = \pm n$ must be satisfied. Similarly, this condition must also hold near the south pole. With the Dirac string, the condition near the north pole becomes $m = Q \pm n$, and near the south pole, it becomes $m = -Q \pm n$. Examining the monopole harmonics $u_j^{Q+m} v_j^{Q-m}$ reveals that orbitals with more complicated phases (i.e., larger $|m|$) contribute more significantly near the poles, which must be captured by the designed wave function ansatz. Additionally, since the largest $|m|$ is proportional to N for a fixed filling, the phase becomes even more complex and challenging to learn compared to the $Q = 0$ case, where the largest $|m|$ is proportional to N^2 . By including LLL monopole harmonics in our DeepHall wave function, the complicated phases of the wave function arising from the Dirac string can be properly handled.

Neural network settings: In the neural network part, $f_k(\mathbf{r}_j; \{\mathbf{r}_{\neq j}\})$, the input feature vector \mathbf{f}_i^0 for the electron i is chosen as the Cartesian coordinates of the electron:

$$\mathbf{f}_i^0 = [\sin \theta_i \cos \phi_i, \sin \theta_i \sin \phi_i, \cos \theta_i]. \quad (\text{A1})$$

The input features are then mapped to the attention inputs \mathbf{h}_i^0 by a linear projection $\mathbf{W}^0 \mathbf{f}_i^0$. After that, \mathbf{h}_i^0 is passed through a sequence of multihead attention layers and fully connected layers with residual connection:

$$\mathbf{f}_i^{l+1} = \mathbf{h}_i^l + \mathbf{W}_o^l \text{concat}_h \times [\text{SELFATTN}_i(\mathbf{h}_1^l, \dots, \mathbf{h}_N^l; \mathbf{W}_q^{lh}, \mathbf{W}_k^{lh}, \mathbf{W}_v^{lh})], \quad (\text{A2})$$

$$\mathbf{h}_i^{l+1} = \mathbf{f}_i^{l+1} + \tanh(\mathbf{W}^{l+1} \mathbf{f}_i^{l+1} + \mathbf{b}^{l+1}), \quad (\text{A3})$$

where l indexes neural network layers and h indexes heads of the attention layer. And the standard self-attention is defined as

$$\text{SELFATTN}_i(\mathbf{h}_1, \dots, \mathbf{h}_N; \mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v) = \frac{1}{\sqrt{d}} \sum_j \sigma_j(\mathbf{q}_i^\top \mathbf{k}_j, \dots, \mathbf{q}_N^\top \mathbf{k}_j) \mathbf{v}_j, \quad (\text{A4})$$

$$\mathbf{k}_i = \mathbf{W}_k \mathbf{h}_i, \quad \mathbf{q}_i = \mathbf{W}_q \mathbf{h}_i, \quad \mathbf{v}_i = \mathbf{W}_v \mathbf{h}_i, \quad (\text{A5})$$

$$\sigma_i(x_1, \dots, x_N) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}, \quad (\text{A6})$$

where d is the output dimension of the key and query weights. The neural network hyperparameters used in this study are listed in Table S1 [41].

Energy corrections: The energy results shown in the main text include the contribution from background charge,

$$E_{\text{bg}} = E_{\text{el-bg}} + E_{\text{bg-bg}} = \kappa \hbar \omega_c \ell \left(-\frac{N^2}{R} + \frac{N^2}{2R} \right) = -\kappa \hbar \omega_c \frac{N^2}{2\sqrt{Q}}. \quad (\text{A7})$$

In addition, we shift the energy by $N\omega_c/2$ and apply a density correction to reduce the dependence of the energy on the system size. The corrected energy E_c is then defined as

$$E_c = \sqrt{\frac{2Q\nu}{N}} \left(E_v + E_{\text{bg}} - \frac{N\omega_c}{2} \right). \quad (\text{A8})$$

Notably, the background contribution to the excited states differs from that of the ground state. Specifically, for the $\nu = 1/3$ state, these contributions are given by

$$E_{\text{bg}}^{\text{qp/qh}} = -\kappa \hbar \omega_c \frac{N^2 - q^2}{2\sqrt{Q^{\text{qp/qh}}}}, \quad (\text{A9})$$

where $|q| = 1/3$ corresponds to the excess charge in the excited states. And the transport gap after correction E_c^{gap} is essentially

$$E_c^{\text{gap}} = \sqrt{\frac{2Q^{\text{qp}}\nu}{N}} \left(E_v^{\text{qp}} + E_{\text{bg}}^{\text{qp}} - \frac{N\omega_c}{2} \right) + \sqrt{\frac{2Q^{\text{qh}}\nu}{N}} \left(E_v^{\text{qh}} + E_{\text{bg}}^{\text{qh}} - \frac{N\omega_c}{2} \right) - 2E_c. \quad (\text{A10})$$

Wave function analyses: The Laughlin wave function for $\nu = 1/m$ state on a sphere is given by

$$\psi^{\text{Laughlin}} = \prod_{i < j} (u_i v_j - u_j v_i)^m, \quad (\text{A11})$$

where $u_i = \cos(\theta_i/2)e^{i\phi_i/2}$ and $v_i = \sin(\theta_i/2)e^{-i\phi_i/2}$ are the spinor coordinates of the i th electron, and m is an odd integer. The overlap between the DeepHall wave function ψ^{DeepHall} and ψ^{Laughlin} is defined in Eq. (6). Given the two wave functions are similar, the overlap can be efficiently sampled using importance sampling from the distribution $|\psi^{\text{DeepHall}}|^2$, expressed as

$$|S|^2 = \frac{|\langle \psi^{\text{Laughlin}} / \psi^{\text{DeepHall}} \rangle|^2}{\langle |\psi^{\text{Laughlin}} / \psi^{\text{DeepHall}}|^2 \rangle}, \quad (\text{A12})$$

where $\langle \cdot \rangle$ denote the expectation value under the distribution $|\psi^{\text{DeepHall}}|^2$.

The number of electrons on the lowest Landau level can be calculated using the trace of the one-body reduced density matrix (1-RDM), with the basis consisting of LLL monopole harmonics φ_i . The 1-RDM for wave function ψ is defined as

$$\Gamma_{ij} = \sum_{a=1}^N \int d^2\mathbf{r}_1 \cdots d^2\mathbf{r}_N d^2\mathbf{r}' [\psi^*(\mathbf{r}_1, \dots, \mathbf{r}_N) \varphi_i(\mathbf{r}_a) \times \varphi_j^*(\mathbf{r}') \psi(\mathbf{r}_1, \dots, \mathbf{r}_{a-1}, \mathbf{r}', \mathbf{r}_{a+1}, \dots, \mathbf{r}_N)]. \quad (\text{A13})$$

In practice, importance sampling based on the distribution $|\psi|^2$ is applied for the integral over $d^2\mathbf{r}_1 \cdots d^2\mathbf{r}_N$, and uniform sampling is used for $d^2\mathbf{r}'$. Assuming the monopole harmonics are normalized, the integral in Eq. (A13) can be calculated with

$$\left\langle \frac{\psi(\mathbf{r}_1, \dots, \mathbf{r}_{a-1}, \mathbf{r}', \mathbf{r}_{a+1}, \dots, \mathbf{r}_N)}{\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)} \varphi_i(\mathbf{r}_a) \varphi_j^*(\mathbf{r}') \right\rangle. \quad (\text{A14})$$

The pair correlation function is defined as

$$g(\mathbf{r}) = \frac{1}{\rho N} \left\langle \sum_{i \neq j} \delta^{(2)}(\mathbf{r} - \mathbf{r}_i + \mathbf{r}_j) \right\rangle, \quad (\text{A15})$$

where ρ is the electron density on the sphere. Since $g(\mathbf{r}) = g(|\mathbf{r}|)$ on the spherical geometry, the PCF depends only on the angular separation θ_{ij} between electrons i and j . Therefore, we can express the PCF as

$$g(\theta) = \frac{1}{\rho N} \left\langle \sum_{i \neq j} \delta(\theta - \theta_{ij}) \right\rangle. \quad (\text{A16})$$

We approximate the δ function as

$$\delta(\theta) \approx \begin{cases} \frac{1}{\Delta S} & \text{if } |\theta| < \frac{\pi}{2K}, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A17})$$

where K is the number of bins, and $\Delta S = 2\pi R^2 \sin \theta \Delta \theta$ with $\Delta \theta = \pi/K$.

Computational cost: The computational cost for each data point in Fig. 1(c) is approximately tens of V100 GPU hours, with detailed per-step training times and total convergence steps provided in Fig. S3 [41]. For the particle numbers studied (6–12 electrons), the per-step computational timescales as $O(N^2)$ (Fig. S3a [41]), governed by local energy evaluations. Concurrently, the number of training steps required for convergence scales linearly with particle numbers (Fig. S3b [41]), leading to an overall $O(N^3)$ time complexity. While theoretical per-step time complexity for QMC ranges between $O(N^3)$ and $O(N^4)$, the $O(N^2)$ scaling dominates in practice here due to system size and prefactor effects.

ED calculations: ED calculations of energy are performed using the DiagHam library [65]. The density profiles of quasiparticle and quasihole excitations are obtained by DMRG calculations in the LLL using the ITensor library for technical convenience [66]. We have verified that the energy values calculated by DMRG are identical to those obtained from DiagHam and the entanglement entropy is also converged to machine precision. These facts guarantee that the results from DMRG calculations are equivalent to ED calculations.

Fp-DMC calculations: Fp-DMC calculations of energy are performed based on the approach in Refs. [13,53,67], where the trial wave functions are specified to be the Laughlin state and the corresponding quasihole and quasiparticle excited states given by the composite fermion theory [43]. In a fp-DMC calculation, the true ground state wave function is a complex function whose phase is approximated by the phase of the given trial wave function, while its amplitude is then optimized during the stochastic process. Similar to the fixed-node DMC, the fp-DMC subjects to the variational principle, because the fp-DMC energy is optimized within the given phase sector.