Artificial Intelligence for Quantum Matter: Finding a Needle in a Haystack

Khachatur Nazaryan,[1,*] Filippo Gaggioli,[1,*] Yi Teng,[1] and Liang Fu[1]

[1] *Department of Physics, Massachusetts Institute of Technology, Cambridge, MA-02139,USA*

(Dated: July 18, 2025)

Neural networks (NNs) have great potential in solving the ground state of various many-body problems. However, several key challenges remain to be overcome before NNs can tackle problems and system sizes inaccessible with more established tools. Here, we present a general and efficient method for learning the NN representation of an arbitrary many-body complex wave function. Having reached overlaps as large as 99.9% for as many as 25 particles, we employ our neural wave function for pre-training to effortlessly solve the fractional quantum Hall problem for 20 electrons with Coulomb interactions and realistic Landau-level mixing.

Introduction – A fundamental challenge in many-body physics is the astronomical size of the Hilbert space: the number of complex amplitudes needed to completely specify a N-particle quantum wave function grows so quickly with N that even modest systems outrun data storage and brute-force algorithms. Quantum computers could in principle solve certain quantum many-body problems efficiently, but with today's noisy intermediate scale quantum processors, much of this promise is yet to be fulfilled. Recently, the artificial intelligence (AI) boom opened a different path [1–11]: representing complex quantum wave functions with neural networks containing a tractable set of parameters and finding accurate approximation to ground states with present-day computing resources.

Can a neural network architecture *accurately* and *efficiently* capture the vast variety of many-body ground states of diverse quantum phases of matter (such as magnets, superconductors and topological materials)? To grasp the scale of the challenge, recall that the complex wave function of a single particle in two spatial dimensions can be rendered as a colourful image whose intensity encodes amplitude $|\psi(\mathbf{r})|$ and hue encodes phase $\varphi(\mathbf{r})$. Learning the wave function of N particles amounts to learning to generate a "hyper-image" that inhabits a $2N$-dimensional configuration space.

As a concrete measure of the expressive power of neural networks, consider the *needle-in-a-haystack* problem: training a neural network to reproduce a target many-body wave function $|\psi_{\text{ref}}\rangle$ that resides in the vastness of the Hilbert space. Success on this task would yield substantial rewards. It can be used for pre-training purpose to initialize networks at physically informed starting point, accelerating subsequent ground state search by energy minimization in neural network variational Monte Carlo (NN-VMC). In addition, training a neural network on a library of reference wavefunctions opens the door to data-driven transfer-learning applications, such as predicting the electronic properties of a novel molecule from existing ones.

While this needle-in-a-haystack task is easy to understand, it is by no means easy to achieve. Even for a small system, almost all N-particle wave functions have vanishing overlap with the target ψ_{ref}, and direct maximization of $|\langle\psi_{\text{ref}}|\psi\rangle|^2$ via gradient descent is extremely challenging. To date, a general method for representing non-trivial target wave functions using neural networks is lacking.

Last but not the least, quantum statistics of identical particles imposes a fundamental constraint in their wave functions $\psi_{\text{ref}}(\mathbf{r}_1, ..., \mathbf{r}_N)$, which must be anti-symmetric under the permutation of any two particles in Fermi systems. To comply with this condition, various Fermi neural network architectures have been introduced for electron systems in continuous space [5, 6, 10–13]. Compared with standard neural networks, their expressive power and training protocol are much less studied or benchmarked. The needle-in-a-haystack task would provide an objective "score" for the performance of Fermi neural network architectures.

In this work, we develop a general and efficient method for learning the neural network representation of many-body wavefunctions. To circumvent the problem plaguing direct overlap maximization, we introduce a new training objective that targets the probability density and probability current of ψ_{ref}. Our method is naturally suited to learning complex-valued wave functions, which appear ubiquitously in magnetic, chiral and spin-orbit-coupled quantum systems.

We test our method on archetypal many-body wave functions: the Laughlin state and the Moore Read state in fractional quantum Hall systems, which represent topological quantum liquids hosting fractionally charged quasiparticles ("anyons") with Abelian and non-Abelian statistics, respectively. A *general-purpose* Fermi neural network architecture based on self-attention is employed for both tasks, without prior knowledge of quantum Hall physics. Remarkably, our *unsupervised* learning method successfully finds neural network representations of these highly-entangled wavefunctions, reaching overlaps as large as 99.9% for as many as 25 particles.

Using these trained neural networks and performing NN-VMC [14, 15] for energy minimization, we effortlessly solve the ground state of the fractional quantum Hall sys-

* These two authors contributed equally.

tem for $N = 20$ particles with Coulomb interaction and realistic Landau-level mixing. This success demonstrates the power of our method for pretraining on physically motivated ansatz, enabling fast and accurate neural network solution of strongly correlated electron systems.

Loss functions – For the needle problem, the key figure of merit is the fidelity (or squared overlap) $F = |\langle\psi_{\rm ref}|\psi_\theta\rangle / \|\psi_{\rm ref}\|\|\psi_\theta\|\,|^2$, with the wave function norm defined as $\|\psi\|^2 = \langle\psi|\psi\rangle$. The fidelity naturally provides us with a simple choice for the loss function $L_{\scriptscriptstyle F} = 1 - F$. In the form of a Monte Carlo expectation value, this reads

$$L_{\scriptscriptstyle F} = 1 - \frac{\left|\int \mathrm{d}\mathbf{R}\,|\psi_\theta(\mathbf{R})|^2 \psi_{\rm ref}(\mathbf{R})/\psi_\theta(\mathbf{R})\right|^2 / \mathcal{N}^2}{\int \mathrm{d}\mathbf{R}\,|\psi_\theta(\mathbf{R})|^2\,|\psi_{\rm ref}(\mathbf{R})/\psi_\theta(\mathbf{R})|^2/\mathcal{N}}, \quad (1)$$

where $\mathcal{N} = \int \mathrm{d}\mathbf{R}\,|\psi_\theta(\mathbf{R})|^2$ and the integration variable $\mathbf{R} = (\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N)$ spans the $\mathbb{R}^{2N}$ coordinate space of N particles in 2D. However, the overlap of $\psi_{\rm ref}$ with another wave function is in general exponentially small, implying that the gradients of $L_{\scriptscriptstyle F}$ will be unable to guide the neural network across the optimization landscape for all but the smallest system size. In the case of real $\psi_{\rm ref}$, the exponentially small gradients can be "amplified" by working with the logarithms of the wave functions [16]. This simple fix, however, is not sufficient in the case of truly complex $\psi_{\rm ref}$.

While the modulus of the wave function represents the N-particle probability density $\rho(\mathbf{R}) = |\psi(\mathbf{R})|^2/\mathcal{N}$ and is closely related to physical observables, the phase φ is a more subtle quantity that cannot be directly accessed experimentally and is only defined up to a constant. The phase gradient $\nabla\varphi$, on the other hand, encodes important information about the current flowing within the system: $\mathbf{j} \propto \rho\nabla\varphi$ represents the probability current density. Motivated by this observation, we introduce a new loss function that consists of two parts, L_ρ and L_j, respectively designed to minimize the difference in the particle density and the phase gradients between the trial and target wave functions.

The density loss function L_ρ is inspired by the Kullback–Leibler divergence [17] that measures the distance between the probability distributions $|\psi_\theta|^2$ and $|\psi_{\rm ref}|^2$, and reads

$$L_\rho = \frac{1}{\mathcal{N}} \int \mathrm{d}\mathbf{R}\,|\psi_\theta(\mathbf{R})|^2 \left(\ln|\psi_\theta(\mathbf{R})/\psi_{\rm ref}(\mathbf{R})|^2\right)^2. \quad (2)$$

As discussed above, this particular choice of L_ρ has the advantage that it retains sensitivity when either of $|\psi_\theta|^2$ and $|\psi_{\rm ref}|^2$ is very small, thanks to the difference between logarithms. The current loss function L_j instead takes the simple form

$$L_j = \frac{1}{\mathcal{N}} \int \mathrm{d}\mathbf{R}\,|\psi_\theta(\mathbf{R})|^2 \sum_\ell |\nabla_\ell\,\varphi_\theta(\mathbf{R}) - \nabla_\ell\,\varphi_{\rm ref}(\mathbf{R})|^2, \quad (3)$$

where ∇_ℓ is the gradient with respect to the ℓ-th particle position $\boldsymbol{r}_\ell$, while φ_θ and $\varphi_{\rm ref}$ are the phases of ψ_θ and $\psi_{\rm ref}$ [18]. Due to the presence of the gradient, the current loss function (3) captures the spatial variation of the phase (which is physically observable) and prevents the fragmentation of φ_θ into local patches that differ by integer multiples of 2π. Moreover, the non-local character of the spatial derivatives allows the loss function to probe the low-density regions otherwise inaccessible to the Monte Carlo sampling. These properties make L_j very well-suited for capturing the phase pattern of wave functions that display singularities such as vortices, as we will show below.

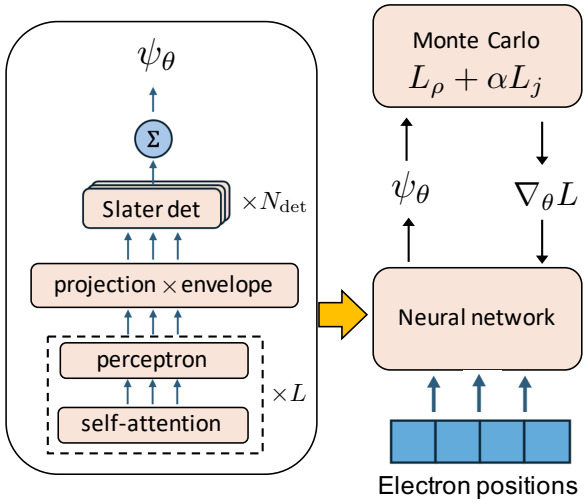

FIG. 1. **Fermionic neural network and VMC:** Illustration of our fermionic attention-based architecture (left), and its role inside the NN variational Monte Carlo (right).

The total loss function is finally obtained by summing Eqs. (2)-(3),

$$L = L_\rho + \alpha L_j. \quad (4)$$

The coefficient $\alpha > 0$ is an important hyperparameter that balances the relative weight of the density- and current loss functions, and needs to be optimized depending on the choice (and normalization) of $\psi_{\rm ref}$.

Our method is applicable to both Bose and Fermi systems. In the rest of this work, we will demonstrate its effectiveness for Fermi systems.

Fermionic neural network – A number of neural network architectures have been developed to represent fermion wave functions in continuous space. Commonly used architectures, such as PauliNet [5] and FermiNet [6] and self-attention based neural networks [12, 13], take the particle coordinates as input, combine them into a set of "orbitals" that depend on the positions of all electrons, and finally assemble these *many-electron orbitals* into Slater determinants to construct an anti-symmetric wave-function that respects Fermi statistics. By incorporating multiparticle correlations into many-electron orbitals, these neural ansatz go beyond Hartree-Fock approximation and can capture the ground states of various correlated electron systems, as demonstrated for atoms, molecules and solids [5–7, 12, 13].

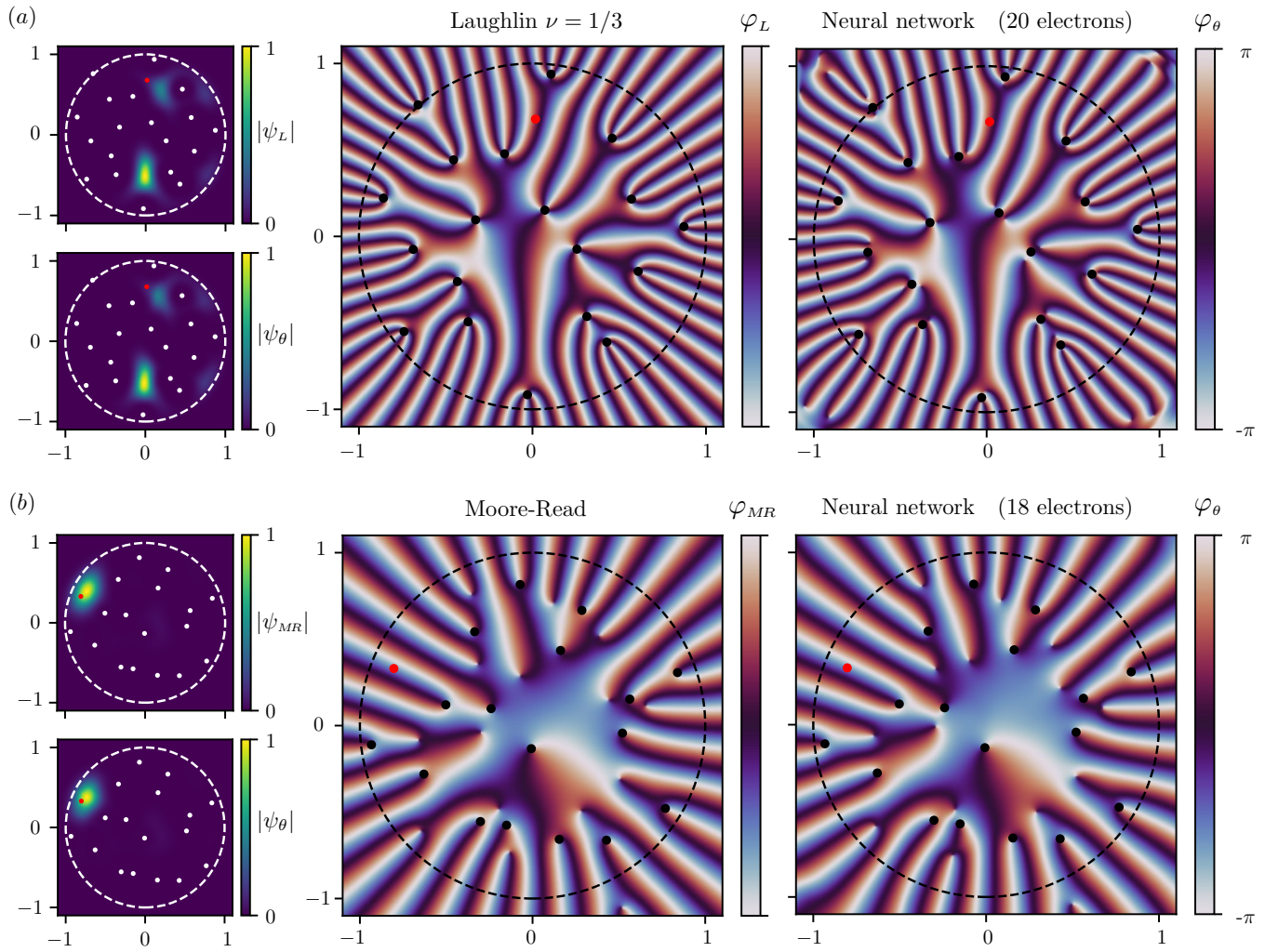

FIG. 2. **Laughlin and MR wave-functions:** Comparison of the wave functions for the Laughlin (a) and MR (b) state, Eqs. (6)-(7), with the output of the neural network ($L = 2$ self-attention layers and $N_{\rm det} = 4$ determinants). These plots are obtained by keeping $N - 1$ particles at fixed positions (black/white dots), obtained from Monte Carlo sampling, and moving the remaining particle away from its "original" position (red dot) across the 2D plane (positions in units of the droplet radius $R_L = \sqrt{6N\ell_M^2}$ (a) and $R_{MR} = \sqrt{4N\ell_M^2}$ (b), with $\ell_M^2 = \phi_0/2\pi H$ the magnetic length associated to the out of plane field H).

Our neural network ansatz is inspired by the transformer architecture originally proposed in the context of large language models [19], and uses the self-attention mechanism to capture electron correlations [12, 13]. As illustrated in Fig. 1, it consists of a stack of self-attention and perceptron layers, repeated L times, that takes the electron positions $\boldsymbol{r}_j$ as input and outputs vectors that, after projection and convolution with a simple Gaussian envelope, create the generalized single-particle orbitals $\phi_i^{(k)}\left(\boldsymbol{r}_j, \{\boldsymbol{r}_{/j}\}\right)$. These are finally combined into $N_{\rm det}$ Slater determinants, whose sum constitutes the antisymmetric fermionic neural wave function

$$\psi_\theta(\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N) = \sum_k^{N_{\rm det}} \det\left[\phi_i^{(k)}\left(\boldsymbol{r}_j, \{\boldsymbol{r}_{/j}\}\right)\right], \qquad (5)$$

that is parametrized by the network weights θ.

Having constructed the wave function, ψ_θ is used as a variational ansatz in the VMC algorithm, where the desired loss function L is evaluated by means of Monte Carlo techniques. The gradients $\nabla_\theta L$ of the loss function are finally passed back to the neural network to update the weights θ via standard backpropagation after each training step.

Results – The fractional quantum Hall effect (FQH) is an archetypal problem of many-body condensed-matter physics, and showcases an intricate interplay between strong electronic correlations and non-trivial topology. Much of the field's progress has come from remarkably insightful trial wave-functions, most famously Laughlin's [20]

$$\psi_L = \prod_{i<j}(z_i - z_j)^3 \exp\left(-|z_i|^2/4\right), \qquad (6)$$

which captures the essential physics of the true ground state at filling 1/3. The Laughlin state supports charge-1/3 quasiparticles that are Abelian anyons. Another celebrated trial wavefunction is the Moore–Read Pfaffian state [21]

$$\psi_{MR} = \mathrm{Pf}\left(\frac{1}{z_i - z_j}\right)\prod_{i<j}(z_i - z_j)^2 \exp\left(-|z_i|^2/4\right), \quad (7)$$

which supports charge-1/4 quasiparticles that have non-Abelian statistics. The Moore-Read wave function (7) can be viewed as a BCS paired state of composite fermions, and hence belongs to a different and more exotic "universality class" than the Laughlin state.

The wave functions ψ_L and ψ_{MR} are shown in Fig. 2 (a) and (b) as a function of the position of a single particle, while the remaining $N-1$ are fixed (white/black dots) in a typical configuration that was sampled from Eqs. (6)-(7) using Monte Carlo methods ($N = 20$ for Laughlin and $N = 18$ for Moore-Read). The absolute values $|\psi_L|$ and $|\psi_{MR}|$ (top left panels) have the spatial profile characteristic of strongly correlated systems: the position of the "last" particle is strongly constrained by every other particle's coordinates. The phases φ_L and φ_{MR} (central panels), on the other hand, display an intricate pattern: the Laughlin state generally features vortices with 6π phase winding where two particles coincide, while the phase pattern of the Moore-Read state is even more subtle. The highly complex nature of these model wavefunctions, which embodies the universal physics of the fractional quantum Hall effect, makes them the ideal "needles" for testing our neural network learning method.

Evaluating $|\psi_\theta|$ and φ_θ on the same pair of electron configurations using our attention-based neural network, we obtained the results shown in the remaining panels of Fig. 2 (a) and (b). For these plots, we trained our NN using the loss function (4), with the hyper-parameter α gradually increasing from zero to unity. The modulus of ψ_θ faithfully captures the strongly correlated electron density. Even more remarkable is the network's phase prediction, φ_θ, which accurately reproduces the intricate patterns of φ_L and φ_{MR} not only in the high-density regions that dominate the Monte-Carlo averages, but also in the low-density areas near the nodes of the target wave functions, where $|\psi_L|^2$ and $|\psi_{MR}|^2$ are vanishingly small. The accuracy of φ_θ close to these points is a beneficial consequence of the non-locality of L_j, as anticipated in the discussion below Eq. (3).

Application to pre-training – Recently, self-attention-based neural networks have demonstrated impressive success in finding the ground state of the fractional quantum Hall problem with Landau-level (LL) mixing for system sizes up to twelve particles, outperforming traditional approaches, such as exact diagonalization (ED) with Landau level truncation [22, 23]. Indeed, while ED is fundamentally limited by the exponential growth of the Hilbert space, neural network based variational method can in principle avoid this bottleneck and attain accurate solution for large systems. However, as the system size increases, the optimization landscape becomes increasingly complex and the neural network training can easily fail to converge, even with substantial computational time and resources.

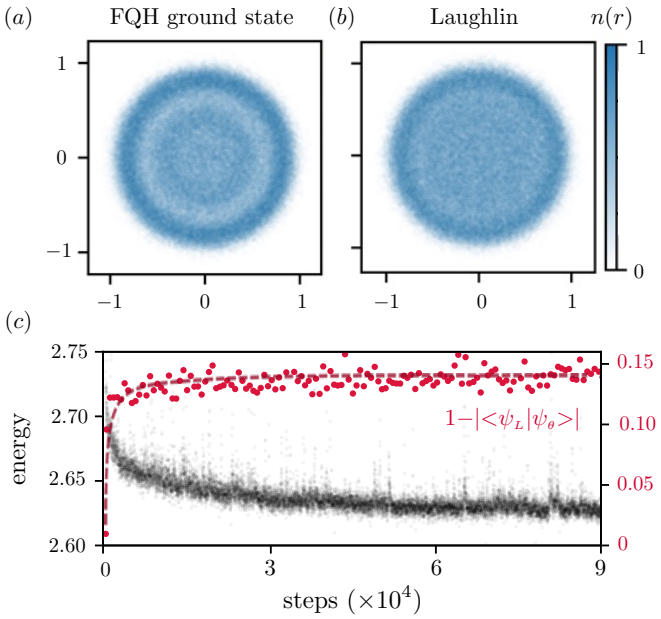

FIG. 3. **FQH ground state:** Spatial density profiles of the FQH ground state for mixing parameter $\lambda = 1$ (a) and Laughlin droplet (b) ($N = 20$ and positions expressed in units of R_L). (c) Evolution of the variational energy (black) and "distance" from the Laughlin state (red), as measured by $1 - |\langle\psi_L|\psi_\theta\rangle|$. As the energy gradually decreases, the wave function ψ_θ diverges away from ψ_L (the dashed line provides a guide for the eye).

By pre-training our neural network to maximize the overlap with ψ_L, we are now able to overcome this problem and efficiently solve the FQH problem with strong LL mixing for an unprecedented system size. For 20 electrons (which is inaccessible to even ED within the lowest Landau level), the corresponding results for mixing parameter $\lambda = e^2/4\pi\varepsilon_0\epsilon\ell_B = 1$ are shown in Fig. 3, where we compare the spatial density profile for the FQH (a) and Laughlin (b) droplet in disk geometry. There, it becomes evident that the long-ranged Coulomb repulsion induces slowly-decaying oscillation in the charge density away from the edge, consistent with previous studies on smaller system sizes [22]. The decrease in energy, of the order of $\approx 4\%$ when compared to the initial Laughlin state, is shown in panel (c) (small black dots), along with the rapid evolution of ψ_θ away from ψ_L as measured by the "distance" $1 - |\langle\psi_L|\psi_\theta\rangle|$ (red dots), which goes from $\approx 1\%$ to $\approx 14\%$. These results clearly show the important distinction between the Laughlin wavefunction and the actual Coulomb ground state. On the other hand, during the entire training process, the total angular momentum of the system remained very close (≈ 569.90) to the integer value of 570 for the Laughlin state with 20

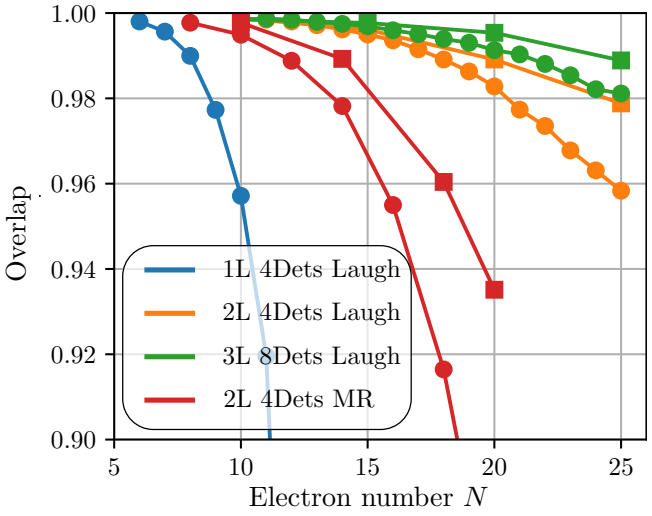

FIG. 4. **Scaling analysis:** Overlap with the Laughlin wave function as a function of the particle number, for three different architectures (blue, orange and green colors) and two different training protocols (circle and square markers). The red curve is for the Moore-Read state.

particles, demonstrating the robust quantization of topological quantum number in the fractional quantum Hall liquid.

Altogether, these results demonstrate that our self-attention NN is capable of solving the FQH problem with realistic LL mixing for large system sizes, once the neural network is appropriately pre-trained. All of this is achieved with very modest computational time and resources (60 thousand steps in $\approx$ 24h time on a single V100 GPU node).

Scaling analysis – To conclude our discussion, we go back to the needle problem and discuss the scaling of the overlap as a function of particle number for three different self-attention architectures, with varying number of layers ($L = 1, 2, 3$) (blue, orange/red and green) and Slater determinants ($N_{\rm det} = 4, 8$), but identical attention heads ($\times 4$ with dimension 64) and MLP hidden dimension (256). At the same time, we compare two different training protocols: a shorter one (circles), where the overlap is learned entirely by minimizing the loss function (4) for a fixed number of steps with the hyper-parameter α gradually increasing from zero to unity (10^{-3} learning rate); and a longer one (squares), where in a second part of the training the fidelity loss (1) is directly minimized (10^{-2} learning rate).

As shown in Fig. 4, the 2- and 3-layer architectures (orange and green, number of parameters $\approx 84 \times 10^4$ and 130×10^4 respectively) excel at reproducing the Laughlin wave function for up to 25 particles, the largest system size studied in this work. At the same time, the Moore-Read state for 20 particles can be faithfully reproduced with $\approx 94\%$ overlap using the 2-layers architecture (red). This favorable scaling highlights the expressive power of self-attention networks for capturing quantum phases of matter and suggests that our method for deep learning a target many-body wavefunction is well-suited to tackle even larger system sizes.

Discussion – The versatility, accuracy and efficiency of neural networks are the crucial ingredients underpinning the rapid development of AI-based methods across different branches of condensed matter physics. Our work expands the AI-for-quantum horizon by introducing a general unsupervised learning method to represent arbitrary wave functions using general neural networks. By targeting the Laughlin and Moore-Read wave functions, which describe archetypal topologically ordered many-body states, we demonstrate high overlaps $> 99\%$ for as many as 25 particles using a simple self-attention NN without prior knowledge of quantum Hall physics. Performing NN-VMC for energy minimization with these pre-trained neural networks, we effortlessly solve the ground state of the fractional quantum Hall system with Coulomb interaction and strong Landau-level mixing for an unprecedented system size.

For comparison, it is known that model FQH wave functions such as the Laughlin and Moore-Read states have exact matrix product states (MPS) representations [24, 25]. This property, which enables efficient density matrix renormalization group calculation on the cylinder geometry, is, however, the consequence of special conformal properties of the model wave functions that do not extend to the ground state of realistic FQH systems with Landau level mixing. As a result, we believe that the success of our NN learning method truly stands out.

Our general method provides an invaluable tool for pre-training neural networks, and this pre-training technique opens the door to many applications of neural networks to quantum condensed matter physics, enabling in particular accurate and efficient solution of many-body systems in continuous space for which traditional methods suffer from band-projection or discretization error. Of particular interest are the study of non-Abelian fractional quantum Hall states, moiré fractional Chern insulators, and chiral superconductivity [26, 27].

Acknowledgments – This work was primarily supported by National Science Foundation (NSF) Convergence Accelerator Award No. 2235945. We acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing computing resources that have contributed to the research results reported within this paper. F.G. is grateful for the financial support from the Swiss National Science Foundation (Postdoc.Mobility Grant No. 222230). K.N. acknowledges the support from the NSF through Award No. PHY-2425180. L.F. was supported by a Simons Investigator Award from the Simons Foundation.

[1] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, Nature Physics **13**, 431 (2017).

[2] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Neural-network quantum states, string-bond states, and chiral topological states, Phys. Rev. X **8**, 011006 (2018).

[3] G. Carleo, K. Choo, D. Hofmann, J. E. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O'Reilly, H. Théveniaut, G. Torlai, F. Vicentini, and A. Wietek, Netket: A machine learning toolkit for many-body quantum systems, SoftwareX **10**, 100311 (2019).

[4] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, Phys. Rev. Lett. **122**, 226401 (2019).

[5] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic schrödinger equation, Nature Chemistry **12**, 891 (2020).

[6] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, Ab initio solution of the many-electron schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).

[7] X. Li, Z. Li, and J. Chen, Ab initio calculation of real solids via neural network ansatz, Nature Communications **13**, 7895 (2022).

[8] C. Roth, A. Szabó, and A. H. MacDonald, High-accuracy variational monte carlo for frustrated magnets with deep neural networks, Phys. Rev. B **108**, 054410 (2023).

[9] G. Cassella, H. Sutterud, S. Azadi, N. D. Drummond, D. Pfau, J. S. Spencer, and W. M. C. Foulkes, Discovering quantum phase transitions with fermionic neural networks, Phys. Rev. Lett. **130**, 036401 (2023).

[10] G. Pescia, J. Nys, J. Kim, A. Lovato, and G. Carleo, Message-passing neural quantum states for the homogeneous electron gas, Phys. Rev. B **110**, 035108 (2024).

[11] D. Luo, D. D. Dai, and L. Fu, Pairing-based graph neural network for simulating quantum materials (2023), arXiv:2311.02143 [cond-mat.str-el].

[12] I. von Glehn, J. S. Spencer, and D. Pfau, A self-attention ansatz for ab-initio quantum chemistry (2023), arXiv:2211.13672 [physics.chem-ph].

[13] M. Geier, K. Nazaryan, T. Zaklama, and L. Fu, Is attention all you need to solve the correlated electron problem? (2025), arXiv:2502.05383 [cond-mat.str-el].

[14] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[15] J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W. M. C. Foulkes, D. Pfau, G. Carleo, and F. Noé, Ab initio quantum chemistry with neural-network wavefunctions, Nature Reviews Chemistry **7**, 692 (2023).

[16] M. J. S. Beach, I. D. Vlugt, A. Golubeva, P. Huembeli, B. Kulchytskyy, X. Luo, R. G. Melko, E. Merali, and G. Torlai, QuCumber: wavefunction reconstruction with neural networks, SciPost Phys. **7**, 009 (2019).

[17] S. Kullback and R. A. Leibler, On Information and Sufficiency, The Annals of Mathematical Statistics **22**, 79 (1951).

[18] Equation (3) is written for electron systems in continuous space, but an analogous expression can be obtained for systems on a lattice.

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need (2023), arXiv:1706.03762 [cs.CL].

[20] R. B. Laughlin, Anomalous quantum hall effect: An incompressible quantum fluid with fractionally charged excitations, Phys. Rev. Lett. **50**, 1395 (1983).

[21] G. Moore and N. Read, Nonabelions in the fractional quantum hall effect, Nuclear Physics B **360**, 362 (1991).

[22] Y. Teng, D. D. Dai, and L. Fu, Solving the fractional quantum hall problem with self-attention neural network, Phys. Rev. B **111**, 205117 (2025).

[23] Y. Qian, T. Zhao, J. Zhang, T. Xiang, X. Li, and J. Chen, Describing landau level mixing in fractional quantum hall states with deep learning, Physical Review Letters **134**, 10.1103/physrevlett.134.176503 (2025).

[24] M. P. Zaletel and R. S. K. Mong, Exact matrix product states for quantum hall wave functions, Physical Review B **86**, 10.1103/physrevb.86.245305 (2012).

[25] B. Estienne, Z. Papić, N. Regnault, and B. A. Bernevig, Matrix product states for trial quantum hall states, Phys. Rev. B **87**, 161112 (2013).

[26] T. Han, Z. Lu, Y. Yao, L. Shi, J. Yang, J. Seo, S. Ye, Z. Wu, M. Zhou, H. Liu, G. Shi, Z. Hua, K. Watanabe, T. Taniguchi, P. Xiong, L. Fu, and L. Ju, Signatures of chiral superconductivity in rhombohedral graphene (2024), arXiv:2408.15233 [cond-mat.mes-hall].

[27] F. Xu, Z. Sun, J. Li, C. Zheng, C. Xu, J. Gao, T. Jia, K. Watanabe, T. Taniguchi, B. Tong, L. Lu, J. Jia, Z. Shi, S. Jiang, Y. Zhang, Y. Zhang, S. Lei, X. Liu, and T. Li, Signatures of unconventional superconductivity near reentrant and fractional quantum anomalous hall insulators (2025), arXiv:2504.06972 [cond-mat.mes-hall].