

Fragmentation is Efficiently Learnable by Quantum Neural Networks

Mikhail Mints¹ and Eric R. Anschuetz²

¹*Department of Computing + Mathematical Sciences,
California Institute of Technology, Pasadena, CA 91125*

²*Institute for Quantum Information and Matter & Walter Burke Institute for Theoretical Physics,
California Institute of Technology, Pasadena, CA 91125*

Quantum algorithms are known to outperform the best known classical algorithms on many computational problems exhibiting some sort of algebraic structure [1–4]. A significant research direction today is focused on trying to understand whether the unique advantages offered by quantum computers can help in machine learning tasks and whether *quantum neural networks* (QNNs) can be superior to classical methods. Classical neural networks are known to behave like Gaussian processes in the asymptotic limit [5], which provides theoretical guarantees on their efficient trainability in a wide range of applications. On the other hand, QNNs are known to have poorly-behaved loss landscapes. They are often dominated by *barren plateaus* [6–8], which generally prevent gradients from being estimated efficiently, and *poor local minima*, which generally prevent gradient descent from reaching the global optimum [9–11]. While there do exist settings where QNNs are efficient to train via gradient descent—such as when they have many symmetries [12–18]—in these settings there often also exist efficient classical simulation algorithms that render the use of a quantum computer for the task unnecessary outside of potentially an initial data acquisition phase [19–21].

Recent work [22] formulated the Jordan Algebraic Wishart System (JAWS) framework, which is the first full theoretical characterization of QNN loss landscapes that provides the conditions under which they can become efficiently trainable using gradient descent. In this work, we use these tools to formally demonstrate a specific setting where QNNs *can* in fact be efficiently trained to solve a physically motivated problem for which no efficient classical simulation algorithm is known. Our approach is to showcase a setting where the QNN has a high-dimensional symmetry group but, crucially, the end-user does not know what that group is and—to the best of our knowledge—has no efficient strategy for learning it. While we have no formal reduction connecting this problem to standard complexity theoretic conjectures, at a high level, this resembles other problems conjectured to be difficult for classical algorithms, including the hidden subgroup problem [23] and the state hidden subgroup problem [24].

I. HILBERT SPACE FRAGMENTATION

The problem that we study in this work is that of classifying states in systems which exhibit *Hilbert space fragmentation* [25–27], a well-studied barrier to ergodicity in quantum many-body systems conceptually similar to the better-known quantum many-body scarring phenomenon [28–30].

We let L be the problem size—physically known as the *system size*—which we will later take to be large. For each L , one is given a set $\mathcal{S}$ of Hamiltonians acting on a Hilbert space $\mathcal{H}$ with a tensor product structure $\mathcal{H} = \bigotimes_{i=1}^L \mathcal{H}_i$. $\mathcal{S}$ is constructed such that all Hamiltonians $\mathbf{H} \in \mathcal{S}$ take the form $\mathbf{H} = \sum_{i=1}^m c_i \mathbf{h}_i$, where all $c_i \in \mathbb{R}$ and the $\mathbf{h}_i$ are *local operators*, i.e., $\mathbf{h}_i$ acts nontrivially only on an L -independent number of tensor product factors $\mathcal{H}_i$. Furthermore, m only grows at most polynomially quickly with L .

Let $\mathcal{A} = \langle \mathbf{h}_1, \dots, \mathbf{h}_m \rangle$ be the associative algebra generated by these operators under addition and matrix multiplication. Let $\mathcal{C}$ be the *commutant algebra*, consisting of all operators in $\mathcal{L}(\mathcal{H})$ that commute with all elements of $\mathcal{A}$. By the Von Neumann bicommutant theorem [31], we know that $\mathcal{A}$ and $\mathcal{C}$ are each other’s centralizers, and we can decompose the Hilbert space as:

$$\mathcal{H} = \bigoplus_{\lambda=1}^{\Lambda} \mathcal{H}_{\lambda}^{(\mathcal{A})} \otimes \mathcal{H}_{\lambda}^{(\mathcal{C})}, \quad (1)$$

where each $\mathcal{H}_{\lambda}^{(\mathcal{A})}$ is an irreducible representation of $\mathcal{A}$ and each $\mathcal{H}_{\lambda}^{(\mathcal{C})}$ is an irreducible representation of $\mathcal{C}$. Let $N_{\lambda} = \dim(\mathcal{H}_{\lambda}^{(\mathcal{A})})$, $N'_{\lambda} = \dim(\mathcal{H}_{\lambda}^{(\mathcal{C})})$. In this decomposition, each $\mathcal{H}_{\lambda}^{(\mathcal{A})}$ is called a *Krylov subspace* of the system, preserved by the action of Hamiltonians constructed from the generators $\mathbf{h}_1, \dots, \mathbf{h}_m$. Each of these subspaces has N'_{λ} degenerate copies. The representation in $\mathcal{H}$ of any operator $A \in \mathcal{A}$ can be written (in a slight abuse of notation identifying the operator with its representation) in the form:

$$\mathbf{A} = \bigoplus_{\lambda=1}^{\Lambda} \mathbf{A}^{\lambda} \otimes \text{Id}_{N'_{\lambda}}. \quad (2)$$

That is, any such operator acts as the identity on the “multiplicity labels” in $\mathcal{H}_\lambda^{(C)}$, meaning that it acts identically on each of the N'_λ degenerate Krylov subspaces isomorphic to $\mathcal{H}_\lambda^{(A)}$.

The phenomenon of *Hilbert space fragmentation* is said to occur when the total multiplicity, $\sum_\lambda N'_\lambda$, is exponential in the system size L [32]. We are interested in the stronger condition where the dimension of $\mathcal{A}$ is polynomial in the system size. If a system does not have this property, we may be able to restrict to a subspace $\mathcal{H}'$ that does, replacing Λ with some potentially smaller Λ' as a function of L .

While this construction may seem ad hoc and technical, surprisingly the phenomenon of Hilbert space fragmentation occurs in many quantum systems. It underlies the ergodicity breaking of the Temperley–Lieb model [33], the t – J_z model [34], and the Pair-Flip model [32].

One can also define an orthonormal basis of $\mathcal{H}$ respecting the decomposition of $\mathcal{H}$ into Krylov subspaces [32]. In analogy with the terminology for permutation-invariant quantum systems [35], we can construct a *Schur basis* for $\mathcal{H}'$ as a set of states $|\lambda, q_\lambda, p_\lambda\rangle = |\lambda, q_\lambda\rangle \otimes |\lambda, p_\lambda\rangle$ such that the states $|\lambda, q_\lambda\rangle$ form an orthonormal basis for $\mathcal{H}_\lambda^{(A)}$ and the states $|\lambda, p_\lambda\rangle$ form an orthonormal basis for $\mathcal{H}_\lambda^{(C)}$.

II. THE CLASSIFICATION TASK

Given a quantum state that is a basis state promised to be in the Schur basis with respect to a Hilbert space fragmented system, we would like to perform some kind of quantum measurement to determine which state exactly it is, ignoring the degeneracy. In some special settings—such as when the system is permutation invariant, and one is interested in the associated Schur basis of that Krylov decomposition—this can be done efficiently using a quantum algorithm known as the Schur transform [35]. However, generally, we know only the generators $\mathbf{h}_1, \dots, \mathbf{h}_m$. Instead, we hope in effect to *learn* the Schur transform associated with $\mathcal{S}$, assuming we are also given copies of Schur basis states $|\lambda, q_\lambda, p_\lambda\rangle$. Specifically, we want to learn a transformation $\mathbf{U}$ such that

$$\text{Tr}_{\mathcal{H}'} \left[\mathbf{U}^\dagger (|\lambda, q_\lambda, p_\lambda\rangle \langle \lambda, q_\lambda, p_\lambda| \otimes |\mathbf{0}\rangle \langle \mathbf{0}|) \mathbf{U} \right] \approx |\boldsymbol{\lambda}, \mathbf{q}_\lambda\rangle \langle \boldsymbol{\lambda}, \mathbf{q}_\lambda|. \quad (3)$$

Here, we have added an ancillary register of n_a qubits with Hilbert space of dimension $N_a = 2^{n_a}$ and initialized it in the computational basis state $|\mathbf{0}\rangle$. We also use the notation $|\boldsymbol{\lambda}, \mathbf{q}_\lambda\rangle$ to label bit strings encoded in the computational basis labeling λ and q_λ . Note that here we are ambivalent to the multiplicity label p_λ ; there are exponentially many such labels, and we cannot hope to learn them efficiently. Instead, we also require that our network *generalizes* its classification of any $|\lambda, q_\lambda, p_\lambda\rangle$ given training examples of $|\lambda, q_\lambda, p'_\lambda\rangle$ for p_λ not necessarily equal to p'_λ . We call this problem *fragment classification*.

III. THE QNN ANSATZ

The quantum learning algorithm we construct will be a *variational quantum algorithm* [36]. Here, one hopes to learn a unitary $\mathbf{U}$ by tuning parameters of a variational ansatz constructed using exponentials of local operators. The variational parameters of the ansatz are then tuned using a gradient-based optimizer to minimize some loss function.

Suppose that we have a dataset $\mathcal{D} = \{(|x\rangle, \mathbf{O}_x)\}_{x \in [M]}$ of size $M \leq N$, where each $|x\rangle$ is of the form $|\lambda, q_\lambda, p_\lambda\rangle \otimes |\mathbf{0}\rangle$ and the corresponding *objective observable* is

$$\mathbf{O}_x = -\text{Id}_{\mathcal{H}'} \otimes |\boldsymbol{\lambda}, \mathbf{q}_\lambda\rangle \langle \boldsymbol{\lambda}, \mathbf{q}_\lambda|. \quad (4)$$

We assume we are given at most one example x for any given λ, q_λ pair—we will see in Theorem 1 that we do not require an example for each p_λ .

We construct our randomized QNN ansatz parameterized by $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$ as follows:

$$\mathbf{U}(\boldsymbol{\theta}) = e^{i\mathbf{H}t''} \left(\prod_{i=1}^p e^{-i\mathbf{H}t_i} e^{i\mathbf{A}\theta_i} e^{i\mathbf{H}t_i} \right) e^{-i\mathbf{H}t'}, \quad (5)$$

where we select $t_1, \dots, t_p, t', t''$ i.i.d. uniformly from $[0, T]$ for some fixed T , we let $\mathbf{H}$ be a Hamiltonian on the system and the ancillary register:

$$\mathbf{H} = \sum_{i=1}^m c_i \mathbf{h}_i \otimes \left(\sum_{j=1}^{n_a} c'_{i,j,x} \sigma_j^x + c'_{i,j,y} \sigma_j^y + c'_{i,j,z} \sigma_j^z \right), \quad (6)$$

and we let $\mathbf{A}$ be a local operator block-diagonal in the Schur basis, which we assume without loss of generality is positive semidefinite to simplify our analysis—we can just take it to be a large multiple of the identity added to $\mathbf{h}_1 \otimes \text{Id}_{N_a}$.

For the purposes of our analysis, we make the physically realistic assumption that the coefficients in Equation (6) are chosen in such a way that the Hamiltonian obeys the *full Eigenstate Thermalization Hypothesis* (ETH) [37] in each Krylov subspace.

We can then use gradient descent to find optimal values of the parameters θ to minimize the loss function:

$$\ell(\theta) = \frac{1}{M} \sum_{x=1}^M \langle x | U(\theta) O_x U(\theta)^\dagger | x \rangle. \quad (7)$$

The training process is illustrated in Figure 1.

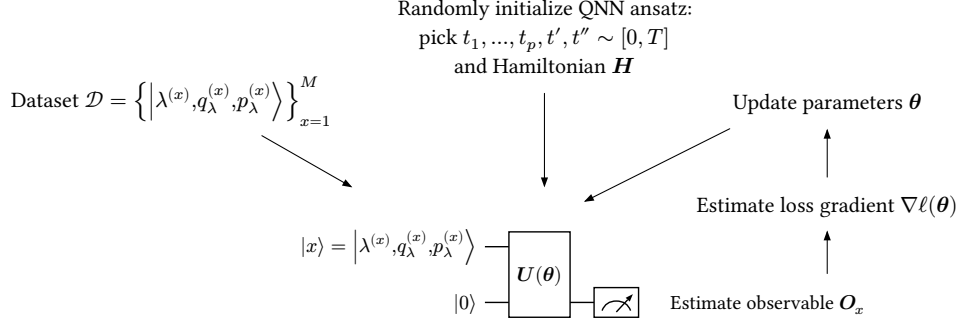


FIG. 1: Diagram of our QNN training process. We are given a dataset of Schur basis states and randomly sample a QNN ansatz architecture and an initial vector of parameters. We then compile the QNN into a quantum circuit and repeatedly run on the input states to estimate the gradient of the loss function, which is then used to adjust the parameters.

IV. MAIN RESULTS

The first result establishes that the algebraic structure of the system allows our QNN construction to generalize to states of the form $|\lambda, q_\lambda, p_\lambda\rangle$ for p_λ not present in the training dataset.

Theorem 1. *If we take the (exponential-size) dataset $\mathcal{D}'$ consisting of every Schur basis state $|\lambda, q_\lambda, p_\lambda\rangle$ for λ, q_λ present in $\mathcal{D}$, the loss $\ell_{\mathcal{D}'}(\theta)$ with respect to that dataset is upper-bounded as $\ell_{\mathcal{D}'}(\theta) \leq \max_{\lambda, x} \ell_x^\lambda(\theta)$.*

Next, we formally establish the standard conditions for QNN trainability [21, 22]. The first condition is the absence of barren plateaus: in the asymptotic limit of large system size, the distribution of the loss landscape over the random initialization of the QNN architecture converges to one where the derivatives only decay polynomially with the system size, thus allowing the gradient to be efficiently estimated on a quantum computer in polynomial time. We formally prove this in the following theorem:

Theorem 2. *We can, up to a vanishing error in Lévy-Prokhorov metric, approximate the distribution of the gradient of the loss function $\nabla \ell$ as some $\widehat{\nabla} \ell$ where*

$$\mathbb{E} \left[\left\| \widehat{\nabla} \ell \right\|^2 \right] = \Omega \left(\frac{1}{\text{poly}(L)} \right). \quad (8)$$

The second condition is the absence of poor local minima. When the number of parameters p is sufficiently large, the loss landscape enters an *overparameterized regime* where all spurious local minima disappear, and the only remaining minima are *degenerate*, existing on some submanifold of the parameter space. We follow the approach of Larocca *et al.* [38] and show the Hessian is not full-rank in order to demonstrate a lack of poor local minima. The value of p needed for this to occur is only polynomially large in the system size, making the QNN in the overparameterized phase efficiently implementable on a quantum computer.

Theorem 3. *For some $p = O(\text{poly}(L))$, the Hessian matrix $\mathbf{Hess}$ at any value of θ (where $\mathbf{Hess}_{i,j} = \partial_i \partial_j \ell(\theta)$) does not have full rank.*

-
- [1] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM review* **41**, 303 (1999).
 - [2] C. Gyurik, A. Schmidhuber, R. King, V. Dunjko, and R. Hayakawa, Quantum computing and persistence in topological data analysis, *arXiv preprint arXiv:2410.21258* (2024).
 - [3] D. W. Berry, Y. Su, C. Gyurik, R. King, J. Basso, A. D. T. Barba, A. Rajput, N. Wiebe, V. Dunjko, and R. Babbush, Analyzing prospects for quantum advantage in topological data analysis, *PRX Quantum* **5**, 010319 (2024).
 - [4] A. W. Harrow, A. Hassidim, and S. Lloyd, Quantum algorithm for linear systems of equations, *Physical review letters* **103**, 150502 (2009).
 - [5] R. M. Neal, Priors for infinite networks, in *Bayesian Learning for Neural Networks* (Springer New York, New York, NY, 1996) pp. 29–53.
 - [6] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, *Nat. Commun.* **9**, 4812 (2018).
 - [7] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, *Nat. Commun.* **12**, 1791 (2021).
 - [8] M. Ragone, B. N. Bakalov, F. Sauvage, A. F. Kemper, C. Ortiz Marrero, M. Larocca, and M. Cerezo, A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits, *Nat. Commun.* **15**, 7172 (2024).
 - [9] E. R. Anschuetz, Critical points in quantum generative models, in *International Conference on Learning Representations* (2022).
 - [10] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, *Nat. Commun.* **13**, 7760 (2022).
 - [11] X. You, S. Chakrabarti, and X. Wu, A convergence theory for over-parameterized variational quantum eigensolvers (2022), *arXiv:2205.12481 [quant-ph]*.
 - [12] J. J. Meyer, M. Mularski, E. Gil-Fuster, A. A. Mele, F. Arzani, A. Wilms, and J. Eisert, Exploiting symmetry in variational quantum machine learning, *PRX Quantum* **4**, 010328 (2023).
 - [13] L. Schatzki, M. Larocca, Q. T. Nguyen, F. Sauvage, and M. Cerezo, Theoretical guarantees for permutation-equivariant quantum neural networks, *npj Quantum Inf.* **10**, 12 (2024).
 - [14] Q. T. Nguyen, L. Schatzki, P. Braccia, M. Ragone, P. J. Coles, F. Sauvage, M. Larocca, and M. Cerezo, Theory for equivariant quantum neural networks, *PRX Quantum* **5**, 020328 (2024).
 - [15] J. Gibbs, Z. Holmes, M. C. Caro, N. Ezzell, H.-Y. Huang, L. Cincio, A. T. Sornborger, and P. J. Coles, Dynamical simulation via quantum machine learning with provable generalization, *Phys. Rev. Res.* **6**, 013241 (2024).
 - [16] E. R. Anschuetz, H.-Y. Hu, J.-L. Huang, and X. Gao, Interpretable quantum advantage in neural sequence learning, *PRX Quantum* **4**, 020338 (2023).
 - [17] M. T. West, J. Heredge, M. Sevir, and M. Usman, Provably trainable rotationally equivariant quantum machine learning, *PRX Quantum* **5**, 030320 (2024).
 - [18] E. R. Anschuetz and X. Gao, Arbitrary polynomial separations in trainable quantum machine learning (2024), *arXiv:2402.08606 [quant-ph]*.
 - [19] E. R. Anschuetz, A. Bauer, B. T. Kiani, and S. Lloyd, Efficient classical algorithms for simulating symmetric quantum systems, *Quantum* **7**, 1189 (2023).
 - [20] M. L. Goh, M. Larocca, L. Cincio, M. Cerezo, and F. Sauvage, Lie-algebraic classical simulations for quantum computing, *Physical Review Research* **7**, 10.1103/3y65-f5w6 (2025).
 - [21] M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, Does provable absence of barren plateaus imply classical simulability?, *Nat. Commun.* **16**, 7907 (2025).
 - [22] E. R. Anschuetz, A unified theory of quantum neural network loss landscapes, in *International Conference on Learning Representations* (OpenReview, 2025).
 - [23] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem (1995), *arXiv:quant-ph/9511026 [quant-ph]*.
 - [24] A. Bouland, T. Giurgică-Tiron, and J. Wright, The state hidden subgroup problem and an efficient algorithm for locating unentanglement, in *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, STOC '25 (Association for Computing Machinery, New York, NY, USA, 2025) p. 463–470.
 - [25] S. Pai, M. Pretko, and R. M. Nandkishore, Localization in fractonic random circuits, *Phys. Rev. X* **9**, 021003 (2019).
 - [26] P. Sala, T. Rakovszky, R. Verresen, M. Knap, and F. Pollmann, Ergodicity breaking arising from Hilbert space fragmentation in dipole-conserving hamiltonians, *Phys. Rev. X* **10**, 011047 (2020).
 - [27] V. Khemani, M. Hermele, and R. Nandkishore, Localization from Hilbert space shattering: From theory to physical realizations, *Phys. Rev. B* **101**, 174204 (2020).
 - [28] H. Bernien, S. Schwartz, A. Keesling, H. Levine, A. Omran, H. Pichler, S. Choi, A. S. Zibrov, M. Endres, M. Greiner, *et al.*, Probing many-body dynamics on a 51-atom quantum simulator, *Nature* **551**, 579 (2017).
 - [29] C. J. Turner, A. A. Michailidis, D. A. Abanin, M. Serbyn, and Z. Papić, Weak ergodicity breaking from quantum many-body scars, *Nat. Phys.* **14**, 745 (2018).
 - [30] S. Moudgalya and O. I. Motrunich, Exhaustive characterization of quantum many-body scars using commutant algebras, *Physical Review X* **14**, 10.1103/physrevx.14.041069 (2024).
 - [31] N. P. Landsman, Lecture notes on C^* -algebras, Hilbert C^* -modules, and quantum mechanics, *arXiv e-prints*, math-ph/9807030 (1998), *arXiv:math-ph/9807030 [math-ph]*.

- [32] S. Moudgalya and O. I. Motrunich, Hilbert space fragmentation and commutant algebras, *Phys. Rev. X* **12**, 011050 (2022).
- [33] M. T. Batchelor and M. N. Barber, Spin-s quantum chains and temperley-lieb algebras, *Journal of Physics A: Mathematical and General* **23**, L15 (1990).
- [34] S. Zhang, M. Karbach, G. Müller, and J. Stolze, Charge and spin dynamics in the one-dimensional t-j_z and t-j models, *Phys. Rev. B* **55**, 6491 (1997).
- [35] D. Bacon, I. L. Chuang, and A. W. Harrow, Efficient quantum circuits for schur and clebsch-gordan transforms, *Phys. Rev. Lett.* **97**, 170502 (2006).
- [36] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, *Nat. Commun.* **5**, 4213 (2014).
- [37] L. Foini and J. Kurchan, Eigenstate thermalization hypothesis and out of time order correlators, *Phys. Rev. E* **99**, 042139 (2019).
- [38] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, Theory of overparametrization in quantum neural networks, *Nature Computational Science* **3**, 542 (2023).