

Scalable Message-Passing Quantum Graph Neural Networks in the Weisfeiler–Leman Hierarchy

Snehal Raj^{1,2}, Brian Coyle^{3,4}, Léo Monbroussou^{3,5}, André J. Ferreira-Martins^{6,1}, Renato M. S. Farias⁶, Elham Kashefi^{1,3}

¹LIP6, CNRS, Sorbonne Université, Paris, France ²QC Ware, Palo Alto, USA and Paris, France

³School of Informatics, University of Edinburgh, Edinburgh, UK ⁴Fujitsu Research of Europe Ltd., United Kingdom

⁵Terra Quantum GmbH, Germany ⁶Quantum Research Center, Technology Innovation Institute, Abu Dhabi, UAE

1 Introduction

Graph neural networks (GNNs) are a central tool in machine learning, driving progress in various fields such as molecular property prediction and drug discovery [1], protein structure modelling (AlphaFold), and combinatorial optimisation on graphs [2, 3]. Their success rests on *message passing*, the rule by which each node updates from its neighbours [1, 4]. Quantum versions of GNNs have been proposed [5, 6] and, like classical ones, made permutation equivariant [7, 8, 9]. Most, however, do not carry out message passing: they derive the circuit topology from the graph, placing gates along its edges [5], while the aggregation step that defines a GNN is left to a classical readout. They also inherit the trainability challenges of variational circuits, whose gradients can vanish as the circuit grows [10, 11, 12]. It is therefore important to show that a quantum graph model can be trained and scaled to large graphs.

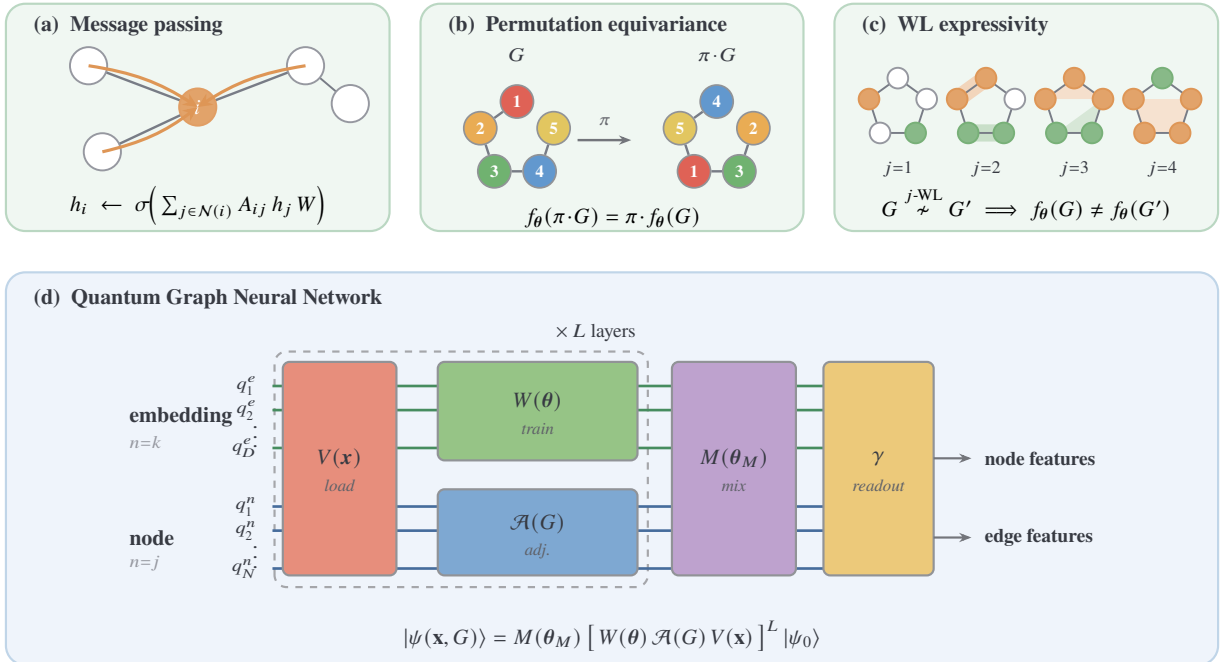


Figure 1: **Graph-learning properties and the quantum graph neural network.** (a) **Message passing:** each node updates from its neighbours, aggregating its L -hop neighbourhood after L rounds. (b) **Permutation equivariance:** relabelling the inputs by π relabels the output by π . (c) **WL expressivity:** the node-register particle number j sets the set-based j -WL level reached, exceeding the 1-WL ceiling for $j \geq 3$. (d) **Architecture:** a loader $V(\mathbf{x})$ encodes the graph; adjacency $\mathcal{A}(G)$ and trainable evolution $W(\theta)$ alternate over L layers; a mixer $M(\theta_M)$ couples the N -qubit node (j) and D -qubit embedding (k) registers; readout γ returns node and edge features.

We summarise results from our companion paper [13] that adapt a GNN’s useful properties to a quantum framework: our quantum GNN performs message passing, is permutation-equivariant, and sits at a chosen level of the Weisfeiler–Leman hierarchy, which measures how finely a model distinguishes graphs. Like classical GNNs, it can be pre-trained on small graphs and read out at a cost that stays low as the graph grows. A sample-complexity analysis, together with an initial gradient variance that does not vanish with subspace size, indicates favourable and scalable training, especially with such pre-training. We validate the framework in simulations of up to 56 qubits across three datasets: synthetic graphs message passing cannot separate, molecular property prediction, and the travelling salesperson problem, paving the way for scalable near-term quantum algorithms with theoretical guarantees.

2 Graph Neural Network framework

A graph neural network processes a graph G on N nodes: an encoder $V(G)$ collects the D_F -dimensional node features into a matrix $X \in \mathbb{R}^{N \times D_F}$, and the connectivity is carried by an adjacency operation $\mathcal{A}(G) \in \mathbb{R}^{N \times N}$. A useful GNN must respect several theoretical properties [1, 14, 15].

Definition 1 (Message passing). *One layer maps $H^{(\ell)} \mapsto H^{(\ell+1)} = \sigma(\mathcal{A}(G) H^{(\ell)} W(\boldsymbol{\theta}))$, with σ a suitable nonlinearity; after L layers, the embedding of node i encodes aggregated features from its L -hop neighbourhood.*

Definition 2 (Exact permutation equivariance). *Let f be the function computed by a GNN. It is permutation equivariant if $f(\pi \cdot G, \pi \cdot X) = \pi \cdot f(G, X)$ for every relabelling $\pi \in S_N$ of the nodes.*

Because graph nodes carry no intrinsic order, equivariance is essential for consistent predictions under relabelling [15, 14]. What limits a model is how finely it separates structurally distinct graphs, measured by the Weisfeiler–Leman (WL) hierarchy.

Definition 3 (Weisfeiler–Leman test). *The j -Weisfeiler–Leman (j -WL) test assigns each j -subset S of the nodes a colour $c^t(S)$, refined over steps $t = 0, 1, 2, \dots$ from the initial colour $c^0(S)$ given by the isomorphism type of the subgraph induced by S . Each step updates*

$$c^{t+1}(S) = \text{HASH}\left(c^t(S), \{\!\!\{ c^t(S') : |S \Delta S'| = 2 \}\!\!\}\right), \quad (1)$$

where HASH is injective, $\{\!\!\{ \cdot \}\!\!\}$ a multiset, and $S \Delta S'$ the symmetric difference; once the colours stabilise, two graphs differ when their subset-colour multisets do [16].

Ordinary message passing reaches only the first level, 1-WL [4, 16], which cannot separate graphs differing in higher-order structure, including molecules with distinct properties [17]. An architecture has j -WL expressivity when it agrees with the j -WL test on every pair of graphs.

We present how to achieve these theoretical guarantees by using subspace preserving circuits [18, 19, 20, 21, 22, 23], and in particular a specific structure for the data (see Figure 1). The N -qubit *node register* indexes the graph’s nodes at Hamming weight j , spanning the subspace $\mathcal{H}_N^j$ of dimension $\binom{N}{j}$: its basis states are individual nodes at $j=1$ and j -subsets of nodes at larger j . The D -qubit *embedding register* holds the corresponding features at Hamming weight k , in $\mathcal{H}_D^k$ with $D_F = \binom{D}{k}$. A hierarchical *data loader* $V(\mathbf{x})$ prepares

$$|\mathbf{x}\rangle = \sum_T \sum_{f=1}^{D_F} x_f^{(T)} |e_T^{\text{node}}\rangle \otimes |e_f^{\text{embedding}}\rangle. \quad (2)$$

The *equivariant adjacency* $\mathcal{A}(G)$ applies Givens rotations between node-qubit pairs connected by an edge, with angles set by the edge weights in a canonical ordering. The *trainable evolution* $W(\boldsymbol{\theta})$ acts on the embedding register only, and re-uploading V each layer makes the map nonlinear [24]. A final *joint-register mixing* $M(\boldsymbol{\theta}_M)$ couples the subspaces, its angles indexed by a canonical S_N -invariant ordering induced by G .

3 Theoretical guarantees

Relabelling the nodes permutes the output features by the same permutation, exactly and at every parameter value: the trainable evolution acts only on the embedding register, and the graph-dependent operations are applied in a canonical ordering fixed by the edge weights. Hence, a relabelling reorders the same gates without changing their action. Earlier equivariant models instead tie parameters across S_N [15, 14, 7], which constrains the function class.

Theorem 1 (Exact S_N -equivariance). *Let $\pi \in S_N$ act on the node register through the permutation operator P_π . For every input $\mathbf{x} = (\mathbf{X}, A)$ consisting of node features $\mathbf{X} \in \mathbb{R}^{N \times d}$ and an adjacency matrix $A \in \mathbb{R}^{N \times N}$, every parameter assignment $\boldsymbol{\theta}$, and every $\pi \in S_N$,*

$$f_{\boldsymbol{\theta}}(\pi \cdot \mathbf{X}, \pi \cdot A) = \pi \cdot f_{\boldsymbol{\theta}}(\mathbf{X}, A). \quad (3)$$

The model performs message passing inside the circuit (Definition 1) with the following operations: the adjacency routes amplitude along the edges, the trainable evolution updates the features, and re-uploading the loader implements the nonlinearity. Each iteration then realises one round of set-based j -WL refinement [16], coupling the j -subsets that differ by a single swap. A message-passing network distinguishes two graphs only when the 1-WL test does, a ceiling that bounds every standard architecture. The node-register particle number j sets how much graph structure the model resolves, and from $j=3$ it separates graphs that 1-WL, and hence every message-passing network, cannot; to our knowledge, it is the first quantum graph model proven to do so.

Theorem 2 (j -WL expressivity, informal). *For $2 \leq j \leq 4$, the L -iteration QGNN has set-based j -WL expressivity at generic parameters. Its basis states are the j -subsets of nodes that j -WL colours (Eq. (2)), one circuit iteration performs one refinement round, and with each subset started from a colour that fixes its induced subgraph up to isomorphism, after L iterations the output separates exactly the graphs that L rounds of j -WL separate,*

$$f_{\boldsymbol{\theta}}(G) = f_{\boldsymbol{\theta}}(G') \iff G \sim_{j\text{-WL}} G'.$$

For $j \geq 3$, this passes the 1-WL ceiling that bounds every standard message-passing network.

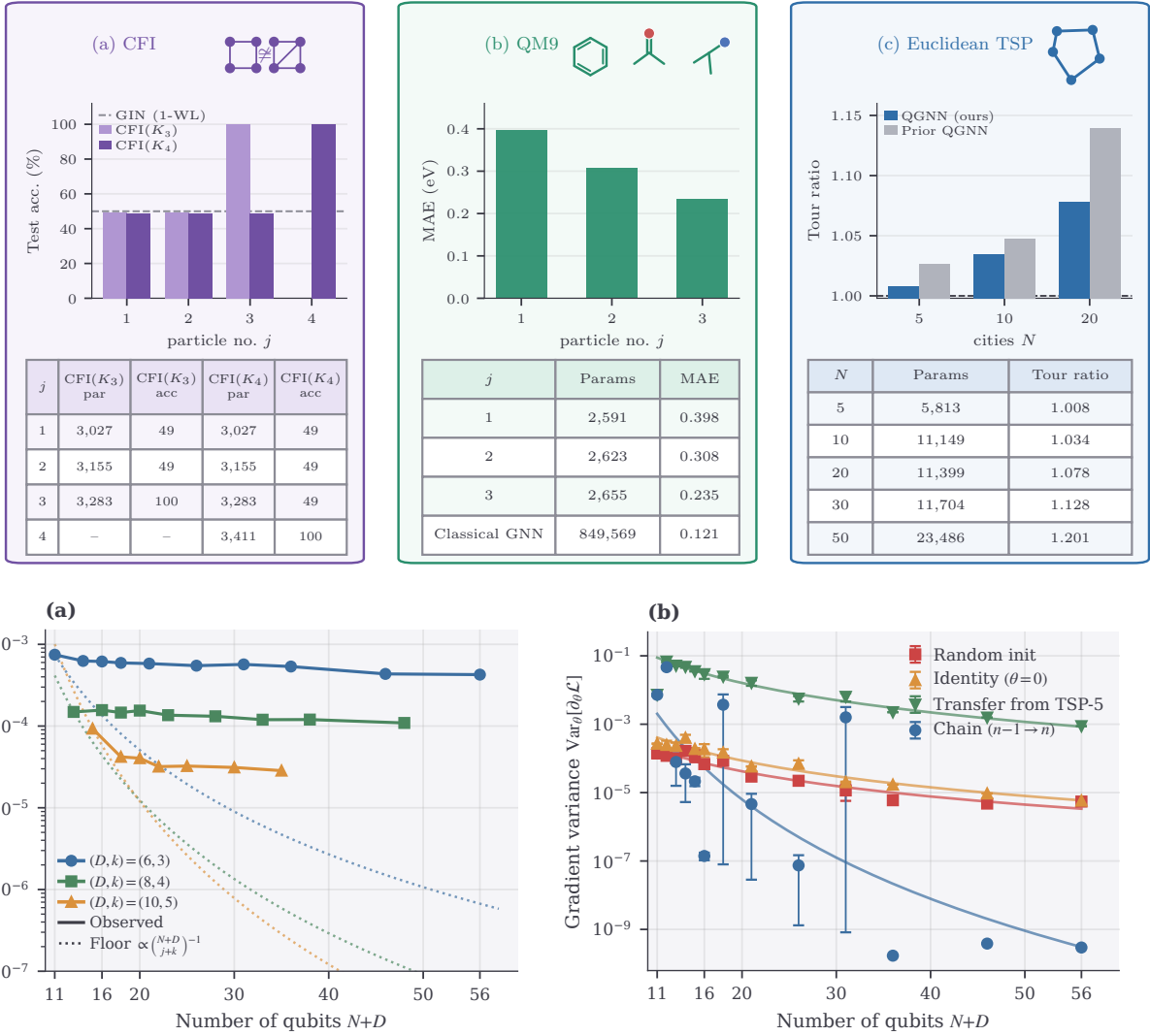


Figure 2: **Numerical validation and trainability.** Top, (a)–(c): CFI discrimination (accuracy jumps to 100% at the predicted j ; a 1-WL GIN stays at chance, a 3-WL PPGN separates both), QM9 HOMO–LUMO error falling with j , and Euclidean TSP tour ratio versus N [7]. Bottom: per-parameter gradient variance at $(D, k) \in \{(6, 3), (8, 4), (10, 5)\}$ versus qubit count up to 56, above the $\binom{N+D}{j+k}^{-1}$ floor (dotted), a log–log fit confirming *polynomial*, not barren-plateau, decay [13].

4 Scalability

Barren plateaus, where gradients vanish as the circuit grows [11, 10], are the main obstacle to scaling variational quantum algorithms; they are mitigated when the dynamics stay in a smaller, structured subspace [12, 25, 21, 26], as the two-register QGNN does.

Proposition 1 (GNN pre-training). *A GNN can be pre-trained on small graph instances: the trainable evolution $W(\theta)$ acts only on the embedding register, whose size does not typically grow with the graph. At the same time, the adjacency layer requires no training.*

As with classical GNNs trained on small graphs and run on larger ones [2, 3], we use a **transferable initialisation**: parameters fitted at one graph size initialise the next [27, 28, 29]. Heuristic but effective, it keeps the gradient signal one to two orders of magnitude above a random start across instances up to 56 qubits (Fig. 2, bottom).

Proposition 2 (Polynomial-cost readout). *The node and edge features are the entries of the embedding-register one-particle density matrix $\gamma_{pq} = \langle \psi | a_p^\dagger a_q | \psi \rangle$ ($p, q \in [D]$), estimable to accuracy ε in $O(D^3/\varepsilon^2)$ shots by a deterministic Hartree–Fock readout [30] or a matchgate shadow [31], at a cost set by the D -qubit feature register, not the number of nodes.*

5 Numerical validation

Figure 2 confirms the guarantees up to 56 qubits. On Cai–Fürer–Immerman graphs, which 1-WL cannot separate [32], accuracy jumps from chance to 100% at the j predicted by Theorem 2, matching a 3-WL network [33]. The ascent also holds on QM9 [34] (HOMO–LUMO error $0.398 \rightarrow 0.235$ eV), and it solves Euclidean TSP end-to-end at $j=1$ up to $N=50$, improving upon Skolik et al. [7].

References

- [1] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1263–1272. PMLR, 2017.
- [2] Chaitanya K Joshi, Thomas Laurent, and Xavier Bresson. An efficient graph convolutional network technique for the travelling salesman problem. *arXiv preprint arXiv:1906.01227*, 2019.
- [3] Wouter Kool, Herke van Hoof, and Max Welling. Attention, learn to solve routing problems!, 2019.
- [4] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks?, 2019.
- [5] Guillaume Verdon, Trevor McCourt, Enxhell Luzhnica, Vikash Singh, Stefan Leichenauer, and Jack Hidary. Quantum graph neural networks. *arXiv preprint arXiv:1909.12264*, 2019.
- [6] Peter Mernyei, Konstantinos Meichanetzidis, and Ismail Ilkan Ceylan. Equivariant quantum graph circuits. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 15401–15420. PMLR, 17–23 Jul 2022.
- [7] Andrea Skolik, Michele Cattelan, Sheir Yarkoni, Thomas Bäck, and Vedran Dunjko. Equivariant quantum circuits for learning on weighted graphs. *npj Quantum Information*, 9(1):47, 2023.
- [8] Louis Schatzki, Martín Larocca, Quynh T. Nguyen, Frédéric Sauvage, and M. Cerezo. Theoretical guarantees for permutation-equivariant quantum neural networks. *npj Quantum Information*, 10:12, 2024.
- [9] Quynh T Nguyen, Louis Schatzki, Paolo Braccia, Michael Ragone, Patrick J Coles, Frederic Sauvage, Martin Larocca, and Marco Cerezo. Theory for equivariant quantum neural networks. *PRX Quantum*, 5(2):020328, 2024.
- [10] Martín Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J. Coles, Lukasz Cincio, Jarrod R. McClean, Zoë Holmes, and M. Cerezo. Barren plateaus in variational quantum computing. *Nature Reviews Physics*, 7(4):174–189, April 2025.
- [11] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812, 2018.
- [12] Michael Ragone, Bojko N Bakalov, Frédéric Sauvage, Alexander F Kemper, Carlos Ortiz Marrero, Martín Larocca, and M Cerezo. A Lie algebraic theory of barren plateaus for deep parameterized quantum circuits. *Nature Communications*, 15:7172, 2024.
- [13] Snehal Raj, Brian Coyle, Léo Monbroussou, André J Ferreira-Martins, Renato Farias, and Elham Kashefi. Scalable message-passing quantum graph neural networks in the weisfeiler-leman hierarchy. *arXiv preprint arXiv:2606.26873*, 2026.
- [14] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.
- [15] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. In *International Conference on Learning Representations*, 2019.
- [16] Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and Leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4602–4609, 2019.
- [17] Zhengdao Chen, Lei Chen, Soledad Villar, and Joan Bruna. Can graph neural networks count substructures? *Advances in neural information processing systems*, 33:10383–10395, 2020.
- [18] Iordanis Kerenidis, Jonas Landman, and Anupam Prakash. Quantum algorithms for deep convolutional neural networks. In *International Conference on Learning Representations*, 2020.
- [19] Jonas Landman, Natansh Mathur, Yun Yvonna Li, Martin Strahm, Skander Kazdaghli, Anupam Prakash, and Iordanis Kerenidis. Quantum methods for neural networks and application to medical image classification. *Quantum*, 6:881, dec 2022.

- [20] El Amine Cherrat, Snehal Raj, Iordanis Kerenidis, Abhishek Shekhar, Ben Wood, Jon Dee, Shouvanik Chakrabarti, Richard Chen, Dylan Herman, Shaohan Hu, Pierre Minssen, Ruslan Shaydulin, Yue Sun, Romina Yalovetzky, and Marco Pistoia. Quantum deep hedging. *Quantum*, 7:1191, 2023.
- [21] Léo Monbroussou, Elliott Z. Mamon, Jonas Landman, Alex B. Grilo, Romain Kukla, and Elham Kashefi. Trainability and expressivity of Hamming-weight preserving quantum circuits for machine learning. *Quantum*, 9:1745, May 2025.
- [22] Léo Monbroussou, Jonas Landman, Letao Wang, Alex B Grilo, and Elham Kashefi. Subspace preserving quantum convolutional neural network architectures. *Quantum Science and Technology*, 10(2):025050, March 2025.
- [23] Snehal Raj and Brian Coyle. QuIC: Quantum-inspired compound adapters for parameter efficient fine-tuning, 2025.
- [24] Adrián Pérez-Salinas, Alba Cervera-Liarta, Elies Gil-Fuster, and José I. Latorre. Data re-uploading for a universal quantum classifier. *Quantum*, 4:226, 2020.
- [25] Enrico Fontana, Dylan Herman, Shouvanik Chakrabarti, Niraj Kumar, Romina Yalovetzky, Jamie Heredge, Shree Hari Sureshababu, and Marco Pistoia. Characterizing barren plateaus in quantum ansätze with the adjoint representation. *Nature Communications*, 15(1):7171, Aug 2024.
- [26] M Cerezo, Martin Larocca, Diego García-Martín, N L Diaz, Paolo Braccia, Enrico Fontana, Manuel S Rudolph, Pablo Bermejo, Aroosa Ijaz, Supanut Thanasilp, Eric R Anschuetz, and Zoë Holmes. Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing. *Nature Communications*, 16:7907, 2025.
- [27] Erik Recio-Armengol, Shah Nawaz Ahmed, and Joseph Bowles. Train on classical, deploy on quantum: scaling generative quantum machine learning to a thousand qubits. *arXiv preprint arXiv:2503.02934*, 2025.
- [28] Manuel S. Rudolph, Jacob Miller, Danial Motlagh, Jing Chen, Atithi Acharya, and Alejandro Perdomo-Ortiz. Synergistic pretraining of parametrized quantum circuits via tensor networks. *Nature Communications*, 14:8367, 2023.
- [29] Bence Bako, Zoltan Kolarovszki, and Zoltan Zimboras. Fermionic born machines: Classical training of quantum generative models based on fermion sampling. *arXiv preprint arXiv:2511.13844*, 2025.
- [30] Frank Arute, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B Buckley, David A Buell, et al. Hartree-Fock on a superconducting qubit quantum computer. *Science*, 369:1084–1089, 2020. arXiv:2004.04174.
- [31] Kianna Wan, William J. Huggins, Joonho Lee, and Ryan Babbush. Matchgate shadows for fermionic quantum simulation. *Communications in Mathematical Physics*, 404:629–700, 2023.
- [32] Jin-Yi Cai, Martin Fürer, and Neil Immerman. An optimal lower bound on the number of variables for graph identification. *Combinatorica*, 12(4):389–410, 1992.
- [33] Haggai Maron, Heli Ben-Hamu, Hadar Serviansky, and Yaron Lipman. Provably powerful graph networks. In *Advances in Neural Information Processing Systems*, 2019.
- [34] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, 2014.