No-Free-Lunch Theories for Tensor-Network Machine Learning Models

Jing-Chuan Wu[a], **Qi Ye**[b], **Dong-Ling Deng**[b], **Li-Wei Yu**[a]

[a] *Theoretical Physics Division, Chern Institute of Mathematics and LPMC, Nankai University, Tianjin 300071, China* wujingchuan@mail.nankai.edu.cn, yulw@nankai.edu.cn

[b] *Center for Quantum Information, IIIS, Tsinghua University, Beijing 100084, China* yeq22@mails.tsinghua.edu.cn, dldeng@tsinghua.edu.cn

1. Introduction

Tensor networks (TNs) have emerged as a powerful tool for studying quantum many-body systems, demonstrating remarkable versatility across various domains of quantum physics [1, 2, 3, 4, 5, 6]. This success has catalyzed a growing interest in harnessing TNs for machine learning applications [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60], where they have shown promise in diverse areas such as dimensionality reduction [9, 10], model compression [29, 54], natural language processing [61, 62, 59], generative models [17, 49]. Despite their promising performance, a comprehensive understanding of the underlying assumptions and limitations of these models is still lacking. Here we focus on the rigorous formulation of their no-free-lunch (NFL) theorem—essential yet notoriously challenging to formalize for specific TN machine learning models. In particular, we rigorously analyze the generalization risks of learning target output functions from input data encoded in TN states. We first prove a NFL theorem for machine learning models based on matrix product states (MPSs), i.e., the one-dimensional TN states. Furthermore, we circumvent the challenging issue of calculating the partition function for two-dimensional Ising model, and prove the no-free-lunch theorem for the case of two-dimensional projected entangled-pair state, by introducing the combinatorial method associated to the "puzzle of polyominoes". This abstract is based on our recent work [63].

2. Quantum NFL

The NFL theorem is one of the most fundamental theorems in the classical machine learning theory [64, 65, 66]. It states that, averaged over all possible problems, every algorithm performs equally well when applied to problems they were not specifically designed for. Inspired by the critical role of NFL theorem in classical machine learning, significant progress has been made in developing NFL theorem for quantum learning models [67, 68, 69, 70, 71, 72, 73]. The quantum NFL theorem establishes straightforward connections between quantum features and the capabilities of quantum learning models. For instance, in practical quantum learning setups with a finite number of measurements, entangled data exhibits a dual effect on prediction errors. With sufficient measurements, highly entangled data can reduce prediction errors. This is consistent with the ideal case of infinite measurements [69]. Conversely, with few measurements, highly entangled data can amplify predicting errors [72]. These results highlight how quantum features contribute to the advantages in quantum machine learning models.

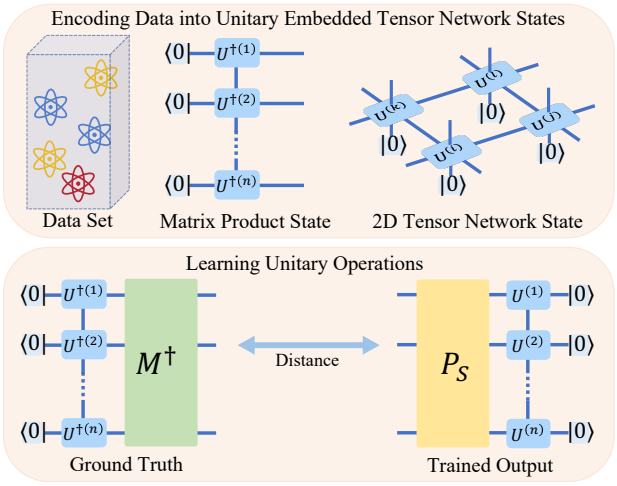

Fig. 1: Upper panel: The encoding strategy. Data samples with labels are encoded into the local tensors $U^{(i)}$ of the unitary embedded tensor network states. Lower panel: The learning strategy. Given the training set $\mathcal{S}$ of samples with the labeled outputs (left), the goal is to minimize the average distance between the learned output states and the ground truth states (acting the unitary $M^\dagger$ on encoding states) over all training samples. Unitary circuit P_S stores the variational parameters.

3. NFL for Tensor network machine learning

We consider a task of learning the unknown unitary operation M based on the input of TN states. Without loss of generality, we take the 1D matrix product state (MPS) for demonstration. The MPS under periodic boundary condition has the form $|\psi\rangle = \sum_{i_1,\ldots,i_n} \mathrm{tr}[A^{(1)}_{i_1} A^{(2)}_{i_2} \cdots A^{(n)}_{i_n}]|i_1, i_2, \ldots, i_n\rangle$, where $A^{(k)}_{i_k}$ denotes the $D \times D$ tensor with D representing the bond dimension, $|i_k\rangle$ denotes the state of k-th physical site with physical dimension d. We define the unitary embedded MPS by converting each $D \times D \times d$ tensor $A^{(k)}$ to a $Dd \times Dd$ unitary $U^{(k)}$

[74, 75, 76, 77, 78, 79], as depicted in Fig. 1. We define the labeled training set $\mathcal{S} = \{(|\psi_j\rangle, |\phi_j\rangle)|j = 1, 2, ...t\}$, where the site size $|\mathcal{S}| = t$, the MPS $|\psi_j\rangle$ belongs to the feature Hilbert space, and the state $|\phi_j\rangle = M|\psi_j\rangle$ belongs to the label Hilbert space. We learn the target unitary M by minimizing the loss function $\mathcal{L} = \sum_{j=1}^{t} |(M|\psi_j\rangle - P_{\mathcal{S}}|\psi_j\rangle)/\langle\psi_j|\psi_j\rangle|^2$, where $P_{\mathcal{S}}$ denotes the variational quantum circuit with sufficient expressivity. If the model is properly trained, then one has $P_{\mathcal{S}}|\psi_j\rangle = M|\psi_j\rangle$, $\forall|\psi_j\rangle \in \mathcal{S}$ up to an overall phase. To quantify the predicting accuracy of our TN model on learning the target M, we define the predicting risk function by the following trace-norm formula

$$R_M(P_{\mathcal{S}}) = \int dx \left\| \left(M|x\rangle\langle x|M^\dagger - P_{\mathcal{S}}|x\rangle\langle x|P_{\mathcal{S}}^\dagger \right) / \langle x|x\rangle \right\|_1^2, \quad (1)$$

where $\|A\|_1 = \frac{1}{2}\mathrm{tr}[\sqrt{A^\dagger A}]$ denotes the trace norm of A, $|x\rangle$ represents the unitary embedded MPS, and the integral is over the Haar measure of all local unitary tensors $\{U^{(i)}|i = 1, 2, \cdots n\}$ in Fig. 1. $R_M(P_{\mathcal{S}})$ represents the prediction error of the trained model. For proper learning without training errors, $R_M(P_{\mathcal{S}})$ is equivalent to the generalization error [80]. We note that the norm $\langle x|x\rangle$ is exponentially concentrated around one [79]. With the above risk function, one can then study the NFL theorem for both 1D and 2D TN-based models.

Theorem 1 (1D MPS). *Define the risk function $R_M(P_{\mathcal{S}})$ in Eq. (1) for learning a target n-qubit unitary M based on the input of MPSs, where $P_{\mathcal{S}}$ represents the hypothesis unitary learned from the training set $\mathcal{S}$. Given a linear independent training set with size $t_k = d^n - d^{n-k}$, the integer $k \in [1, n-1]$, d is the physical dimension of MPS, and n denotes the qubit number of the system. The average risk is lower bounded by*

$$\mathbb{E}_{M,\mathcal{S}}\left[R_M(P_{\mathcal{S}})\right] \geq 1 - (1 - \frac{2}{d^k})(1 + (dAB)^n) - (\frac{1}{d^n} + \frac{1}{d^k})(A^k + B^k)(1 + (dAB)^{n-k}), \quad (2)$$

where $A = \frac{D+1}{Dd+1}$, $B = \frac{D-1}{Dd-1}$, and D is the bond dimension of MPS.

Theorem 2 (2D PEPS). *Define the risk function $R_M(P_{\mathcal{S}})$ in Eq. (1) for learning a target L^2-qubit unitary M based on the input of PEPS, where $P_{\mathcal{S}}$ represents the hypothesis unitary learned from the training set $\mathcal{S}$. Given a linear independent training set with size $t_k = d^{L^2} - d^{L^2-k}$, the integer $k \in [1, L^2 - 1]$, d represents the physical dimension and virtual dimension of PEPS, and L^2 denotes the qubit number of the system. The average risk is lower bounded by*

$$\mathbb{E}_{M,\mathcal{S}}\left[R_M(P_{\mathcal{S}})\right] \geq 1 - (1 + c(0.7)^L)\left[1 - \frac{2}{d^k} + (1 + \frac{1}{d^{L^2-k}}) \cdot \left(\frac{2D^4d - 2}{D^4d^3 - d}\right)^k \left(\frac{1+D}{2D}\right)^{2k} (1 + G(1/d, 1/D^2))^{2l}\right], \quad (3)$$

where $l = \lceil\sqrt{k}\rceil$, $G(q,p) = \frac{p}{2}\left(\sqrt{\frac{(1+q)(1+q-qp)}{1-q(2+p)+q^2(1-p)}} - 1\right)$, D is the bond dimension of the PEPS and c is a constant.

The above two Theorems establish the analytical lower bounds for the average risk of the TN-based machine learning model, and thus quantify the capability of the model in learning an arbitrary target unitary with an arbitrary training set. Rigorous proofs show that the average risk is lower bounded by zero for the full training set, whereas the average risk is lower bounded by one for the empty training set. These results formalize the NFL theorems for MPS and PEPS-based machine learning models.

4. Conclusion

Our results provide a fundamental understanding on the generalization limits of TN-based models, extract how the performance of these specifically structured models would be limited by the NFL theorem, and analytically unveil that the lower bound of the average risk depends on both the bound and physical dimensions of TNs. Our findings would inspire further research on the learning capabilities of TN-based models with quantum computer, where TNs are employed as efficient representations of quantum circuit models. One potential direction is to incorporate the issues of practical quantum computing hardware, such as the noise and finite measurement times [72], into the analytical study of generalization ability for TN-based learning models. From the perspective of experiments, future research could focus on experimentally validating NFL bounds in practical quantum computing environments. As quantum hardware continues to advance, testing these theoretical predictions on real quantum systems will be crucial to understanding how NFL constraints manifest in noisy, resource-limited settings. Such experiments would also help refine our theoretical models, potentially revealing new strategies for optimizing TN-based machine learning models for practical applications.

In summary, we have rigorously formulated the NFL theorems in the TN-based machine learning models. Particularly, we consider the supervised task of learning arbitrary target unitary based on the TN models, and then present the analytical lower bounds for the average risk of the models. Our results reveal the intrinsic limitations in learning arbitrary unitaries from input states encoded via TNs. The risk bounds, which depend on both the bond and physical dimensions, provide a quantitative understanding of the connections between model generalization and training set size. Our results offer valuable guidelines for designing more efficient models, and open promising research directions aimed at improving the generalization capabilities of quantum-inspired TN machine learning systems.

Acknowledgments

We thank Wenjie Jiang, Zhide Lu and Weikang Li for helpful discussions. This work was supported by the National Natural Science Foundation of China (Grant No. 12375060, No. T2225008, and No. 12075128), the Fundamental Research Funds for the Central Universities (Grant No. 63243070), the Shanghai Qi Zhi Institute Innovation Program SQZ202318, the Innovation Program for Quantum Science and Technology (No. 2021ZD0302203), and the Tsinghua University Dushi Program.

References

[1] Román Orús. Tensor networks for complex quantum systems. *Nat. Rev. Phys.*, 1(9):538–550, 2019.

[2] Jacob Biamonte. Lectures on quantum tensor networks. *arXiv:1912.10049*, 2019.

[3] J. Ignacio Cirac, David Pérez-García, Norbert Schuch, and Frank Verstraete. Matrix product states and projected entangled pair states: Concepts, symmetries, theorems. *Rev. Mod. Phys.*, 93:045003, Dec 2021.

[4] Mari Carmen Bañuls. Tensor Network Algorithms: A Route Map. *Annu. Rev. Condens. Matter Phys.*, 14(1):173–191, 2023.

[5] Hans-Martin Rieser, Frank Köster, and Arne Peter Raulf. Tensor networks for quantum machine learning. *Proc. Roy. Soc. A*, 2023.

[6] Maolin Wang, Yu Pan, Zenglin Xu, Xiangli Yang, Guangxi Li, and Andrzej Cichocki. Tensor Networks Meet Neural Networks: A Survey and Future Perspectives. 2023.

[7] Andrzej Cichocki. Tensor networks for big data analytics and large-scale optimization problems. *arXiv:1407.3124*, 2014.

[8] Alexander Novikov, Dmitrii Podoprikhin, Anton Osokin, and Dmitry P Vetrov. Tensorizing Neural Networks. In *Advances in NeuralIPS*, volume 28, 2015.

[9] Andrzej Cichocki, Namgil Lee, Ivan Oseledets, Anh-Huy Phan, Qibin Zhao, and Danilo P. Mandic. Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 1 Low-Rank Tensor Decompositions. *Foundations and Trends in Machine Learning*, 9(4-5):249–429, 2016.

[10] Andrzej Cichocki, Anh-Huy Phan, Qibin Zhao, Namgil Lee, Ivan V Oseledets, Masashi Sugiyama, and Danilo Mandic. Tensor Networks for Dimensionality Reduction and Large-scale Optimization: Part 2 Applications and Future Perspectives. *Foundations and Trends in Machine Learning*, 9(6):431–673, 2016.

[11] Edwin Stoudenmire and David J Schwab. Supervised learning with tensor networks. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in NeuralIPS*, pages 4799–4807. Curran Associates, Inc., 2016.

[12] Alexander Novikov, Mikhail Trofimov, and Ivan Oseledets. Exponential machines. *arXiv:1605.03795*, 2016.

[13] Yuhan Liu, Xiao Zhang, Maciej Lewenstein, and Shi-Ju Ran. Entanglement-guided architectures of machine learning by quantum tensor network. *arXiv:1803.09111*, 2018.

[14] Zhongming Chen, Kim Batselier, Johan A. K. Suykens, and Ngai Wong. Parallelized Tensor Train Learning of Polynomial Classifiers. *IEEE Trans. Neural Netw. Learn. Syst.*, 29(10):4621–4632, 2018.

[15] Yoav Levine, David Yakira, Nadav Cohen, and Amnon Shashua. Deep learning and quantum entanglement: Fundamental connections with implications to network design. In *ICLR*, 2018.

[16] E Miles Stoudenmire. Learning relevant features of data with multi-scale tensor networks. *Quantum Sci. Technol.*, 3(3):034003, April 2018.

[17] Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, and Pan Zhang. Unsupervised generative modeling using matrix product states. *Phys. Rev. X*, 8(3):031012, July 2018.

[18] Ding Liu, Shi-Ju Ran, Peter Wittek, Cheng Peng, Raul Blázquez García, Gang Su, and Maciej Lewenstein. Machine learning by unitary tensor network of hierarchical tree structure. *New J. Phys.*, 21(7):073059, July 2019.

[19] Kohei Hayashi, Taiki Yamaguchi, Yohei Sugawara, and Shin-ichi Maeda. Exploring Unexplored Tensor Network Decompositions for Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[20] Jing Liu, Sujie Li, Jiang Zhang, and Pan Zhang. Tensor networks for unsupervised machine learning. *arXiv:2106.12974v1*, 2021.

[21] Jing Chen, Song Cheng, Haidong Xie, Lei Wang, and Tao Xiang. Equivalence of restricted Boltzmann machines and tensor network states. *Phys. Rev. B*, 97(8):085104, February 2018.

[22] Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua. Quantum entanglement in deep learning architectures. *Phys. Rev. Lett.*, 122:065301, Feb 2019.

[23] Amandeep Singh Bhatia, Mandeep Kaur Saggi, Ajay Kumar, and Sushma Jain. Matrix product state–based quantum classifier. *Neural Comput.*, 31(7):1499–1517, 2019.

[24] Stavros Efthymiou, Jack Hidary, and Stefan Leichenauer. Tensornetwork for machine learning. *arXiv:1906.06329*, 2019.

[25] William Huggins, Piyush Patil, Bradley Mitchell, K Birgitta Whaley, and E Miles

Stoudenmire. Towards quantum machine learning with tensor networks. *Quantum Sci. Technol.*, 4(2):024001, jan 2019.

[26] Ivan Glasser, Nicola Pancotti, and J. Ignacio Cirac. From probabilistic graphical models to generalized tensor networks for supervised learning. *IEEE Access*, 8:68169–68182, 2020.

[27] Jiahao Su, Wonmin Byeon, Jean Kossaifi, Furong Huang, Jan Kautz, and Anima Anandkumar. Convolutional Tensor-Train LSTM for Spatio-Temporal Learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 13714–13726, 2020.

[28] Zheng-Zhi Sun, Shi-Ju Ran, and Gang Su. Tangent-space gradient optimization of tensor network for machine learning. *Phys. Rev. E*, 102:012152, Jul 2020.

[29] Xingwei Sun, Ze-Feng Gao, Zhong-Yi Lu, Junfeng Li, and Yonghong Yan. A Model Compression Method With Matrix Product Operators for Speech Enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 28:2837–2847, 2020.

[30] Samuel Yen-Chi Chen, Chih-Min Huang, Chia-Wei Hsing, and Ying-Jer Kao. Hybrid quantum-classical classifier based on tensor network and variational quantum circuit. 2020.

[31] Zheng-Zhi Sun, Cheng Peng, Ding Liu, Shi-Ju Ran, and Gang Su. Generative tensor network classification model for supervised machine learning. *Phys. Rev. B*, 101(7):075135, 2020.

[32] Jinhui Wang, Chase Roberts, Guifre Vidal, and Stefan Leichenauer. Anomaly detection with tensor networks. *arXiv:2006.02516*, 2020.

[33] Ze-Feng Gao, Song Cheng, Rong-Qiang He, Z. Y. Xie, Hui-Hai Zhao, Zhong-Yi Lu, and Tao Xiang. Compressing deep neural networks by matrix product operators. *Phys. Rev. Res.*, 2(2):023300, 2020.

[34] Michael L. Wall, Matthew R. Abernathy, and Gregory Quiroz. Generative machine learning with tensor networks: Benchmarks on near-term quantum computers. *Phys. Rev. Res.*, 3:023010, Apr 2021.

[35] Song Cheng, Lei Wang, and Pan Zhang. Supervised learning with projected entangled pair states. *Phys. Rev. B*, 103:125117, Mar 2021.

[36] M Nazareth da Costa, Romis Attux, Andrzej Cichocki, and João MT Romano. Tensor-train networks for learning predictive modeling of multidimensional data. *arXiv:2101.09184*, 2021.

[37] Andrey Kardashin, Alexey Uvarov, and Jacob Biamonte. Quantum Machine Learning Tensor Network States. *Front. Phys.*, 8:644, 2021.

[38] Timo Felser, Marco Trenti, Lorenzo Sestini, Alessio Gianelle, Davide Zuliani, Donatella Lucchesi, and Simone Montangero. Quantum-inspired machine learning on high-energy physics data. *npj Quantum Inf.*, 7(1):1–8, 2021.

[39] Kunkun Wang, Lei Xiao, Wei Yi, Shi-Ju Ran, and Peng Xue. Experimental realization of a quantum image classifier via tensor-network-based machine learning. *Photon. Res.*, 9(12):2332, 2021.

[40] Cole Hawkins and Zheng Zhang. Bayesian tensorized neural networks with automatic rank selection. *Neurocomputing*, 453:172–180, 2021.

[41] Cong Chen, Kim Batselier, Wenjian Yu, and Ngai Wong. Kernelized support tensor train machines. *Pattern Recognit.*, 122:108337, 2022.

[42] Tom Vieijra, Laurens Vanderstraeten, and Frank Verstraete. Generative modeling with projected entangled-pair states. 2022.

[43] Xiao Shi, Yun Shang, and Chu Guo. Clustering using matrix product states. *Phys. Rev. A*, 105(5):052424, 2022.

[44] Ian Convy, William Huggins, Haoran Liao, and K. Birgitta Whaley. Mutual information scaling for tensor network machine learning. *Mach. Learn.: Sci. Technol.*, 3(1):015017, 2022.

[45] Friederike Metz and Marin Bukov. Self-correcting quantum many-body control using reinforcement learning with tensor networks. *Nat. Mach. Intell.*, 5(7):780–791, 2023.

[46] Haoran Liao, Ian Convy, Zhibo Yang, and K. Birgitta Whaley. Decohering tensor network quantum machine learning models. *Quantum Mach. Intell.*, 5(1):7, 2023.

[47] Shi-Ju Ran and Gang Su. Tensor Networks for Interpretable and Efficient Quantum-Inspired Machine Learning. *Intell. Comput.*, 2:0061, 2023.

[48] Dian Wu, Riccardo Rossi, Filippo Vicentini, and Giuseppe Carleo. From tensor-network quantum states to tensorial recurrent neural networks. *Phys. Rev. Res.*, 5(3):L032001, 2023.

[49] Javier Lopez-Piqueres, Jing Chen, and Alejandro Perdomo-Ortiz. Symmetric tensor networks for generative modeling and constrained combinatorial optimization. *Mach. Learn.: Sci. Technol.*, 4(3):035009, 2023.

[50] Ye-Ming Meng, Jing Zhang, Peng Zhang, Chao Gao, and Shi-Ju Ran. Residual matrix product state for machine learning. *SciPost Phys.*, 14(6):142, 2023.

[51] Seongwook Shin, Yong Siah Teo, and Hyunseok Jeong. Dequantizing quantum machine learning models using tensor networks. *Phys. Rev. Res.*, 6(2):023218, 2024.

[52] Frederiek Wesel and Kim Batselier. Tensor Network-Constrained Kernel Machines as Gaussian Processes. 2024.

[53] Amandeep Singh Bhatia and David E. Bernal Neira. Federated Hierarchical Tensor Networks: A Collaborative Learning Quantum AI-Driven Framework for Healthcare. 2024.

[54] Andrei Tomut, Saeed S. Jahromi, Sukhbinder Singh, Faysal Ishtiaq, Cesar Muñoz, Prabdeep Singh Bajaj, Ali Elborady, Gianni del Bimbo, Mehrazin Alizadeh, David Montero, Pablo Martin-Ramiro, Muhammad Ibrahim, Oussama Tahiri Alaoui, John Malcolm, Samuel Mugel, and Roman Orus. CompactifAI: Extreme Compression of Large Language Models using Quantum-Inspired Tensor Networks. 2024.

[55] Yanting Teng, Rhine Samajdar, Katherine Van Kirk, Frederik Wilde, Subir Sachdev, Jens Eisert, Ryan Sweke, and Khadijeh Najafi. Learning topological states from randomized measurements using variational tensor network tomography. 2024.

[56] Pablo Bermejo, Paolo Braccia, Manuel S. Rudolph, Zoë Holmes, Lukasz Cincio, and M. Cerezo. Quantum Convolutional Neural Networks are (Effectively) Classically Simulable. 2024.

[57] Heitor P. Casagrande, Bo Xing, William J. Munro, Chu Guo, and Dario Poletti. Tensor-Networks-based Learning of Probabilistic Cellular Automata Dynamics. 2024.

[58] Hao Chen and Thomas Barthel. Machine learning with tree tensor networks, CP rank constraints, and tensor dropout. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46:7825, 2024.

[59] Zhan Su, Yuqin Zhou, Fengran Mo, and Jakob Grue Simonsen. Language Modeling Using Tensor Trains. 2024.

[60] Alejandro Pozas-Kerstjens, Senaida Hernández-Santana, José Ramón Pareja Monturiol, Marco Castrillón López, Giannicola Scarpa, Carlos E. González-Guillén, and David Pérez-García. Privacy-preserving machine learning with tensor networks. *Quantum*, 8:1425, July 2024.

[61] Chu Guo, Zhanming Jie, Wei Lu, and Dario Poletti. Matrix product operators for sequence-to-sequence learning. *Phys. Rev. E*, 98:042114, Oct 2018.

[62] Konstantinos Meichanetzidis, Stefano Gogioso, Giovanni De Felice, Nicolò Chiappori, Alexis Toumi, and Bob Coecke. Quantum natural language processing on near-term quantum computers. *arXiv:2005.04147*, 2020.

[63] Jing-Chuan Wu, Qi Ye, Dong-Ling Deng, and Li-Wei Yu. No-Free-Lunch Theories for Tensor-Network Machine Learning Models.

[64] Cullen Schaffer. A Conservation Law for Generalization Performance. In William W. Cohen and Haym Hirsh, editors, *Machine Learning Proceedings 1994*, pages 259–265. San Francisco (CA), 1994.

[65] David H. Wolpert. The Lack of A Priori Distinctions Between Learning Algorithms. *Neur. Comput.*, 8(7):1341–1390, 1996.

[66] D.H. Wolpert and W.G. Macready. No free lunch theorems for optimization. *IEEE Trans. Evol. Computat.*, 1(1):67–82, 1997.

[67] Kyle Poland, Kerstin Beer, and Tobias J Osborne. No free lunch for quantum machine learning. 2020.

[68] Tyler Volkoff, Zoë Holmes, and Andrew Sornborger. Universal Compiling and (No-)Free-Lunch Theorems for Continuous-Variable Quantum Learning. *PRX Quantum*, 2(4):040327, 2021.

[69] Kunal Sharma, M. Cerezo, Zoë Holmes, Lukasz Cincio, Andrew Sornborger, and Patrick J. Coles. Reformulation of the no-free-lunch theorem for entangled datasets. *Phys. Rev. Lett.*, 128:070501, Feb 2022.

[70] Richard Brierley. No free lunch for Schrödinger's cat. *Nat. Phys.*, 18(4):373–373, 2022.

[71] Haimeng Zhao, Laura Lewis, Ishaan Kannan, Yihui Quek, Hsin-Yuan Huang, and Matthias C. Caro. Learning quantum states and unitaries of bounded gate complexity. 2023.

[72] Xinbiao Wang, Yuxuan Du, Zhuozhuo Tu, Yong Luo, Xiao Yuan, and Dacheng Tao. Transition role of entangled data in quantum machine learning. *Nat. Commun.*, 15(1):3716, 2024.

[73] Xinbiao Wang, Yuxuan Du, Kecheng Liu, Yong Luo, Bo Du, and Dacheng Tao. Separable Power of Classical and Quantum Learning Protocols Through the Lens of No-Free-Lunch Theorem. 2024.

[74] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac. Matrix product state representations. *Quantum Inf. Comput.*, 7(5):401–430, July 2007.

[75] D. Gross and J. Eisert. Quantum computational webs. *Phys. Rev. A*, 82:040303(R), Oct 2010.

[76] Silvano Garnerone, Thiago R. de Oliveira, Stephan Haas, and Paolo Zanardi. Statistical properties of random matrix product states. *Phys. Rev. A*, 82:052312, Nov 2010.

[77] Silvano Garnerone, Thiago R. de Oliveira, and Paolo Zanardi. Typicality in random matrix product states. *Phys. Rev. A*, 81:032336, Mar 2010.

[78] Benoît Collins, Carlos E González-Guillén, and David Pérez-García. Matrix product states, random matrix theory and the principle of maximum entropy. 2012.

[79] Jonas Haferkamp, Christian Bertoni, Ingo Roth, and Jens Eisert. Emergent statistical mechanics from properties of disordered random matrix product states. *PRX Quantum*, 2:040308, Oct 2021.

[80] Matthias C. Caro, Hsin-Yuan Huang, M. Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J. Coles. Generalization in quantum machine learning from few training data. *Nat. Commun.*, 13(1):4919, 2022.

[81] Matthew Fishman, Steven R. White, and E. Miles Stoudenmire. The ITensor Software Library for Tensor Network Calculations. *SciPost Phys. Codebases*, page 4, 2022.

[82] Matthew Fishman, Steven R. White, and E. Miles Stoudenmire. Codebase release 0.3 for ITensor. *SciPost Phys. Codebases*, pages 4–r0.3, 2022.

[83] Zidu Liu, Qi Ye, Li-Wei Yu, L.-M. Duan, and Dong-Ling Deng. Theory on variational high-dimensional tensor networks. 2023.

Appendix A. Numerical details

In our previous theorems, we have analytically obtained the lower bound of the average generalization risk. To show how these theorems perform in practice, we carry out numerical simulations based on the open-source package `ITensors.jl` [81, 82] in the Julia programming language. Based on the MPS machine learning model, we consider the supervised task of learning target unitaries U with the labeled training samples $\{(|\psi_j\rangle, U|\psi_j\rangle)\}$, where $|\psi_j\rangle$ denotes the normalized MPS. We then plot the average generalization risks of the trained MPS-based learning models with respect to different qubit size n of learning models, as depicted in Fig. A1. For example, in $n = 4$, we randomly generate a 16-dimensional target unitary and a training set of MPSs, and conduct the MPS-based supervised task. By repeatedly conducting learning tasks for different target unitaries, one obtains the average generalization risk for different size of training sets. We see from Fig. A1 that the average error risks decrease with respect to the training set size. This is consistent with the analytical lower bound of the average risks predicted in our theorems.

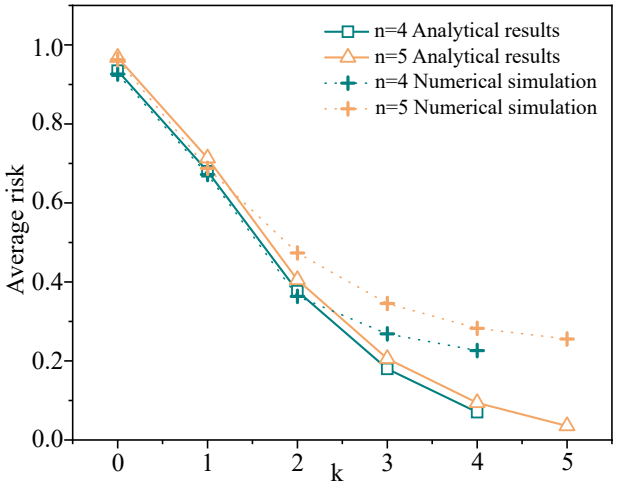

Fig. A1: Average risk of the trained MPS-based machine learning models with respect to the training set size $t_k = 2^n - 2^{n-k}$, where the system qubit size n varies from four to five. The physical dimension $d = 2$, and the bond dimension $D = 2$. The solid lines represent the analytical lower bounds of the average risk predicted by Theorem 1. And the dotted lines denote the average risk of the trained MPS-based machine learning models for predicting target unitaries.

Appendix B. The polynominoes

To address the 2D TN problem, we initially introduce a statistical model for enumerating lattice configurations on a two-dimensional plane, known as the polyomino. The focus of this model is on directed figures that encompass a specific number of sites within an infinite square grid. We will find that the partition function of the two-dimensional Ising model can be translated into a problem sets on a periodic plane of an infinite square lattice grid, making the application of polyomino calculation method particularly effective. We first provide some definitions related to polyominoes [83], and then map our problem to this framework. The formal definition of the directed polyomino(shown in Fig. A2) is listed as follows:

Definition 1. *(Directed polyomino). A directed polyomino is a finite subset $\vec{\tau} \subseteq \mathbb{Z} \times \mathbb{Z}$ ($\mathbb{Z}$ denotes the set of integers) in the Euclidean plane rooted at (x_0, y_0) such that*

- $(x_0, y_0) \in \vec{\tau}$,
- *for $(x, y) \in \vec{\tau}$ and $(x, y) \neq (x_0, y_0)$, at least one of $(x + 1, y) \in \vec{\tau}$ and $(x, y + 1) \in \vec{\tau}$ are in $\vec{\tau}$ as well.*

The polyominoes can be characterized by their areas, perimeters and upper perimeters as their properties, whose formal definitions are as follows:

Definition 2. *(area, perimeter and upper perimeter). Let $\vec{\tau}$ be a directed polyomino.*

- *The area m of $\vec{\tau}$ is the size of the polyomino, i.e.$|\vec{\tau}|$.*
- *The perimeter p of $\vec{\tau}$ is the number of edges on the boundary of $\vec{\tau}$. Formally speaking, $p = |\{(x, y) \in \mathbb{Z} \times \mathbb{Z} : \tau_{x,y} \neq \tau_{x+1,y}\}| + |\{(x, y) \in \mathbb{Z} \times \mathbb{Z} : \tau_{x,y} \neq \tau_{x,y+1}\}|$.*
- *The upper perimeter n of $\vec{\tau}$ is the number of horizontal edges on the top boundary of $\vec{\tau}$. Formally speaking,$n = |\{(x, y) \in \mathbb{Z} \times \mathbb{Z} : \tau_{x,y} = 0, \tau_{x+1,y} = 1\}|$.*

Then we can enumerate directed polyominoes exactly via their generating function by the following lemma:

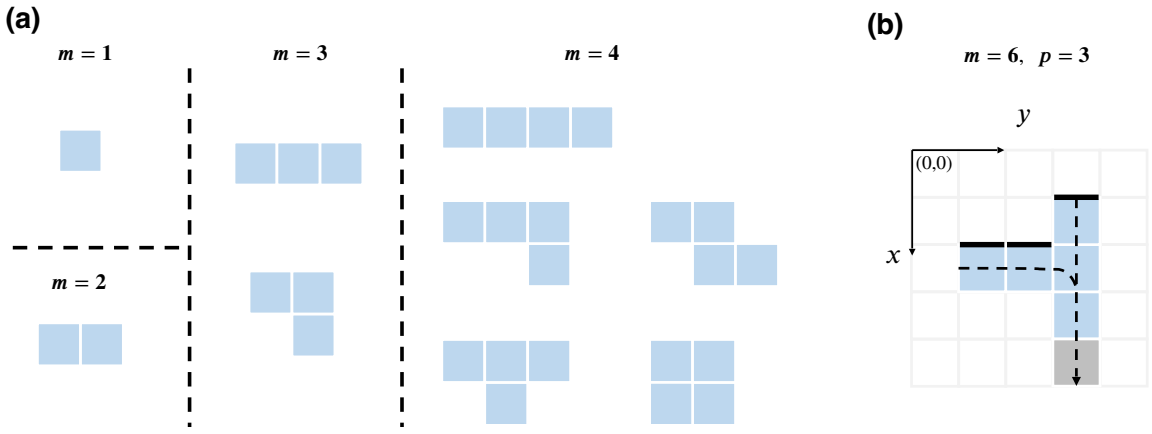

Fig. A2: Illustrations of the general polyomino and the directed polyomino. (a) Illustration of the various types of general polyominoes with the area ranging from 1 to 4. Here we regard the different rotations of a polyomino as being of the same type. (b) Illustration of a directed polyomino rooted at the gray site with area m = 6, perimeter p = 14, and upper perimeter n = 3 (depicted as the black lines).

Lemma 1. *Let $D_{m,n}$ be the number of directed polyominoes rooted at (0,0) with area and upper perimeter m, n respectively. p and q are two variables. We introduce the generating function of $D_{m,n}$ as*

$$G(q,p) = \sum_{m,n} D_{m,n} q^m p^n. \tag{A1}$$

When $p, q \in (0,1]$ such that $|q(2+p) - q^2(1-p)| < 1$, the power series converges to a finite value

$$G(q,p) = \frac{p}{2}\left(\sqrt{\frac{(1+q)(1+q-qp)}{1-q(2+p)+q^2(1-p)}} - 1\right). \tag{A2}$$

This lemma implies that when $q, p \in (0,1]$ such that $|q(2+p) + q^2(1-p)| \leq 1$, the power series $\sum_{m,n} D_{m,n} q^m p^n$ converges to a finite value $G(q,p)$. We will see the problem of partition function for two-dimensional lattice having a similar power-series form reduces to determining the number of configurations $D_{m,n}$.