JuliaCN 2024 冬季见面会

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

陈省身数学研究所
Chern Institute of Mathematics

于立伟

陈省身数学研究所
南开大学

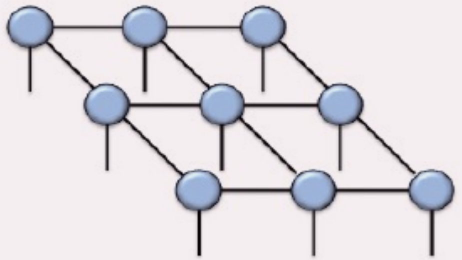
2024-11-02

Outline

- Introduction
- No-Free-Lunch theorem in MPS based ML models
- No-Free-Lunch theorem in PEPS based ML models
- Discussion & Conclusion

Introduction

Tensor Network

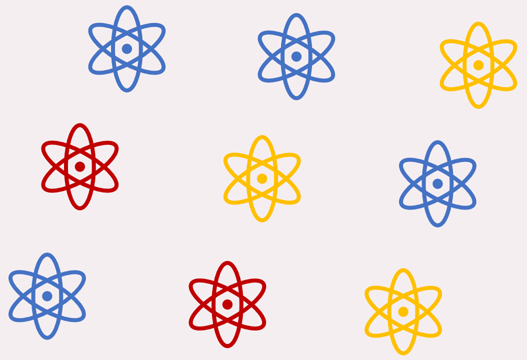


- ✓ DMRG
- ✓ Describe quantum entanglement
- ✓ Understand QEC
- ✓ Ads/CFT
- ✓ …

- ✓ Efficent representation power: $O(MD^2\ d)$ parameters for $O(d^M\ )$-D
- ✓ Data compression
- ✓ Show quantum advantages in QML
- ✓ …

Quantum inspired

Quantum Physics



Quantum-enhanced ML

ML for quantum physics:
- ✓ ML phases of matter
- ✓ Optimize quantum circuits
- ✓ High energy physics
- ✓ …

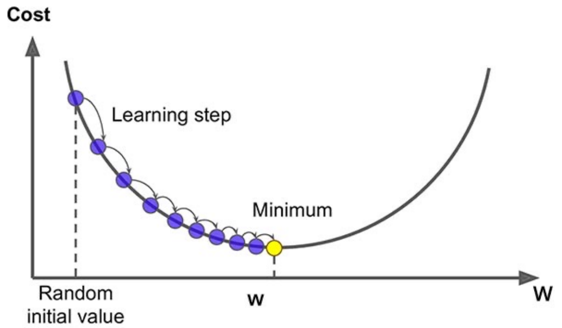
Machine Learning

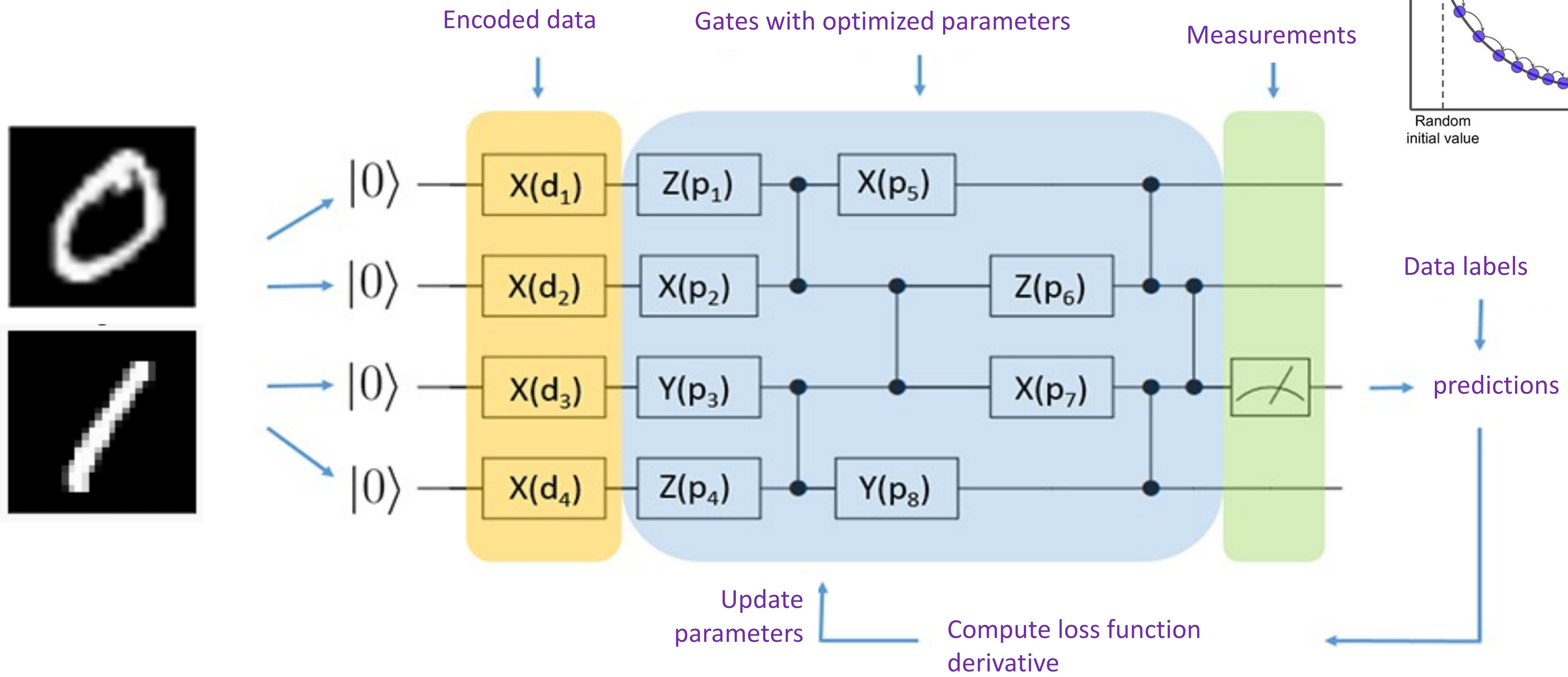
Theoretical properties in machine learning models

- Trainability
- Generalization ability
- Learnability
- Expressibility
- ….

Quantum enhanced machine learning model

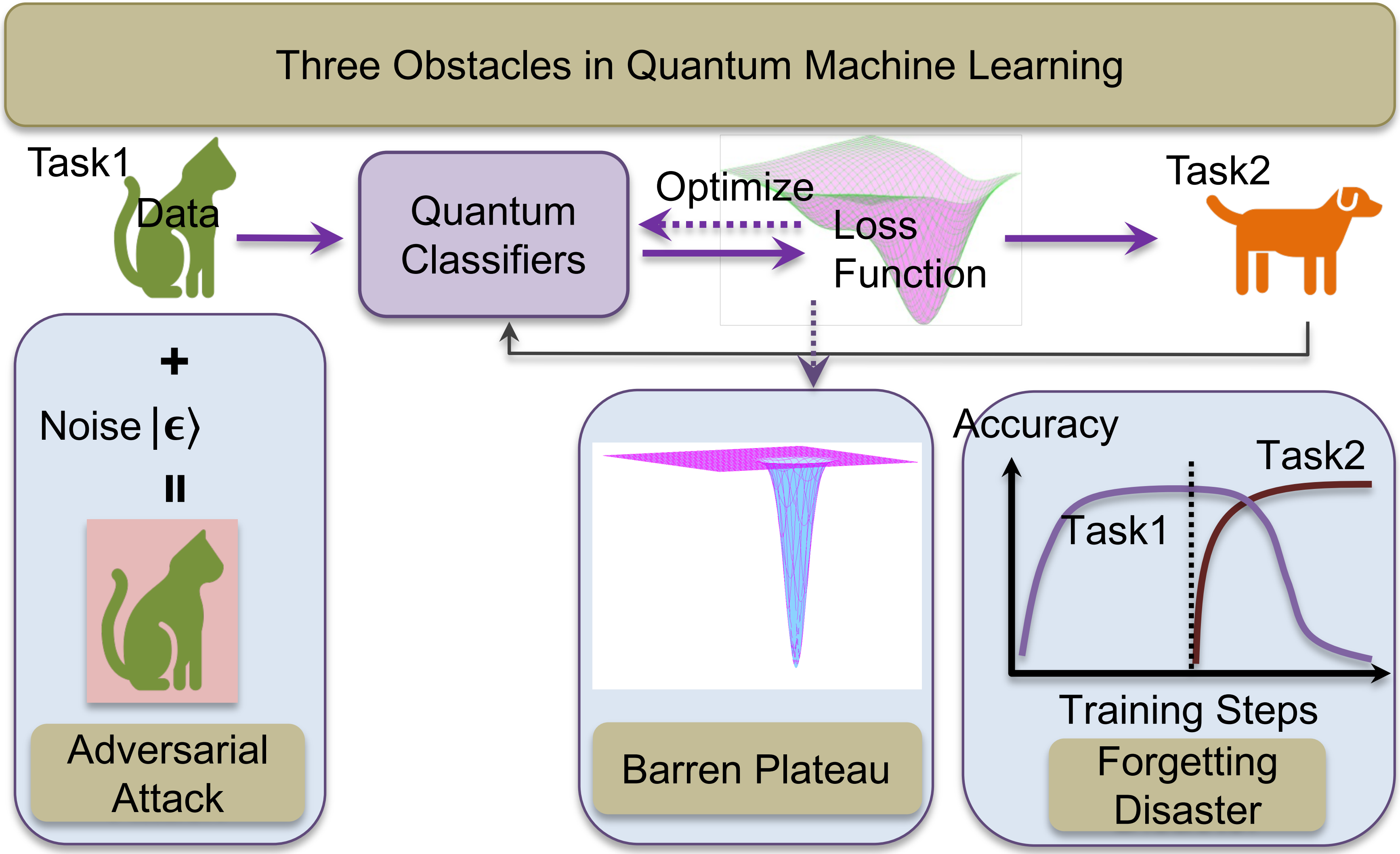
Variational quantum circuit for identifying MINST data





$$\mathcal{L} = \frac{1}{2}\sum_i \left(\langle \vec{x}_i | U^\dagger(\vec{\theta}) \hat{O} U(\vec{\theta}) | \vec{x}_i \rangle - y_i \right)^2$$

Obstacles in Quantum-enhanced ML



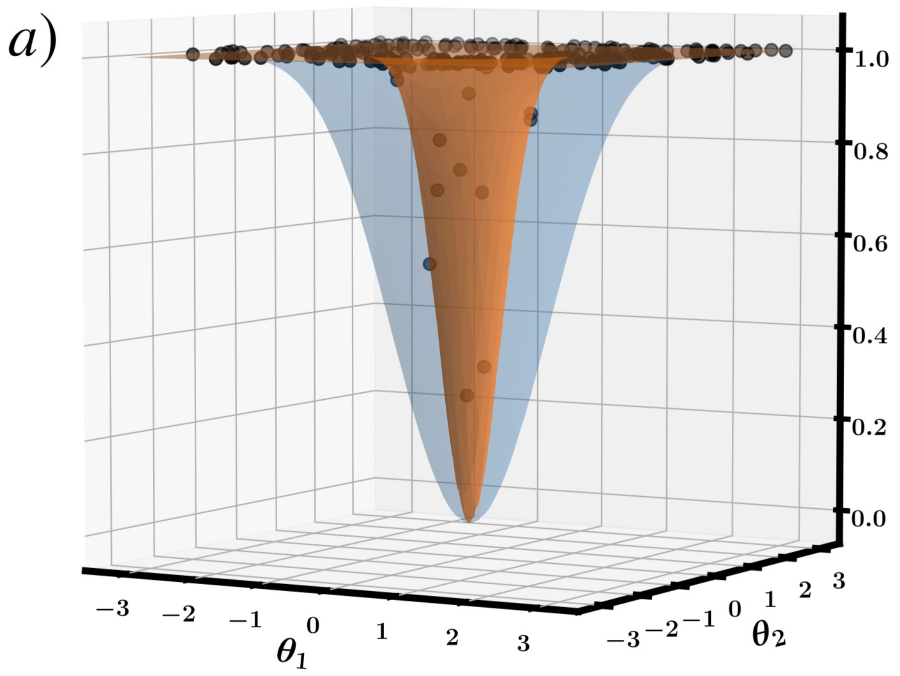
Sun, **YLW**, Deng, 量子人工智能简介, CCCF 2023

Obstacles in Quantum enhanced ML

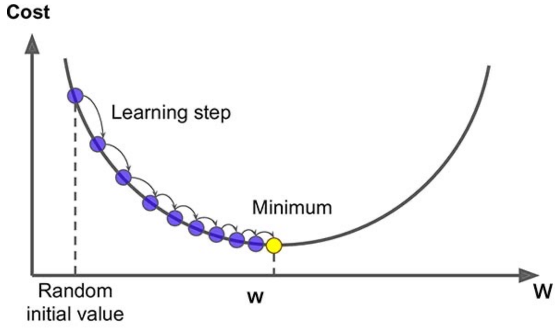
Barren plateau problem

Barren Plateau: $\mathbb{E}[\partial_k E] = 0;\ \mathrm{Var}\,[\partial_k E] \sim O(2^{-n})$

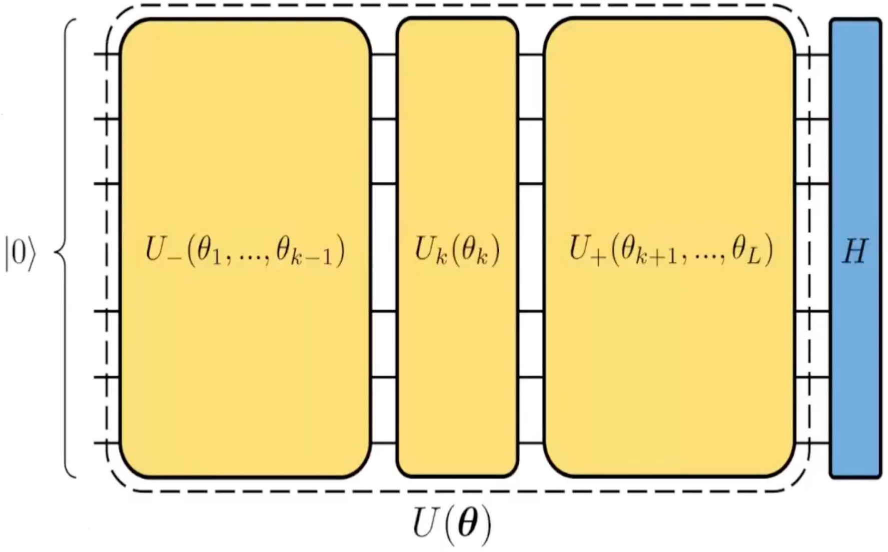
$$\Pr(|\partial_k E| \geq \epsilon) \leq \epsilon^{-2} O(2^{-n})$$



Machine learning task ~ optimize cost function



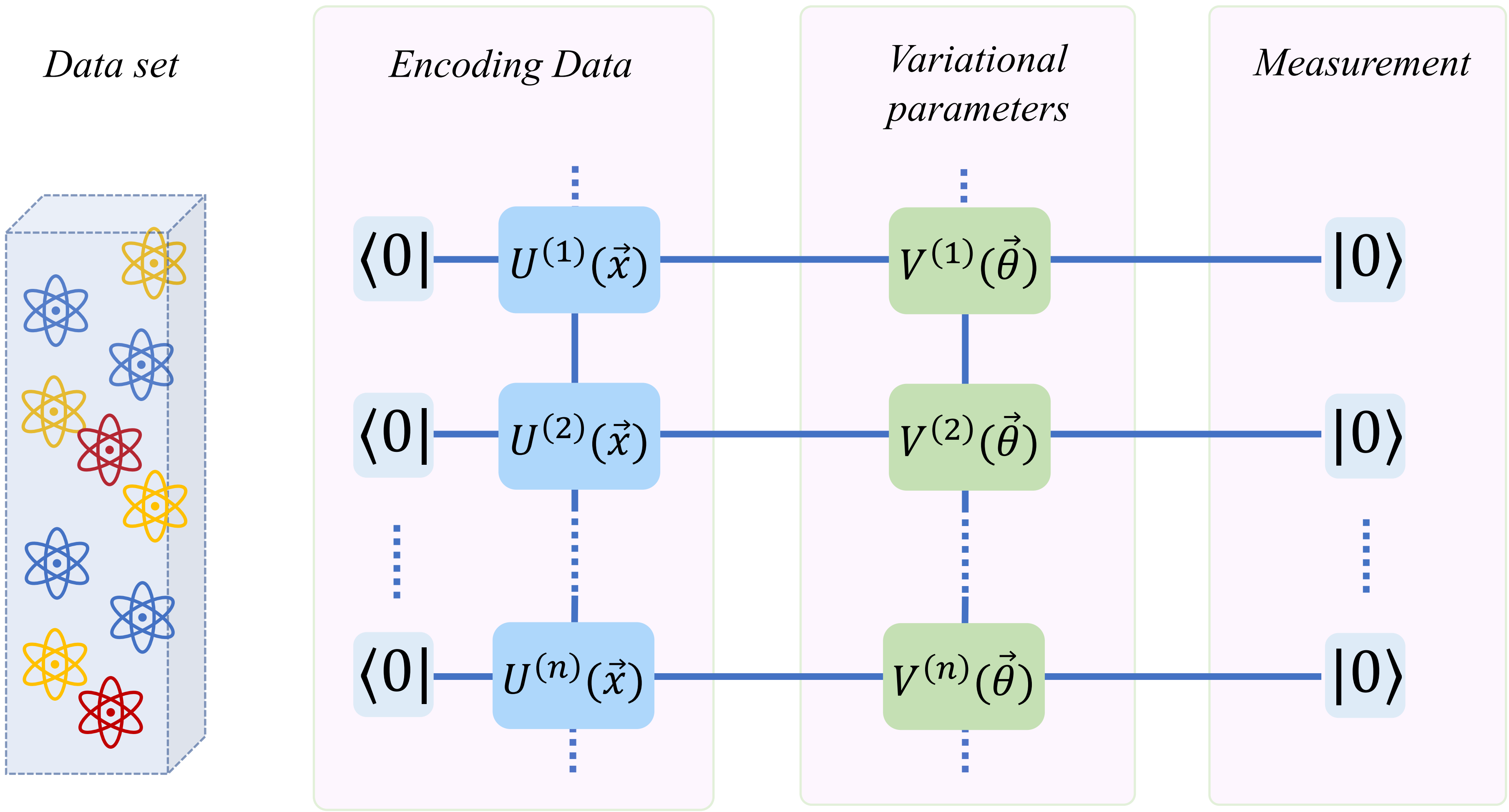
Variational quantum circuit



McClean, J.R. et al. Nat. Commun. 9, 4812 (2018).

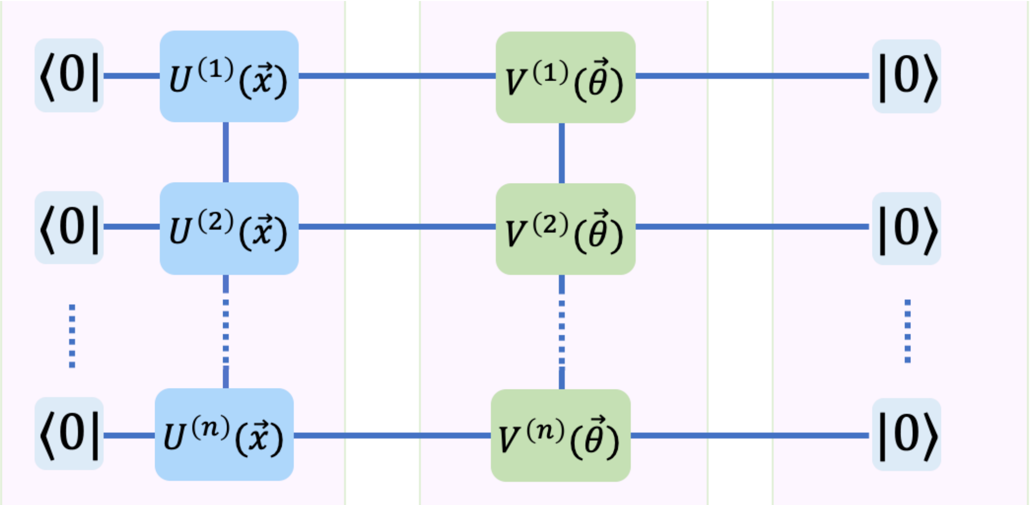
Tensor Network based Machine Learning

A toy example: *Machine learning model based on Unitary Embedded Matrix Product States*



Barren Plateaus in TN-based ML

MPS-based machine learning



Training process: Randomly initialized

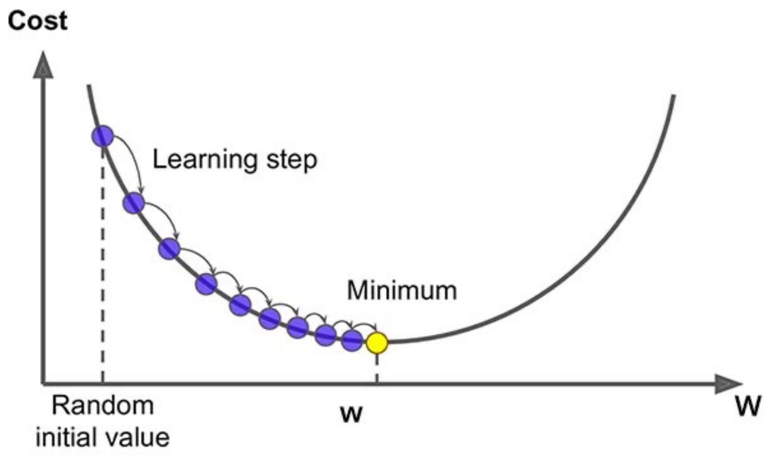
Optimize loss function L

$$L = 1 - \frac{\langle \psi(\vec{\theta}) | \hat{H} | \psi(\vec{\theta}) \rangle}{\langle \psi(\vec{\theta}) | \psi(\vec{\theta}) \rangle}$$

Gradient based

$$\partial_{\vec{\theta}} L = -\partial_{\vec{\theta}} \left(\frac{\langle \psi(\vec{\theta}) | \hat{H} | \psi(\vec{\theta}) \rangle}{\boxed{\langle \psi(\vec{\theta}) | \psi(\vec{\theta}) \rangle}} \right)$$

Concentrate to 1

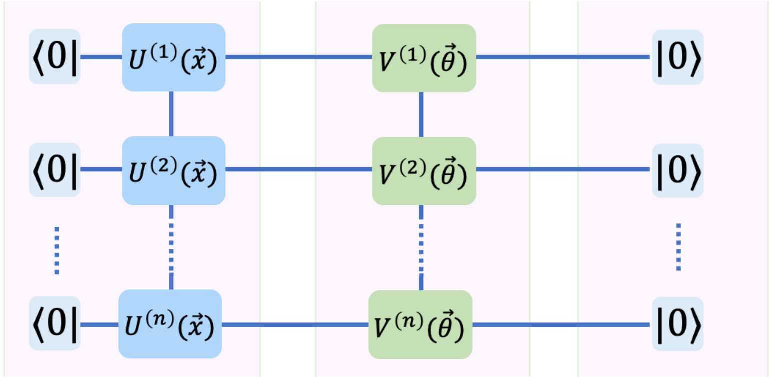


Z Liu, **YLW,** LM Duan, DL Deng, PRL 129, 270501(2022)

Presence of Barren Plateaus for global loss

$$L_g = 1 - \frac{\langle \psi(\vec{\theta}) | \hat{H} | \psi(\vec{\theta}) \rangle}{\langle \psi(\vec{\theta}) | \psi(\vec{\theta}) \rangle}$$

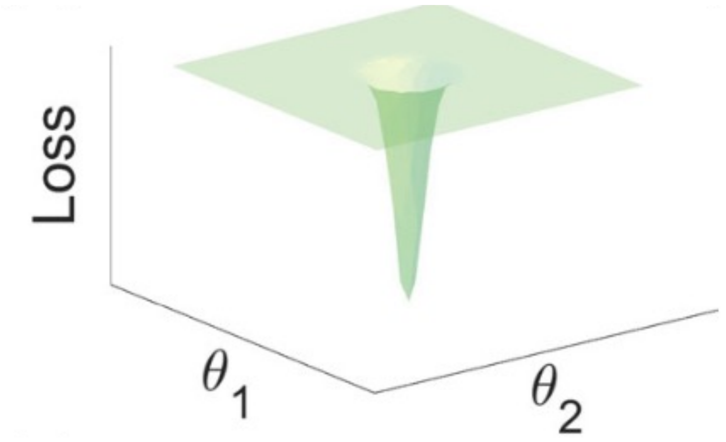
$\hat{H}$: Density matrices of states,
e.g. $\hat{H} = (|0\rangle\langle 0|)^{\otimes N}$.



Theorem 1. *Define the derivative of the global loss function with respect to the variational parameter $\theta_\lambda^{(k)}$ by $\partial_\lambda^{(k)} \mathcal{L}_g$. Then $\forall\ \theta_\lambda^{(k)} \in \Theta$, $\partial_\lambda^{(k)} \mathcal{L}_g$ obeys the following inequality:*

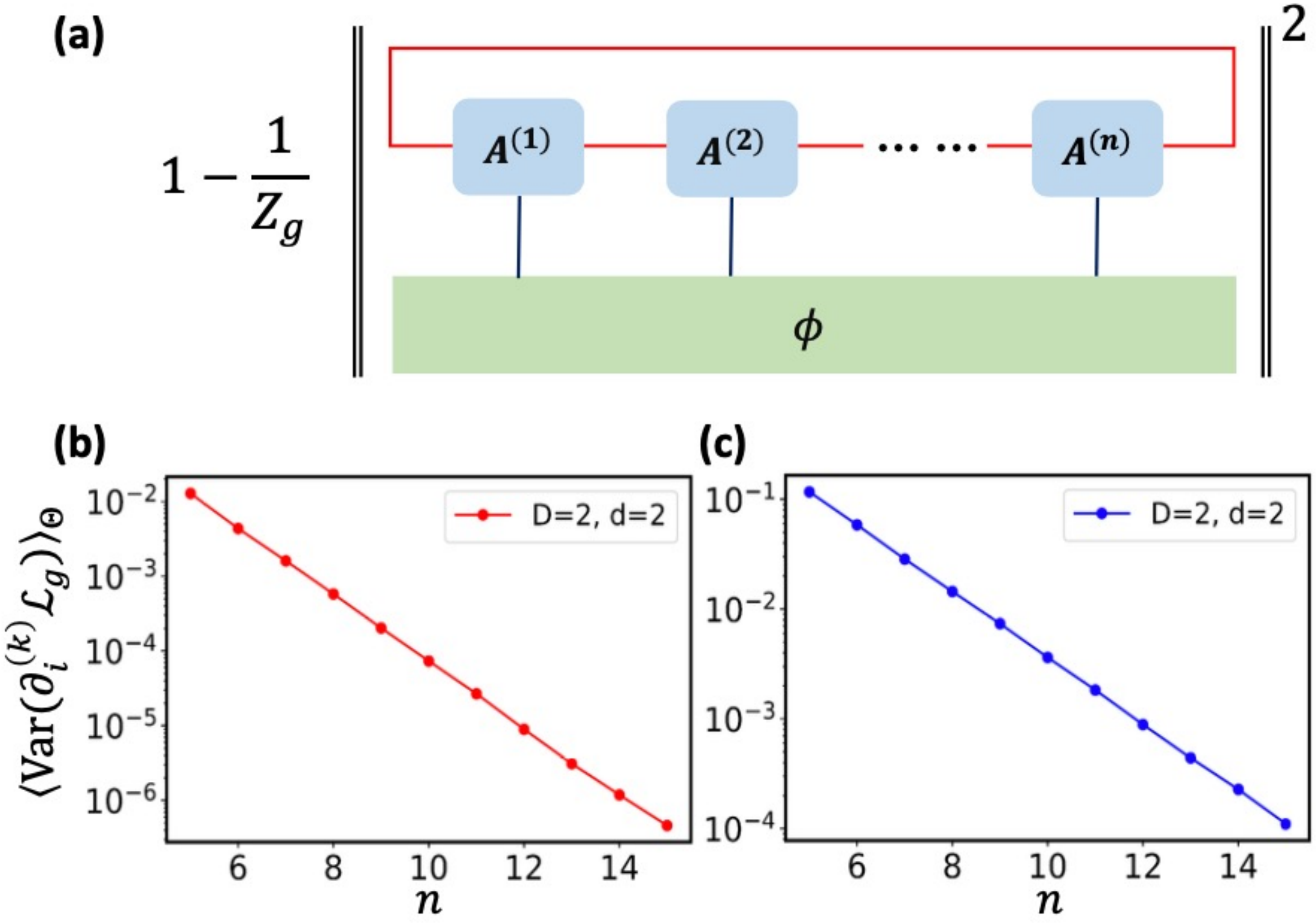
$$\Pr\left(|\partial_\lambda^{(k)} \mathcal{L}_g| > \epsilon \right) \leq \epsilon^{-2} \mathcal{O}(d^{-n}),$$

where $\Pr(\cdot)$ *represents the probability.*



Z Liu, **YLW,** LM Duan, DL Deng, PRL 129, 270501(2022)

Presence of Barren Plateaus for global loss



Chebyshev inequality:

$$\Pr(|\partial_k L - \mathbb{E}[\partial_k L]| \geq \epsilon) \leq \frac{\mathrm{Var}\,[\partial_k L]}{\epsilon^2}$$

$$\mathbb{E}[\partial_k L] = 0$$

$$\mathrm{Var}\,[\partial_k L] \sim O(d^{-n})$$

Z Liu, **YLW,** LM Duan, DL Deng, PRL 129, 270501(2022)

Absence of Barren Plateaus for local loss

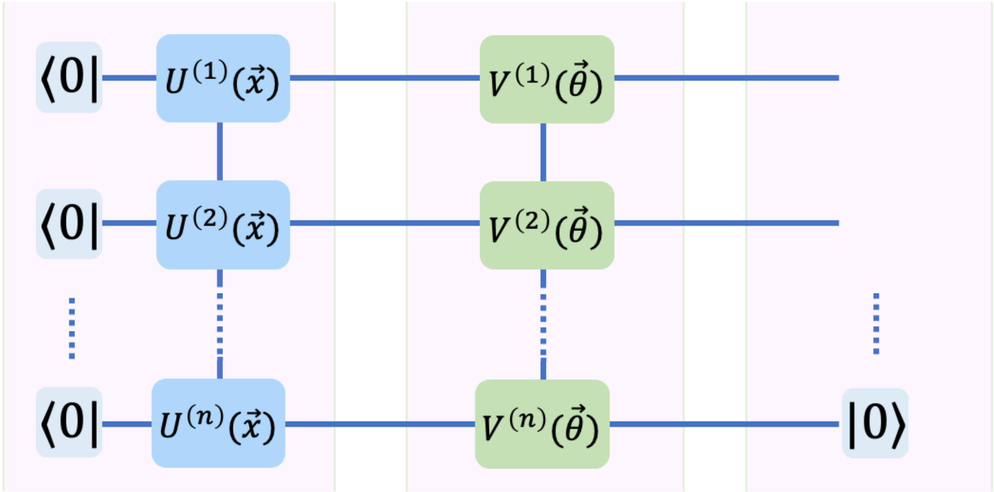
$$\mathcal{L}_l = \frac{\langle\psi(\Theta)|\hat{O}_m|\psi(\Theta)\rangle}{\langle\psi(\Theta)|\psi(\Theta)\rangle}$$

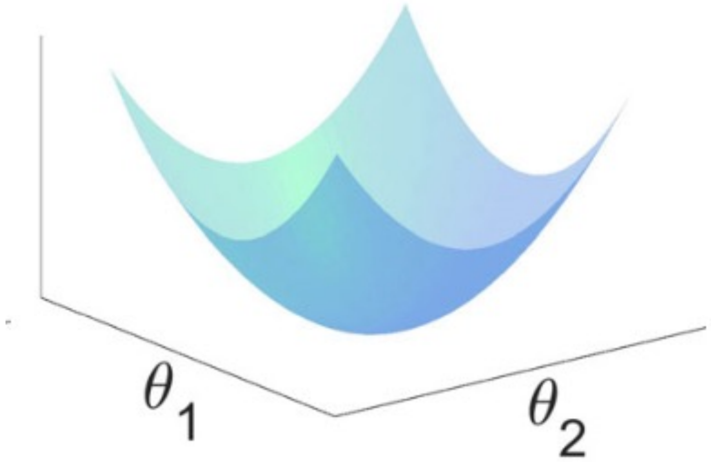
Theorem 2. *Define the derivative of the local loss function $\mathcal{L}_l$ with respect to the parameter $\theta_\lambda^{(k)}$ by $\partial_\lambda^{(k)}\mathcal{L}_l$. Then $\forall\ \theta_\lambda^{(k)} \in \Theta$, the expectation value of the gradient $\partial_\lambda^{(k)}\mathcal{L}_l$ vanishes, and as the system size approaches infinity, the variance of gradient $\partial_\lambda^{(k)}\mathcal{L}_l$ approaches a finite value:*

$$\mathrm{Var}(\partial_\lambda^{(k)}\mathcal{L}_l) \sim \mathcal{O}\left(\mathrm{Tr}(\hat{O}_m^2)\frac{P(D,d)}{Q(D,d)}\right),$$

where $P(D,d)$ and $Q(D,d)$ are certain polynomials of D and d with constant degrees [94]. In addition, the upper bound of the variance of $\partial_\lambda^{(k)}\mathcal{L}_l$ decays exponentially with respect to the distance $\Delta \equiv |k-m|$:

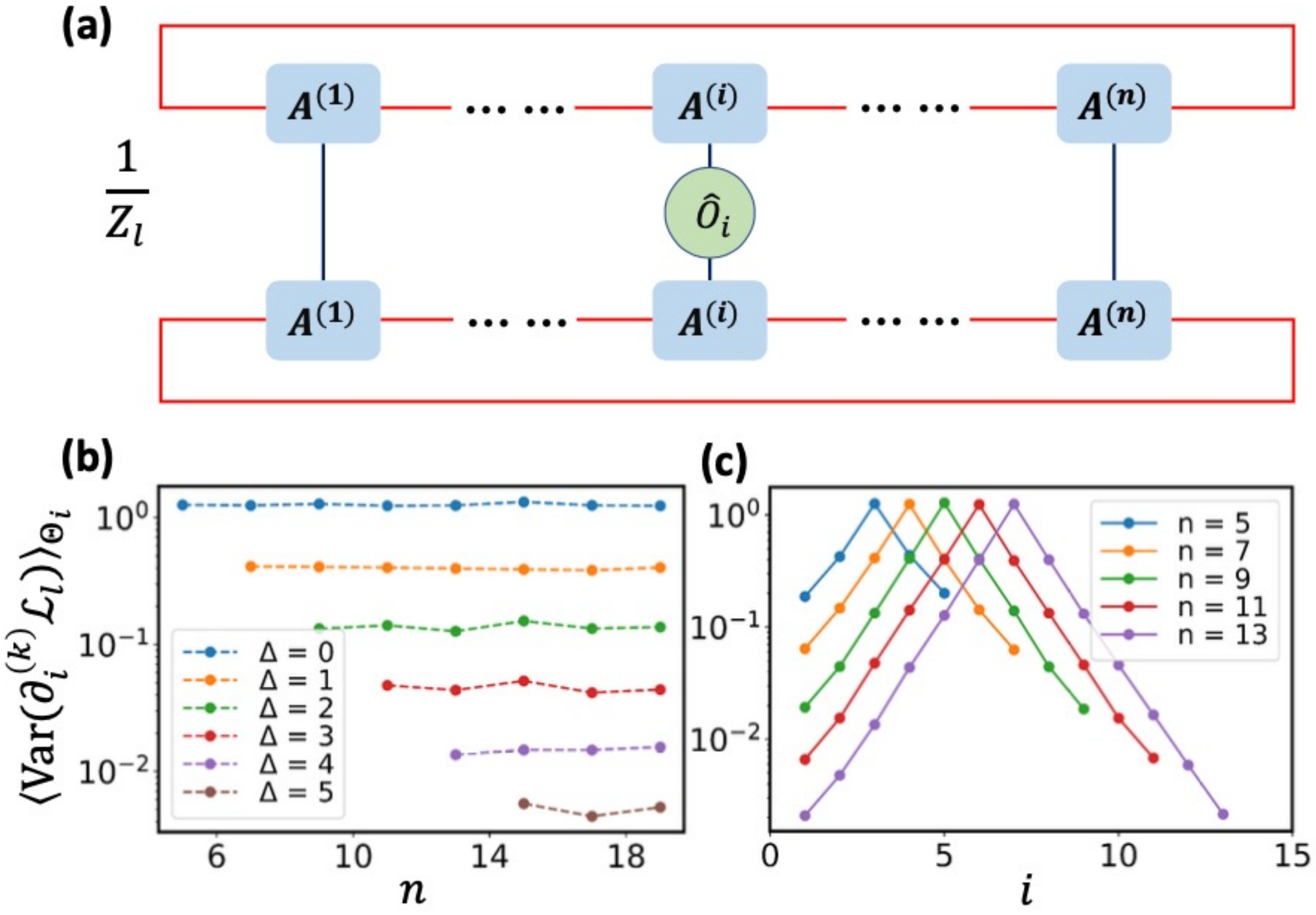
$$\mathrm{Var}(\partial_\lambda^{(k)}\mathcal{L}_l) \le \mathcal{O}(d^{-\Delta}).$$





Z Liu, **YLW,** LM Duan, DL Deng, PRL 129, 270501(2022)

Absence of Barren Plateaus for local loss



Chebyshev inequality:

$$\Pr(|\partial_k L - \mathbb{E}[\partial_k L]| \geq \epsilon) \leq \frac{\mathrm{Var}\,[\partial_k L]}{\epsilon^2}$$

$$\mathbb{E}[\partial_k L] = 0$$

$$\mathrm{Var}\,[\partial_k L] \sim Poly(D, d)$$

Z Liu, **YLW,** LM Duan, DL Deng, PRL 129, 270501(2022)

Theoretical Properties in TN-based ML

- Trainability
- Generalization ability
- Learnability
- Expressibility
- ….

Generalization bounded by Quantum NFL Theorem

Goal: Learn unitary U

Training set: $S_Q = \{(|\psi_i\rangle, U|\psi_i\rangle)\}_{i=1}^{r*t}$

Variational circuit: V_{S_Q}

Entangled training set: $\{|\phi_\alpha\rangle \rightarrow \sum_{j=1}^{r}|\psi_j\rangle \otimes |\eta_j\rangle\}_{\alpha=1}^{t}$

Error risk: $R_U(V_{S_Q}) := \frac{1}{4}\int dx \left\| U|x\rangle\langle x|U^\dagger - V_{S_Q}|x\rangle\langle x|V_{S_Q}^\dagger \right\|_1$

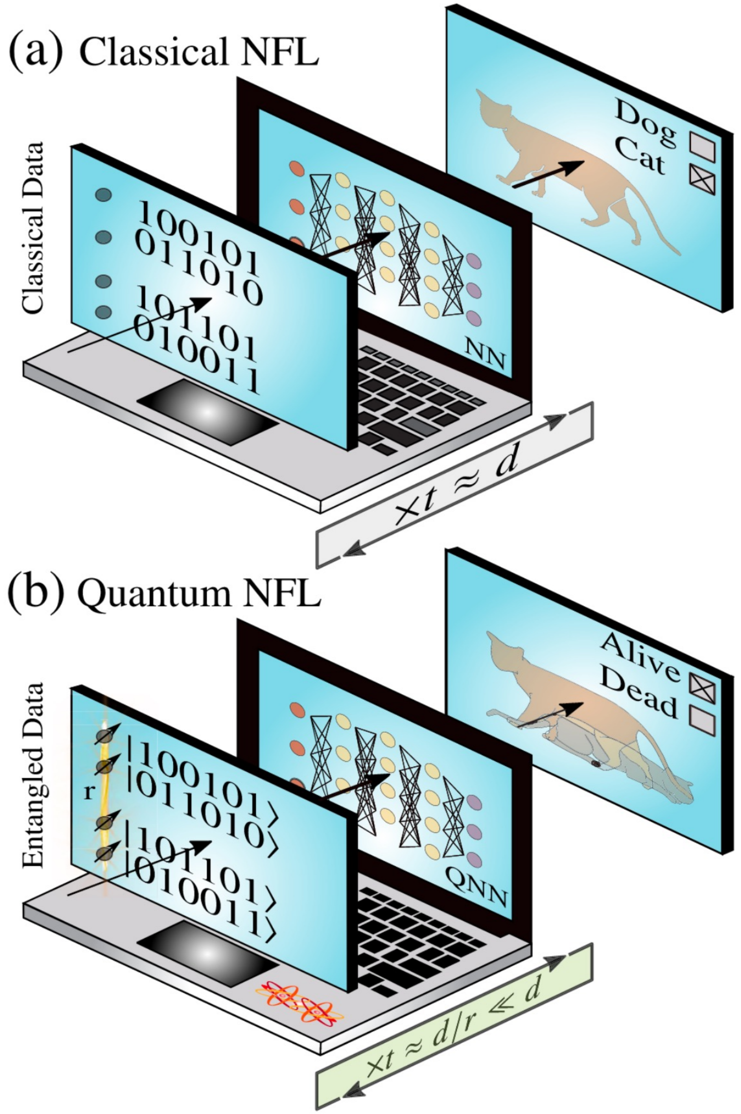
Testing set: $\{|x\rangle\}$

Generalization risk for proper learning

$$\mathbb{E}_U[\mathbb{E}_{S_Q}[R_U(V_{S_Q})]] \geq 1 - \frac{r^2t^2 + d + 1}{d(d+1)} \quad \text{(QNFL)}$$

Quantum entanglement assists to reduce the required size of training set to achieve the same lower bound of average risks.

r : Schmidt number, entangled degree with ancillary system
t : Required training samples
d: Dim. of Hilbert space
U: Target unitary
V: Learned unitary



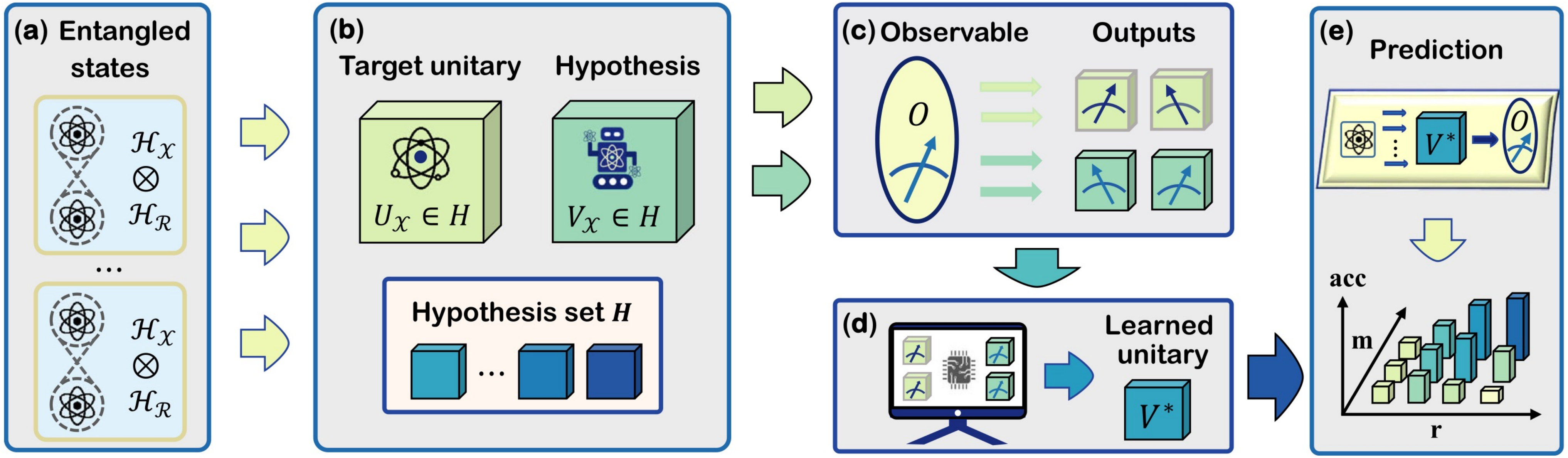
Sharma, et.al., PRL 128, 070501 (2022)
Wang, et.al., Nat. Comm. 15, 3716 (2024)

Predicting Risk Bounded by Quantum NFL Theorem

Transition role of entangled data in quantum machine learning

Goal: Learn $f_U(\psi) = \mathrm{Tr}(OU|\psi\rangle\langle\psi|U^\dagger)$ Hypothesis: $h_S(\psi) = \mathrm{Tr}(OV_S|\psi\rangle\langle\psi|V_S^\dagger)$

Error risk: $R_U(V_S) = \int d\psi (f_U(\psi) - h_S(\psi))^2$ m: Measurement times   r: Schmidt number

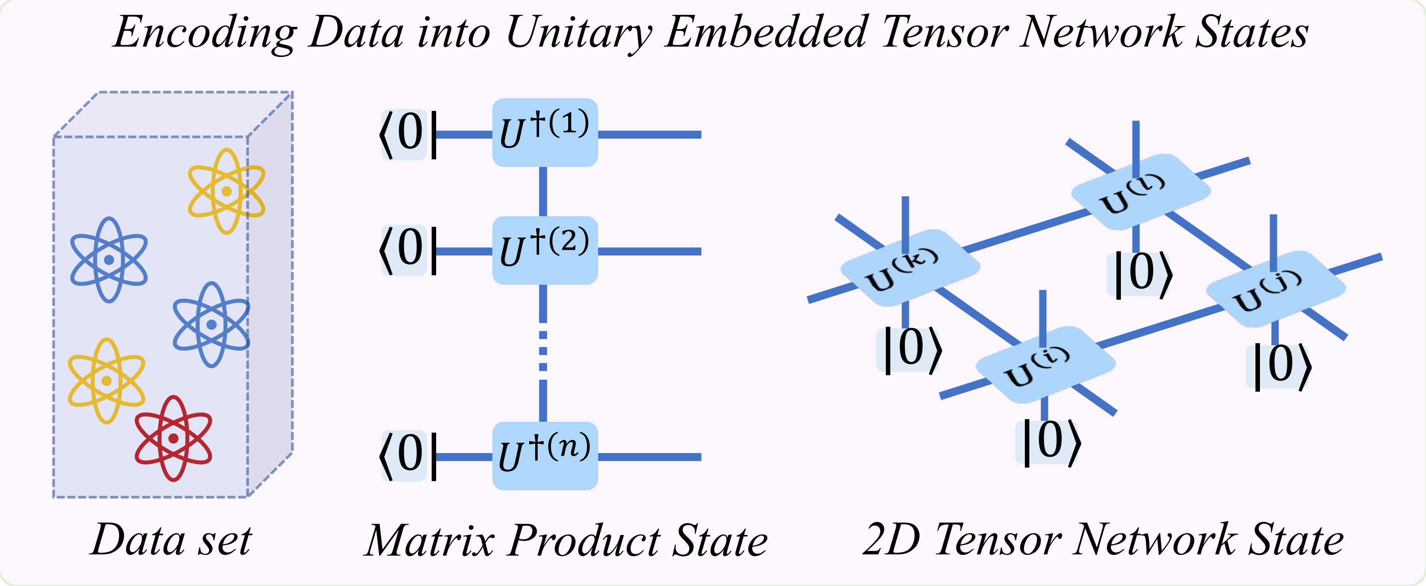


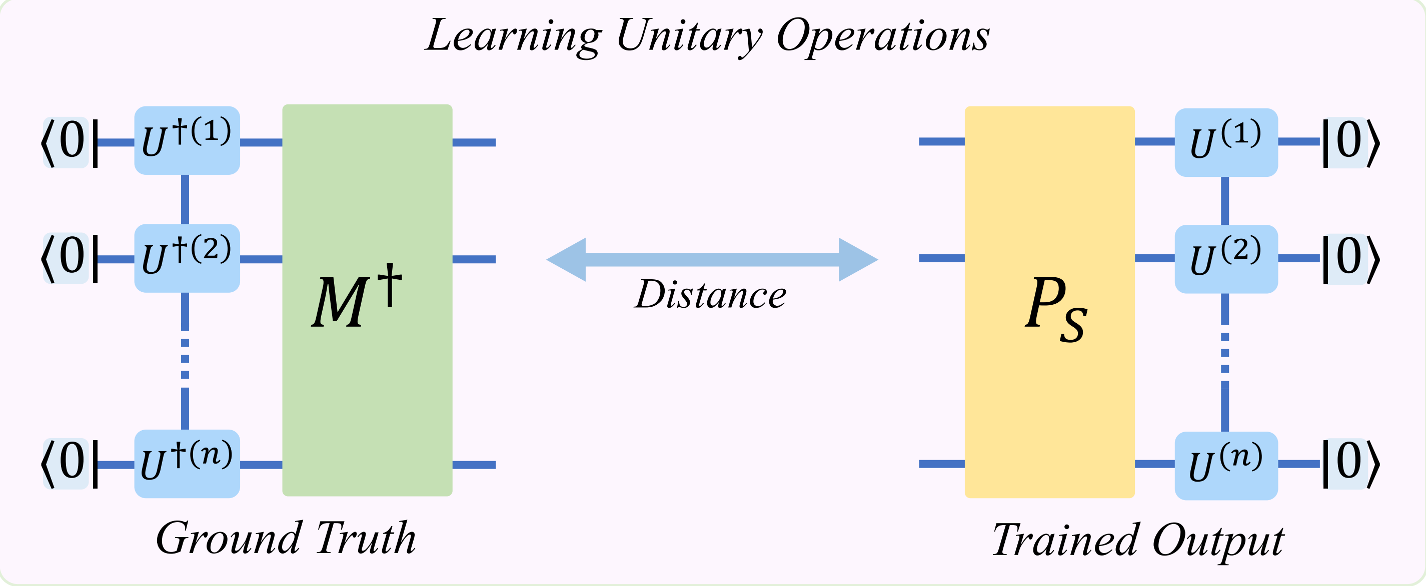
Wang, et.al., Nat. Comm. 15, 3716 (2024)

No-free-lunch theorem for TN based ML

Classical NFL theory

- ✓ averaged over all possible problems, every algorithm performs equally when applied to problems they were not specifically designed for.
- ✓ the average performance of model across all target functions dependents on the size of the input set.





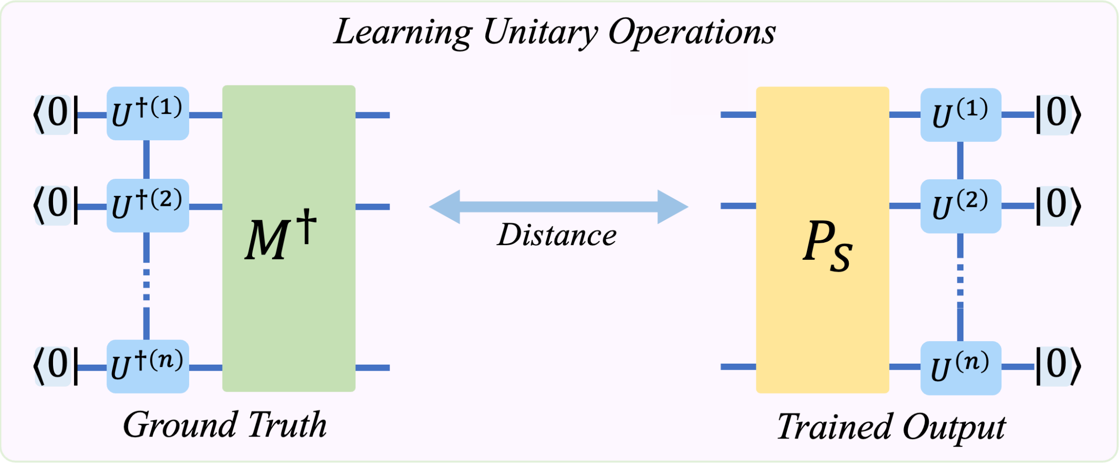
Results

No-free-lunch theorem for TN based ML

Goal: Learn unitary M Variational circuit: P_S

Training set (MPSs): $S = \{(|\psi_i\rangle, M|\psi_i\rangle)\}_{i=1}^{t}$

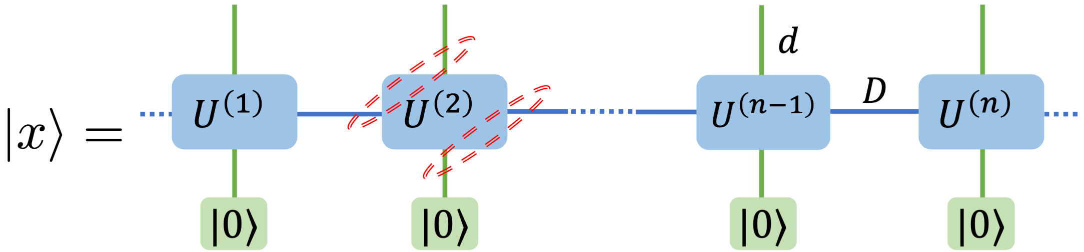
Testing set (MPSs): $\{|x\rangle\}$



Error risk:

$$R_M(P_S) \to \frac{1}{4}\int dx \left\| M|x\rangle\langle x|M^\dagger - P_S|x\rangle\langle x|P_S^\dagger \right\|_1^2$$
$$\approx 1 - \int dx \left|\langle x|M^\dagger P_S|x\rangle\right|^2 .$$

Generalization risk for proper learning

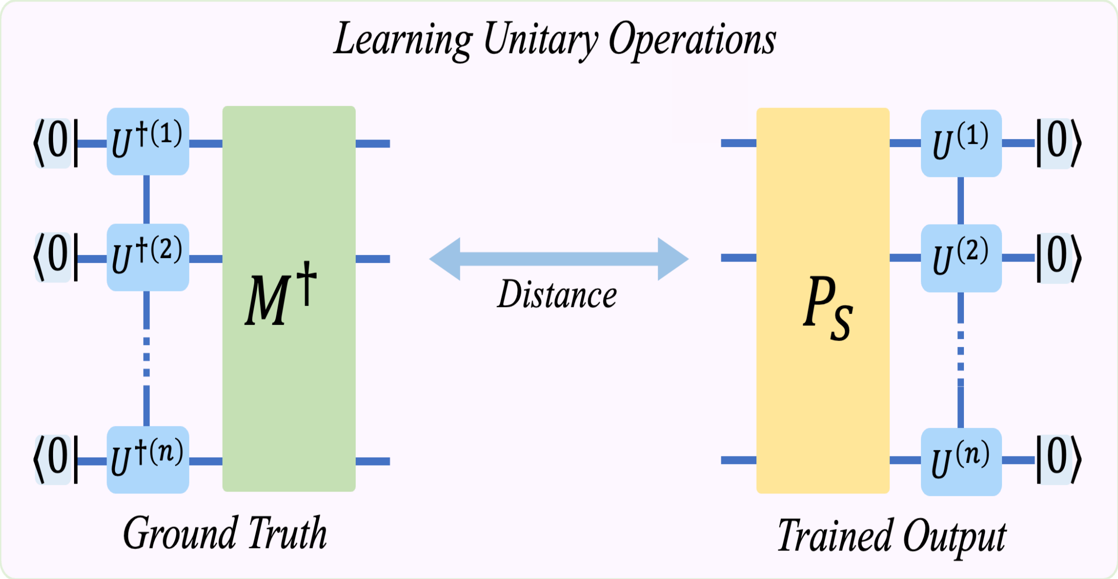


No-free-lunch theorem for TN based ML

Challenges

- Calculate the variance of random TN states. This can be mapped to calculating the partition function of a high-dimensional Ising model;

- The second one is to embed the learned information from the training set (the diagonal terms of $M^{\dagger} P_S$) into the higher-order tensor properly.



No-free-lunch theorem for TN based ML

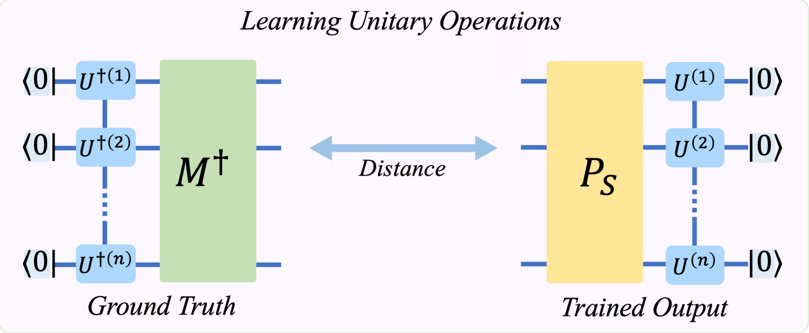
MPS

Theorem 1. *Define the error risk function $R_M(P_S)$ in Eq. (1) for learning a target n-qubit unitary M based on the input of MPSs, where P_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, one thus has the following lower bond for the averaged error risk*

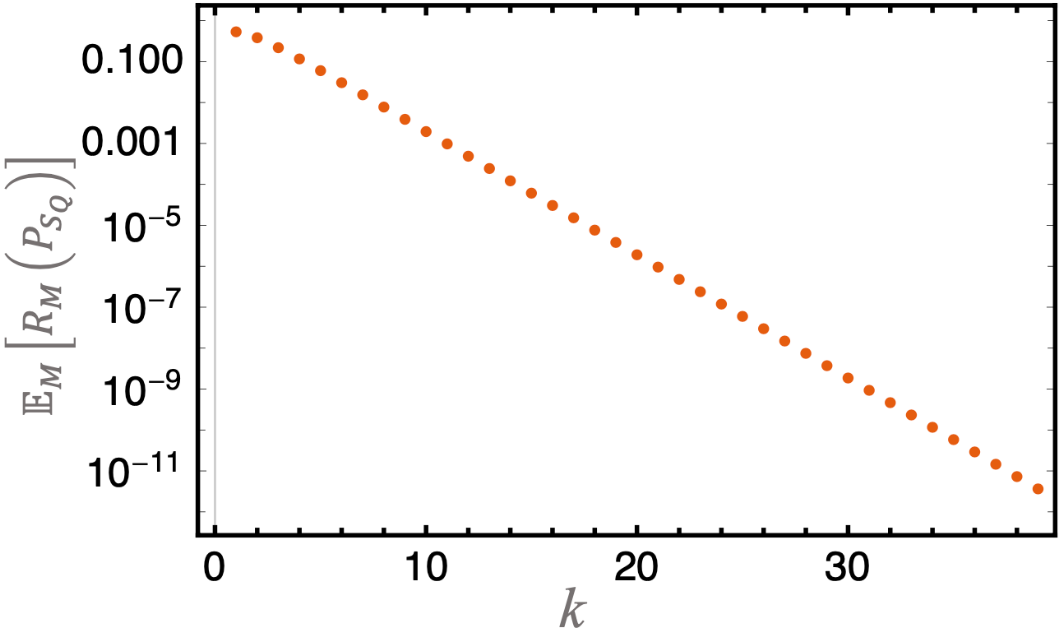
$$\begin{aligned} \mathbb{E}_{M,\mathcal{S}}\left[R_M(P_S)\right] \geq\, & 1 - \left(1 - \frac{2}{d^k}\right)\left(1 + (dAB)^n\right) \\ & - \left(\frac{1}{d^n} + \frac{1}{d^k}\right)\left(A^k + B^k\right)\left(1 + (dAB)^{n-k}\right), \end{aligned} \tag{2}$$

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

Error risk function



$$R_M(P_S) = \frac{1}{4}\int dx \left\| \left(M|x\rangle\langle x|M^\dagger - P_S|x\rangle\langle x|P_S^\dagger \right) / \langle x|x\rangle \right\|_1^2$$

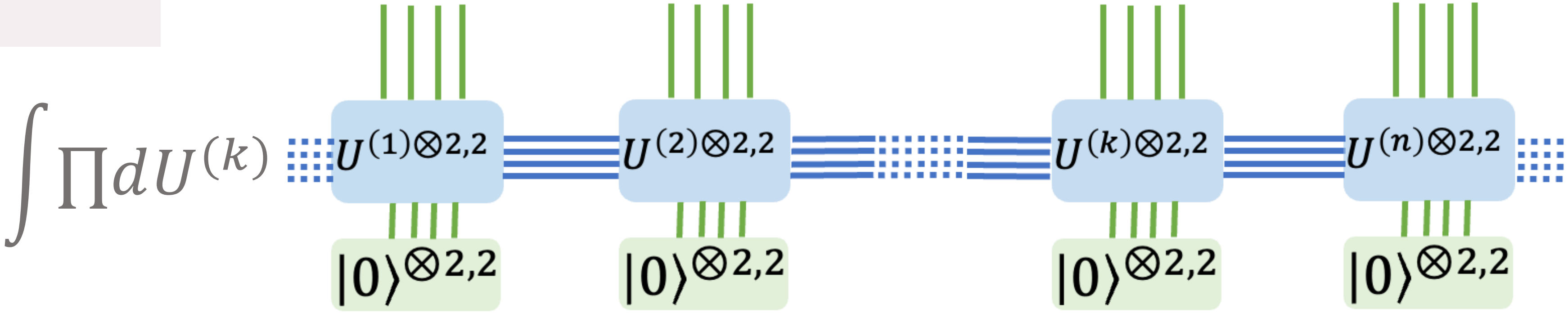


J. Wu, Q. Ye, D. Deng, **YLW**, NFL in TN based ML, to appear

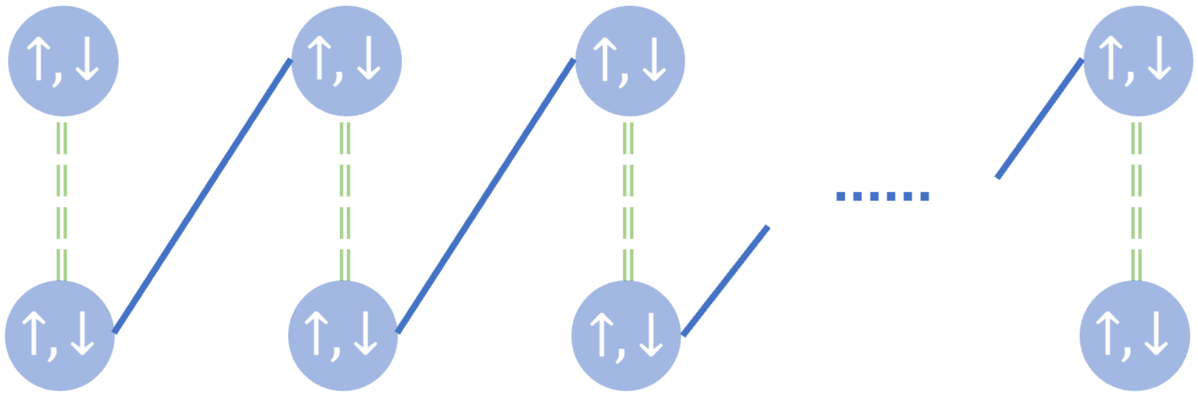
Methods

Integrating the random TNs (MPSs)

Random tensor network states



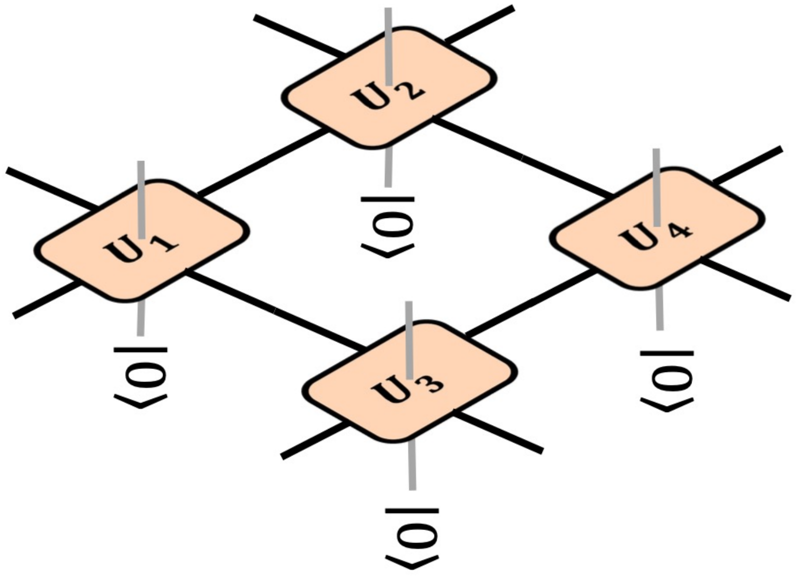
Classical Ising model
(Partition function)



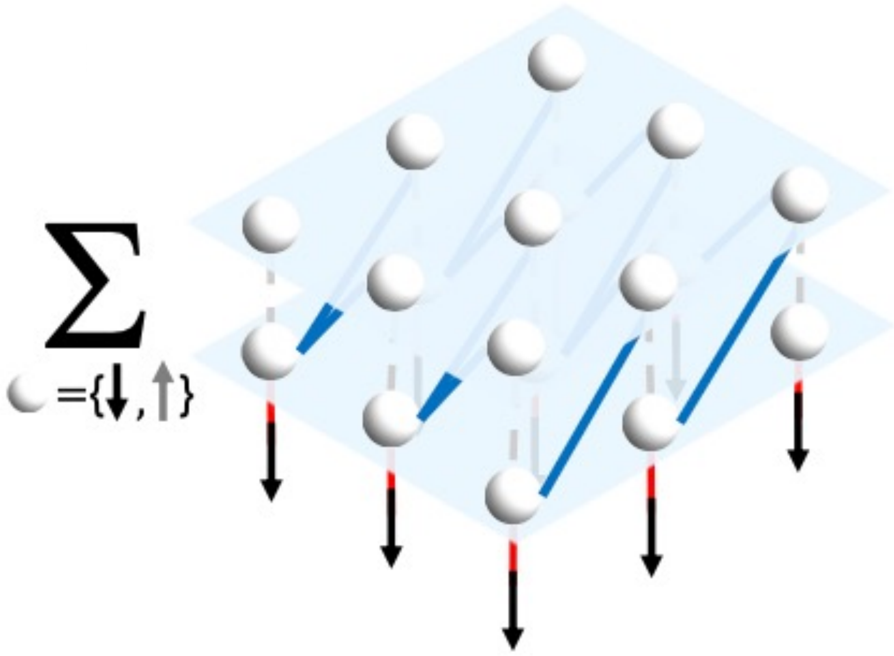
Z Liu, **YLW,** LM Duan, DL Deng, PRL 129, 270501(2022)

From 1D to higher dimension ?

PEPS



2D Ising partition

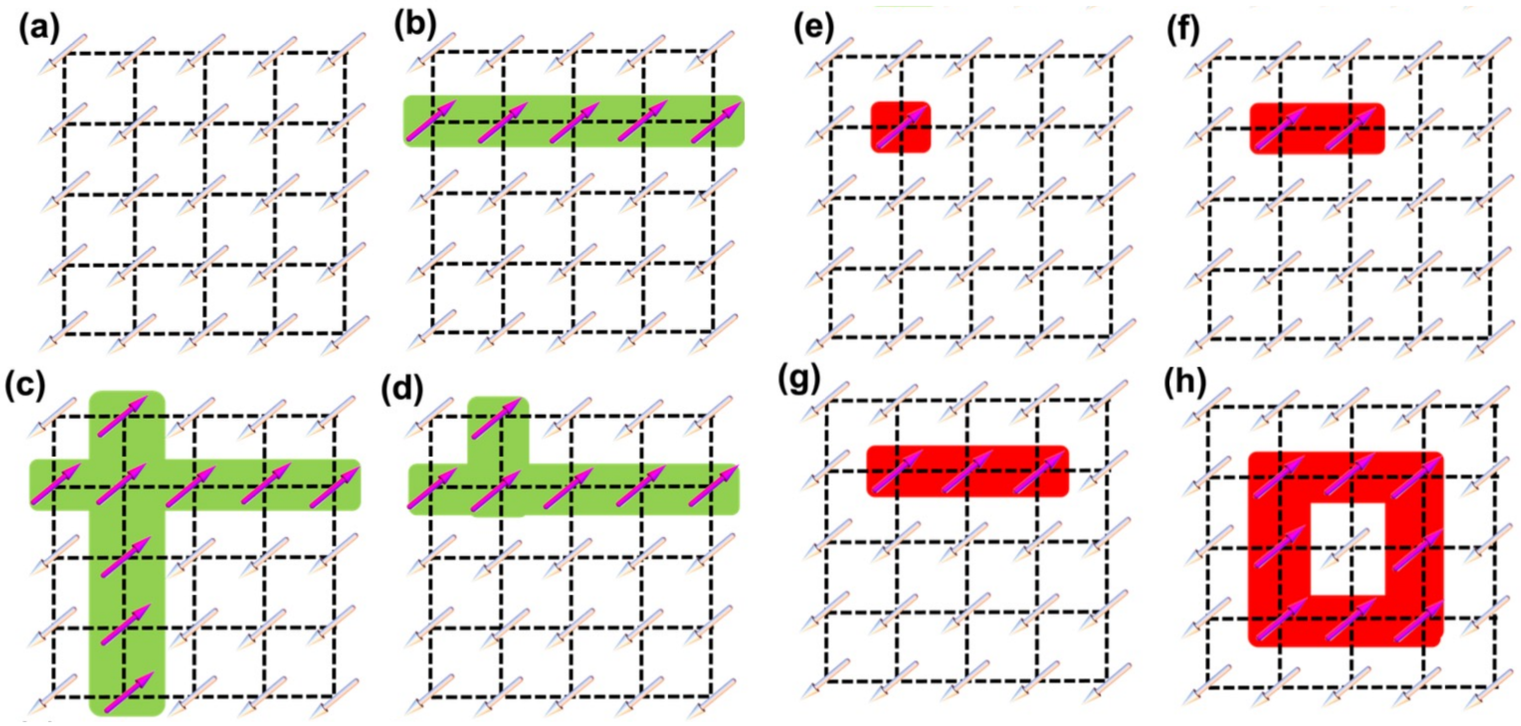


Z Liu, Q Ye, **YLW**, LM Duan, DL Deng, arXiv:2303.17452

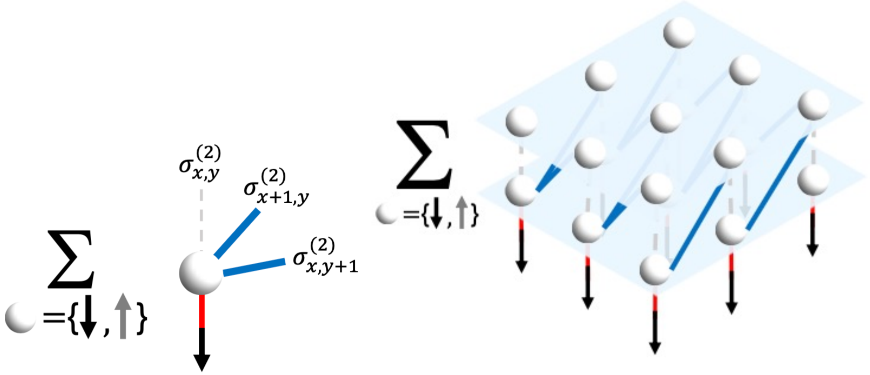
Results

From 1D to higher dimension ?

Permitted | Forbidden



Polyominos



$$f(\downarrow,\downarrow,\downarrow) = 1,$$

$$f(\downarrow,\downarrow,\uparrow) = f(\downarrow,\uparrow,\downarrow) = \frac{D^3d^2-D}{D^4d^2-1},$$

$$f(\downarrow,\uparrow,\uparrow) = \frac{D^2d^2-D^2}{D^4d^2-1},$$

$$f(\uparrow,\downarrow,\downarrow) = 0,$$

$$f(\uparrow,\downarrow,\uparrow) = f(\uparrow,\uparrow,\downarrow) = \frac{D^3d-Dd}{D^4d^2-1},$$

$$f(\uparrow,\uparrow,\uparrow) = \frac{D^4d-d}{D^4d^2-1}.$$

Z Liu, Q Ye, **YLW**, LM Duan, DL Deng, arXiv:2303.17452

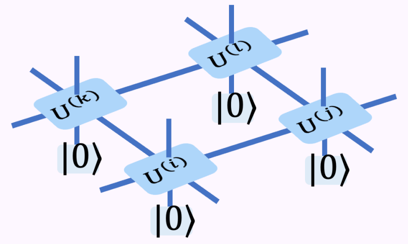
No-free-lunch theorem for TN based ML

PEPS

Theorem 2. *Define the risk function $R_M(P_S)$ in Eq. (1) for learning a target n-qubit unitary M based on the input of PEPS, where P_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 represents the physical dimension and virtual dimension of PEPS, and n denotes the qubit number of the system. One thus has the following lower bond for the averaged risk*

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

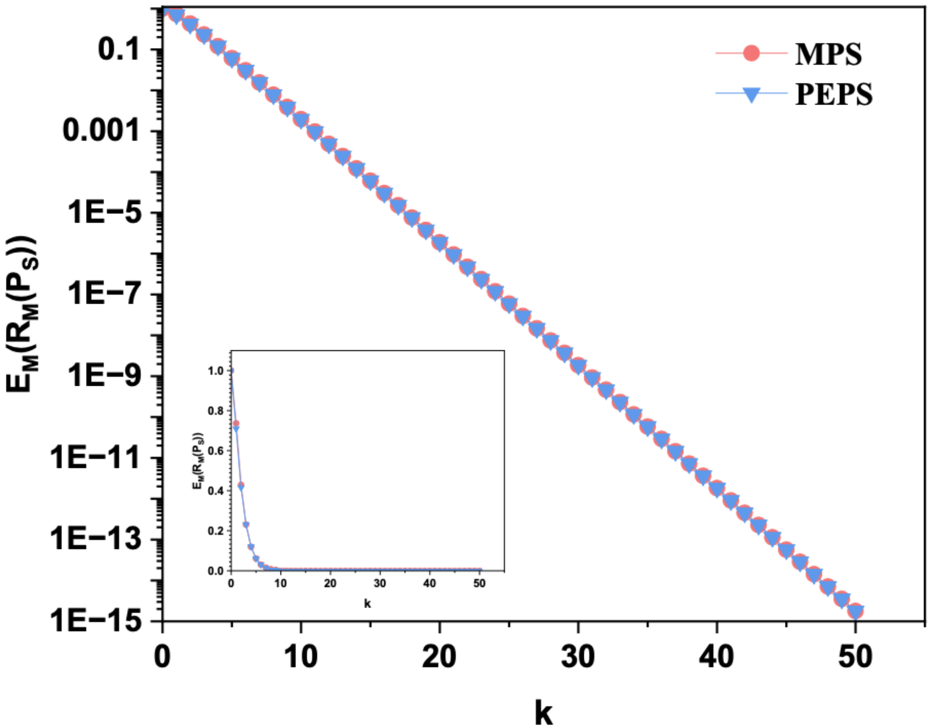
where $l = \lceil \sqrt{k} \rceil$, $G(q,p) = \frac{p}{2}\left(\sqrt{\frac{(1+q)(1+q-qp)}{1-q(2+p)+q^2(1-p)}} - 1\right)$, $L^2 = n$, *and* D *is the bond dimension of the PEPS.*



2D Tensor Network State

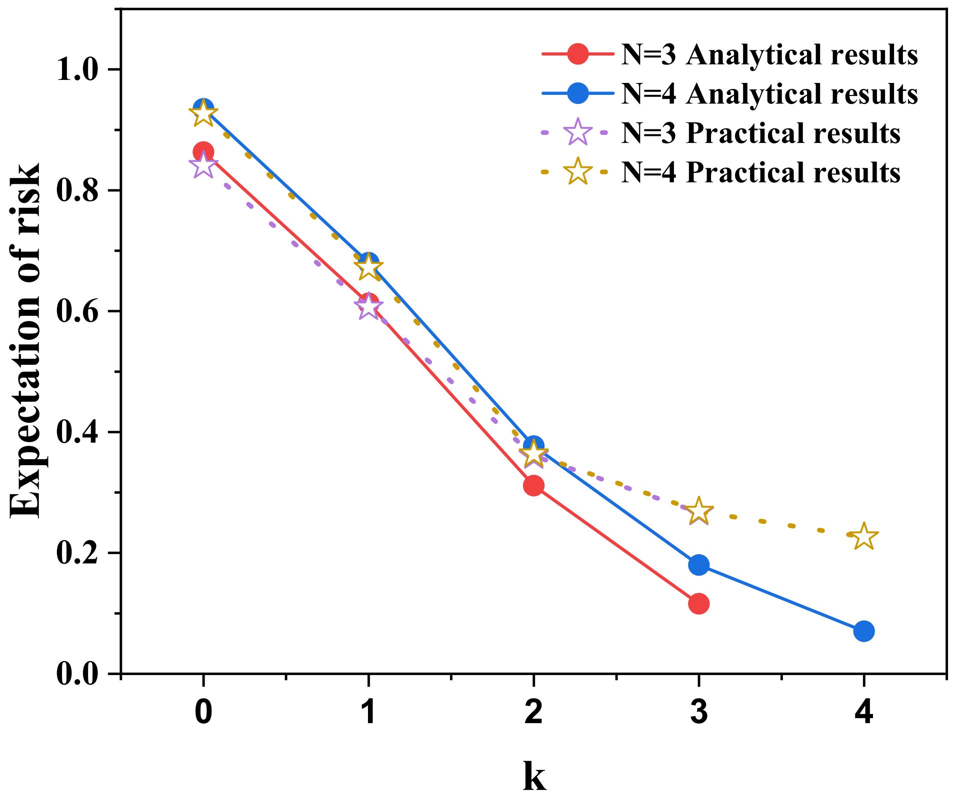
Error risk function

$$R_M(P_S) = \frac{1}{4}\int dx \left\| \left(M|x\rangle\langle x|M^\dagger - P_S|x\rangle\langle x|P_S^\dagger\right) / \langle x|x\rangle \right\|_1^2$$



J. Wu, Q. Ye, D. Deng, **YLW,** NFL in TN based ML, to appear

No-free-lunch theorem for TN based ML

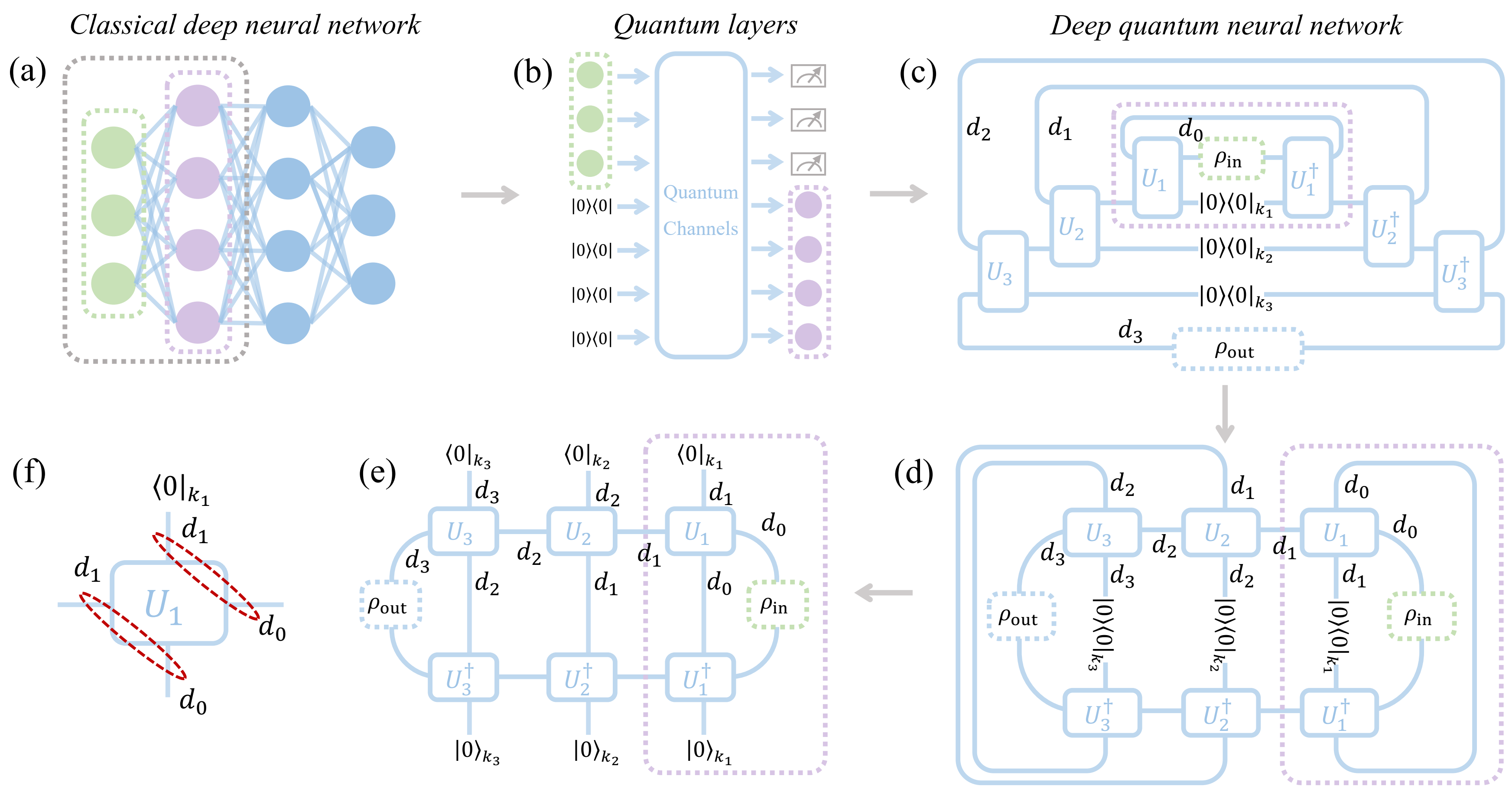


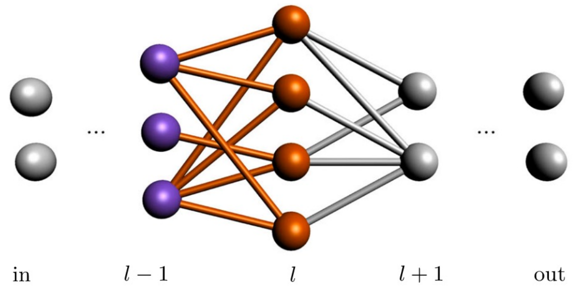
Numerical simulations based on **ITensor.jl**

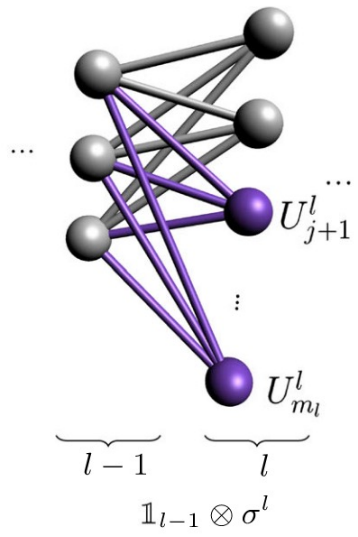
Fig. 2. Averaged risk of the trained MPS-based machine learning models with respect to the training set size $t_k = 2^n - 2^{n-k}$, where the system qubit size n varies from three to four. The solid circle lines represent the analytical lower bounds of the average risk predicted by Theorem 1. And the dashed star lines denote the average risk of the trained MPS-based machine learning models for predicting target unitaries.

J. Wu, Q. Ye, D. Deng, **YLW,** NFL in TN based ML, to appear

Apply to Deep Quantum Neural Network





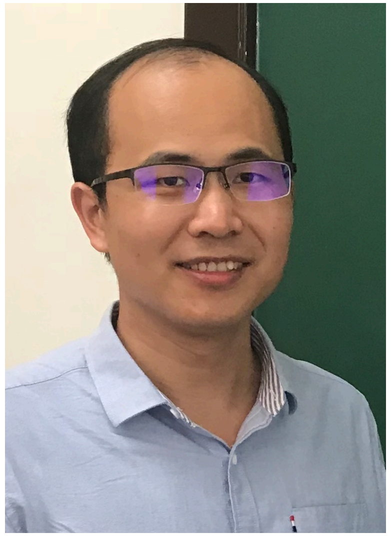


Take-home message

- ✓Challenges in rigorous formulation of NFL for TN-based ML;
- ✓Lower bound of generalization depends on both the physical and bond dimensions;
- ✓An efficient framework for studying generalization abilities of TN-based ML models;
- ✓Apply to deep QNN …

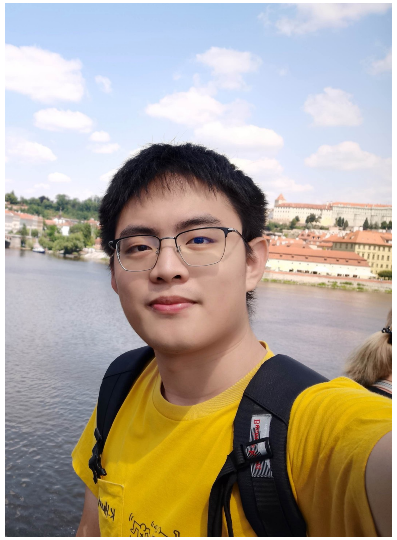
相关工作的合作者

段路明 教授
THU

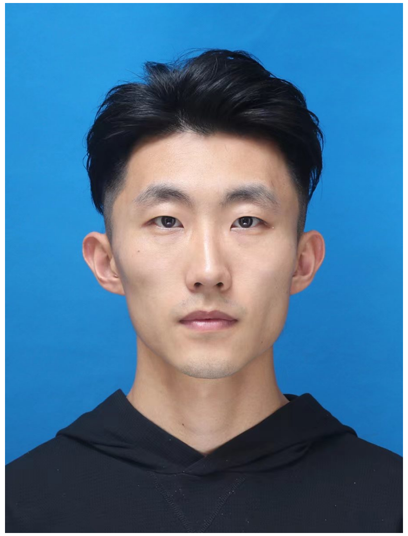


邓东灵 研究员
THU

刘子都 博士
Max-Planck



叶奇
THU&Harvard



武靖川
NKU

谢谢大家！欢迎交流合作

Email: yulw@nankai.edu.cn