

パイロット講義: 計算科学・量子計算における情報圧縮

Data Compression in Computational Science and Quantum Computing

2022.01.25

# #4 Quantum-Classical Hybrid Algorithms and Tensor Network

理学系研究科 量子ソフトウェア寄付講座

藤堂眞治 Synge Todo

Lecture materials: <https://github.com/utokyo-qsw/data-compression>

# Outline

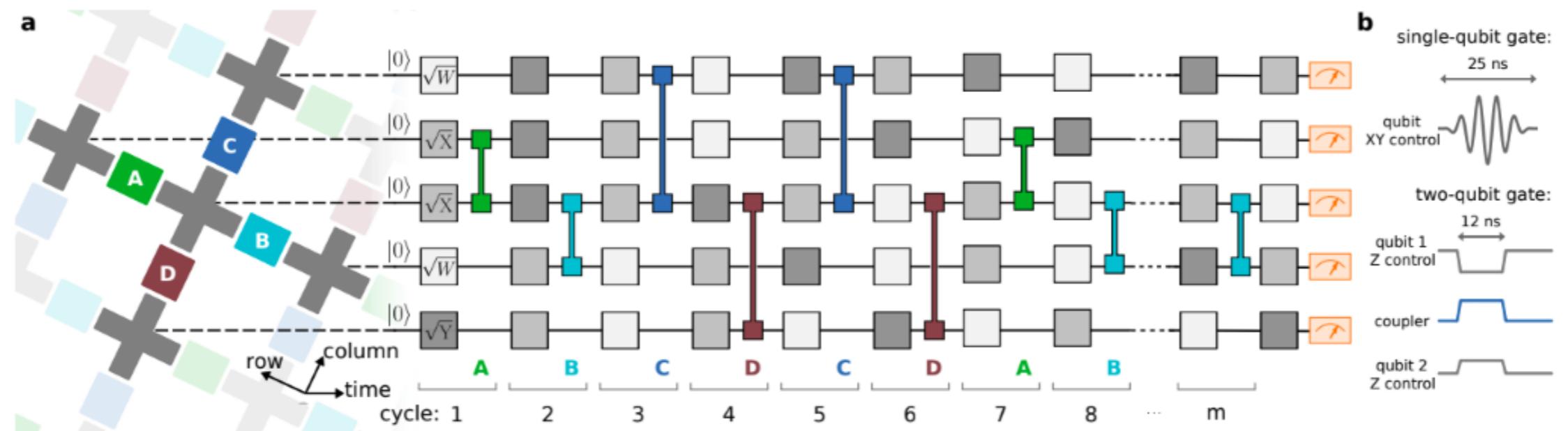
---

- Quantum Computer and Quantum Computation [review]
  - quantum circuits; quantum operations in quantum circuits; quantum measurement
- Quantum-classical Hybrid Algorithms
  - quantum computation on NISQ devices; variational quantum eigensolver; universal quantum gates; variational wave function
- Optimization of Quantum Circuit
  - gradient calculation; sequential optimization
- Machine Learning Based on Tensor Network
  - supervised learning with tensor network
- Qubit Efficient Implementation of Tensor-Network Machine Learning
  - converting MPS into quantum circuit; qubit-efficient implementation

# Quantum Computer and Quantum Computation

# Quantum Circuits

- Prepare a set of quantum bits (qubits)
- A number of quantum gates (typically one- or two-qubit gates) are applied to qubits in order
  - quantum gates are unitary operations
  - combination of quantum gates are also unitary operation
- Finally, perform measurements to extract information



Arute, F., Arya, K., Babbush, R. et al. Quantum supremacy using a programmable superconducting processor. Nature 574, 505–510 (2019). <https://doi.org/10.1038/s41586-019-1666-5>

# Typical Quantum Gates

- One-qubit gates (2x2 matrix)

- X gate (NOT)

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$



- H gate (Hadamard)

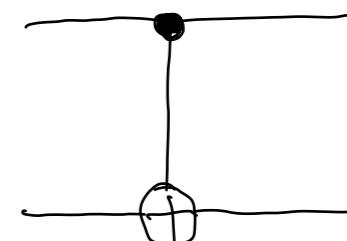
$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$



- Two-qubit gates (4x4 matrix)

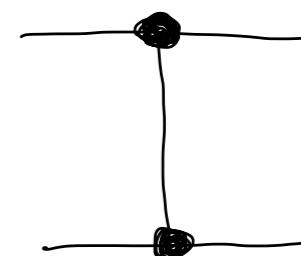
- CX gate (controlled-NOT)

$$CX = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$



- CZ gate (controlled-Z)

$$CZ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$



# Parameterized Quantum Gates

- One-qubit rotation gates (2x2 matrix)

- RX gate (rotation around X axis)

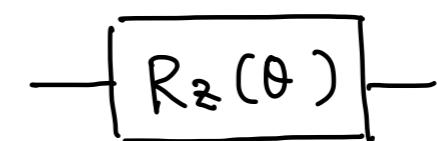
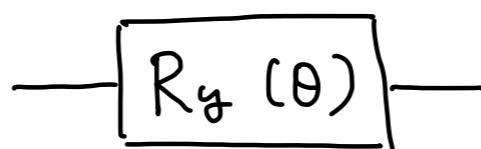
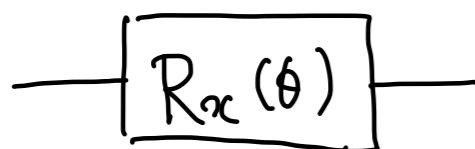
$$R_x(\theta) = e^{-i\theta\sigma_x/2} = \begin{bmatrix} \cos \frac{\theta}{2} & -i \sin \frac{\theta}{2} \\ -i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}$$

- RY gate (rotation around Y axis)

$$R_y(\theta) = e^{-i\theta\sigma_y/2} = \begin{bmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{bmatrix}$$

- RZ gate (rotation around Z axis)

$$R_z(\theta) = e^{-i\theta\sigma_z/2} = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}$$



# Quantum Operations in Quantum Circuits

---

- Parallelism
  - inputting a state of superposition produces a superposition of the corresponding outputs.
- Bifurcation
  - states bifurcate (in terms of computational basis) when H-gates, etc. are applied
- Interference
  - superposition coefficients of the states are complex, and they may cancel each other out and vanish
- Collapse
  - collapse to one of the states in the basis used for the measurement

# Quantum Measurement

- Pure state
  - in a pure state, entropy is zero
    - density matrix:  $\rho = |\Psi\rangle\langle\Psi|$  is rank-1
    - the largest eigenvalue is unity, all other eigenvalues are zero
    - Von Neumann entropy:  $S(\rho) = - \text{tr} (\rho \log \rho) = 0$
  - Manipulation of states by quantum circuits
    - entropy does not change since operation are unitary transformations
      - if the initial state is a pure state, the entropy stays zero
      - $S(U\rho U^\dagger) = - \text{tr} (U\rho U^\dagger \log U\rho U^\dagger) = - \text{tr} (\rho \log \rho) = S(\rho) = 0$



# Quantum Measurement

- Non-selective projective measurement (= measure and don't see the result)
  - entropy always increases
    - projection operators  $\{P_i\}$  ( $\sum P_i = 1, (P_i)^2 = P_i$ )
    - by projective measurement:  $\rho \rightarrow \rho' = \sum_{\geq} P_i \rho P_i$
    - $S(\rho') \geq S(\rho)$  (Klein's theorem)
  - a pure state is converted into a mixed state
    - off-diagonal elements vanishes
- Selective projective measurement (measure and see the result)
  - a pure state is sampled according to the measurement probability
  - selection causes the state collapse and the entropy to be zero again

# Essence of Quantum Algorithm

- Prepare a superposition of many states, using Hadamard gates, etc.
  - if non-selective projective measurements is performed at this stage, the entropy is extensive (proportional to the number of qubits)
- Manipulate and interfere the states so that desired answer has large amplitude
  - entropy of the quantum state remains zero
    - interference reduces the entropy after non-selective projective measurement
    - a kind of data compression?
- Selective projective measurement (measure and see the result) to get the correct answer with high probability
  - sampling from the state with entropy as small as possible
    - a kind of information extraction?

# Quantum-classical Hybrid Algorithms

# Quantum Computation on NISQ Devices

---

- Fault-tolerant quantum computing with quantum error correction
  - quantum phase estimation
    - prime factorization, eigenvalue problems, simultaneous linear equations (HHL)
    - exponential speed improvement over classical computers
  - many *physical* qubits (100-1000) are needed for one *logical* qubits
- Today's NISQ (noisy intermediate-scale quantum) devices
  - only 50-100 qubits (not enough to implement quantum error correction)
  - at most 1000 quantum operations (due to large noise)

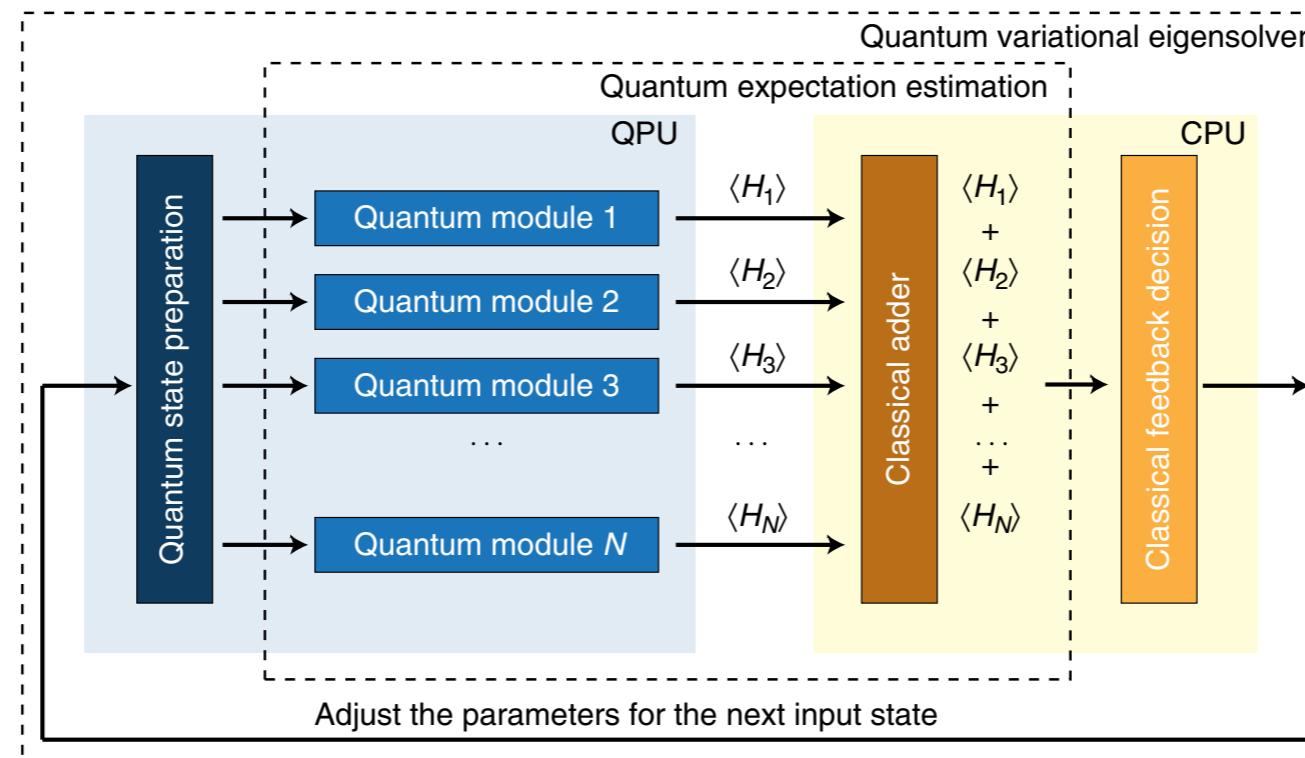
# Quantum-classical Hybrid Algorithms

---

- Combination of quantum and classical computation
  - evaluate "loss" by using quantum computer
  - optimize parameters of quantum circuit by using classical computer
- Examples of quantum-classical hybrid algorithms
  - variational quantum eigensolver (VQE)
  - quantum approximate optimization algorithm (QAOA)
  - variational quantum simulator (VQS)
  - variational quantum linear solver (VQLS)
  - quantum circuit learning (QCL)
  - ...

# Variational Quantum Eigensolver

- Calculate ground state of quantum many-body system
  - represent variational wave function by parameterized quantum circuit
  - evaluate expectation value of Hamiltonian by using quantum computer
  - optimize parameters by using classical computer



$$\langle H_1 \rangle = \langle \sigma_1^x \sigma_2^x \rangle$$

$$\langle H_2 \rangle = \langle \sigma_1^y \sigma_2^y \rangle$$

$$\langle H_3 \rangle = \langle \sigma_1^z \sigma_2^z \rangle$$

- Peruzzo, A., et al (2014). A variational eigenvalue solver on a photonic quantum processor. *Nature Comm.* 5, 4213. <https://doi.org/10.1038/ncomms5213> (arXiv:1304.3061)

# An Example - Spin Dimer



- Interacting two spins

- Hamiltonian:  $H = JS_1 \cdot S_2 = \frac{J}{4}(\sigma_1^x\sigma_2^x + \sigma_1^y\sigma_2^y + \sigma_1^z\sigma_2^z)$  ( $J > 0$ )

- ground state:  $|\Psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$

$$E = -\frac{3}{4}J$$

*singlet*

$$\begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \begin{matrix} 00 \\ 01 \\ 10 \\ 11 \end{matrix}$$

- Variational wave function:  $|\phi(\{\theta_i\})\rangle = U(\{\theta_i\})|00\rangle$

- Expectation value of Hamiltonian (energy):

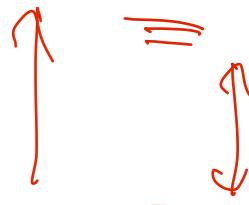
$$\langle \phi(\{\theta_i\}) | \underline{H} | \phi(\{\theta_i\}) \rangle = \frac{J}{4} (\langle \phi(\{\theta_i\}) | \underline{\sigma_1^x\sigma_2^x} | \phi(\{\theta_i\}) \rangle + \langle \phi(\{\theta_i\}) | \sigma_1^y\sigma_2^y | \phi(\{\theta_i\}) \rangle + \langle \phi(\{\theta_i\}) | \underline{\sigma_1^z\sigma_2^z} | \phi(\{\theta_i\}) \rangle)$$

*triplet*

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$S_1 = \frac{1}{2} (\sigma_1^x, \sigma_1^y, \sigma_1^z)$$

$$H = \frac{J}{4} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



$$E = \frac{J}{4}$$

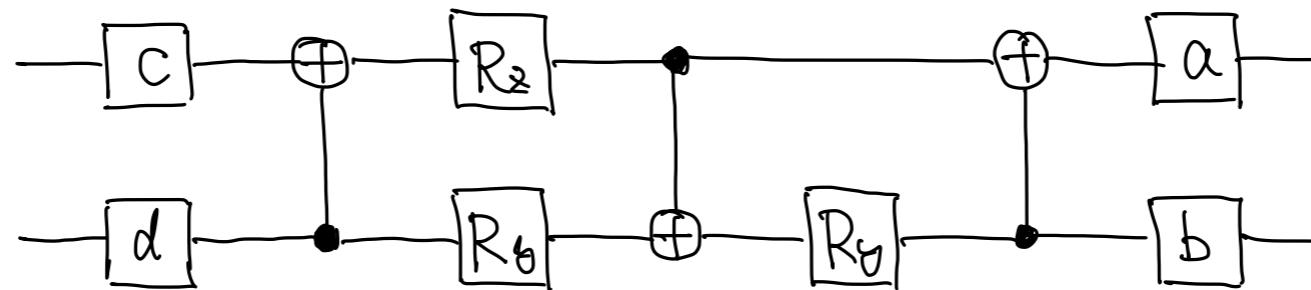
# Universal Quantum Gates

- *Universal* one-qubit unitary operation

$$U = R_x(\theta_1)R_y(\theta_2)R_x(\theta_3)$$

R

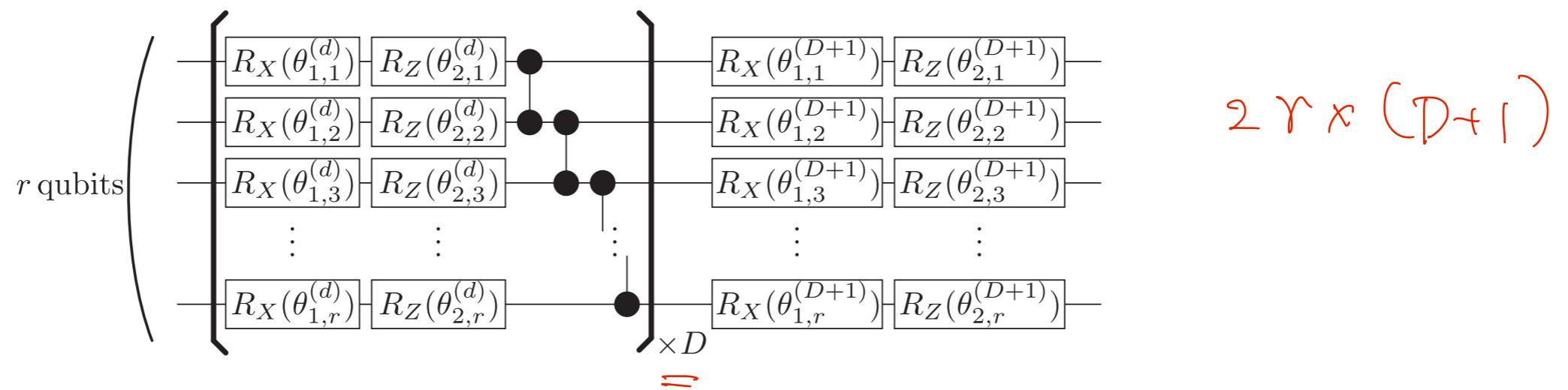
- *Universal* two-qubit unitary operation



- [a]-[d]: universal one-qubit unitary operations
- 15 one-qubit rotations + 3 CNOTs
- Shende, V. V., Markov, I. L., & Bullock, S. S. (2004). *Minimal universal two-qubit controlled-NOT-based circuits*. Phys, Rev. A, 69, 1. <https://doi.org/10.1103/PhysRevA.69.062321> (arXiv:quant-ph/0308033)

# Variational Wave Function

- Universal multi-qubit unitary operation
  - number of elemental (one- and two-qubit) operations increases rapidly → the number of parameters that must be optimized also increases
- In practical VQE algorithms
  - use some heuristic parameterized quantum circuits, e.g.,

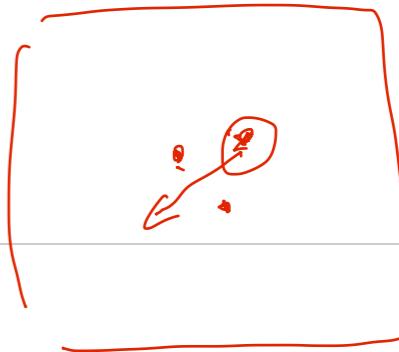


- For quantum chemical calculations
  - various high-precision variational wave functions have been proposed
  - ref) 杉崎「量子コンピュータによる量子化学計算入門」(講談社, 2020)

# Optimization of Quantum Circuit

# Optimization of Quantum Circuit

- Optimization is performed on classical computer
- Standard optimization algorithms can be used
  - gradient-free algorithms
    - Nelder-Mead, Powell method
  - optimization using gradient
    - gradient descent, Adam, etc
- Gradient-based algorithms generally show better performance than gradient-free methods
  - How to calculate gradient of cost function using quantum computer?



# Gradient Calculation

- Finite difference method

- sensitive to noise

- Using special property of loss function

- fix  $(\theta_1, \theta_2, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_n)$

- unitary transformation for generating variational wave function can be  $\mathcal{V}$   
written as

$$\frac{\partial C}{\partial \theta} \approx \frac{C(\theta + \Delta) - C(\theta - \Delta)}{2\Delta}$$

$$U(\theta_k) = \underbrace{V U_k(\theta_k) W}_{\sigma^2 \delta^2}$$

$$U_k(\theta_k) = \underbrace{e^{-i\sigma_k \theta_k / 2}}$$

- expectation value of operator  $\mathcal{O}$   $\sigma^2 \delta^2$

$$\begin{aligned}
 \langle \mathcal{O}(\theta_k) \rangle &= \langle 0 | U^\dagger(\theta_k) \mathcal{O} U(\theta_k) | 0 \rangle \\
 &= \langle 0 | W^\dagger U_k^\dagger(\theta_k) V^\dagger \mathcal{O} V U_k(\theta_k) W | 0 \rangle \\
 &= \langle \Psi | \underbrace{U_k^\dagger(\theta_k) Q}_{\sigma^2 \delta^2} \underbrace{U_k(\theta_k)}_{\sigma^2 \delta^2} | \Psi \rangle \quad |\Psi\rangle = W|0\rangle, Q = V^\dagger \mathcal{O} V
 \end{aligned}$$

# Derivative of Cost Function

- Using  $U_k(\theta_k) = e^{-i\sigma_k \theta_k / 2}$

$$\frac{\partial}{\partial \theta_k} \langle \mathcal{O}(\theta_k) \rangle = \frac{i}{2} \langle \Psi | \underline{\sigma_k} U_k^\dagger(\theta_k) Q U_k(\theta_k) | \Psi \rangle - \frac{i}{2} \langle \Psi | U_k^\dagger(\theta_k) Q U_k(\theta_k) \underline{\sigma_k} | \Psi \rangle \quad \text{A}$$

- On the other hand, since  $(\sigma_k)^2 = I$

$$\underline{e^{i\sigma_k \theta/2}} = I \cos \frac{\theta}{2} + i\sigma_k \sin \frac{\theta}{2} \quad \Rightarrow \quad e^{\pm i\sigma_k \pi/4} = \boxed{\frac{\sqrt{2}}{2} (I \pm i\sigma_k)}$$

even      odd.

- Stable derivative estimation

$$\frac{\partial}{\partial \theta_k} \langle \mathcal{O}(\theta_k) \rangle = \frac{1}{2} \left( \overbrace{\langle \mathcal{O}(\theta_k + \frac{\pi}{2}) \rangle}^{\text{U}_k(\theta_k)} - \overbrace{\langle \mathcal{O}(\theta_k - \frac{\pi}{2}) \rangle}^{\text{U}_k(\theta_k)} \right)$$

$$e^{-i\sigma_k (\theta_k + \frac{\pi}{2}) / 2} = \underbrace{e^{-i\sigma_k \theta_k / 2}}_{\text{U}_k(\theta_k)} e^{-i\sigma_k \frac{\pi}{4}}$$

- Mitarai, K., Negoro, M., Kitagawa, M., & Fujii, K. (2018). *Quantum circuit learning*. Physical Review A, 98, 1. <https://doi.org/10.1103/PhysRevA.98.032309>  
(arXiv:1803.00745)

# Sequential Optimization

$$\nabla \mathcal{C}(\theta_1, \dots, \theta_k, \underline{\theta_m})$$

- Using  $e^{i\sigma_k\theta/2} = I \cos \frac{\theta}{2} + i\sigma_k \sin \frac{\theta}{2}$

$$\begin{aligned}\langle \mathcal{O}(\theta'_k) \rangle &= \langle \Psi | (I \cos \frac{\theta'_k}{2} + i\sigma_k \sin \frac{\theta'_k}{2}) Q (I \cos \frac{\theta'_k}{2} + i\sigma_k \sin \frac{\theta'_k}{2}) | \Psi \rangle \\ &= a_1 \cos(\theta'_k - a_2) + a_3\end{aligned}$$

- Coefficients  $a_1, a_2, a_3$  can be determined from  $\langle \mathcal{O}(\theta_k) \rangle$ ,  $\langle \mathcal{O}(\theta_k + \frac{\pi}{2}) \rangle$ ,

$$\langle \mathcal{O}(\theta_k - \frac{\pi}{2}) \rangle$$

- $\langle \mathcal{O}(\theta'_k) \rangle$  can be minimized by choosing  $\theta'_k = a_2$  or  $\theta'_k = -a_2$  (depending on the sign of  $a_1$ )

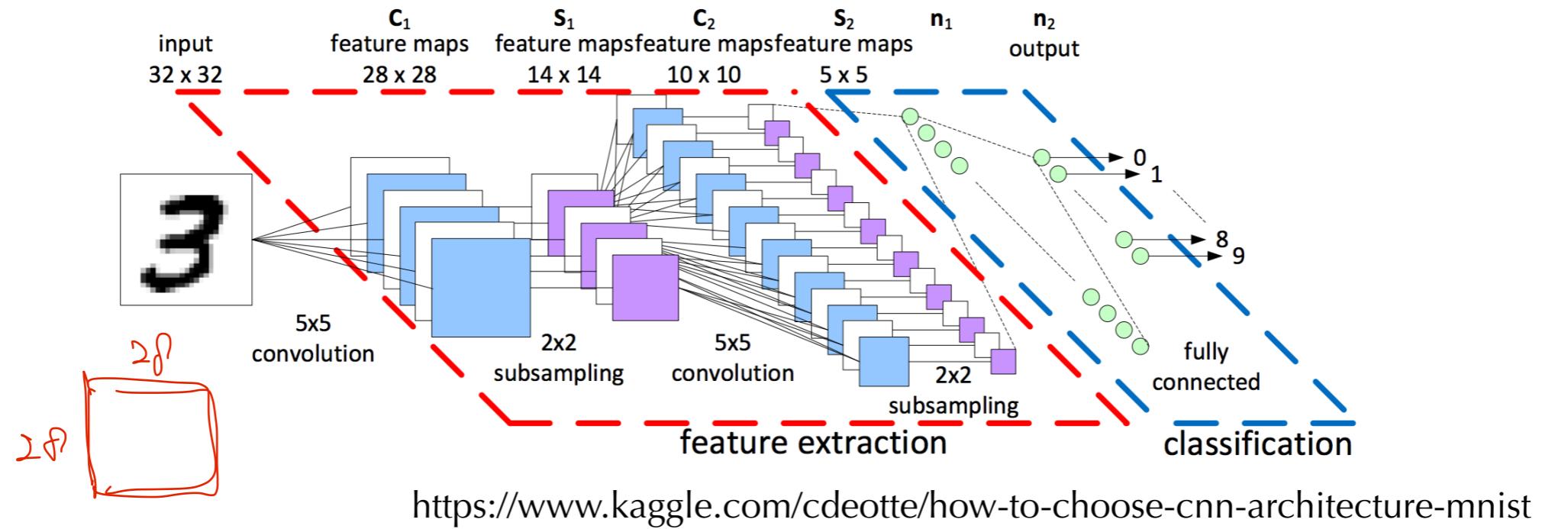
$$a_2 + \pi$$

- Nakanishi, K. M., Fujii, K., & Todo, S. (2020). Sequential minimal optimization for quantum-classical hybrid algorithms. Phys. Rev. Research, 2, 1. <https://doi.org/10.1103/PhysRevResearch.2.043158> (arXiv:1903.12166)

# Machine Learning Based on Tensor Network

# Handwritten Character Recognition

- Machine learning based on convolutional neural network (CNN)

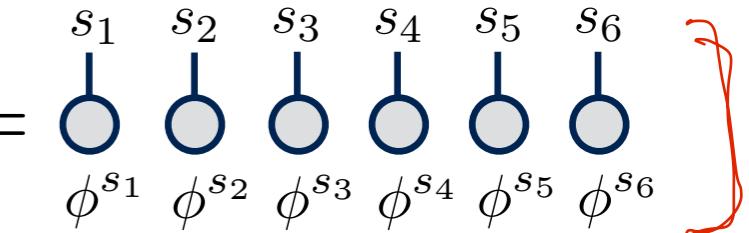


- input: gray scale image
- output: 10-dimensional vector
- Accuracy > 99.5% for MNIST
- c.f. Gao, Z. F., Cheng, S., He, R. Q., Xie, Z. Y., Zhao, H. H., Lu, Z. Y., & Xiang, T. (2020). Compressing deep neural networks by matrix product operators. *Phys. Rev. Research*, 2(2), 23300. <https://doi.org/10.1103/PhysRevResearch.2.023300> (arXiv:1904.06194)

# Supervised Learning with Tensor Networks

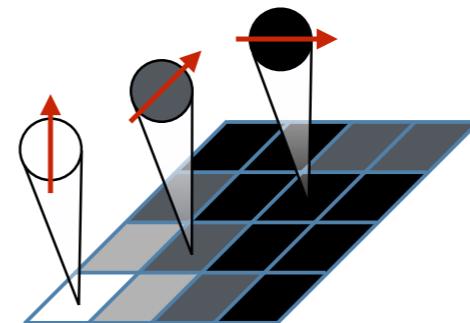
- Tensor network has multi-linear property
  - how to introduce non-linearity and increase expressive power?
- Encode of input data
  - convert  $N$ -dimensional vector ( $N$ -pixel image) into  $2^N$ -dimensional feature map  $28 \times 28$

$$\Phi^{s_1 s_2 \cdots s_N}(\mathbf{x}) = \phi^{s_1}(x_1) \otimes \phi^{s_2}(x_2) \otimes \cdots \otimes \phi^{s_N}(x_N)$$



- $\Phi(x)$  is a tensor product of (2-dimensional) local feature map

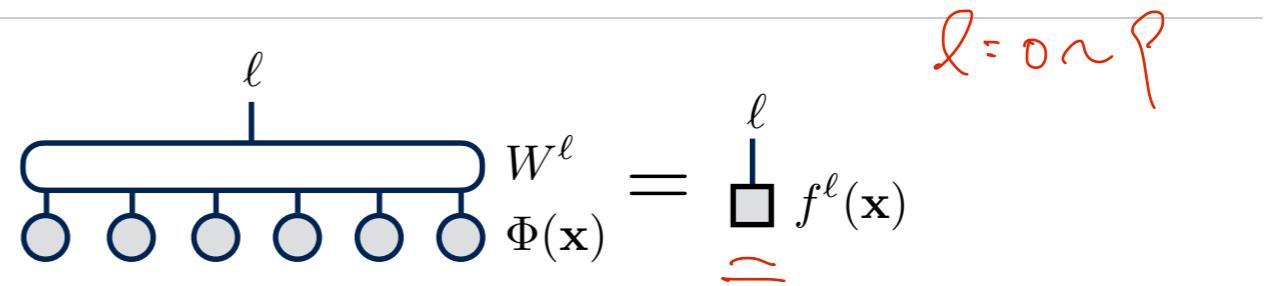
$$\phi^{s_j}(x_j) = \left[ \cos\left(\frac{\pi}{2}x_j\right), \sin\left(\frac{\pi}{2}x_j\right) \right]$$



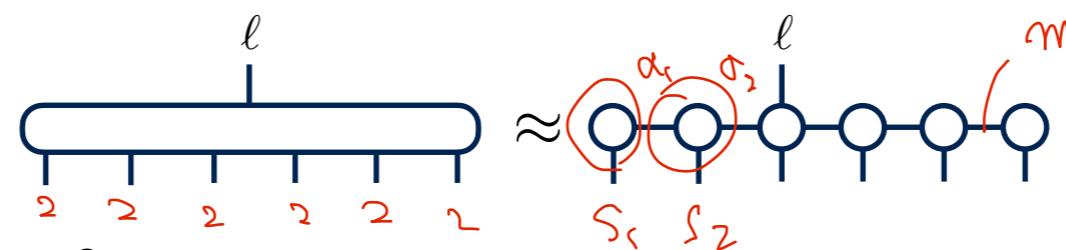
- Stoudenmire, E. M., & Schwab, D. J. (2016). Supervised Learning with Quantum-Inspired Tensor Networks. *Advances in Neural Information Processing Systems*, 29, 4799. (arXiv:1605.05775)

# Classification Model

- $f^\ell(\mathbf{x}) = \underbrace{W^\ell}_{\text{800}} \cdot \Phi(\mathbf{x})$
- $W$  is a huge  $(10 \times 2^N)$  matrix
- Decompose  $W$  into tensor network (matrix product)



$$W_{s_1 s_2 \dots s_N}^\ell = \sum_{\{\alpha\}} A_{s_1}^{\alpha_1} A_{s_2}^{\alpha_1 \alpha_2} \dots A_{s_j}^{\ell; \alpha_j \alpha_{j+1}} \dots A_{s_N}^{\alpha_{N-1}}$$



- each tensor has  $2m^2$  elements ( $m$ : bond dimension between tensors)

$$2m^2 \ll 2^{N \cdot 10}$$

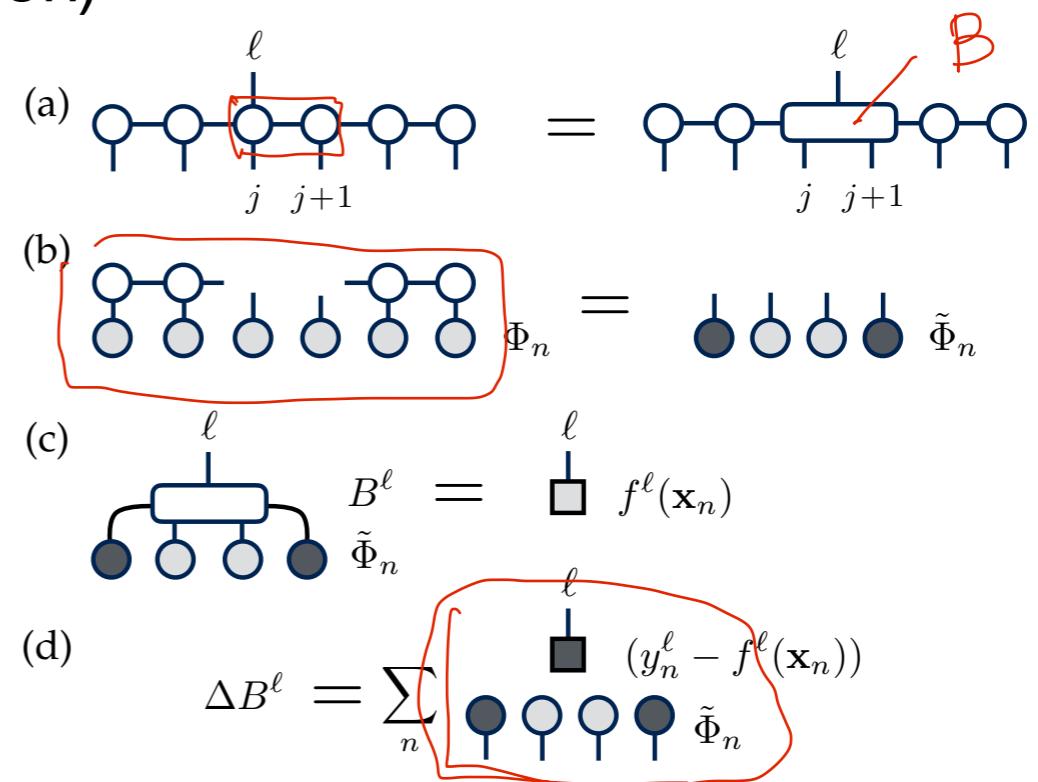
- Stoudenmire, E. M., & Schwab, D. J. (2016). Supervised Learning with Quantum-Inspired Tensor Networks. *Advances in Neural Information Processing Systems*, 29, 4799. (arXiv:1605.05775)

# Cost Function and Gradient

- Cost:  $C = \frac{1}{2} \sum_{n=1}^{N_T} \sum_{\ell} (\underline{f^{\ell}(\mathbf{x}_n)} - \underline{y_n^{\ell}})^2$ 
    - $N_T$ : number of training data
    - $f^{\ell}(\mathbf{x}_n)$ : contraction result of tensor network (10-dimensional vector)
    - $y_n^{\ell}$ : correct label (one-hot representation)

- Gradient:

$$\Delta B^\ell = -\frac{\partial C}{\partial B^\ell} = \sum_{n=1}^{N_T} (y_n^\ell - f^\ell(\mathbf{x}_n)) \tilde{\Phi}_n$$



- Stoudenmire, E. M., & Schwab, D. J. (2016). Supervised Learning with Quantum-Inspired Tensor Networks. *Advances in Neural Information Processing Systems*, 29, 4799. (arXiv:1605.05775)

# Qubit-efficient Implementation of Tensor-network Machine Learning

# Tensor Network Representation of Quantum Circuits

- Tensor representation of quantum gates and states

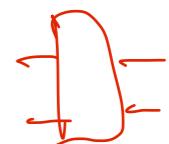
- 1-qubit gate:

- $2 \times 2$  matrix  $\rightarrow$  2-leg tensor



- 2-qubit gate:

- $4 \times 4$  matrix  $\rightarrow 2 \times 2 \times 2 \times 2$  tensor  $\rightarrow$  4-leg tensor



- $r$ -qubit gate:

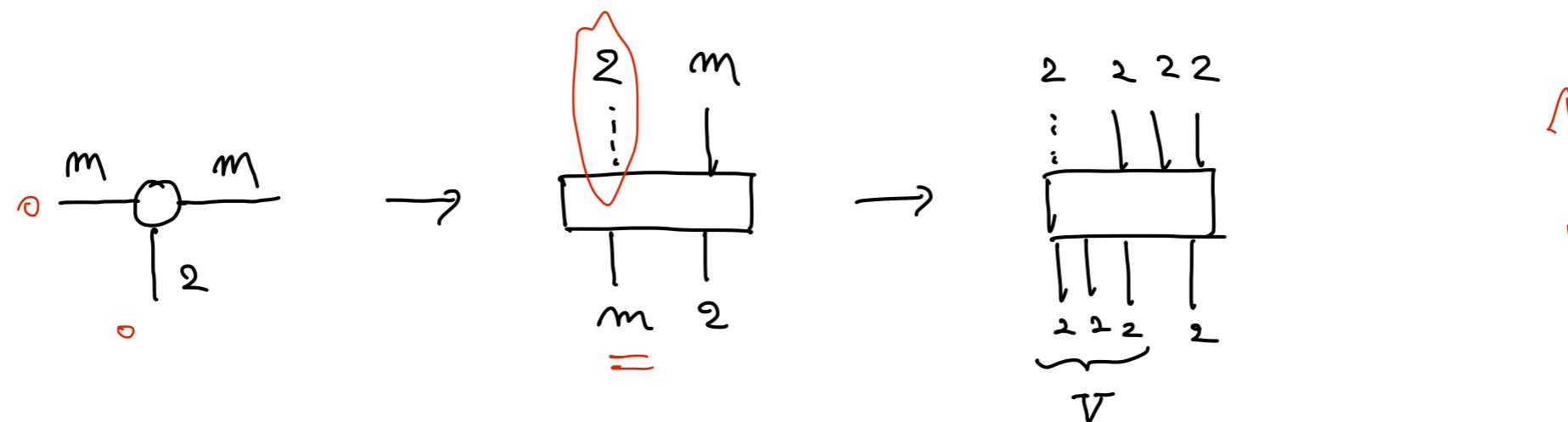
- $2^r \times 2^r$  matrix  $\rightarrow 2 \times 2 \times \dots \times 2$  tensor  $\rightarrow 2r$ -leg tensor

- initial quantum states

- product state, e.g.,  $|0\rangle \otimes |0\rangle \otimes \dots \otimes |0\rangle$   
 $\rightarrow$  a set of  $n$  1-leg tensors (vectors)

# Converting Tensors into Quantum Gates

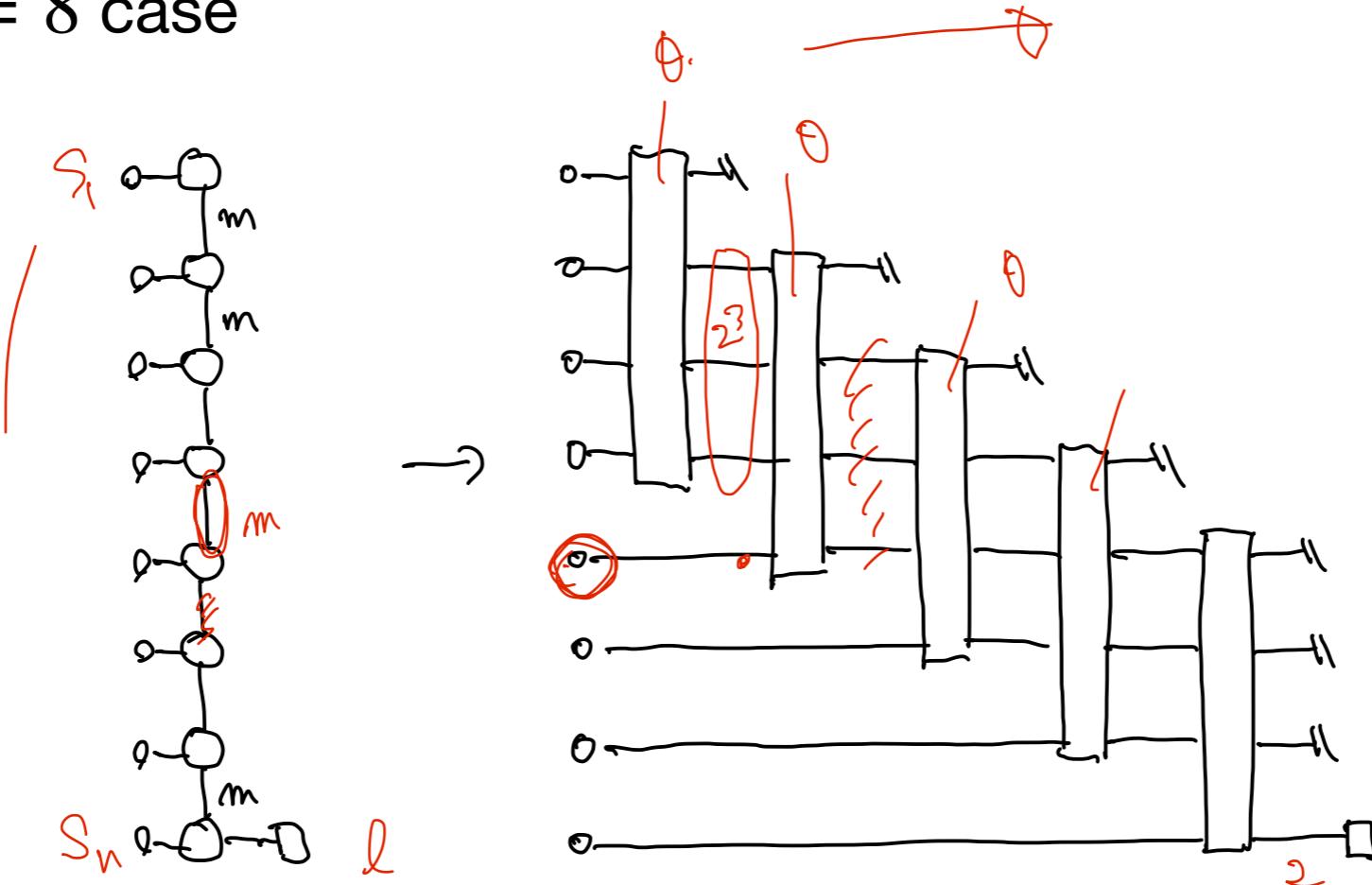
- Quantum gate is tensor
  - Tensor is quantum gate?
  - Yes! But, quantum gate should be a unitary operator
- $m = 2^V$  ( $V = 3$ ) case



- Huggins, W., Patil, P., Mitchell, B., Birgitta Whaley, K., & Miles Stoudenmire, E. (2019). Towards quantum machine learning with tensor networks. *Quantum Science and Technology*, 4, 1. <https://doi.org/10.1088/2058-9565/aaea94> (arXiv:1803.11537)

# Converting MPS into Quantum Circuit

- $N = 8, m = 8$  case



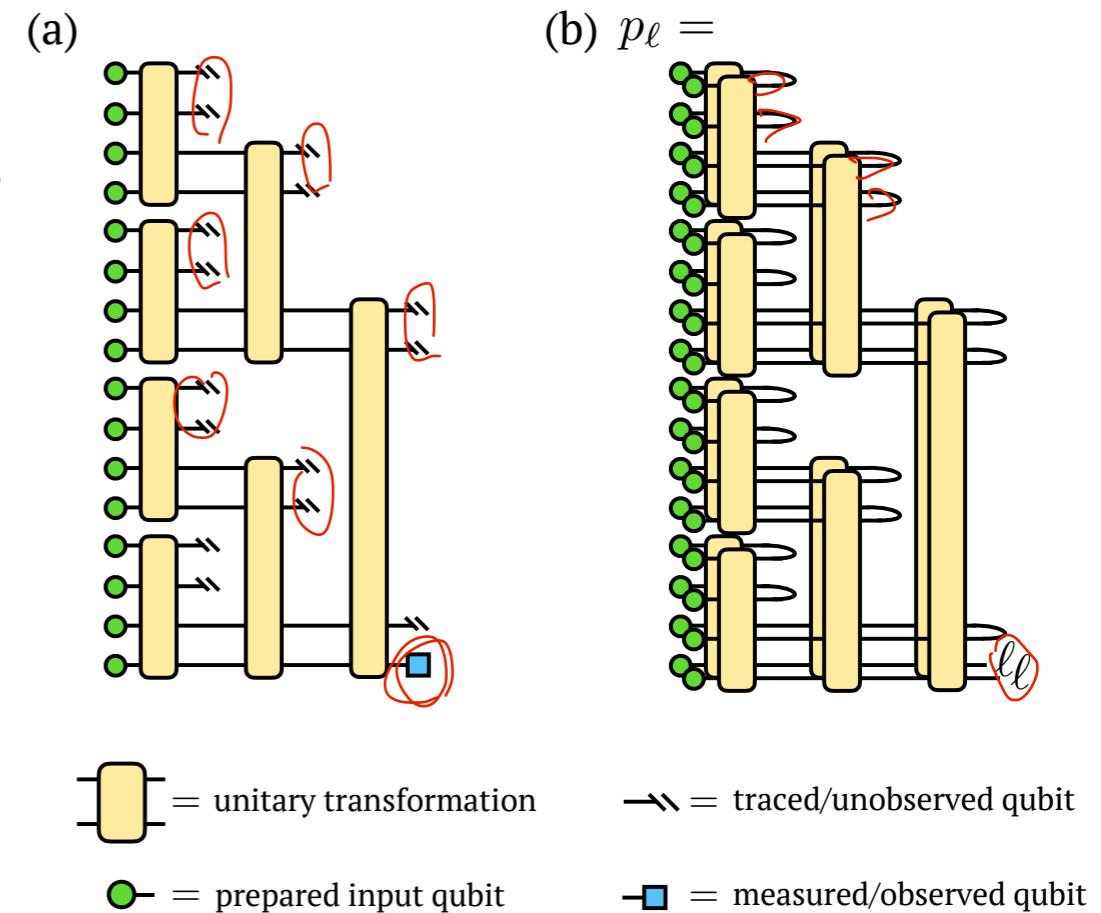
- Optimization of tensors → optimization of parameters in parameterized quantum circuit → quantum-classical hybrid algorithm

- Huggins, W., Patil, P., Mitchell, B., Birgitta Whaley, K., & Miles Stoudenmire, E. (2019). Towards quantum machine learning with tensor networks. *Quantum Science and Technology*, 4, 1. <https://doi.org/10.1088/2058-9565/aaea94> (arXiv:1803.11537)

# Unobserved Qubits

$$\boxed{P} = \text{tr} \left[ \Psi \otimes \bar{\Psi} \right]$$

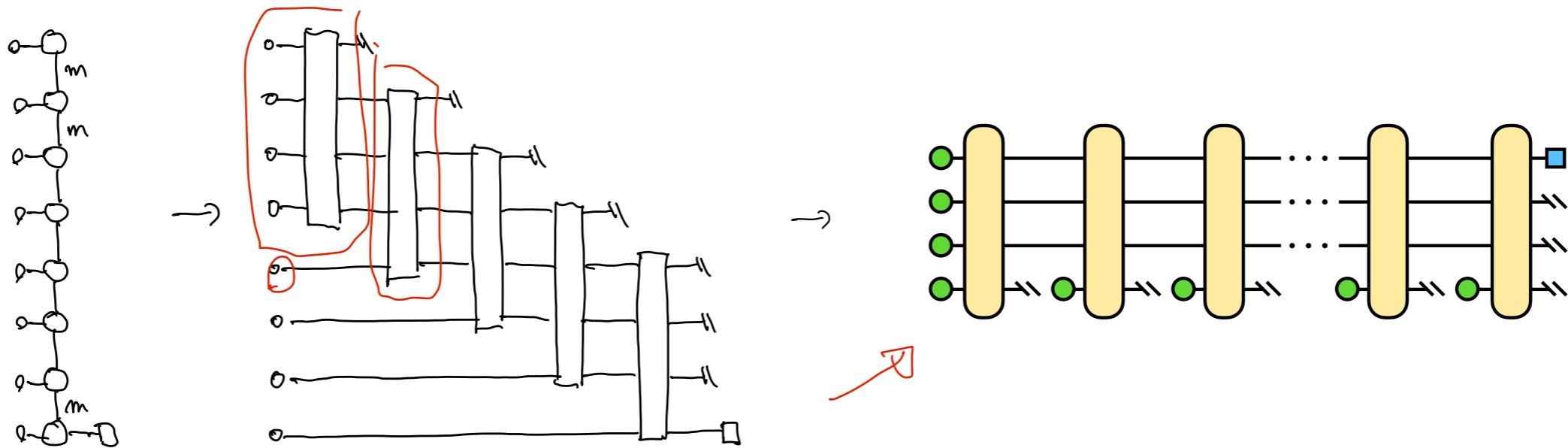
- In classical simulation
  - take a partial trace on unobserved qubits
- On quantum computer
  - just ignore or reset qubits
  - we can reuse discarded qubits!



- Huggins, W., Patil, P., Mitchell, B., Birgitta Whaley, K., & Miles Stoudenmire, E. (2019). Towards quantum machine learning with tensor networks. *Quantum Science and Technology*, 4, 1. <https://doi.org/10.1088/2058-9565/aaea94> (arXiv:1803.11537)

# Qubit-efficient Implementation

- Reset and reuse discarded qubits



- Number of physical qubits required is independent of input size  $N$

- Huggins, W., Patil, P., Mitchell, B., Birgitta Whaley, K., & Miles Stoudenmire, E. (2019). Towards quantum machine learning with tensor networks. *Quantum Science and Technology*, 4, 1. <https://doi.org/10.1088/2058-9565/aaea94> (arXiv:1803.11537)
- Liu, J.-G., Zhang, Y.-H., Wan, Y., & Wang, L. (2019). Variational quantum eigensolver with fewer qubits. *Phys. Rev. Research*, 1, 23025. <https://doi.org/10.1103/physrevresearch.1.023025> (arXiv:1902.02663)