計算科学・量子計算における情報圧縮

Data compression in computational science and quantum computing

2023.01.12

#12: 量子古典ハイブリッドアルゴリズムとテンソルネットワーク

Quantum-Classical Hybrid Algorithms and Tensor Network

Synge Todo / 藤堂眞治

Department of Phys., UTokyo / 東京大学大学院理学系研究科

Today's topic

1. Computational science, quantum computing, and data compression
2. Review of linear algebra
3. Singular value decomposition
4. Application of SVD and generalization to tensors
5. Entanglement of information and matrix product states
6. Application of MPS to eigenvalue problems
7. Tensor network representation
8. Data compression in tensor network
9. Tensor network renormalization
10. Quantum mechanics and quantum computation
11. Simulation of quantum computers
12. Quantum-classical hybrid algorithms and tensor network
13. Quantum error correction and tensor network

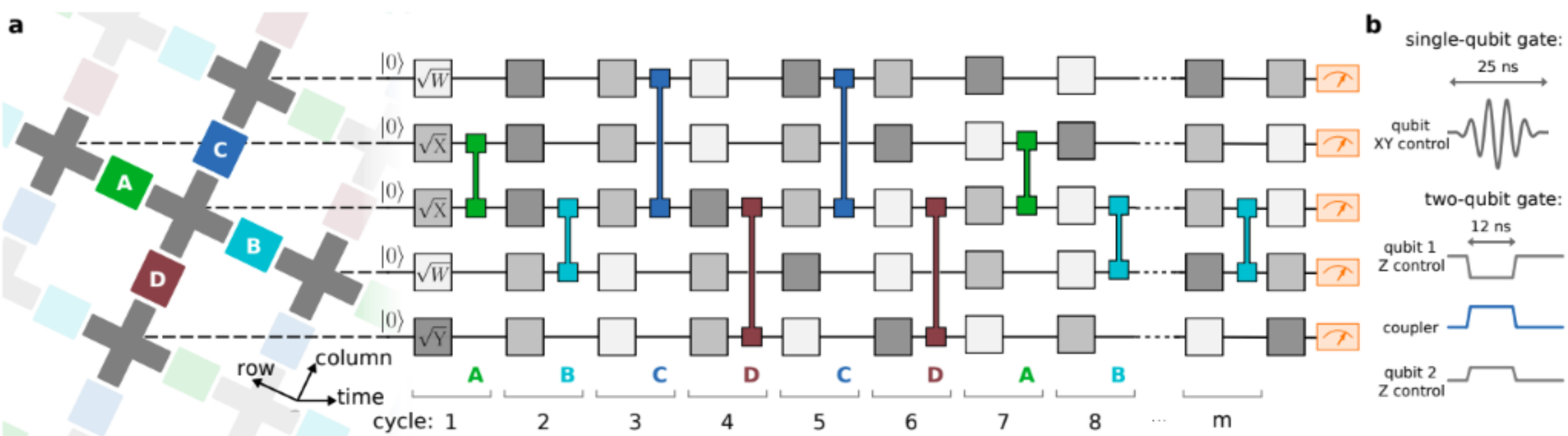
Outline

- Quantum Computation
- Quantum-classical Hybrid Algorithms
  - Variational Quantum Eigensolver (VQE)
  - QAOA
- Optimization of Quantum Circuit
  - Gradient Calculation
  - Sequential Optimization
- Machine Learning Based on Tensor Network
- Qubit-efficient Implementation of Tensor-network Machine Learning
  - Converting Tensors into Quantum Gates

Quantum Computation

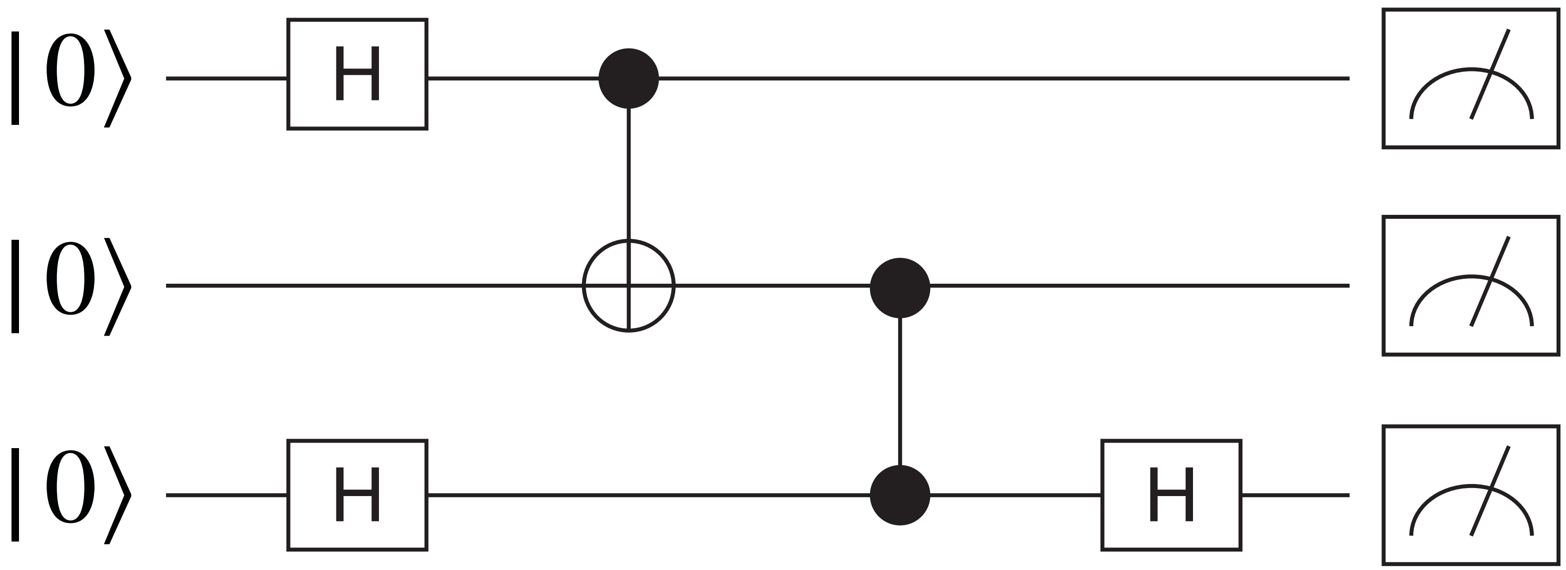
Quantum circuits

- Prepare a set of quantum bits (qubits)
- A number of quantum gates (typically 1-qubit or 2-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

State bifurcation and interference



- $|000\rangle \Rightarrow \frac{1}{2}(|0\rangle + |1\rangle)|0\rangle(|0\rangle + |1\rangle) = \frac{1}{2}(|000\rangle + |001\rangle + |100\rangle + |101\rangle)$

$$\Rightarrow \frac{1}{2}(|000\rangle + |001\rangle + |110\rangle + |111\rangle) \Rightarrow \frac{1}{2}(|000\rangle + |001\rangle + |110\rangle - |111\rangle)$$

$$\Rightarrow \frac{1}{2\sqrt{2}}(|00\rangle + |11\rangle)(|0\rangle + |1\rangle) + (|00\rangle - |11\rangle)(|0\rangle - |1\rangle)$$

$$= \frac{1}{2\sqrt{2}}(|000\rangle + |001\rangle + |110\rangle + |111\rangle + |000\rangle - |001\rangle - |110\rangle + |111\rangle)$$

$$= \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$$

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

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 have 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

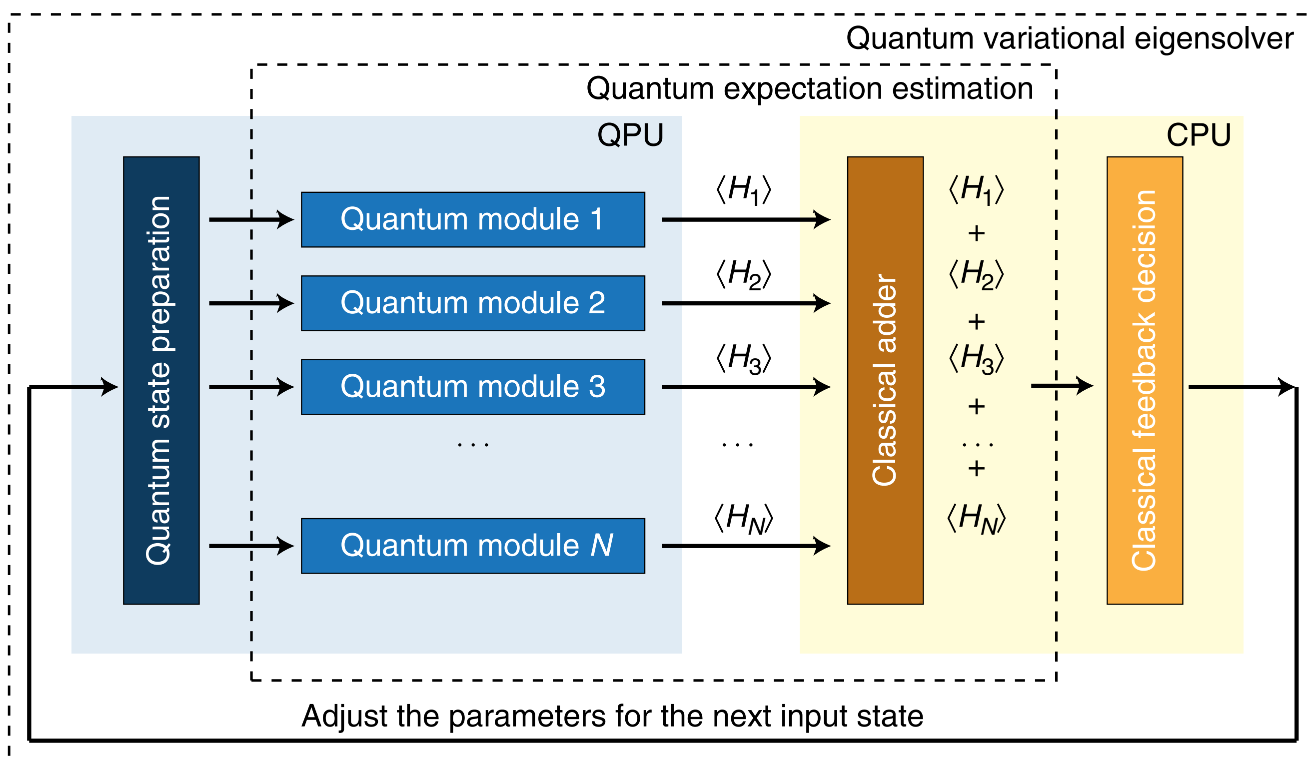
- Fault-tolerant quantum computing with quantum error correction
  - Grover's algorithm (or amplitude amplification)
    - unstructured database search, optimization problem
    - quadratical speed improvement: $O(N) \to O(\sqrt{N})$
  - quantum phase estimation
    - prime factorization, eigenvalue problems, simultaneous linear equations (HHL)
    - exponential speed improvement over classical computers
  - many physical qbits (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 (VQE)

- 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



- 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: $H = JS_1 \cdot S_2 = \frac{J}{4}(\sigma_1^x \sigma_2^x + \sigma_1^y \sigma_2^y + \sigma_1^z \sigma_2^z) \quad (J > 0)$
  - ground state: $|\Psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$
- Variational wave function: $|\phi(\{\theta_i\})\rangle = U(\{\theta_i\})|00\rangle$
- Expectation value of Hamiltonian (energy):

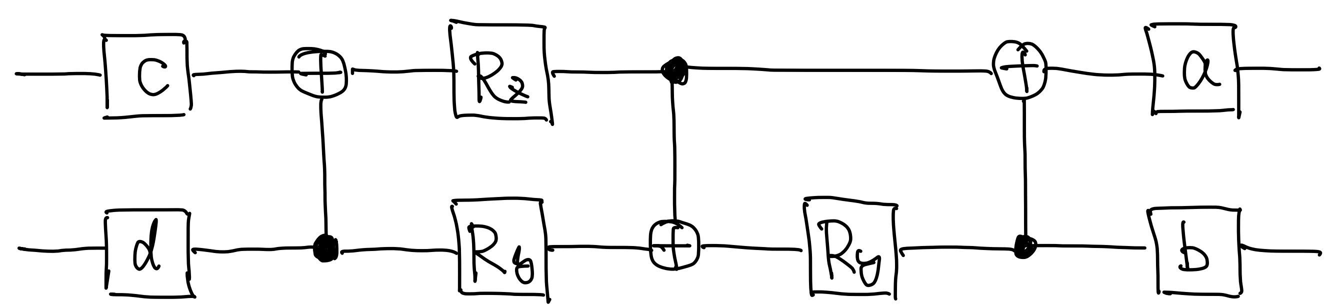
$$\langle\phi(\{\theta_i\})|H|\phi(\{\theta_i\})\rangle = \frac{J}{4}(\langle\phi(\{\theta_i\})|\sigma_1^x\sigma_2^x|\phi(\{\theta_i\})\rangle + \langle\phi(\{\theta_i\})|\sigma_1^y\sigma_2^y|\phi(\{\theta_i\})\rangle$$
$$+\langle\phi(\{\theta_i\})|\sigma_1^z\sigma_2^z|\phi(\{\theta_i\})\rangle$$

Universal quantum gates

- Universal one-qubit unitary operation

$$U = \mathrm{R}_x(\theta_1)\mathrm{R}_y(\theta_2)\mathrm{R}_x(\theta_3)$$

- Universal two-qubit unitary operation

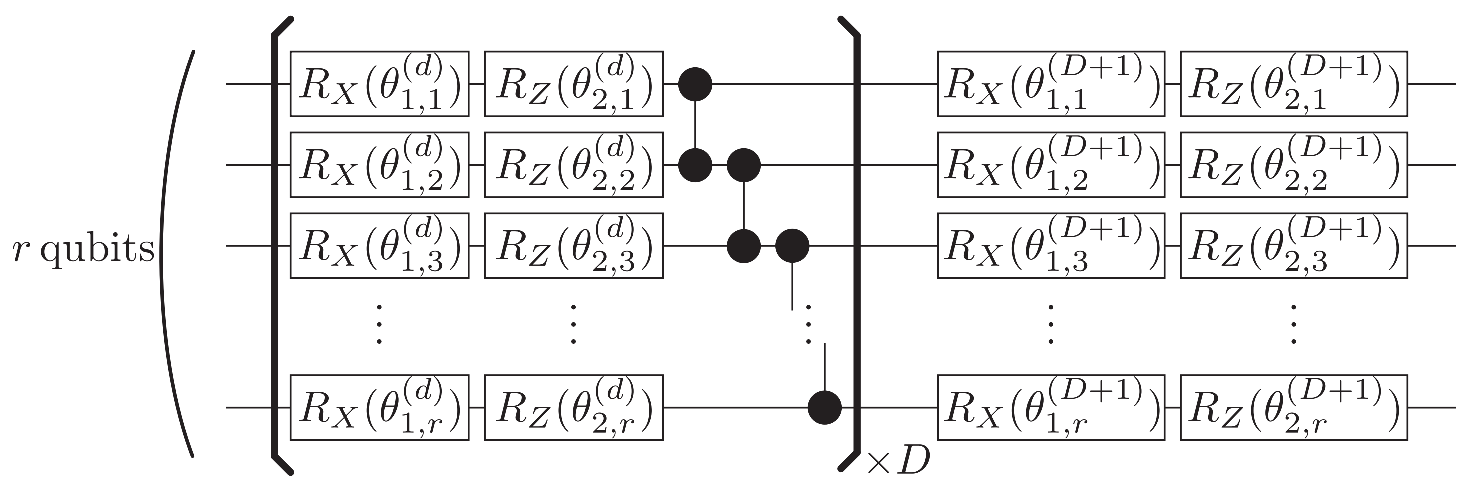


  - [a]-[d]: universal one-qubit unitary operations
  - 15 one-qubit rotations + 3 CNOTs

- 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.,

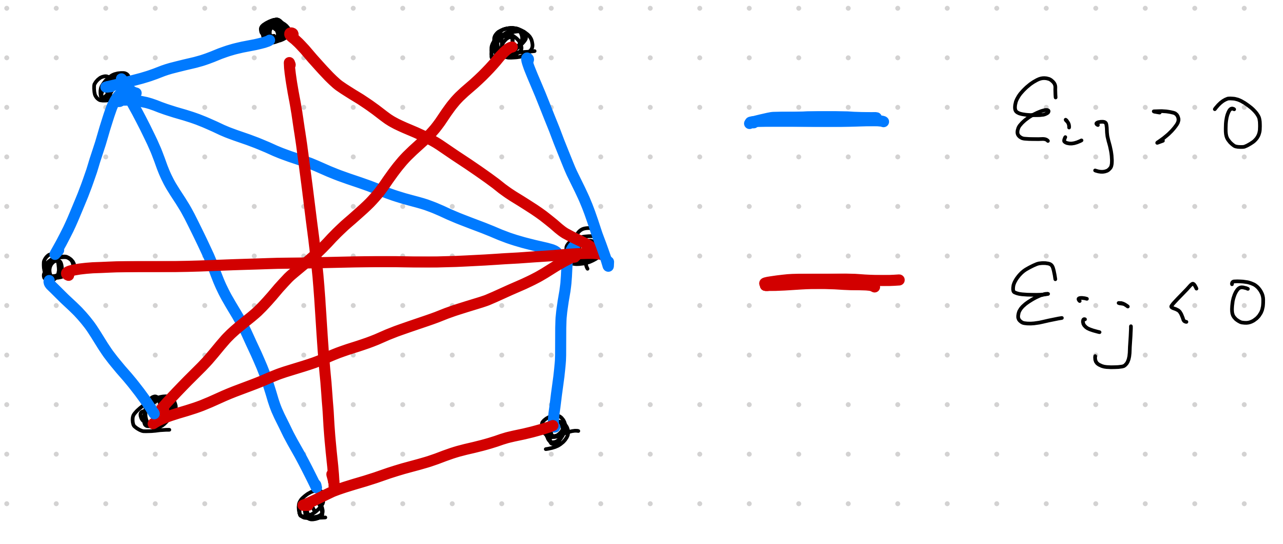


- For quantum chemical calculations
  - various high-precision variational wave functions have been proposed
  - ref) 杉崎「量子コンピュータによる量子化学計算入門」(講談社, 2020)

Combinatorial optimization problem

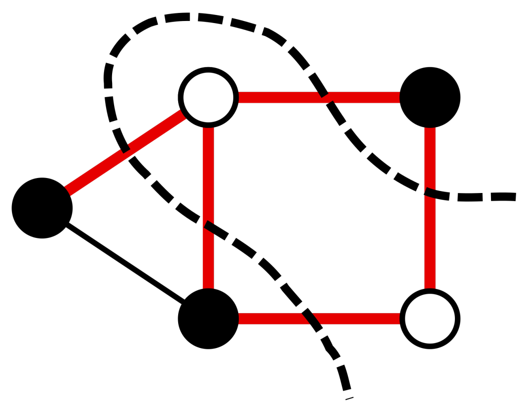
- aka. discrete optimization problem
  - many combinatorial optimization problem → finding the ground state of (frustrated) Ising model

$$H = -J \sum_{i<j} \epsilon_{ij} \sigma_i^z \sigma_j^z$$



Max-cut problem

- Maximum cut
  - → partition of graph vertices into two complementary sets, such that the number of edges between them is as large as possible
  - generalized version → weighted max-cut
- NP-hard problem
  - → no polynomial-time algorithms



https://en.wikipedia.org/wiki/Maximum_cut

- Mapping to classical Ising model
  - if we assign $x_i = 0$ for vertices in the first set and $x_i = 1$ for the others, the number of cut is

$$\sum_{(i,j)\in E} x_i(1-x_j) = \sum_{(i,j)\in E} \frac{1}{2}(x_i(1-x_j) + x_j(1-x_i))$$

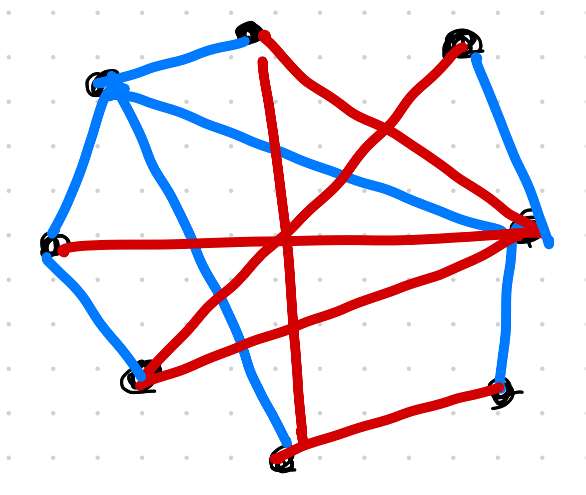
  - by setting $x_i = (1-\sigma_i^z)/2$, max-cut is given by the ground state of

$$H = \sum_{(i,j)\in E} \sigma_i^z \sigma_j^z + C$$

Quantum annealing

- Introduce transverse-field (magnetic field to x direction)

$$H = \sum_{(i,j)\in E} \sigma_i^z \sigma_j^z - \sum_i \sigma_i^x$$



  - classical limit: $\Gamma = 0$
    - ground state gives max-cut
  - quantum limit: $\Gamma \to \infty$
    - uniform superposition of all 2^n states
  - quantum annealing
    - adiabatic time evolution from $\Gamma \to \infty$ to $\Gamma = 0$ over time T
    - can reach max-cut ground state with probability 1 in the limit of $T \to \infty$
- Existing quantum annealer: e.g) D-Wave

Quantum approximate optimization algorithm (QAOA)

- Implements real-time dynamics using quantum gate operations
  - decomposition of Hamiltonian and time-evolution operator

$$H = \sum_{(i,j)\in E} \sigma_i^z \sigma_j^z - \sum_i \sigma_i^x = H_z + H_x$$

$$e^{-itH} \approx e^{-itH_0} e^{-itH_x}$$

  - trial (variational) state

$$|\Psi(\vec{\gamma}, \vec{\beta})\rangle = e^{-i\beta_p H_x} e^{-i\gamma_p H_z} \cdots e^{-i\beta_1 H_x} e^{-i\gamma_1 H_z} | + \rangle^n$$

  - optimize $\vec{\gamma} = (\gamma_1, \cdots, \gamma_p)$ and $\vec{\beta} = (\beta_1, \cdots, \beta_p)$ so that

$$\langle \Psi(\vec{\gamma}, \vec{\beta}) | H_z | \Psi(\vec{\gamma}, \vec{\beta}) \rangle$$

is minimized

Farhi, Edward, Jeffrey Goldstone, and Sam Gutmann. "A quantum approximate optimization algorithm." arXiv preprint arXiv:1411.4028 (2014).

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 $(\theta_1, \theta_2, \cdots, \theta_{k-1}, \theta_{k+1}, \cdots, \theta_n)$
  - unitary transformation for generating variational wave function can be written as

$$U(\theta_k) = VU_k(\theta_k)W \qquad\qquad U_k(\theta_k) = e^{-i\sigma_k\theta_k/2}$$

  - expectation value of operator $\mathcal{O}$

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

Derivative of Cost Function

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

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

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

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

- Stable derivative estimation

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

- 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

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

$$\langle \mathcal{O}(\theta_k') \rangle = \langle \Psi | (I\cos\frac{\theta_k'}{2} + i\sigma_k \sin\frac{\theta_k'}{2}) Q (I\cos\frac{\theta_k'}{2} + i\sigma_k \sin\frac{\theta_k'}{2}) | \Psi \rangle$$
$$= a_1 \cos(\theta_k' - a2) + a_3$$

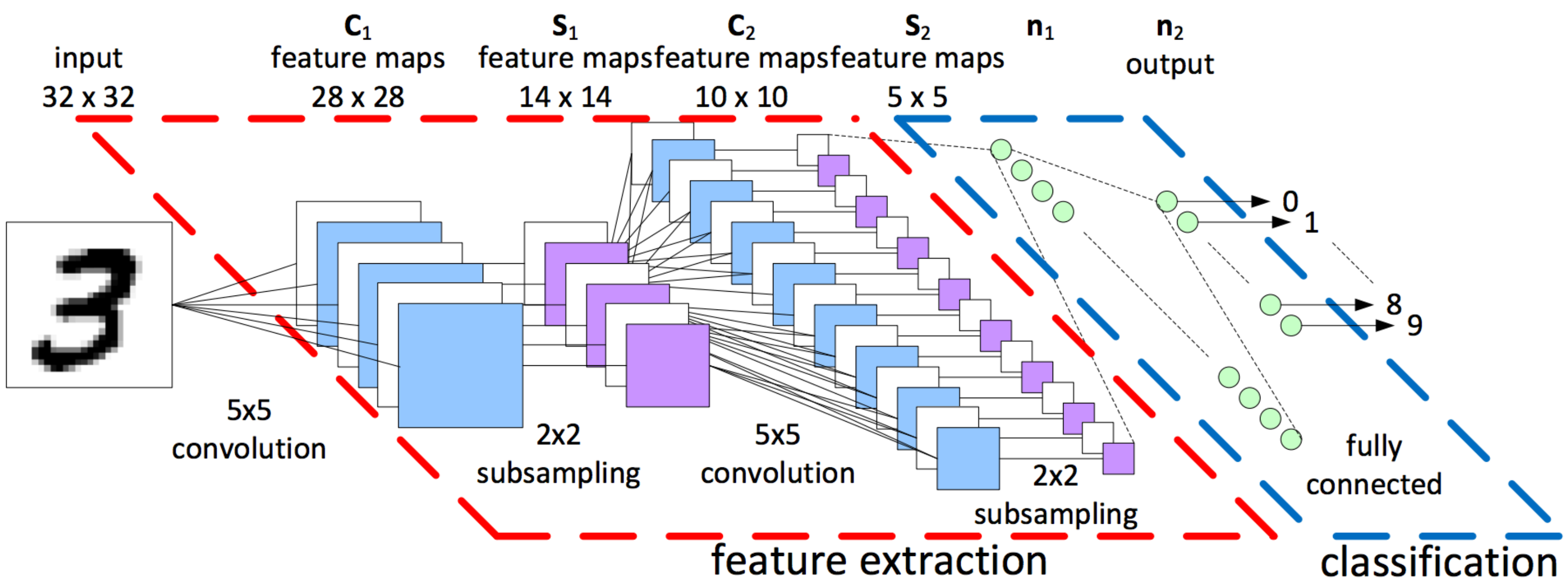
- Coefficients a_1, a_2, a_3 can be determined from $\langle \mathcal{O}(\theta_k) \rangle$, $\langle \mathcal{O}(\theta_k + \frac{\pi}{2}) \rangle$,$\langle \mathcal{O}(\theta_k - \frac{\pi}{2}) \rangle$
- $\langle \mathcal{O}(\theta_k') \rangle$ can be minimized by choosing $\theta_k' = a_2$ or $\theta_k' = -a_2$ （depending on the sign of a_1)

- 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)



https://www.kaggle.com/cdeotte/how-to-choose-cnn-architecture-mnist

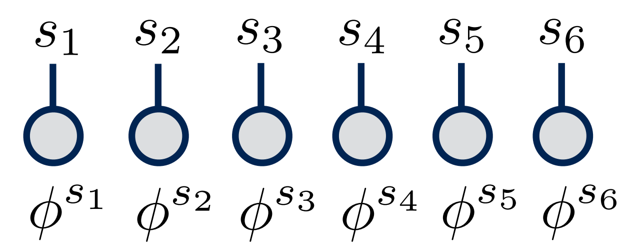
  - input: gray scale image
  - output: 10-dimensional vector
- Accuracy > 99.5% for MNIST

  - 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 Leaning with Tensor Networks

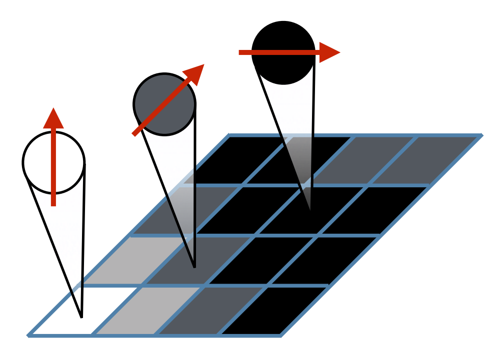
- Tensor network has multi-linear property
  - how to introduce non-linearity and increase expressive power?
- Encode of input data
  - convert N-dimensional vector (N-pixel image) into 2^N-dimensional feature map

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



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

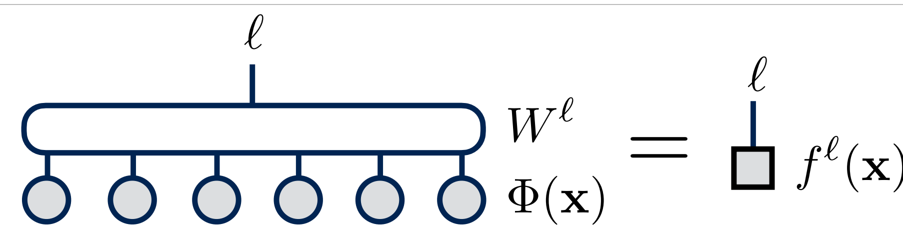
$$\phi^{s_j}(x_j) = \left[\cos\left(\frac{\pi}{2} x_j\right),\ \sin\left(\frac{\pi}{2} x_j\right)\right]$$



- 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)

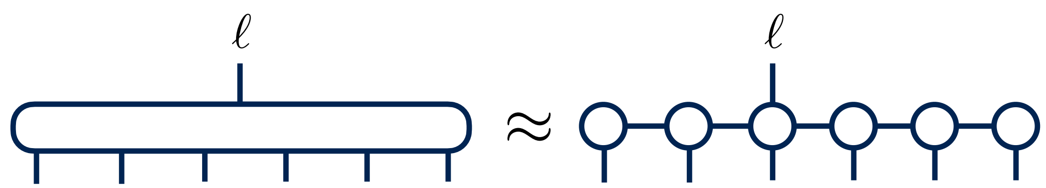
Classification Model

- $f^{\ell}(\mathbf{x}) = W^{\ell} \cdot \Phi(\mathbf{x})$
  - W is a huge (10×2^N) matrix



- Decompose W into tensor network (matrix product)

$$W^{\ell}_{s_1 s_2 \cdots s_N} = \sum_{\{\alpha\}} A^{\alpha_1}_{s_1} A^{\alpha_1 \alpha_2}_{s_2} \cdots A^{\ell;\alpha_j \alpha_{j+1}}_{s_j} \cdots A^{\alpha_{N-1}}_{s_N}$$



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

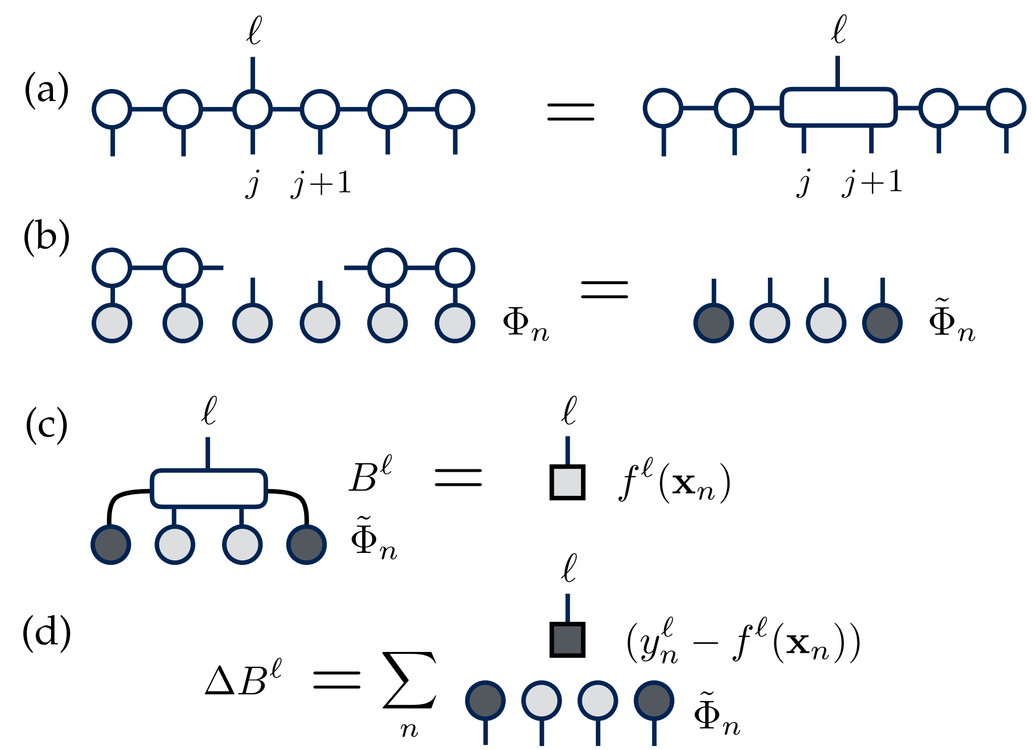
- Stoudenmire, E. M., & Schwab, D. J. (2016). Supervised Learning with Quantum-Inspired Tensor Networks. Advances in Neural Information Processing Systems, 29, 4799. (arXiv:1605.05775)

Cost Function and Gradient

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

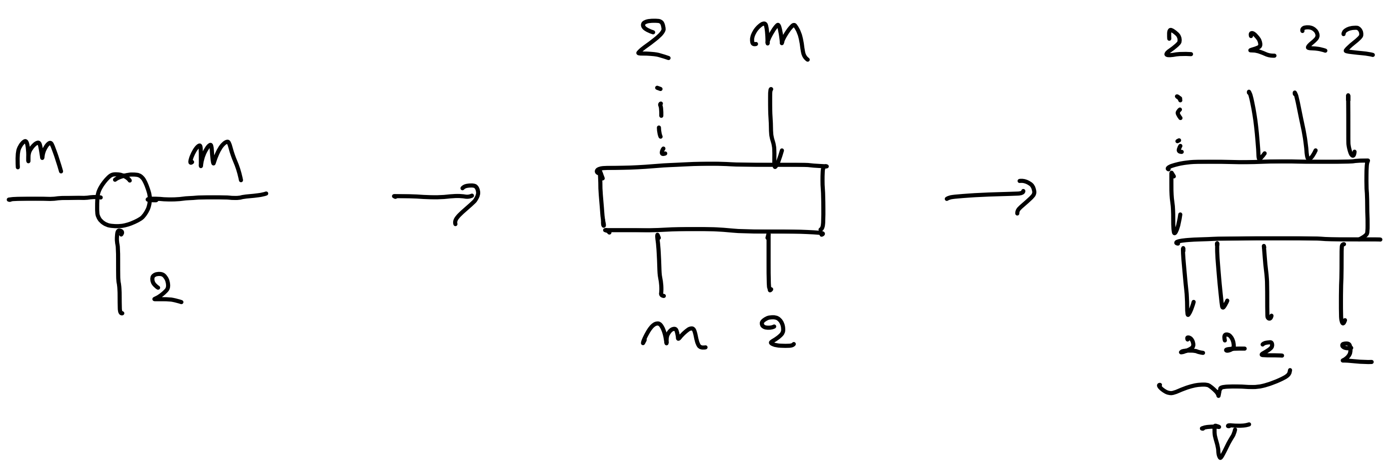
Qubit-efficient Implementation of Tensor-network Machine Learning

Tensor Network Representation of Quantum Circuits

- Tensor representation of quantum gates and states
  - 1-qubit gate:
    - 2×2 matrix → 2-leg tensor
  - 2-qubit gate:
    - 4×4 matrix → $2 \times 2 \times 2 \times 2$ tensor → 4-leg tensor
  - r-qubit gate:
    - $2^r \times 2^r$ matrix → $2 \times 2 \times \cdots \times 2$ tensor → $2r$-leg tensor
  - initial quantum states
    - product state, e.g., $|0\rangle \otimes |0\rangle \otimes \cdots \otimes |0\rangle$
      → a set of n 1-leg tensors (vectors)

Converting Tensors into Quantum Gates

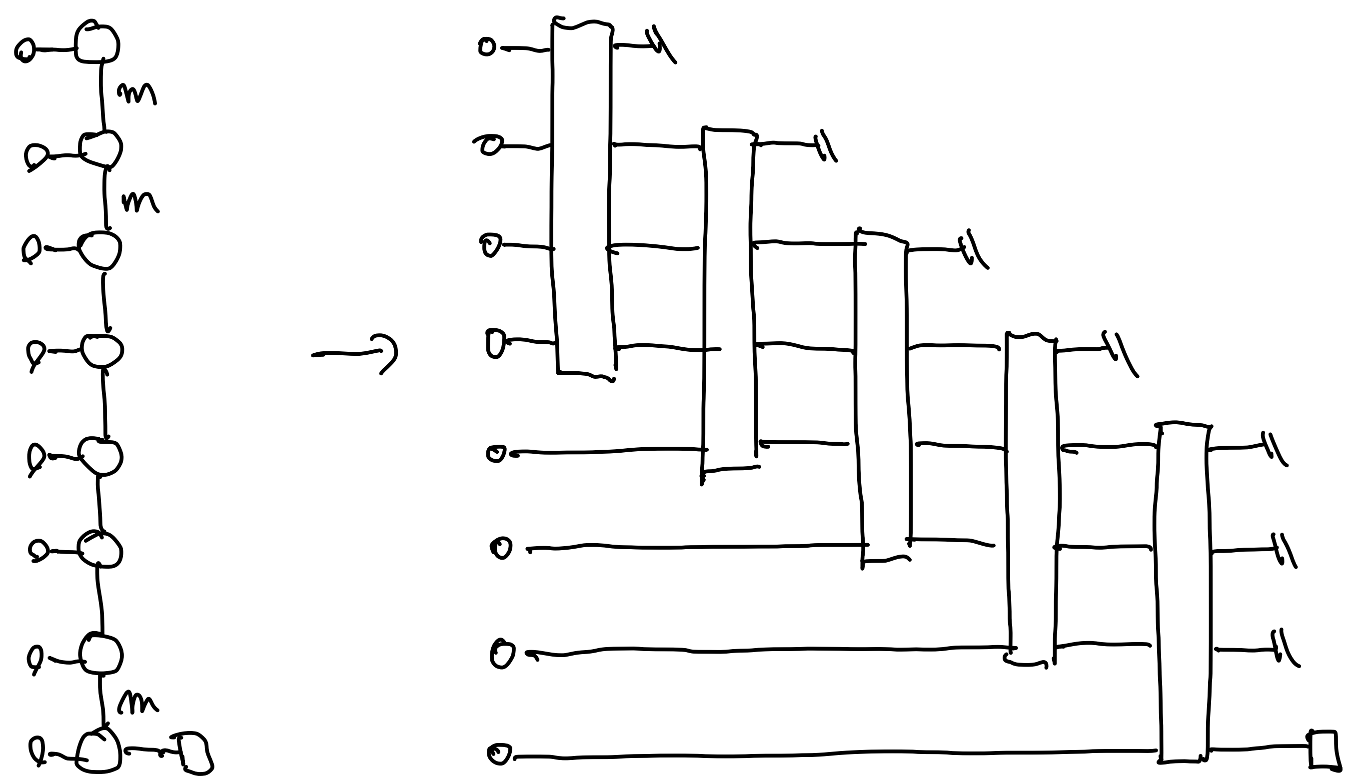
- Quantum gate is tensor
  - Tensor is quantum gate?
  - Yes! But, quantum gate should be a unitary operator
- $m = 2^V$ $(V = 3)$ case



- 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

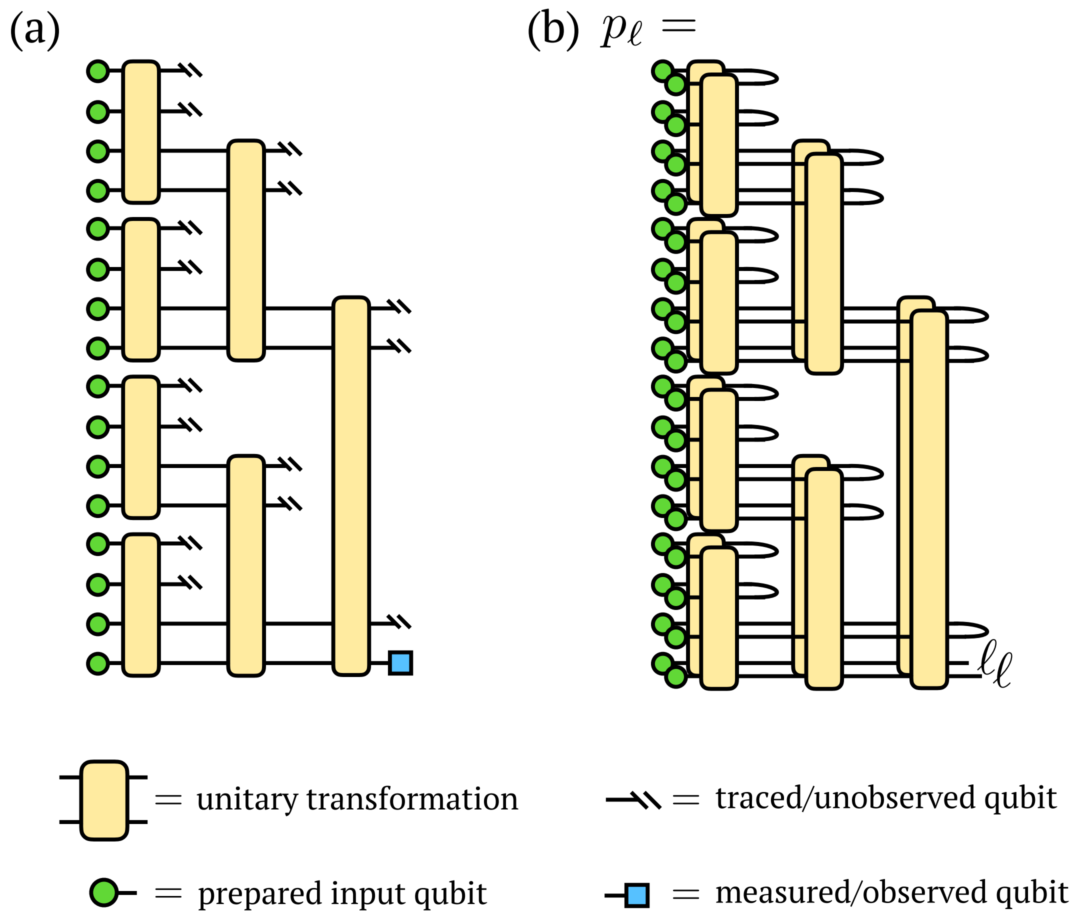
- $N = 8$, $m = 8$ case



- 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

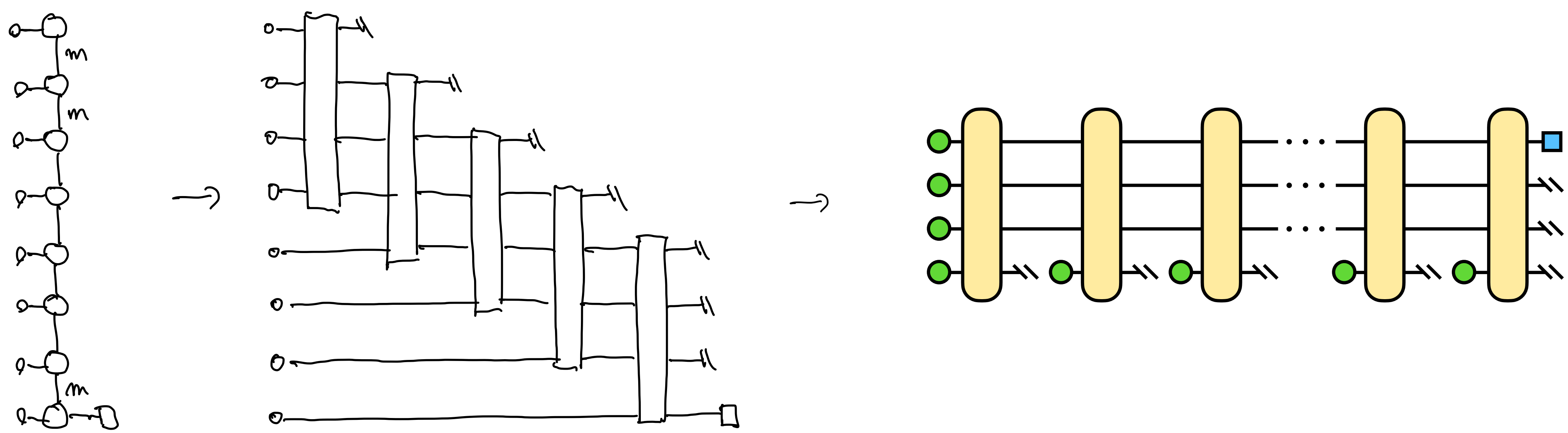
- In classical simulation
  - take a partial trace on unobserved qubits
- On quantum computer
  - just ignore or reset qubits
  - we can reuse discarded qubits!



- 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)

Next (Jan. 19)

1. Computational science, quantum computing, and data compression
2. Review of linear algebra
3. Singular value decomposition
4. Application of SVD and generalization to tensors
5. Entanglement of information and matrix product states
6. Application of MPS to eigenvalue problems
7. Tensor network representation
8. Data compression in tensor network
9. Tensor network renormalization
10. Quantum mechanics and quantum computation
11. Simulation of quantum computers
12. Quantum-classical hybrid algorithms and tensor network
13. Quantum error correction and tensor network