Quantum Machine Learning: Opportunities and Challenges

AI and Quantum Information Applications in Fundamental Physics | 2023.02.17

Daniel Kyungdeock Park
Yonsei University
Department of Applied Statistics
Department of Statistics and Data Science

q-DNA

Quantum Data Science & AI @ Yonsei University

Welcome to the Quantum Data Science & AI (a.k.a q-DNA) lab at Yonsei University. We are a multi-disciplinary group where expertise in physics, computer science, statistics, and applied mathematics intersect.

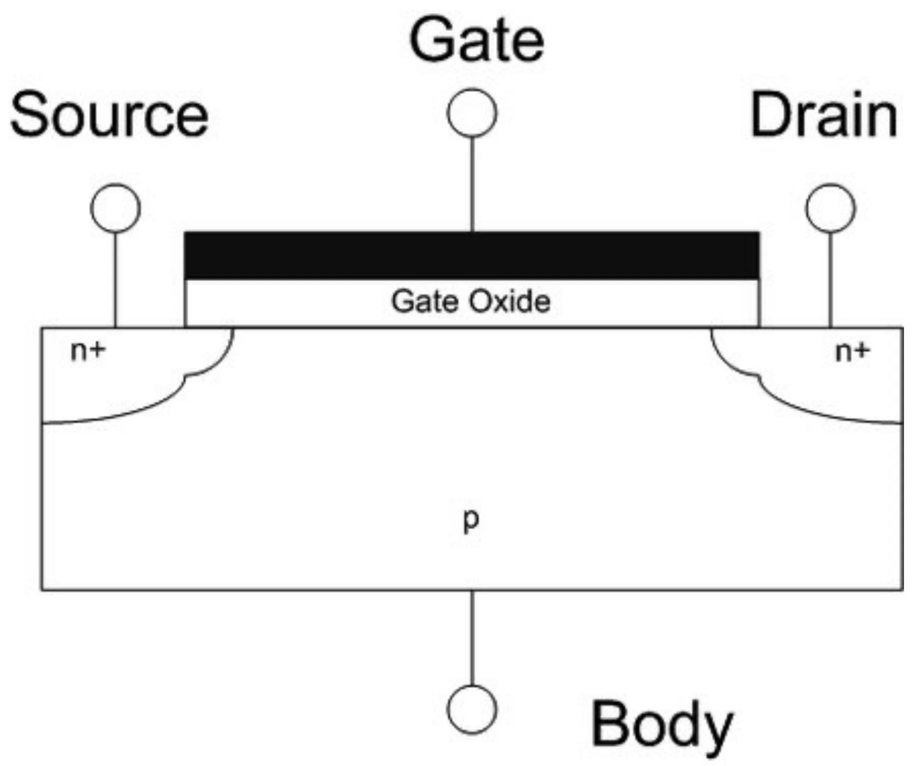
In brief, here is what we do:

- Harness quantum information theory to solve various challenges in data science and AI.
- Develop industrial applications of quantum computing empowered by quantum optimization techniques.
- Develop statistical and machine learning methods that combat noise and imperfections in quantum information processing tasks.

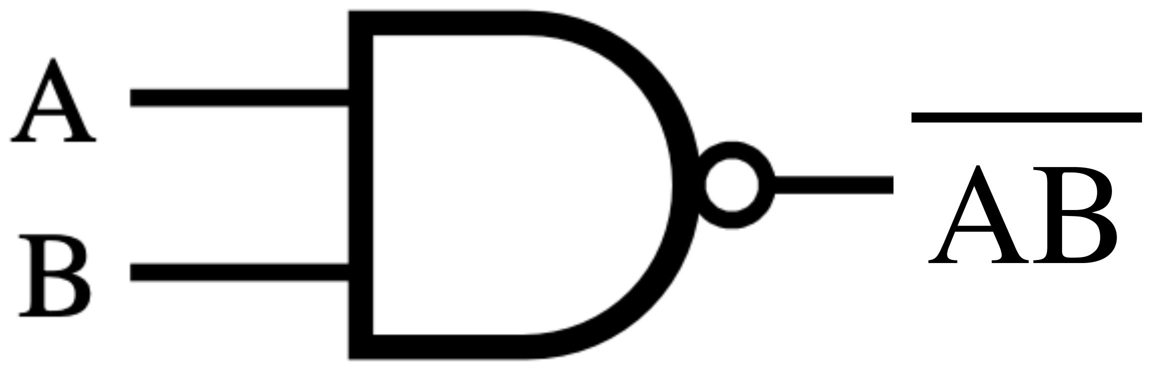
Introduction

What is Digital Computing?

Digital computation with n bits: $\{0,1\}^n \rightarrow \{0,1\}^m,\ m \leq n$



+



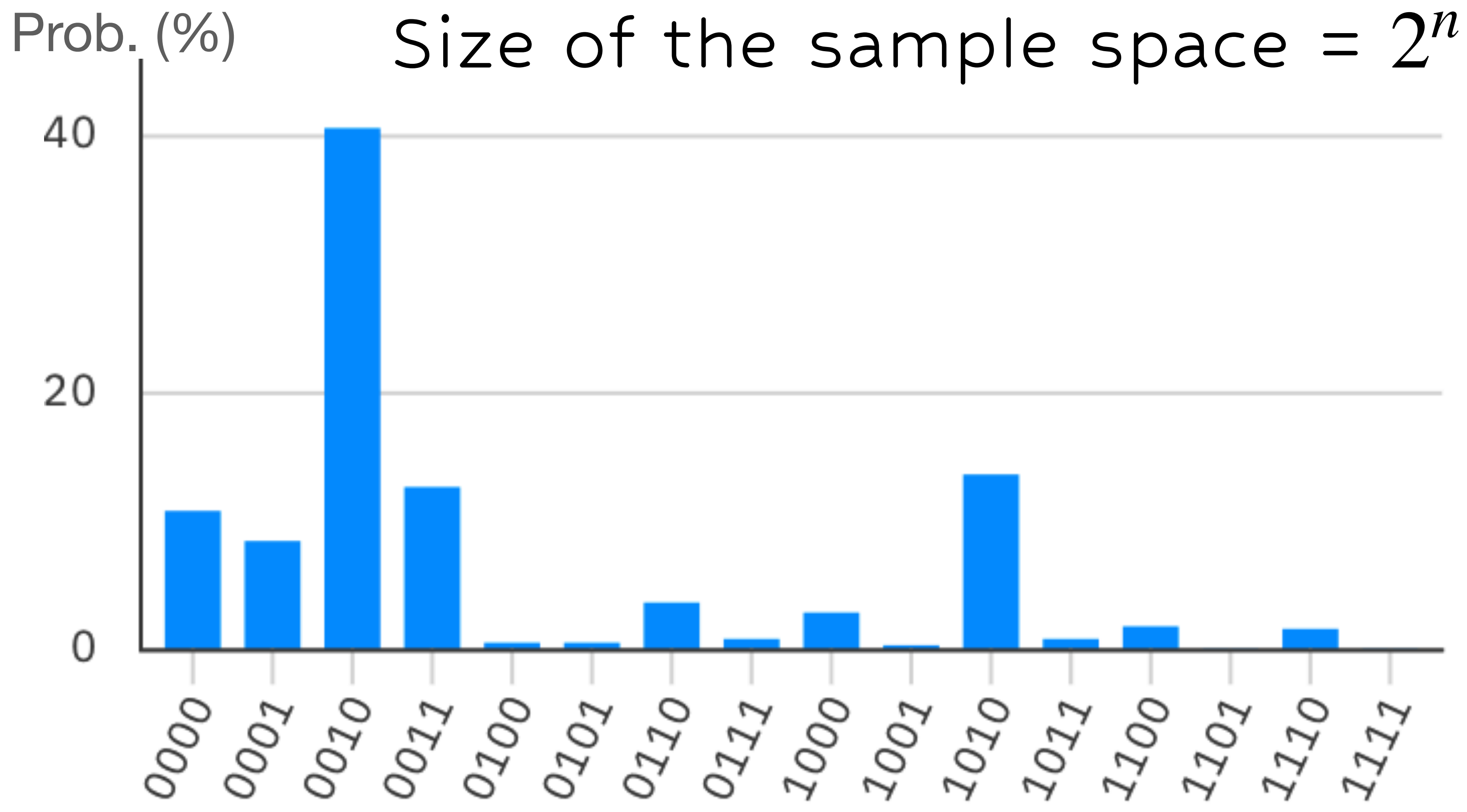
A	B	$\overline{AB}$
0	0	1
0	1	1
1	0	1
1	1	0

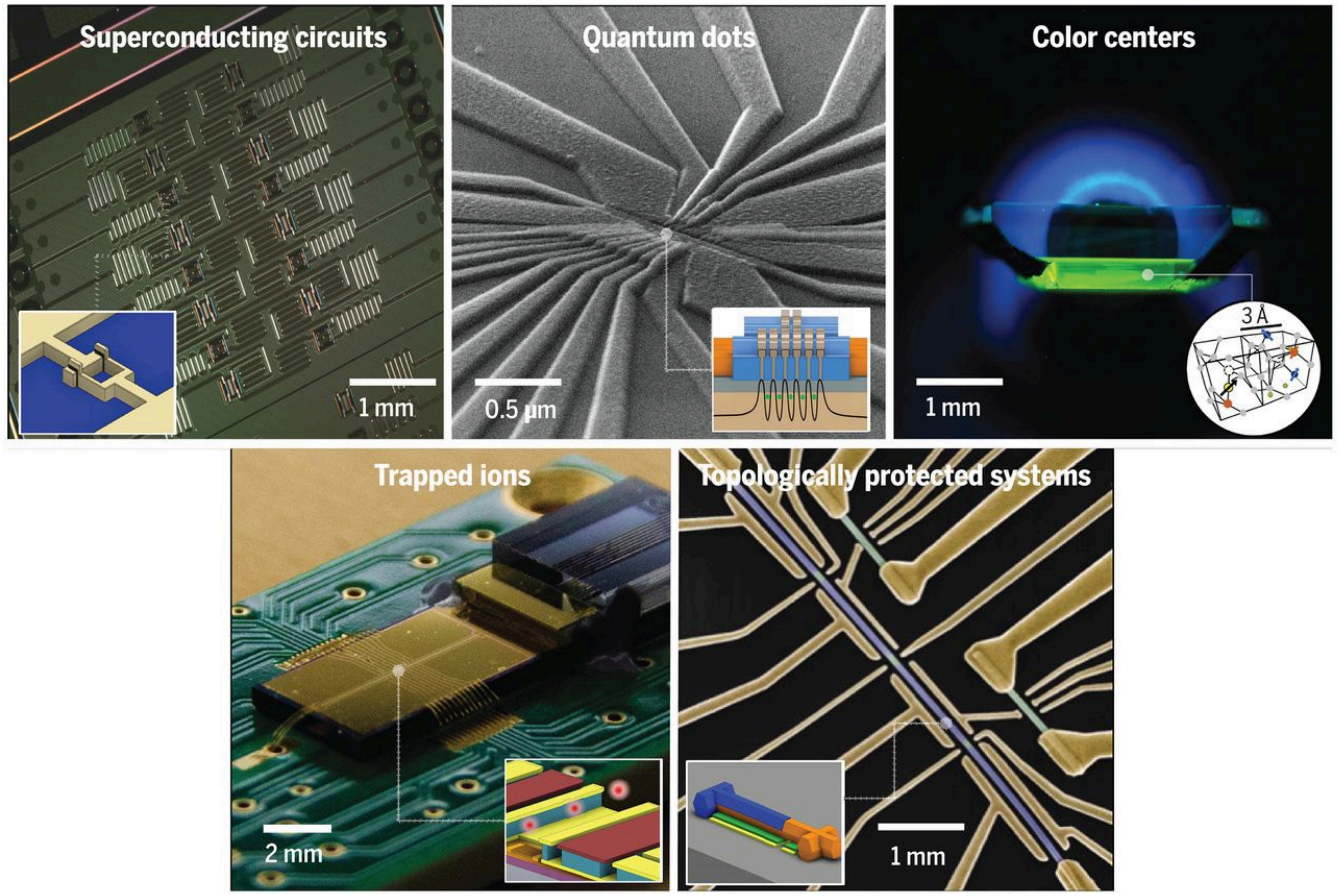
Introduction

What is Quantum Computing?

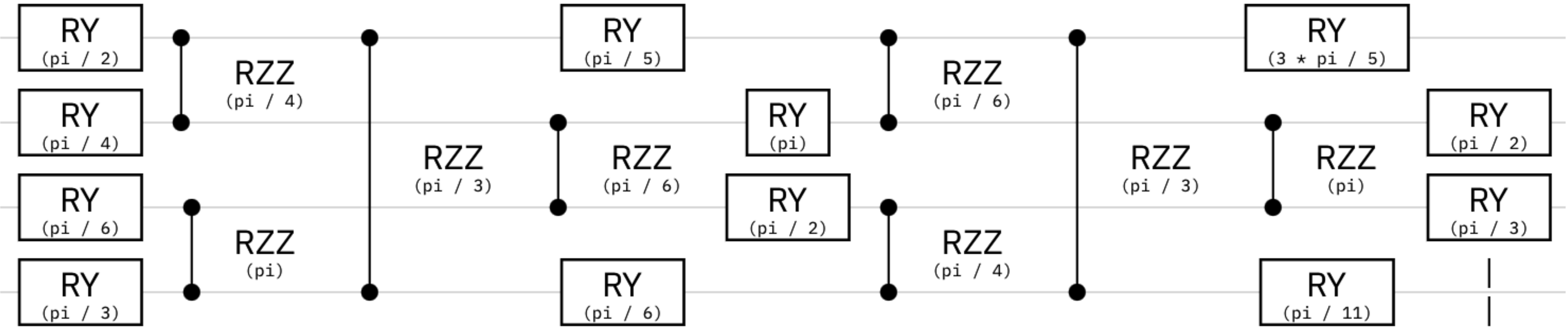
Quantum computation with n qubits: $\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{2^n} \end{pmatrix} \in \mathbb{C}^{2^n} \rightarrow$

Size of the sample space = 2^n





+



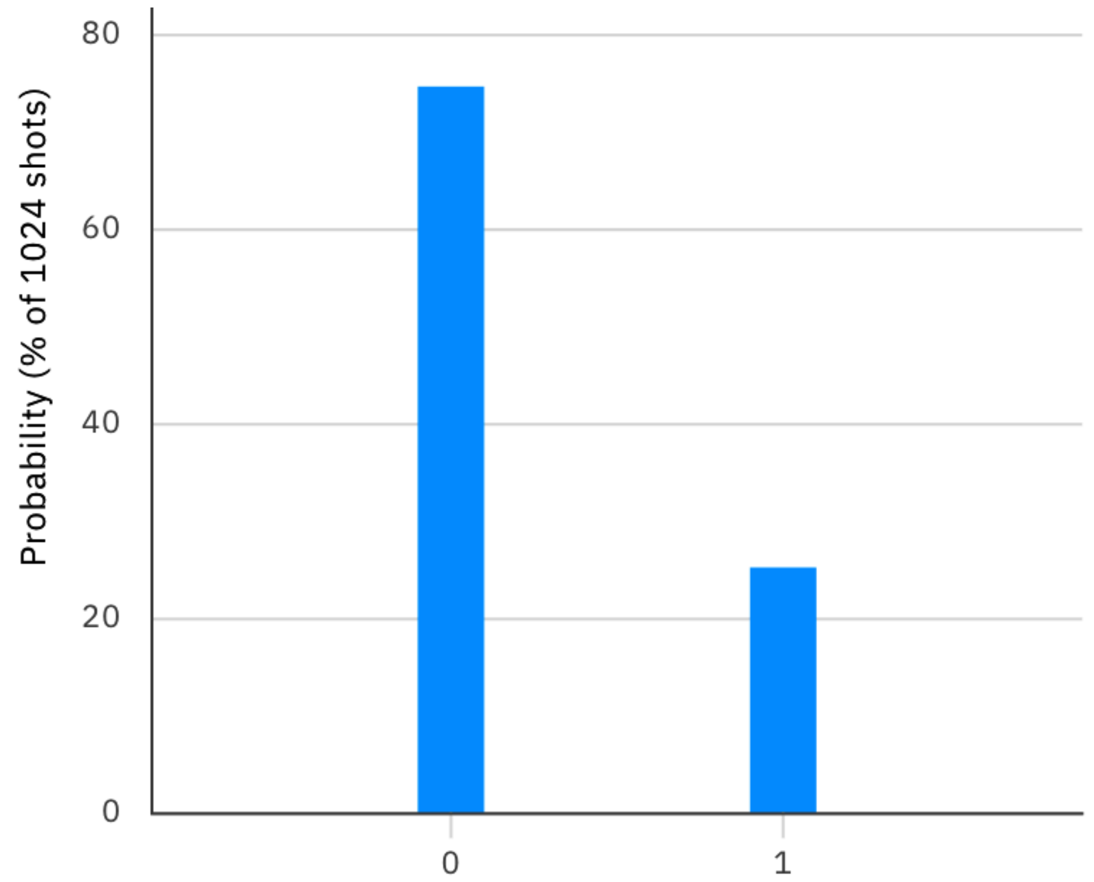
Linear transformation under unitary matrix

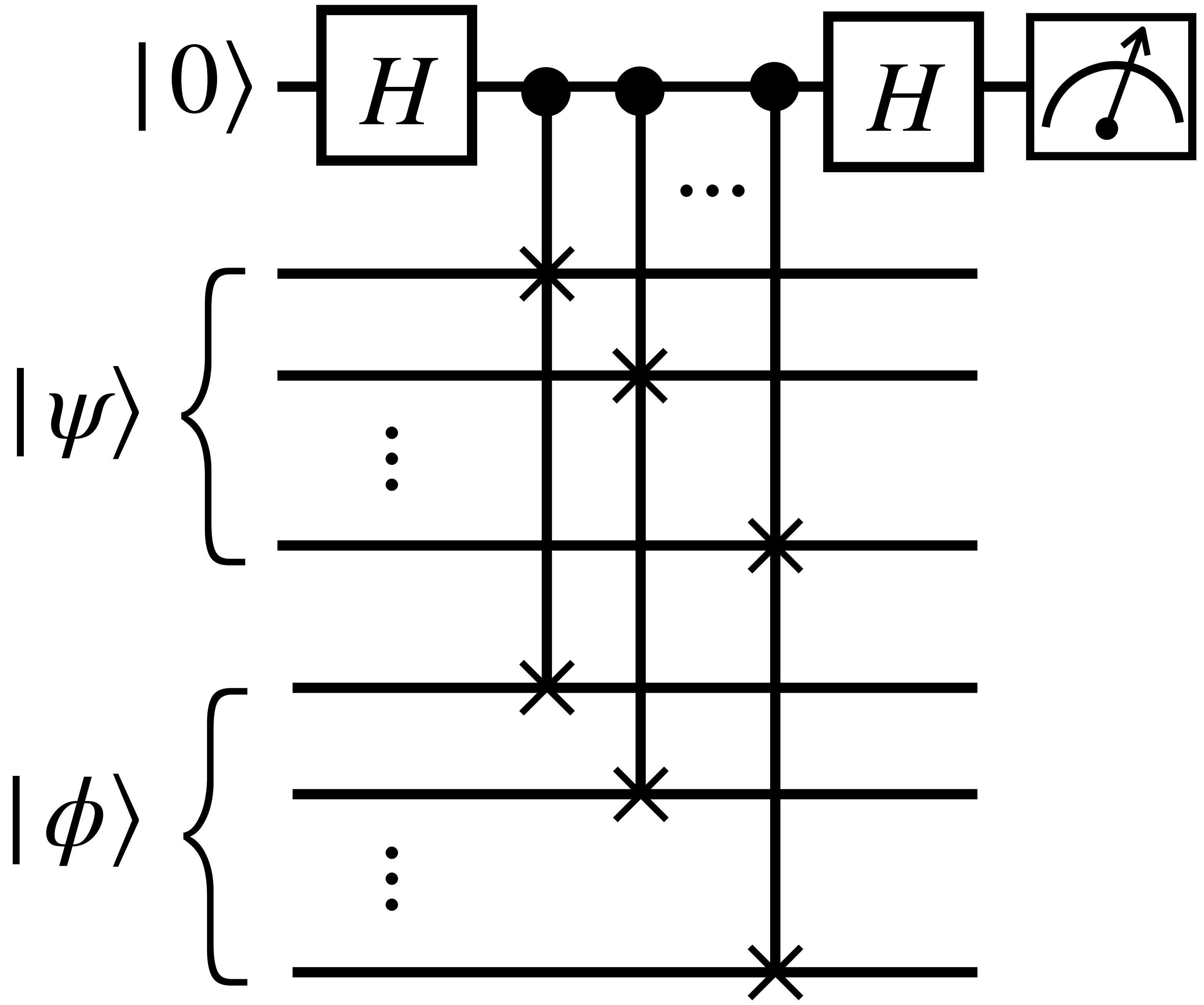
Source: https://www.science.org/doi/10.1126/science.abb2823

Example

Inner Product Calculation

- Let $|\psi\rangle, |\phi\rangle \in \mathbb{C}^{2^n}$ be two vectors. How to compute the magnitude of the inner product $|\langle\phi|\psi\rangle|^2$?
- Digital:
  - 2^n multiplications & additions
  - Decompose multiplications & additions as NAND gate
- Quantum:
  - Run the following circuit with $2n+1$ qubits and $n+2$ gates
  - $\Pr(0) - \Pr(1) = |\langle\phi|\psi\rangle|^2$





The First Wave of Quantum Machine Learning

PRL 103, 150502 (2009) PHYSICAL REVIEW LETTERS week ending 9 OCTOBER 2009

Quantum Algorithm for Linear Systems of Equations

Aram W. Harrow,[1] Avinatan Hassidim,[2] and Seth Lloyd[3]

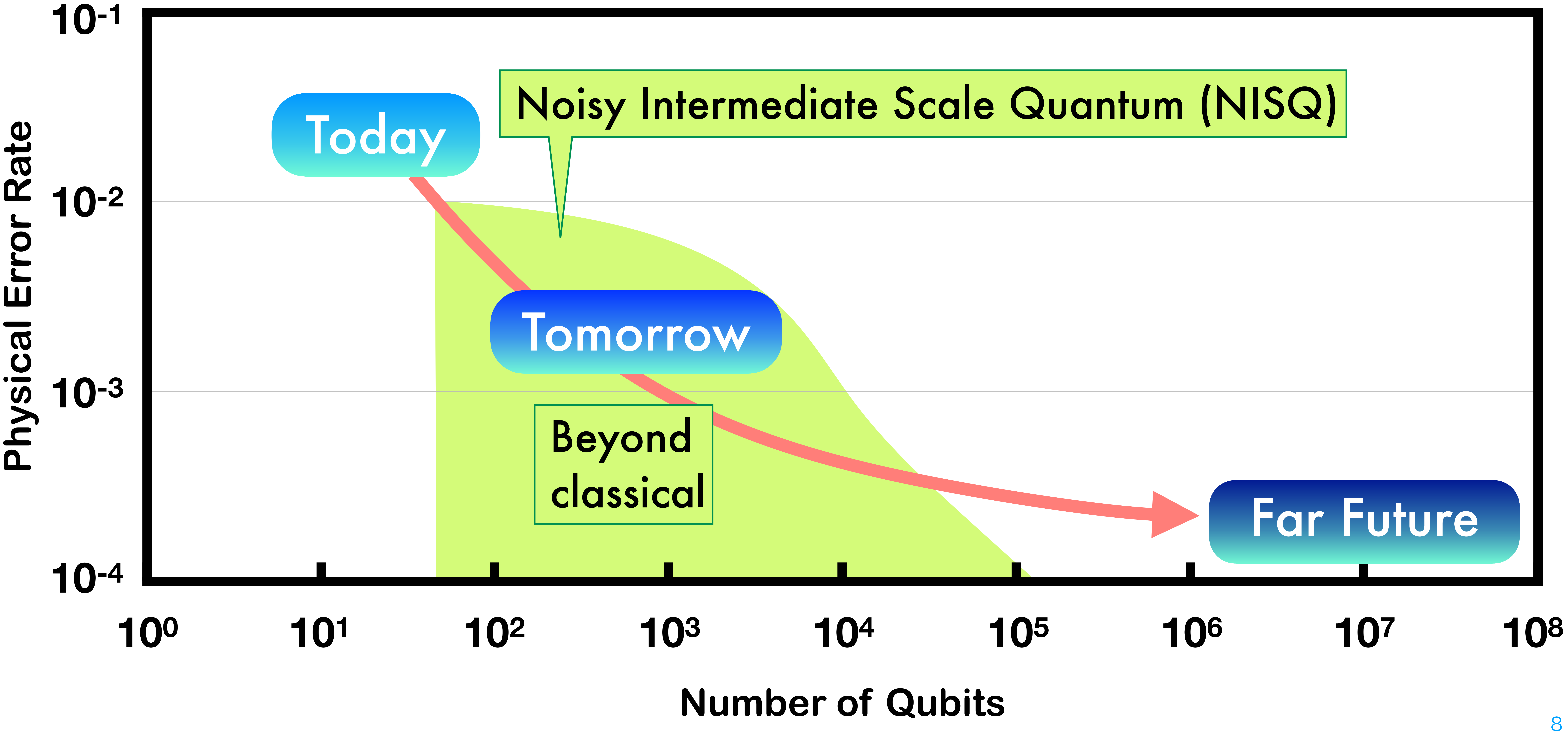
[1]*Department of Mathematics, University of Bristol, Bristol, BS8 1TW, United Kingdom*
[2]*Research Laboratory for Electronics, MIT, Cambridge, Massachusetts 02139, USA*
[3]*Research Laboratory for Electronics and Department of Mechanical Engineering, MIT, Cambridge, Massachusetts 02139, USA*

(Received 5 July 2009; published 7 October 2009)

Solving linear systems of equations is a common problem that arises both on its own and as a subroutine in more complex problems: given a matrix A and a vector $\vec{b}$, find a vector $\vec{x}$ such that $A\vec{x} = \vec{b}$. We consider the case where one does not need to know the solution $\vec{x}$ itself, but rather an approximation of the expectation value of some operator associated with $\vec{x}$, e.g., $\vec{x}^\dagger M\vec{x}$ for some matrix M. In this case, when A is sparse, $N \times N$ and has condition number κ, the fastest known classical algorithms can find $\vec{x}$ and estimate $\vec{x}^\dagger M\vec{x}$ in time scaling roughly as $N\sqrt{\kappa}$. Here, we exhibit a quantum algorithm for estimating $\vec{x}^\dagger M\vec{x}$ whose runtime is a polynomial of $\log(N)$ and κ. Indeed, for small values of κ [i.e., poly $\log(N)$], we prove (using some common complexity-theoretic assumptions) that any classical algorithm for this problem generically requires exponentially more time than our quantum algorithm.

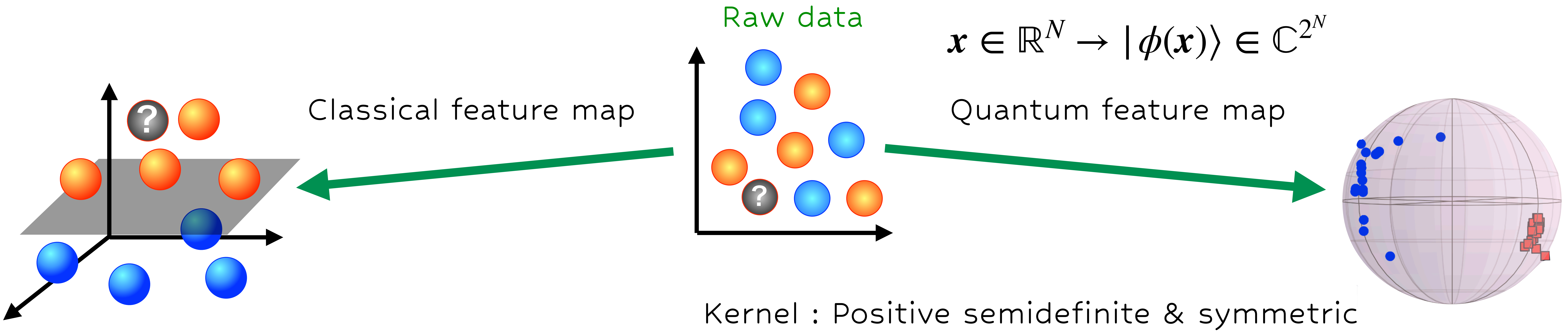
Quantum Computer is Hard to Build

Quantum Hardware Roadmap



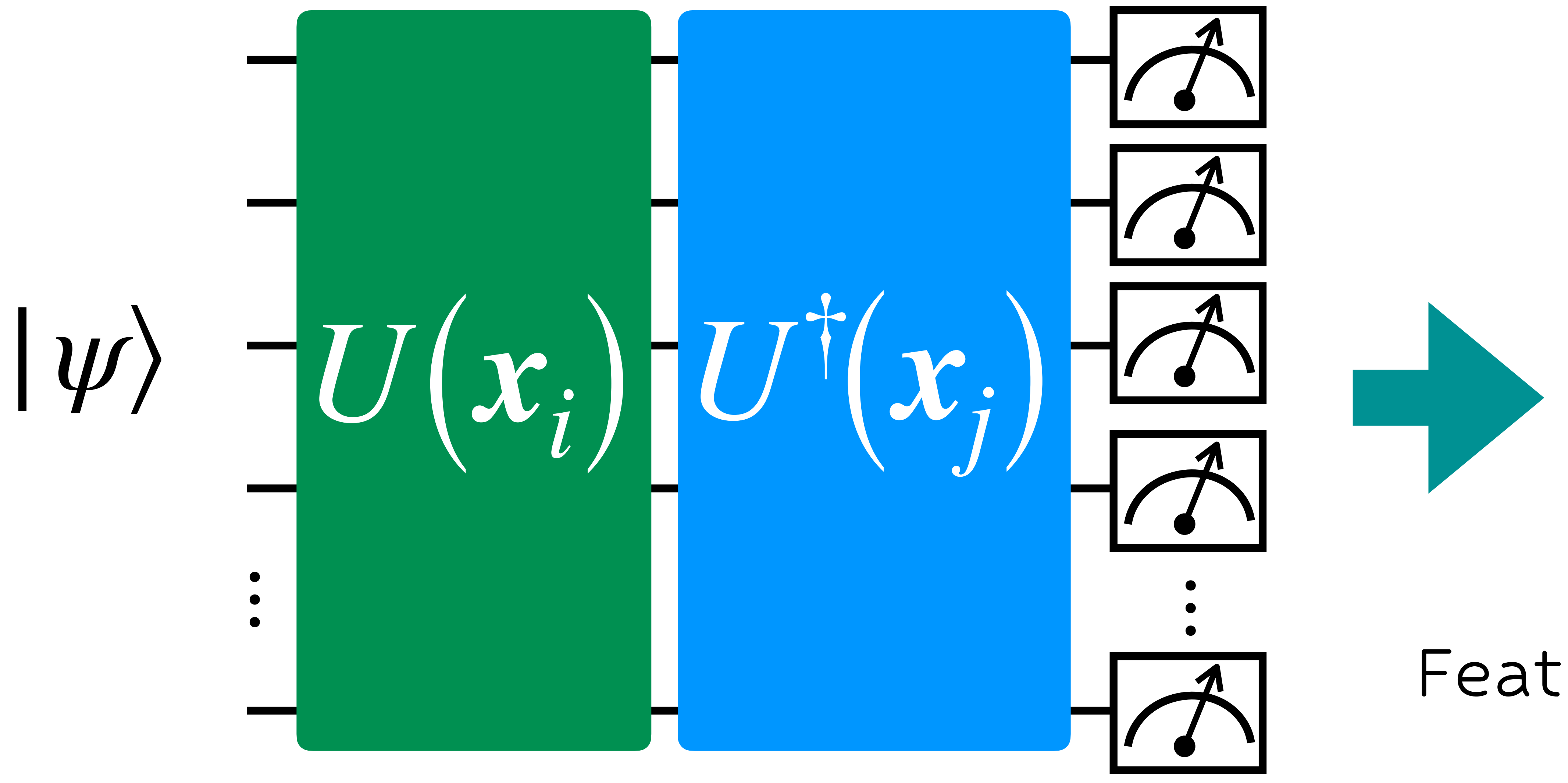
Kernel Method

Quantum Feature Map



$$\boldsymbol{x} \in \mathbb{R}^N \rightarrow |\phi(\boldsymbol{x})\rangle \in \mathbb{C}^{2^N}$$

Kernel : Positive semidefinite & symmetric functions that quantify the similarity between data points



$$|\langle\psi|\, U^\dagger(\boldsymbol{x}_j)\, U(\boldsymbol{x}_i)\, |\psi\rangle|^2 = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$$

Feature maps and kernels are automatically given in QC

Quantum Feature Map

Simple Examples

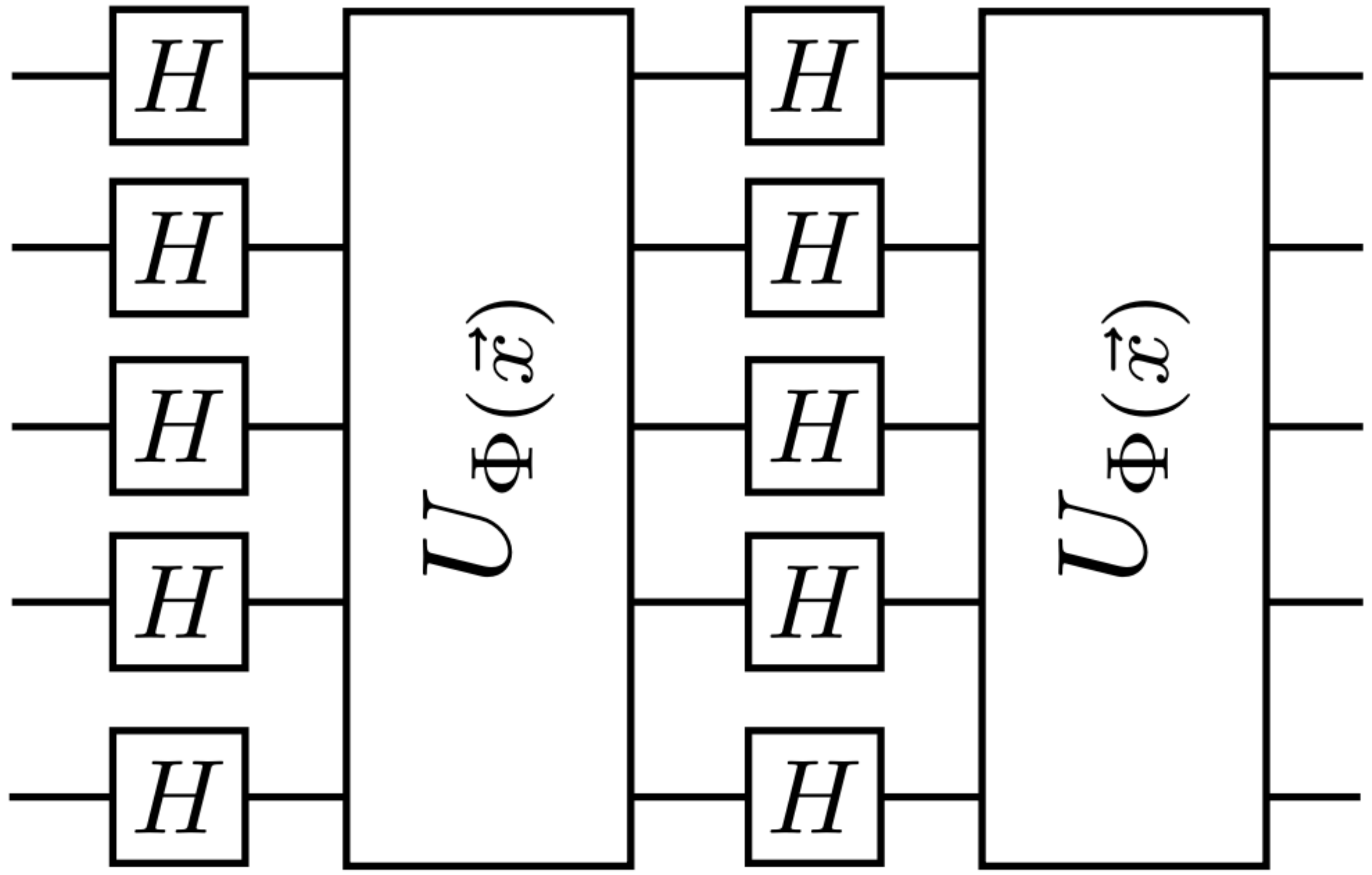
- Amplitude encoding with multiple copies
  - $\boldsymbol{x} \in \mathbb{R}^N \rightarrow |\phi(\boldsymbol{x})\rangle^{\otimes d} = \left(\sum_{i=1}^{N} x_i |i\rangle\right)^{\otimes d}$ using $d\lceil \log_2(N) \rceil$ qubits.
  - $k(\boldsymbol{x}, \boldsymbol{y}) = \langle \phi(\boldsymbol{x}) | \phi(\boldsymbol{y}) \rangle \cdots \langle \phi(\boldsymbol{x}) | \phi(\boldsymbol{y}) \rangle = \left(\boldsymbol{x}^\top \boldsymbol{y}\right)^d$
- Qubit (product) encoding
  - $\boldsymbol{x} \in \mathbb{R}^N \rightarrow |\phi(\boldsymbol{x})\rangle = \bigotimes_{i=1}^{N} \left(\cos\left(x_i\right)|0\rangle + \sin\left(x_i\right)|1\rangle\right)$ using N qubits.
  - $k(\boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{N} \cos(x_i - y_i)$

Quantum Feature Map

Simple Examples

- General strategy for a quantum advantage:
  - Use a quantum feature map that is hard to simulate classically.
  - Popular example:

$\mathscr{U}(\boldsymbol{x}) =$



- $U_\Phi(\vec{x}) = \exp\left[i\left(\sum_j \phi_1(x_j)Z_j + \sum_{j<k} \phi_2(x_j, x_k)Z_jZ_k\right)\right]$

  with some functions ϕ_1 and ϕ_2

- Typically, ϕ_2 is a nonlinear function

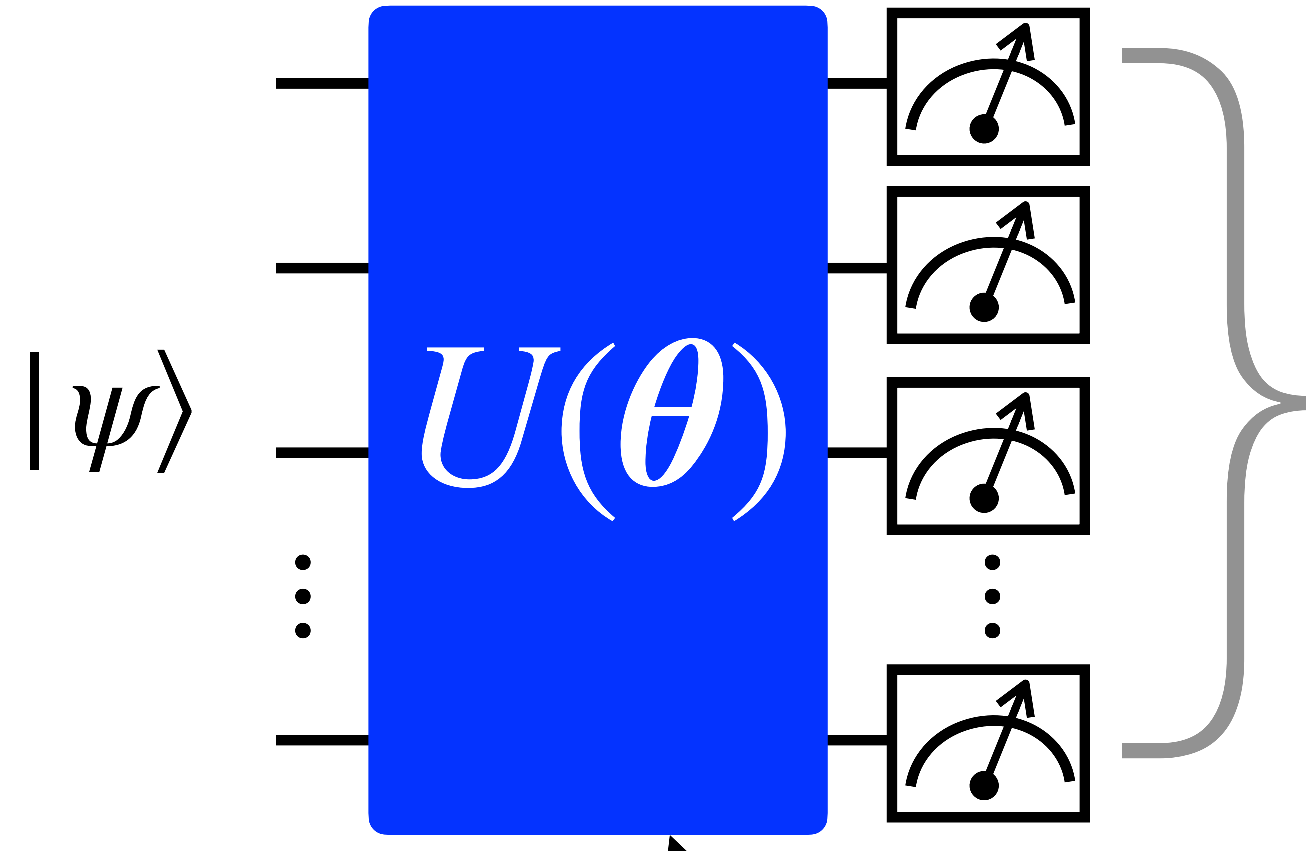
$$k(\boldsymbol{x}, \boldsymbol{y}) = |\langle 0|\left(\mathscr{U}^\dagger(\boldsymbol{y})\right)^d\left(\mathscr{U}(\boldsymbol{x})\right)^d|0\rangle|^2$$

Havlíček et al. Nature 567, 209–212 (2019)

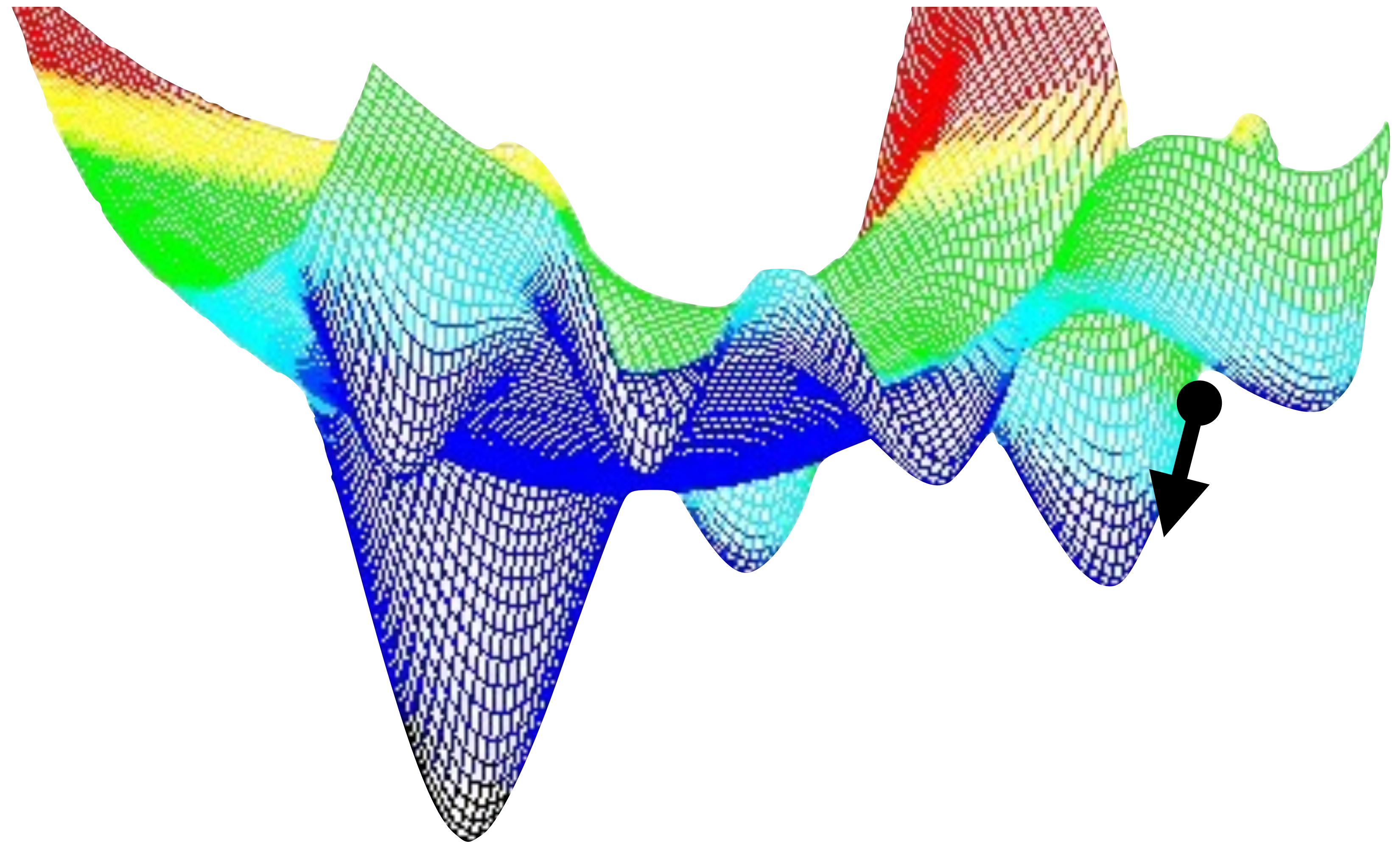
Parameterized Quantum Circuit & Analytical Gradient

Basic Idea

$$h(\boldsymbol{\theta}) = \mathrm{Tr}\left(HU(\boldsymbol{\theta})|\psi\rangle\langle\psi|U^{\dagger}(\boldsymbol{\theta})\right),\ H = H^{\dagger}$$

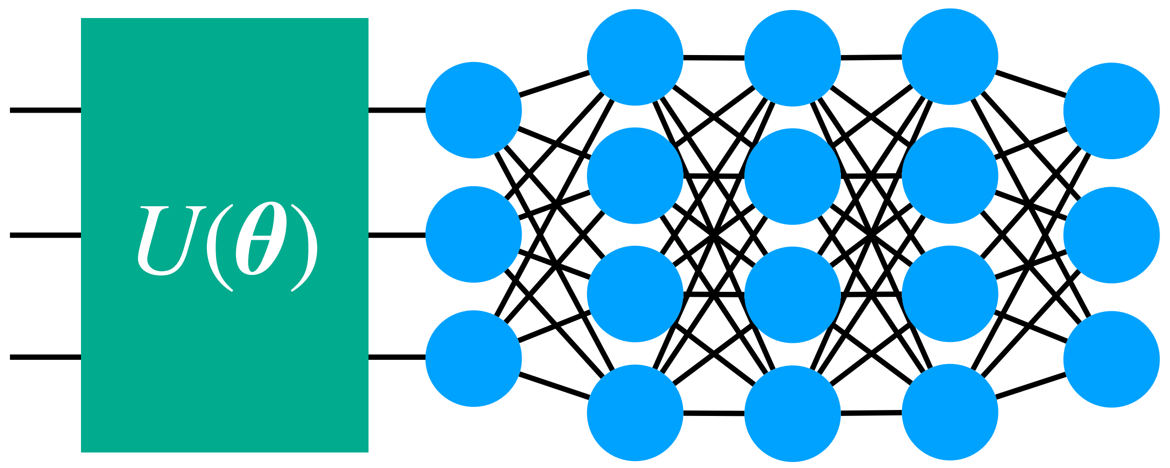


$$h(\boldsymbol{\theta}) \longrightarrow L(\boldsymbol{\theta}) = f\left(h(\boldsymbol{\theta})\right)$$



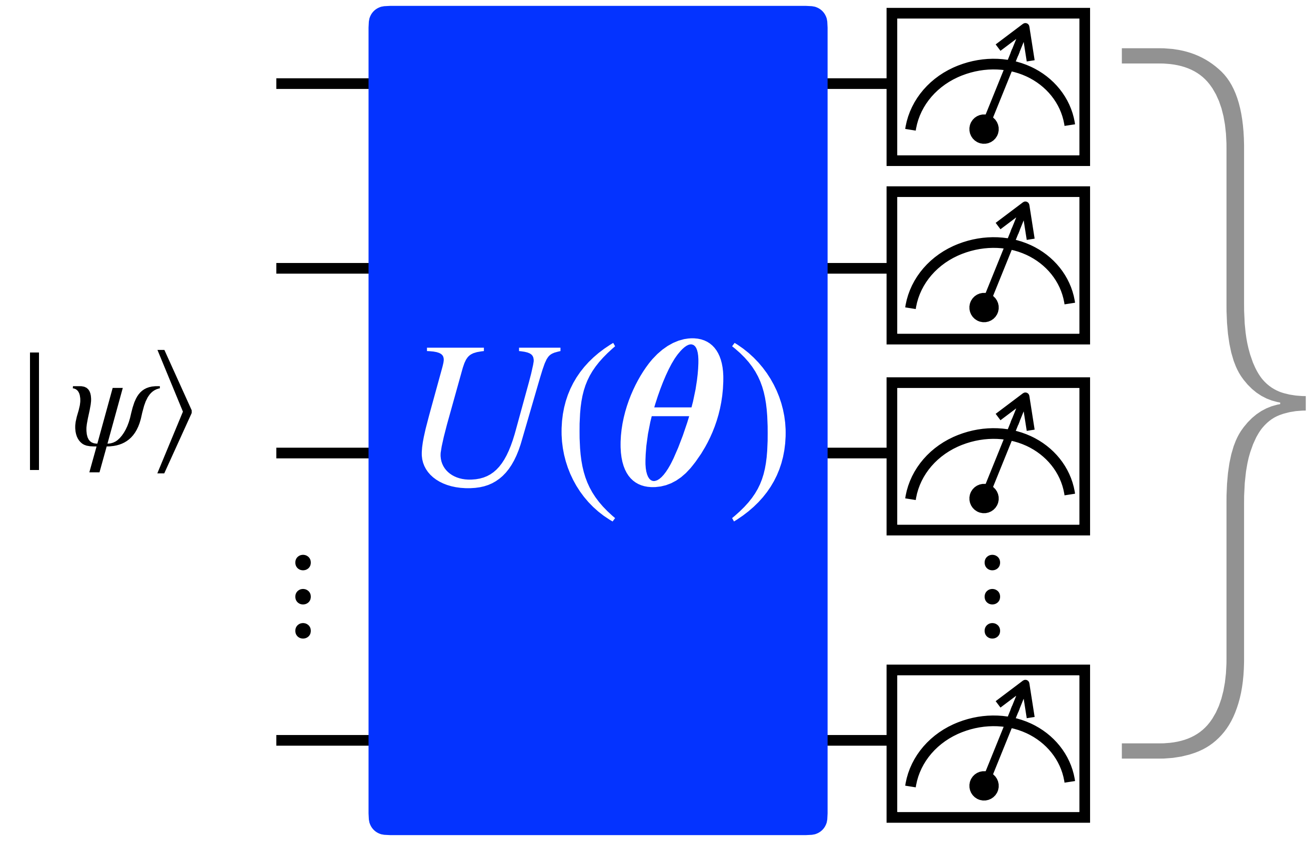
Update $\boldsymbol{\theta} \leftarrow \arg\min_{\boldsymbol{\theta}'} f(h(\boldsymbol{\theta}'))$

$\dfrac{\partial h(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ can be computed directly on a quantum computer!



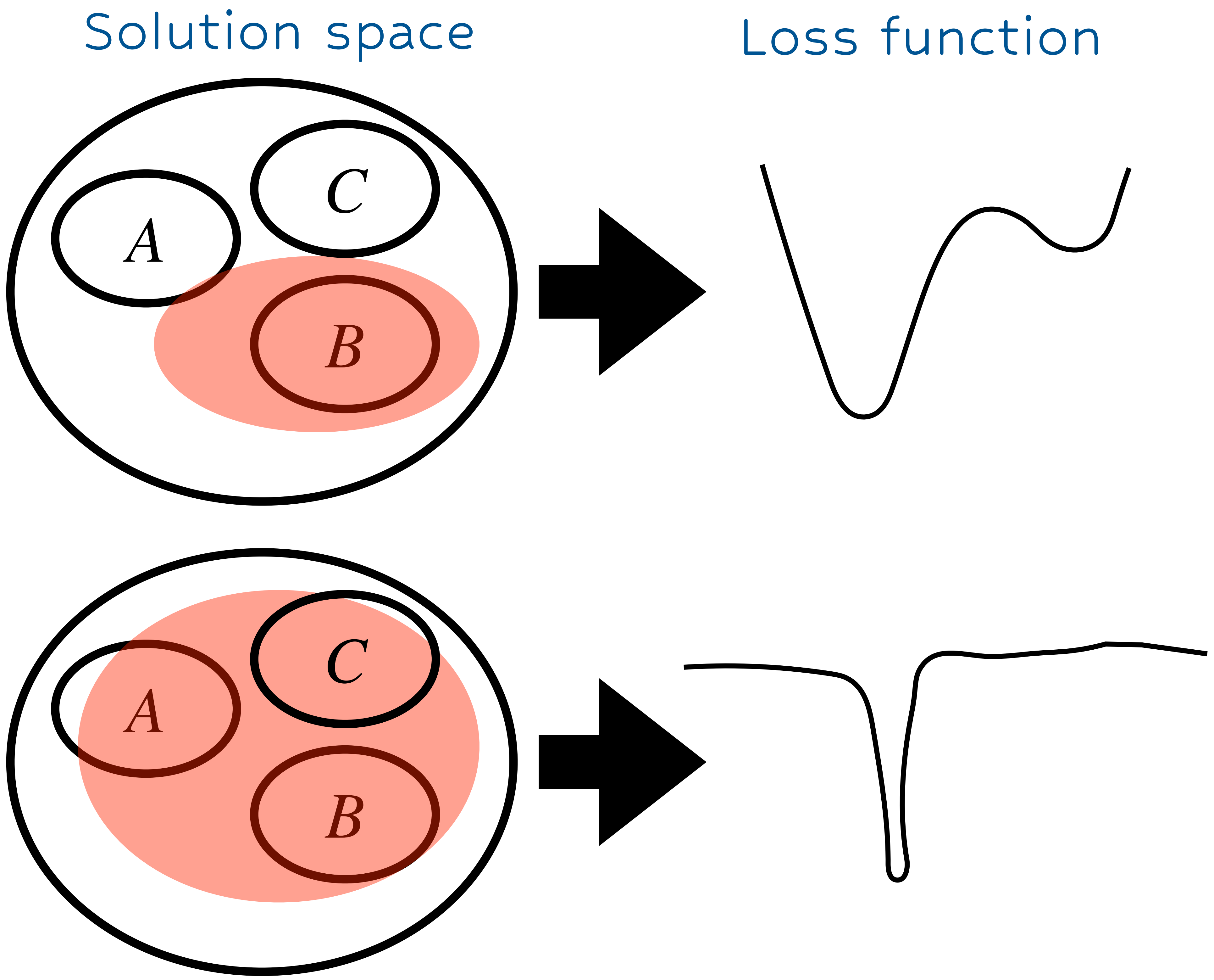
Parameterized Quantum Circuit & Analytical Gradient

The Curse of Barren Plateaus



$$h(\boldsymbol{\theta}) = \mathrm{Tr}\left(HU(\boldsymbol{\theta})\,|\psi\rangle\langle\psi|\,U^{\dagger}(\boldsymbol{\theta})\right),\ H = H^{\dagger}$$

$$h(\boldsymbol{\theta}) \longrightarrow L(\boldsymbol{\theta}) = f\left(h(\boldsymbol{\theta})\right)$$



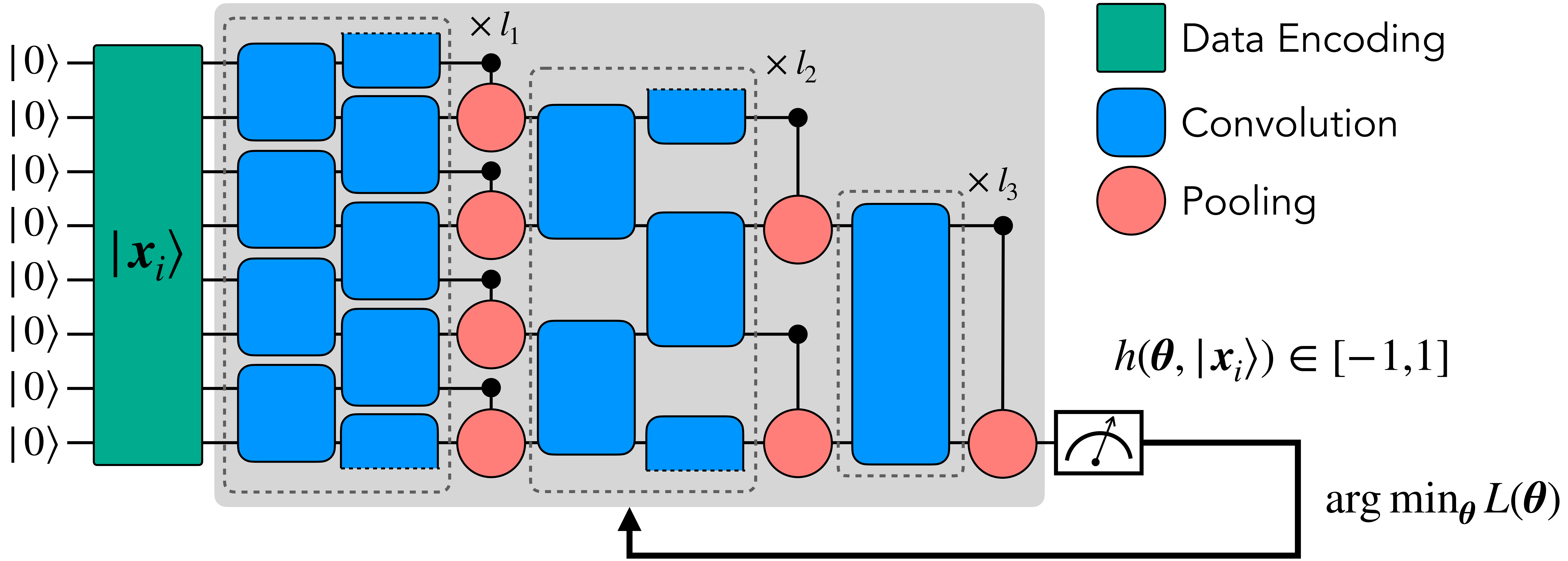
Barren plateaus in layman's terms:

If parameterized quantum circuit is able to sample uniformly from all unitary operations (a.k.a Haar random), then the gradient decreases exponentially with the number of qubits

Can we design a quantum circuit that avoids this?

Quantum Convolutional Neural Network

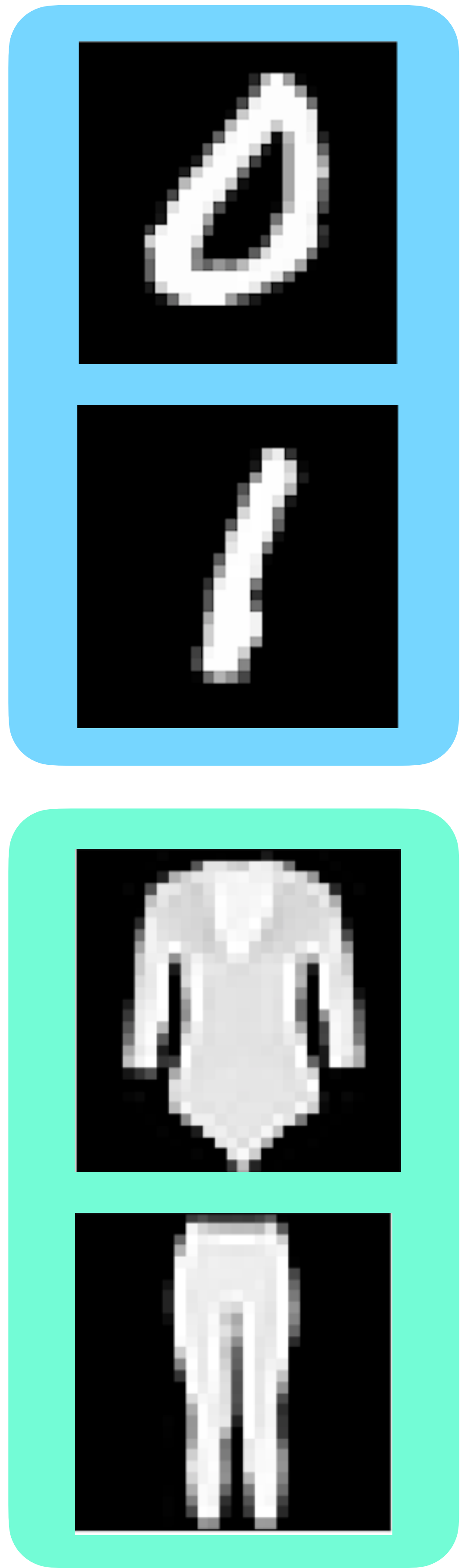
Basic Framework

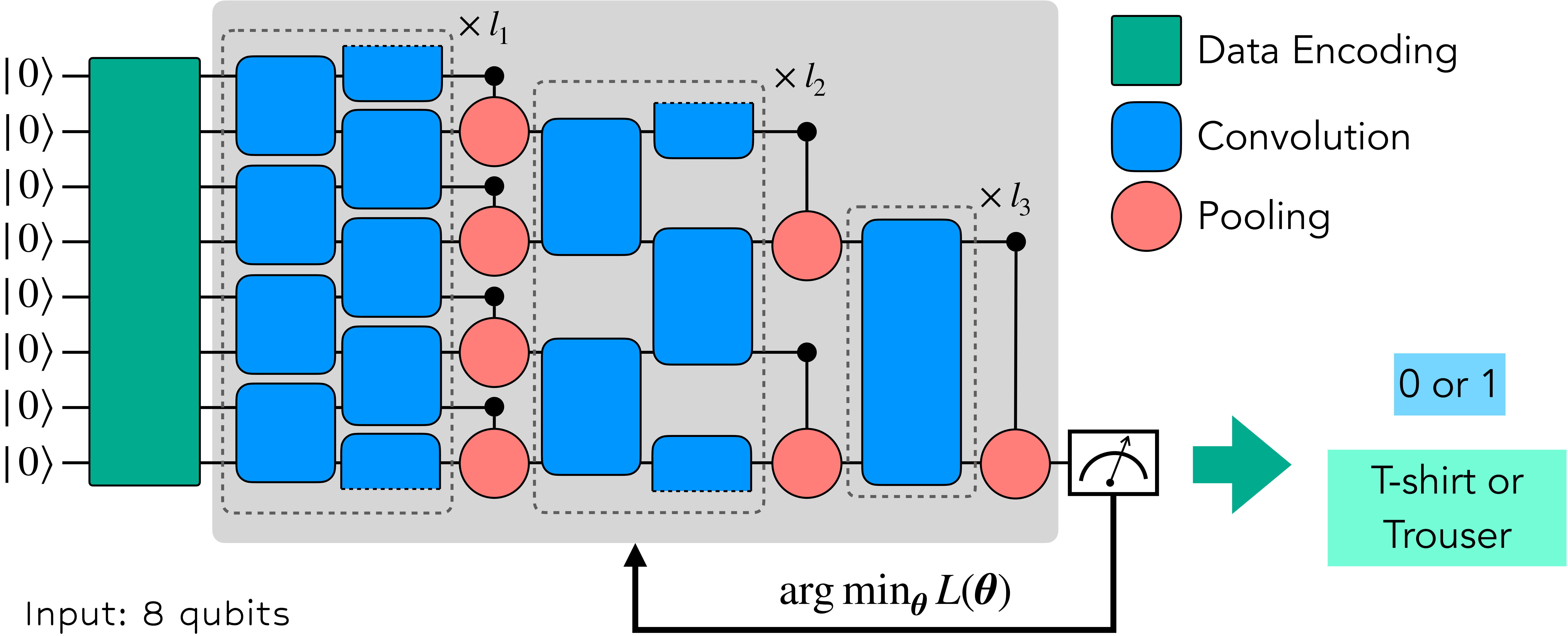


- MSE loss as an example: $L(\boldsymbol{\theta}) = \sum_{i=1}^{m} \left(y_i - h(\boldsymbol{\theta}, |\boldsymbol{x}_i\rangle)\right)^2$
- With n input qubits: 2^n features & $\log(n)$ layers, $O(\log(n))$ parameters
- Quantum entanglement beyond local correlation

Quantum Convolutional Neural Network

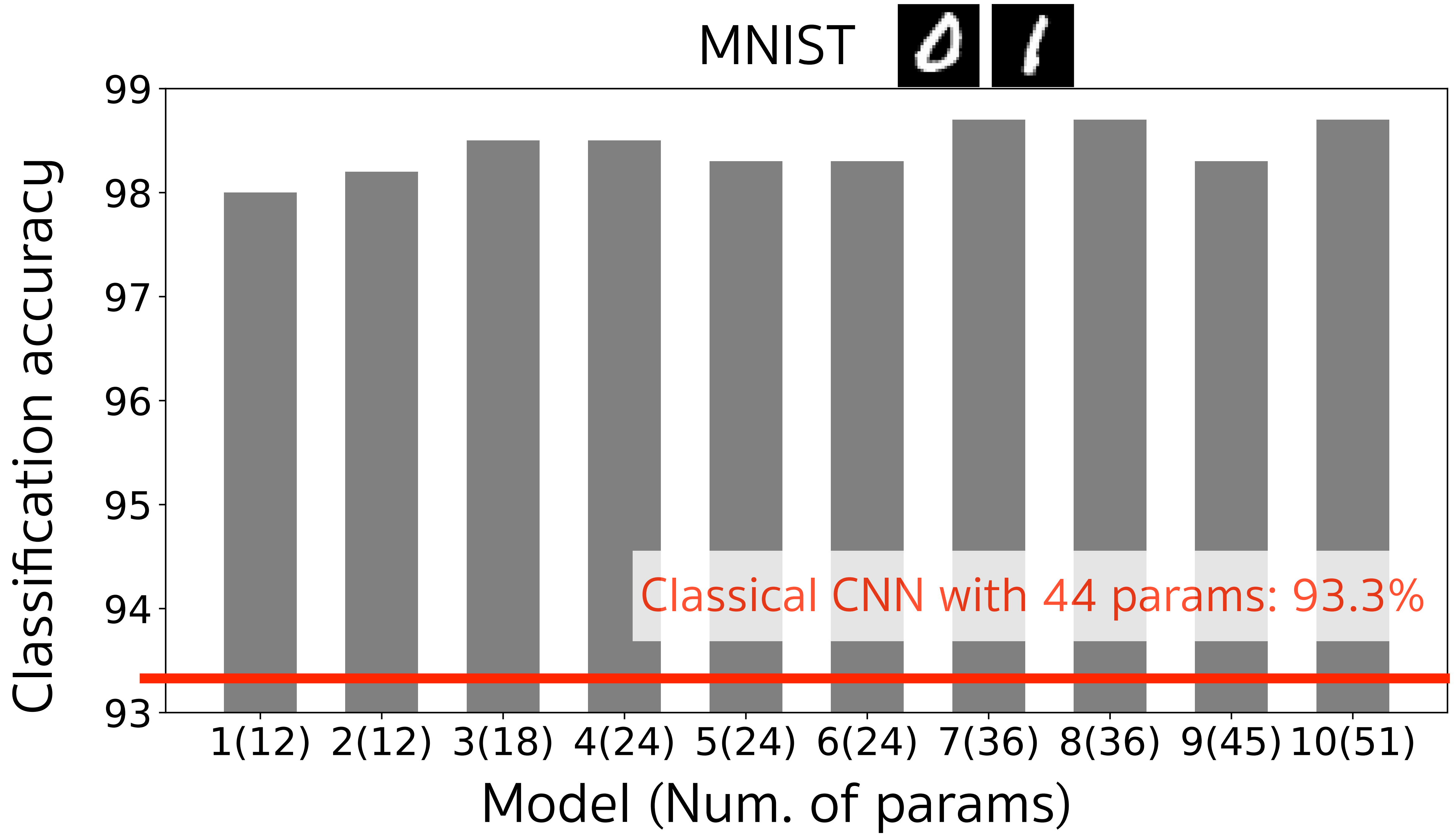
Classical Data Classification

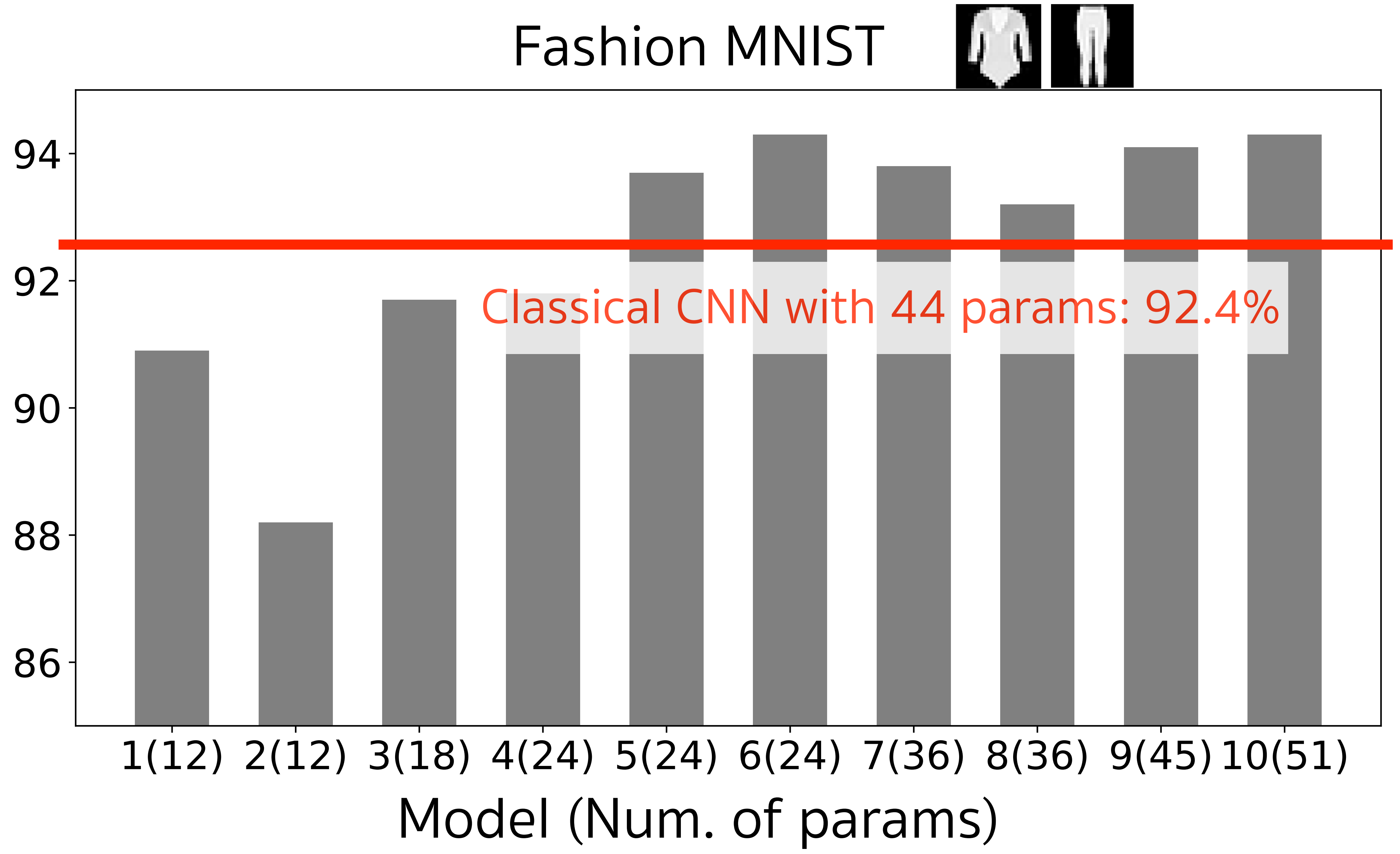


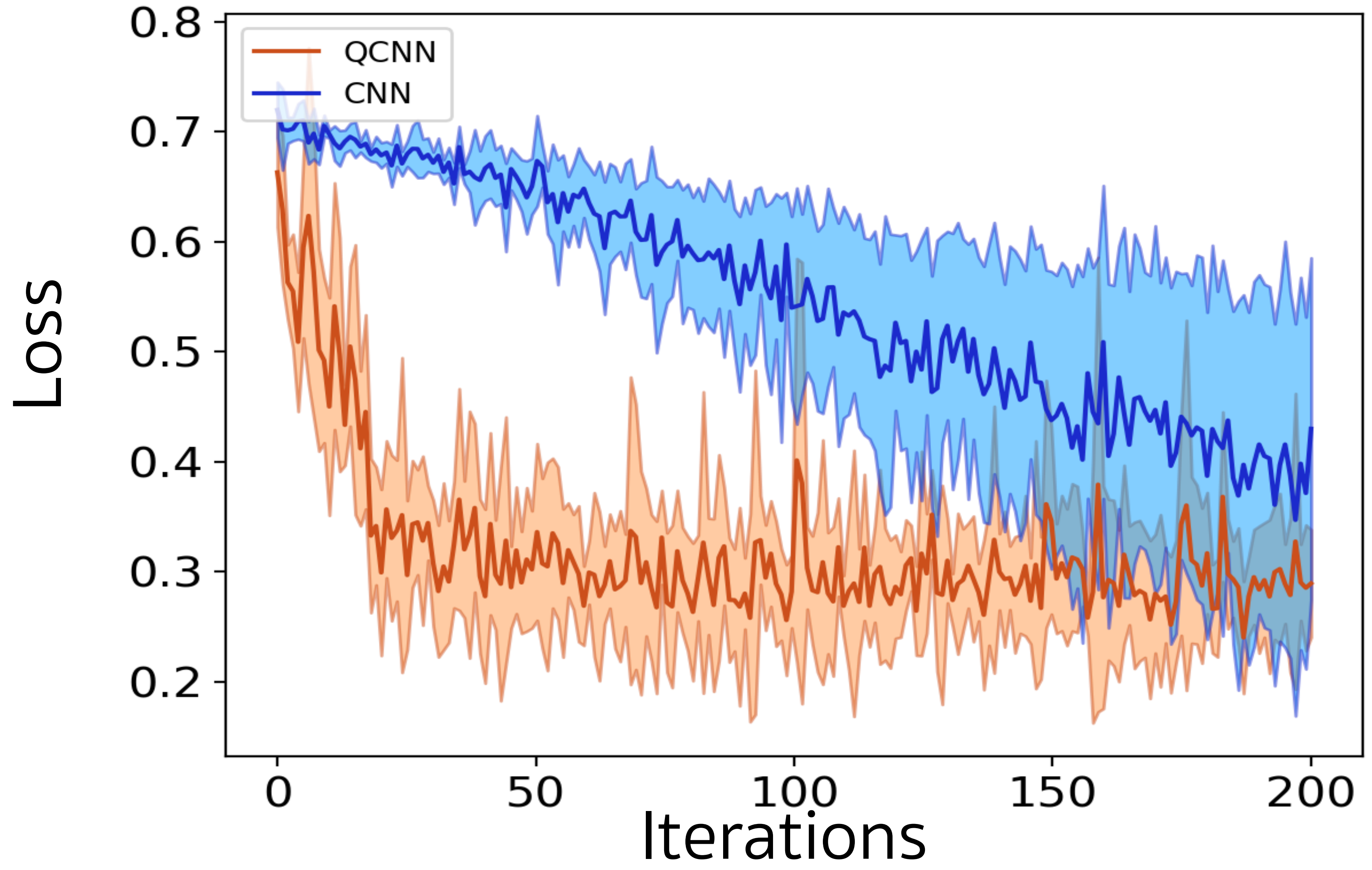


- Input: 8 qubits
- Total number of parameters: 12 ~ 51
- Training: 12000 / Test: 2000

QCNN for Classical Data Classification





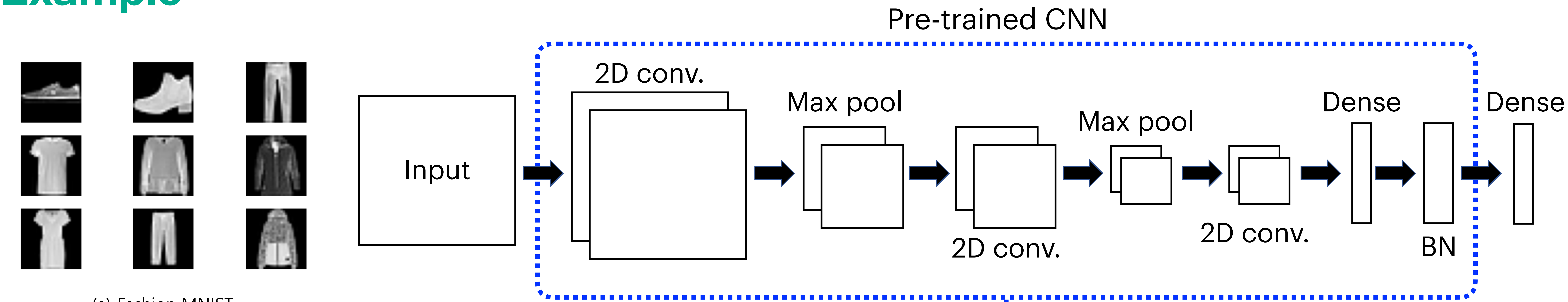


- Higher accuracy in the <u>few parameter regime</u>
- Trains faster
- Will these improvements continue to hold in larger systems? To be verified.

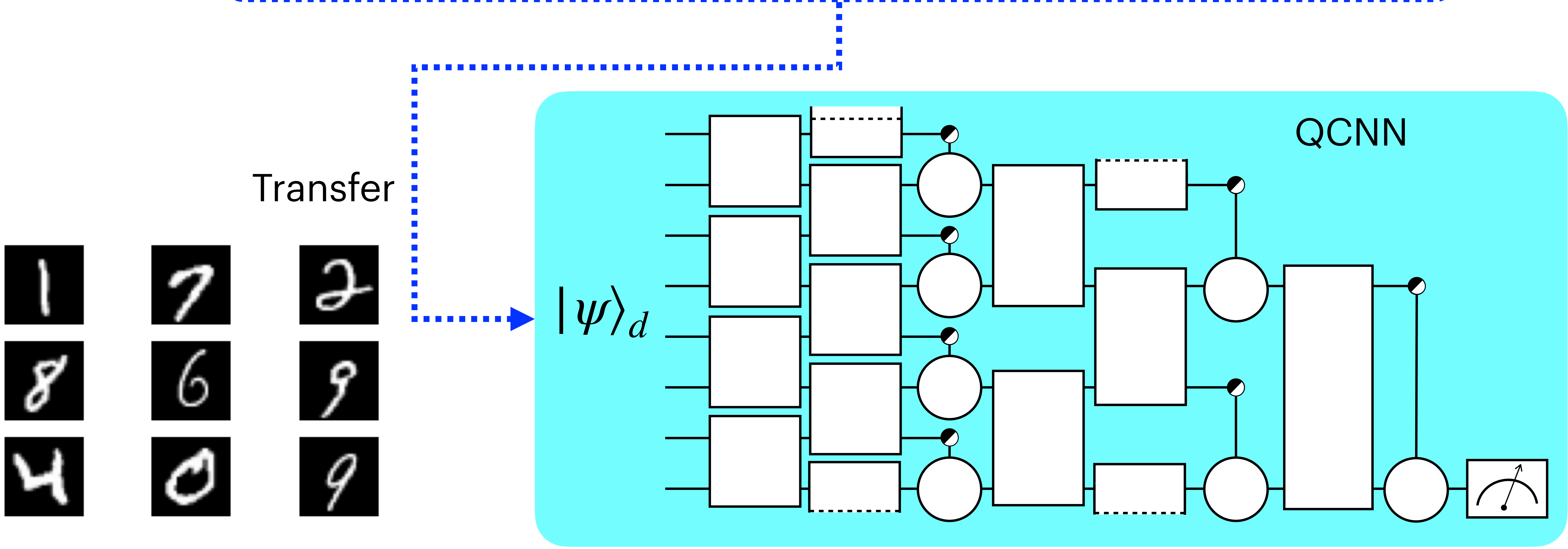
T. Hur, L. Kim, D. K. Park. Quantum Machine Intelligence **4** 3 (2022)

Classical-to-Quantum Transfer Learning

Example



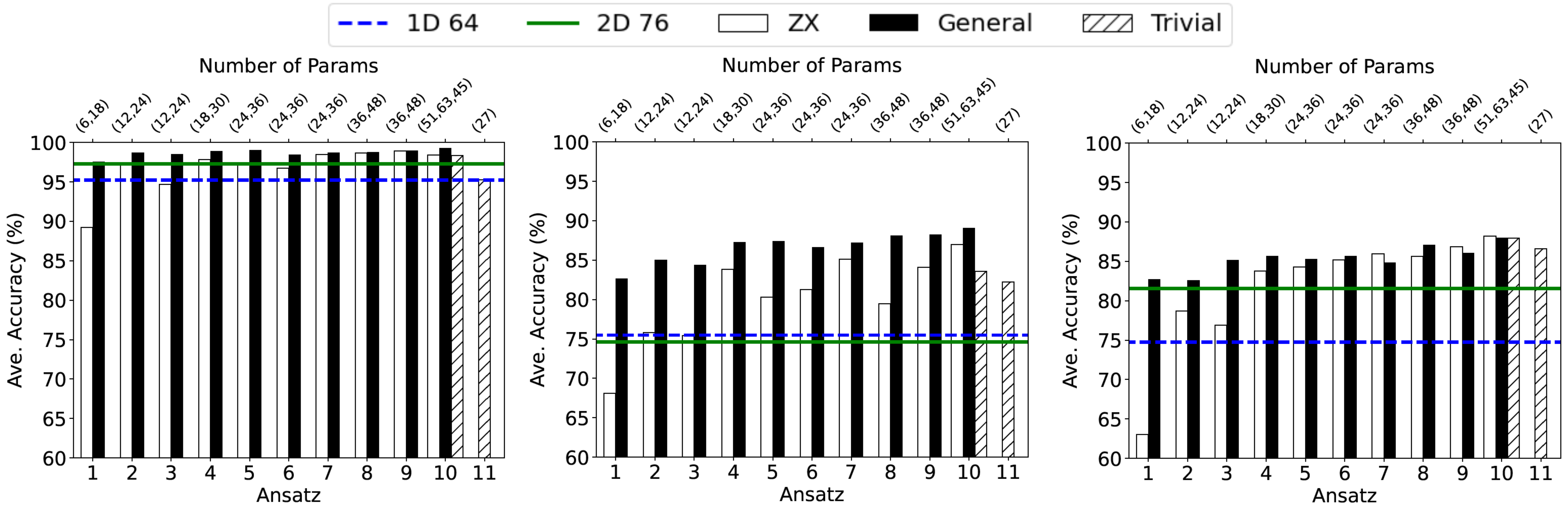
(a) Fashion MNIST



(b) MNIST

Classical-to-Quantum Transfer Learning

Benchmarking Example



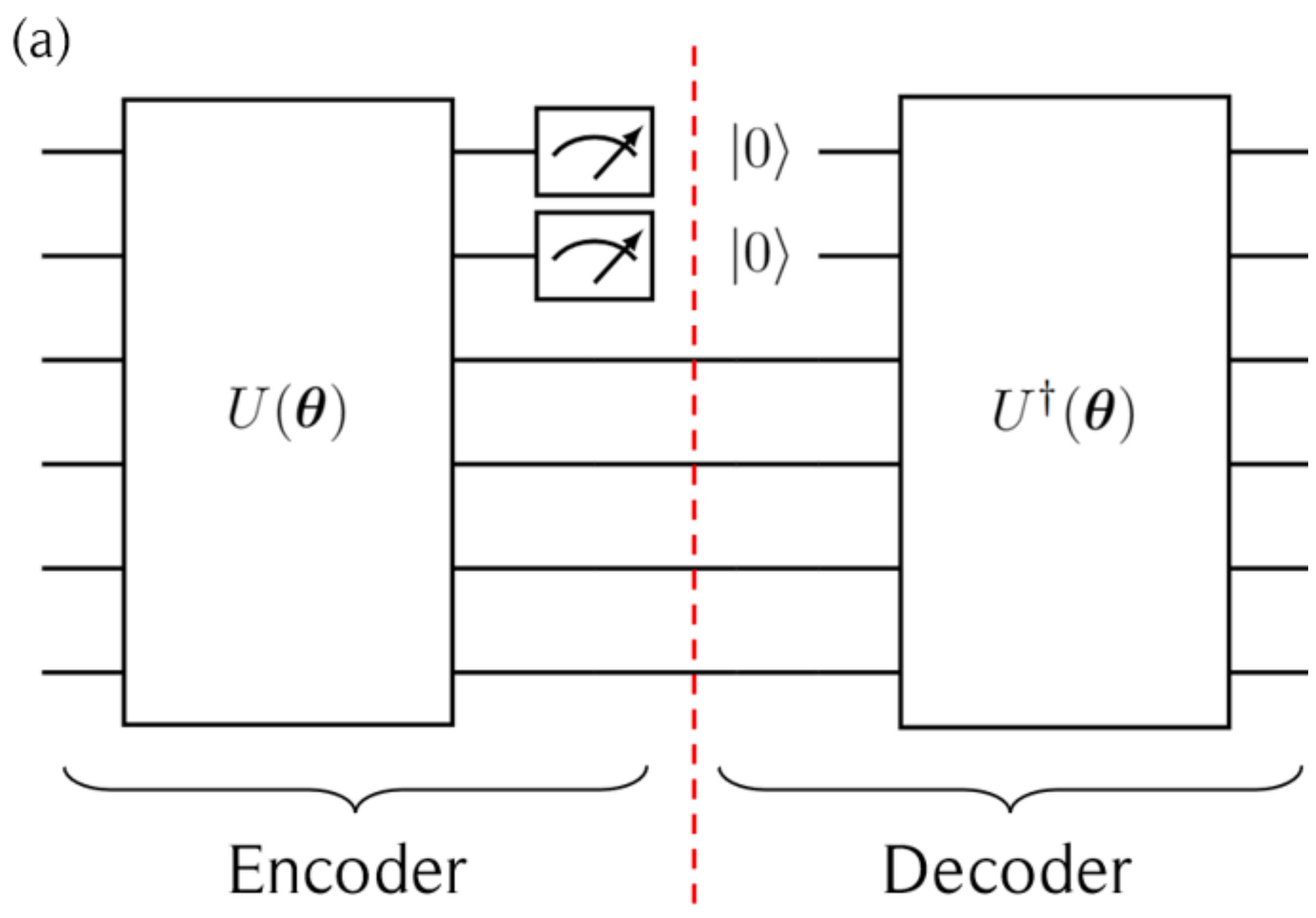
(a) 0 vs 1 (b) 2 vs 3 (c) 8 vs 9

J. Kim, J. Huh, D. K. Park arXiv:2208.14708 **[quant-ph]**

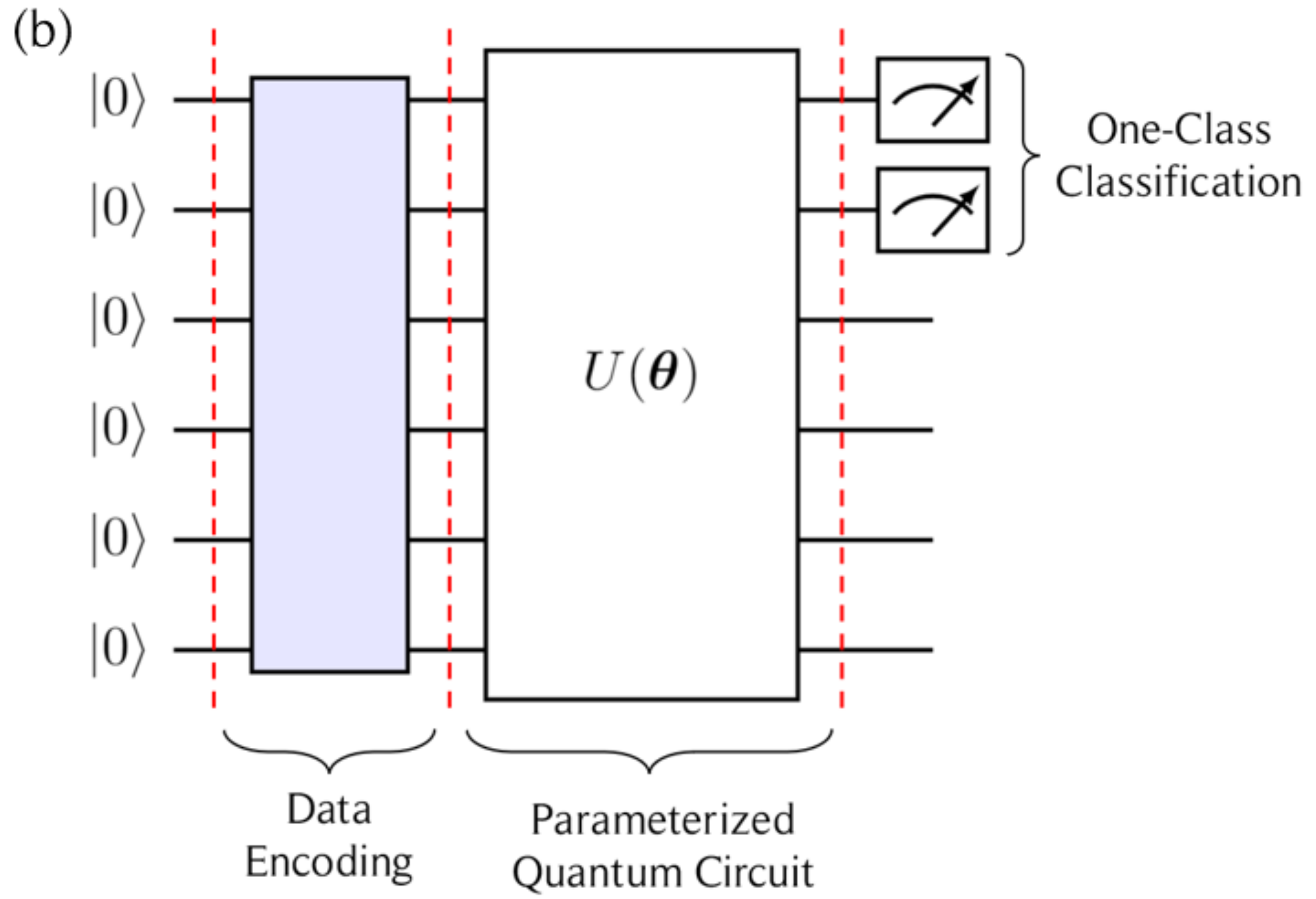
Quantum Semi-Supervised Learning

One-Class Classification

- A typical dataset in supervised classification: $\mathscr{D} = \left\{(\vec{x}_1, y_1), \ldots, (\vec{x}_M, y_M)\right\} \subset \mathbb{C}^N \times \mathbb{Z}_l$
- One-class classification: $\mathscr{D} = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_M\}$
- Problem: Is a new data point $\tilde{x}$ normal or not?



Quantum Autoencoder



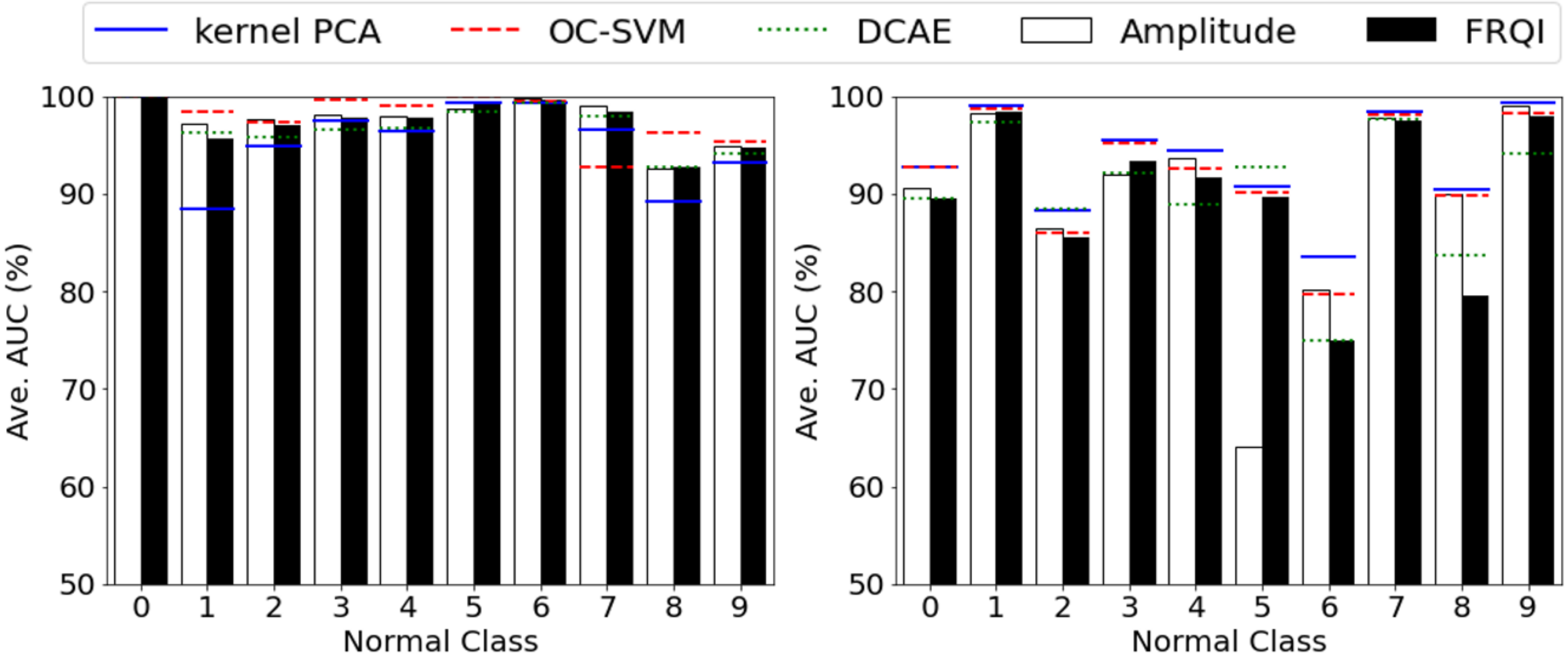
Quantum OCC

Quantum Semi-Supervised Learning

One-Class Classification

(b) MNIST

(a) Fashion MNIST



(a) Average AUC for Handwritten digit

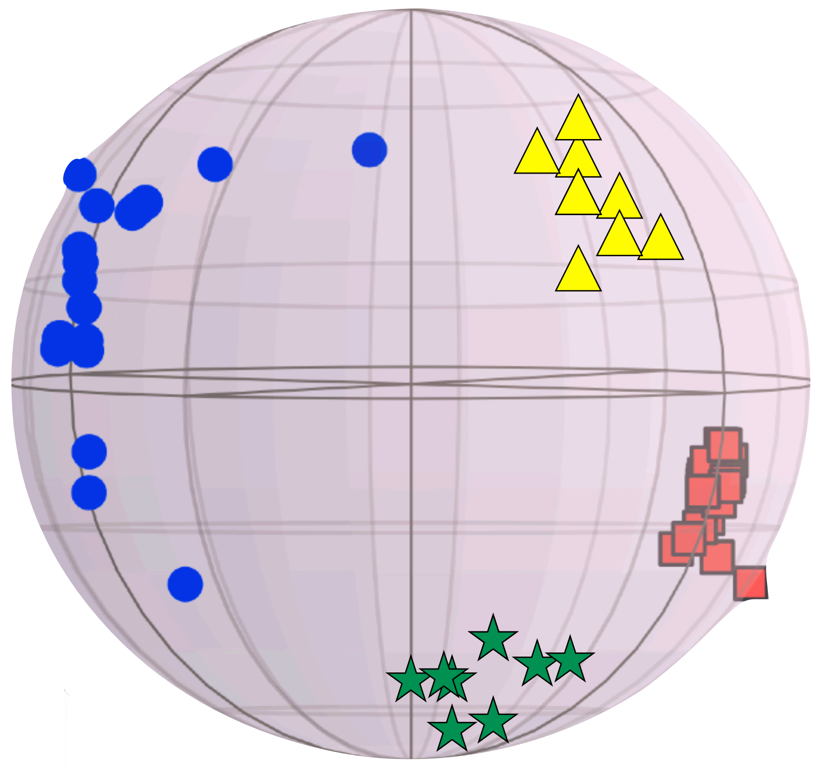
(b) Average AUC for Fashion-MNIST

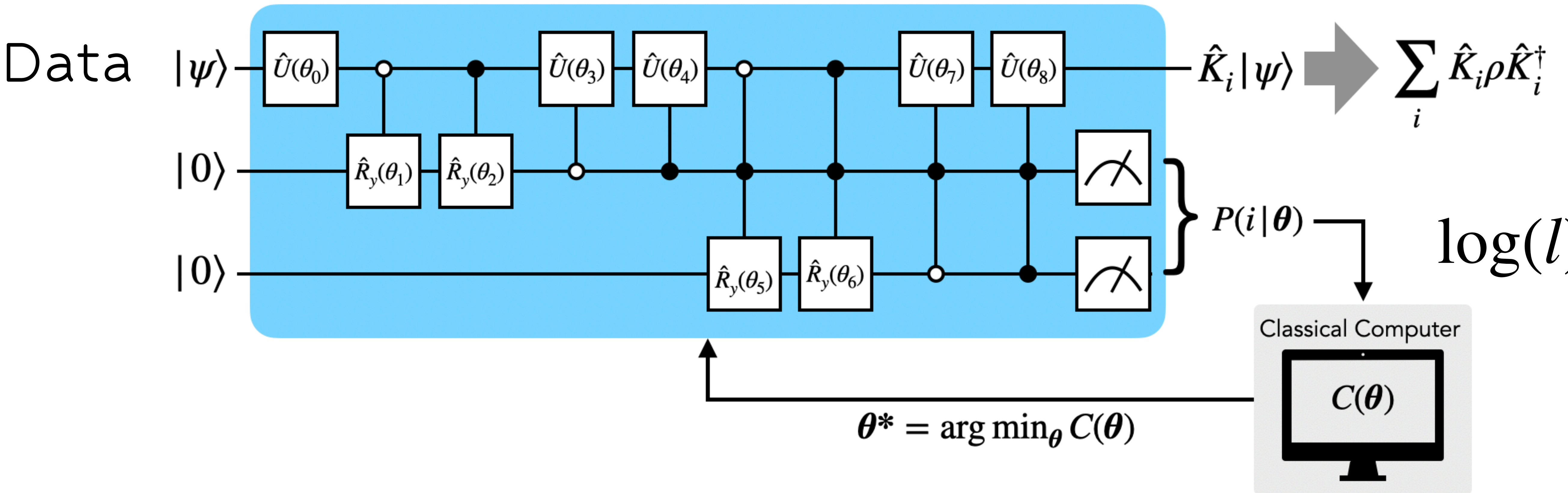
The Quantum model uses an exponentially smaller number of parameters subject to training.

G. Park, J. Huh, D. K. Park. MLST 4 015006 (2023)

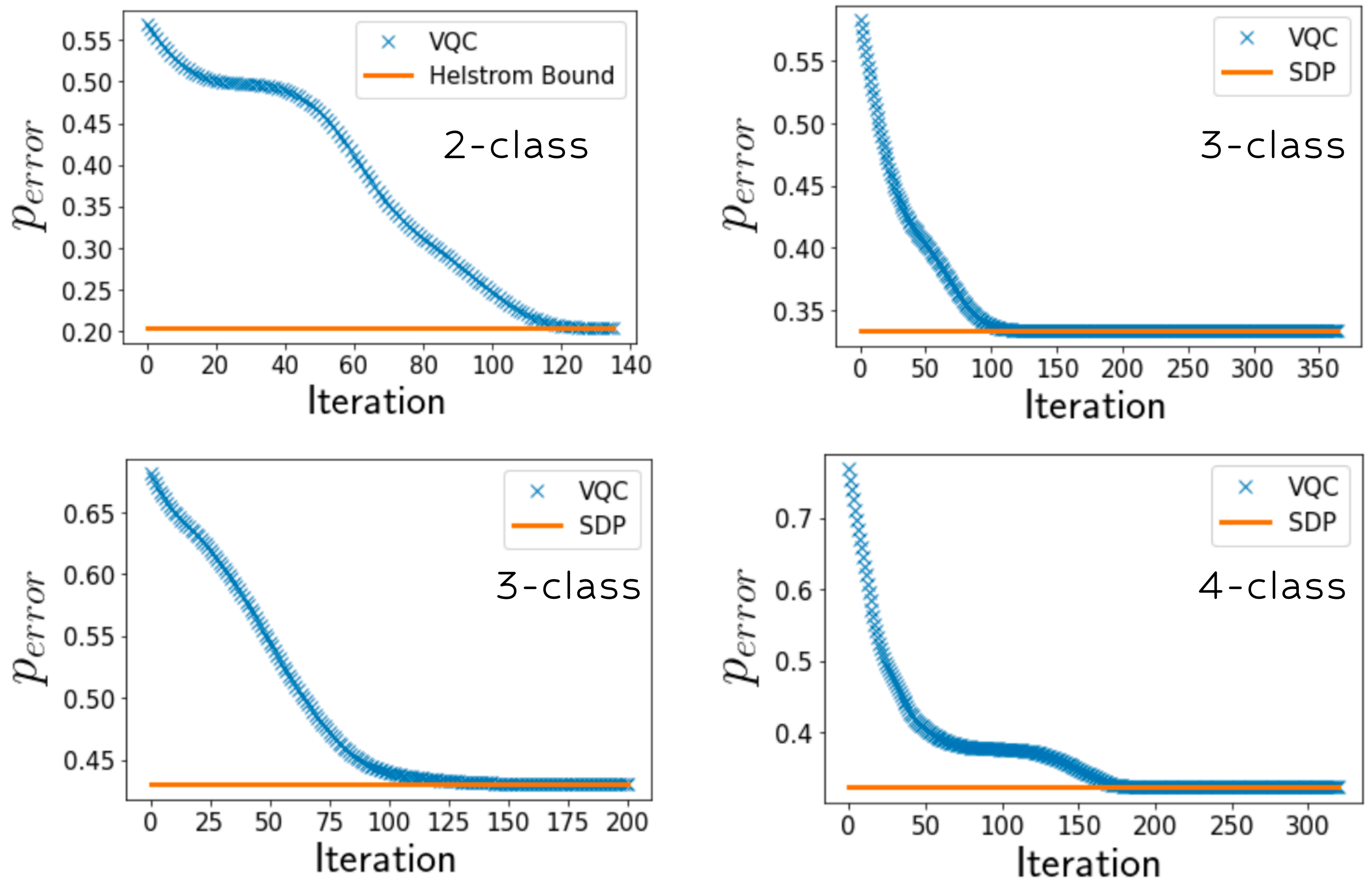
Variational Quantum State Discriminator

Positive Operator-Valued Measurement (POVM) can distinguish non-orthogonal states!

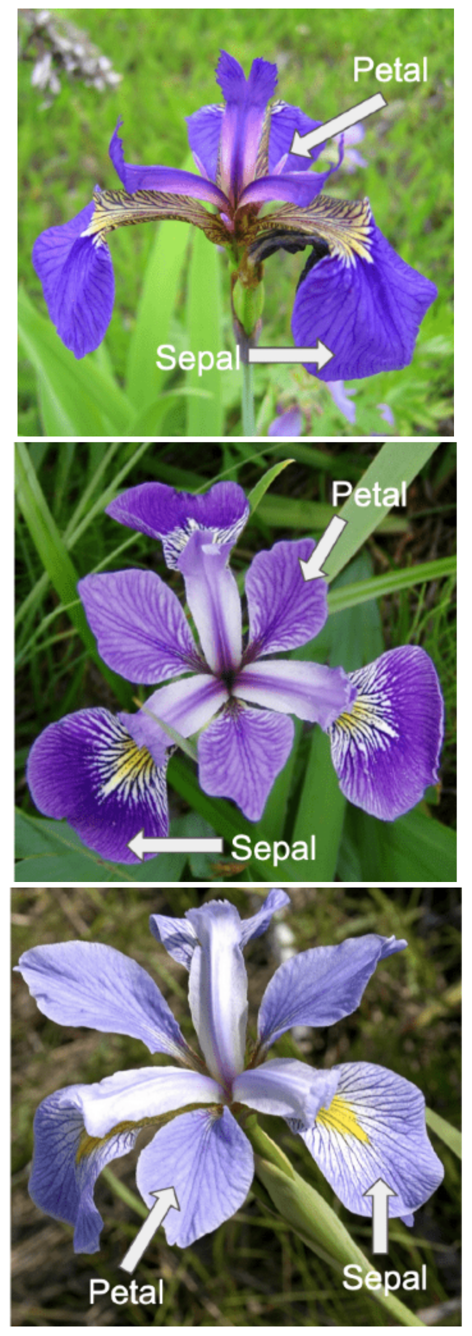


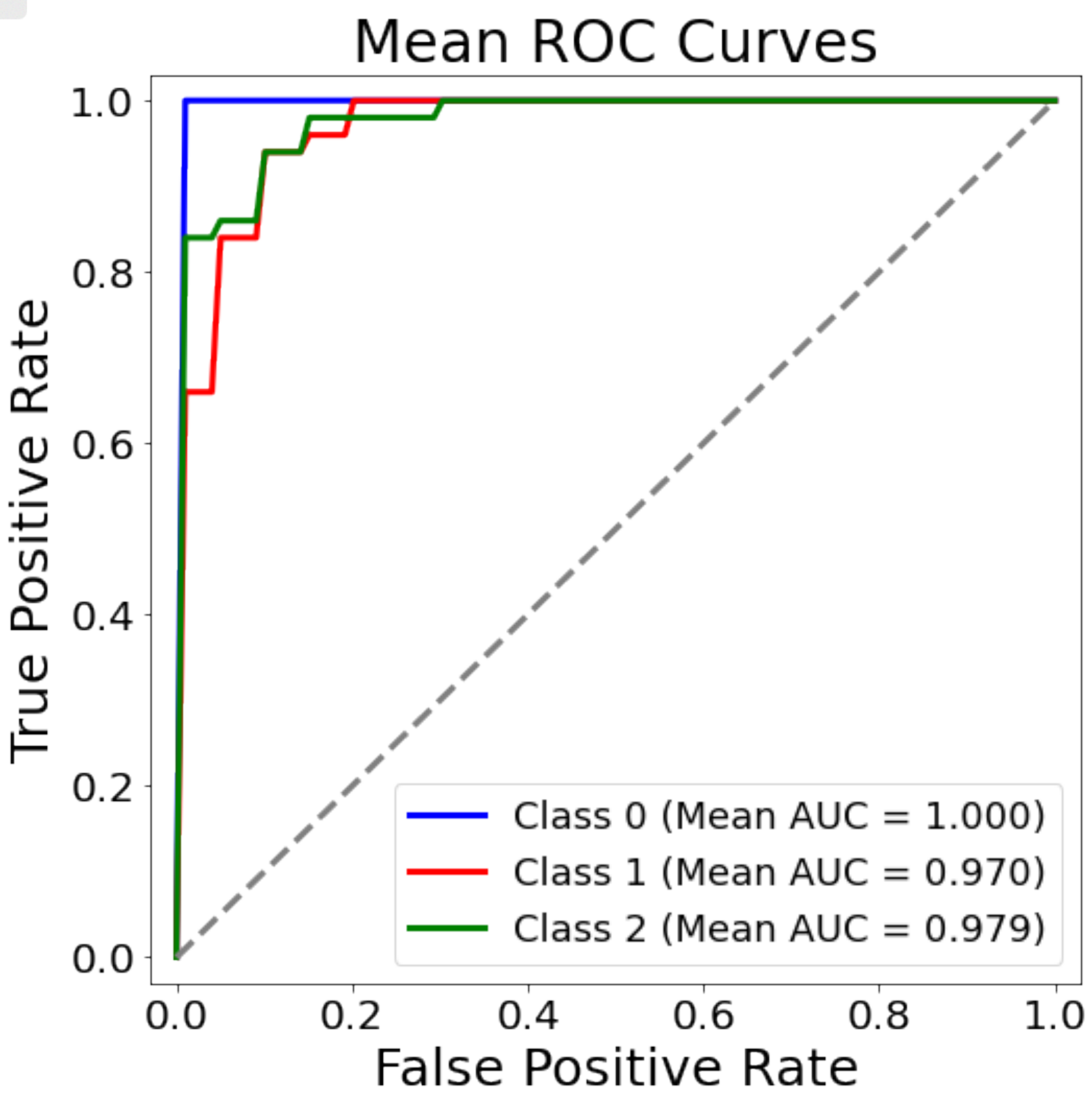


$\log(l)$ ancilla qubits for l classes



Iris data classification





Variational Quantum Classifier

Interpretation

- What does a variational quantum classifier do? → Quantum state discrimination!
  - $\max_{\boldsymbol{\theta}} \sum_{i=1}^{l} \mathrm{Tr}\left[M_i^\dagger M_i U(\boldsymbol{\theta})\rho_i U(\boldsymbol{\theta})\right]$, where ρ_i encodes the training data in class i
  - Optimal measurement for the binary case: Helstrom measurement.
  - Helstrom bound: $p_{\text{success}} = \frac{1}{2} + \frac{1}{2}\|p_0\rho_0 - p_1\rho_1\|_1 = \frac{1}{2} + \frac{1}{2}T(p_0\rho_0, p_1\rho_1)$
  - The trace norm is contrastive under CPTP: $T(\Phi(\rho_0), \Phi(\rho_1)) \leq T(\rho_0, \rho_1)$
  - The training accuracy is determined by the classical-to-quantum encoding method! VQC simply tries to reach the Helstrom bound.
- We can optimize the data encoding on a quantum computer! (In progress)

Gradient-Free Quantum Optimization

Quantum Annealing

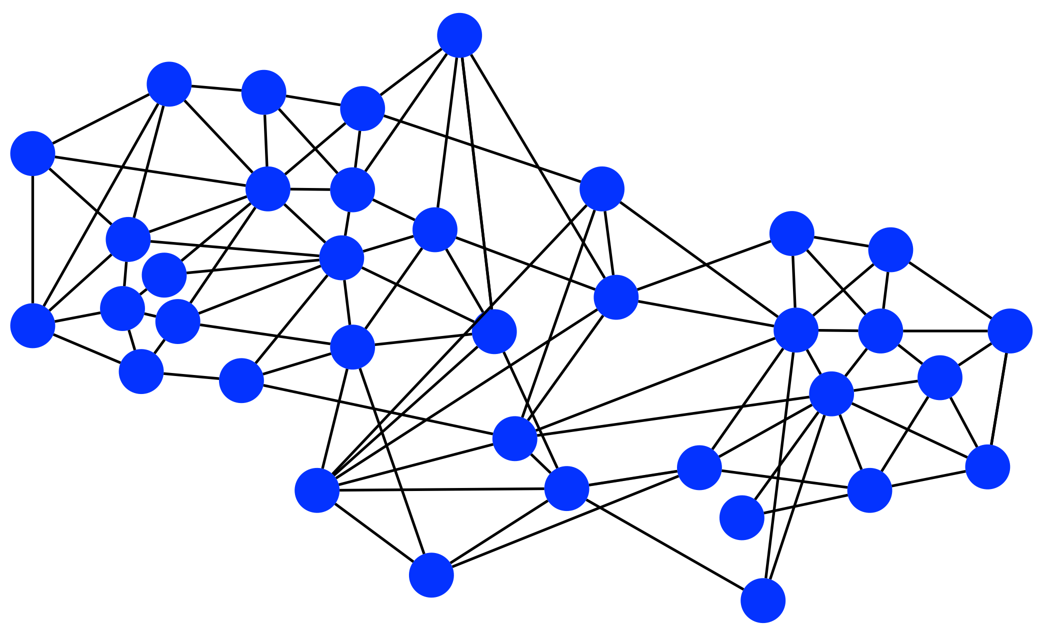
- Quantum computer is governed by the differential equation

$$i\hbar \frac{d|\psi(t)\rangle}{dt} = H(t)|\psi(t)\rangle$$

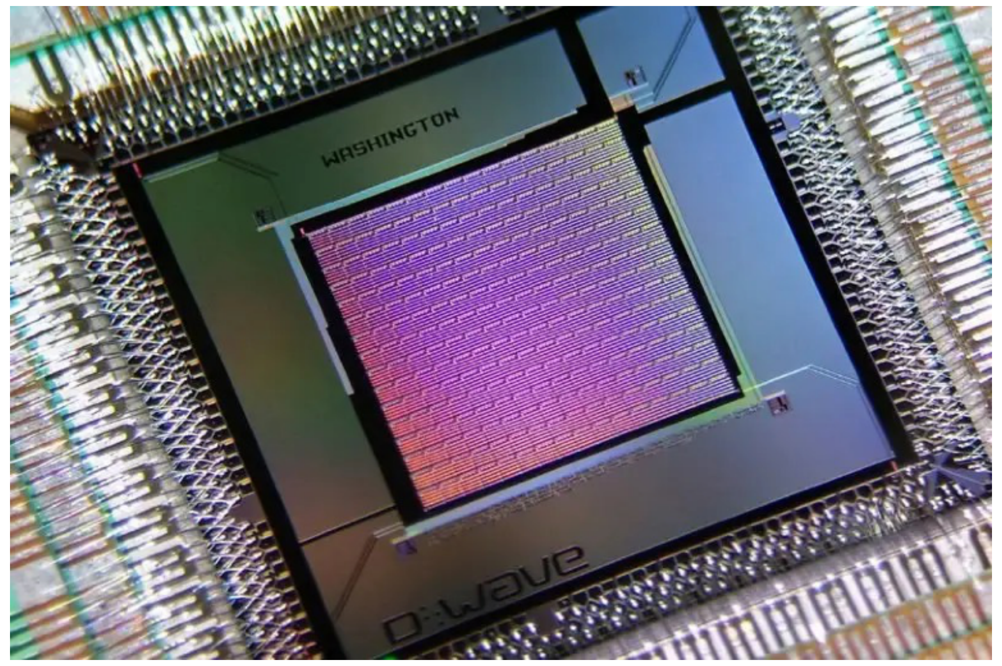
$H(t) = H^{\dagger}(t)$ and the size of $|\psi(t)\rangle$ and $H(t)$ grows exponentially with the number of qubits

This condition is usually impractical...

- Under certain condition (called adiabatic), this can be used to find the lowest eigenvalues and corresponding eigenstates (a.k.a ground state) of $H(t)$
  - Equivalent to solving certain NP-hard combinatorial optimization problems!



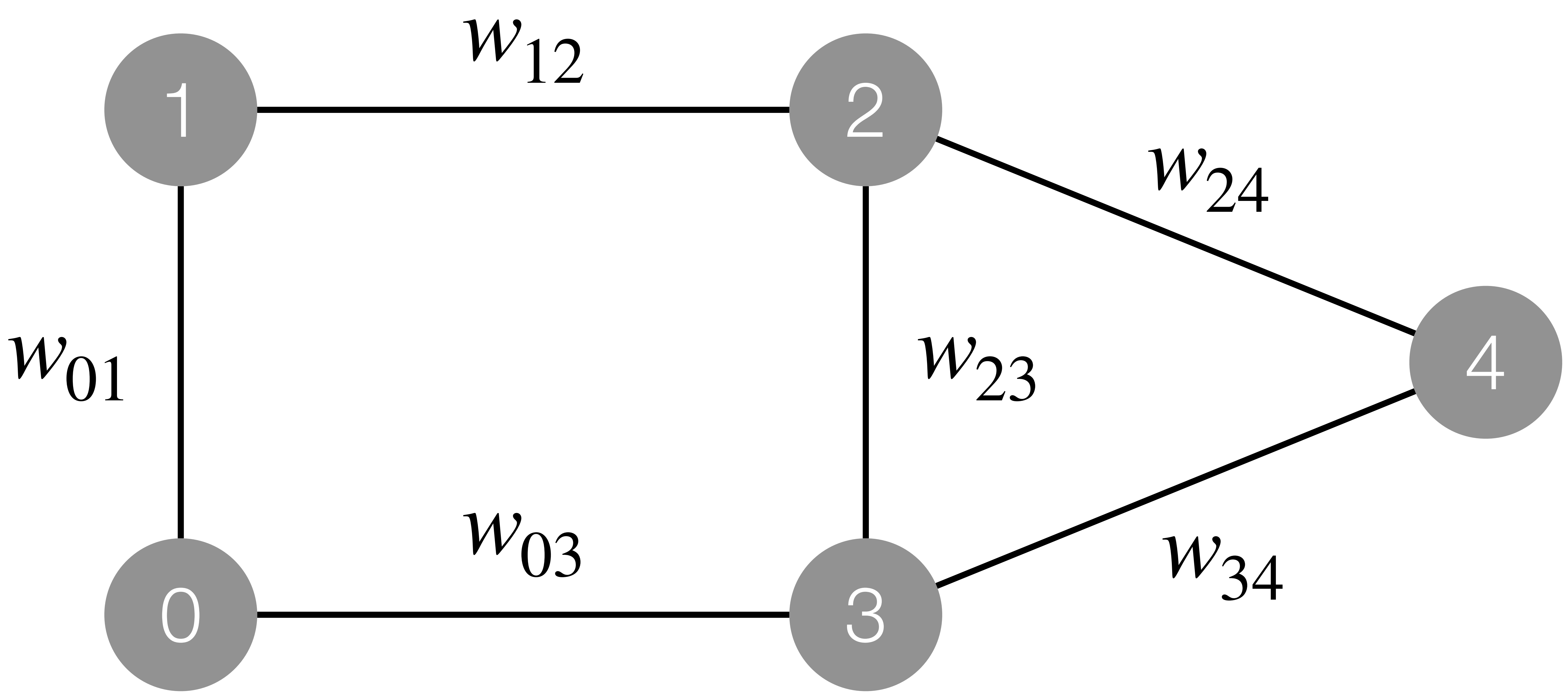
$$H_F = \sum_i h_i Z_i + \sum_{i<j} J_{ij} Z_i Z_j$$



Quadratic Unconstrained Binary Optimization (QUBO)

QUBO Example

Weighted Max-cut

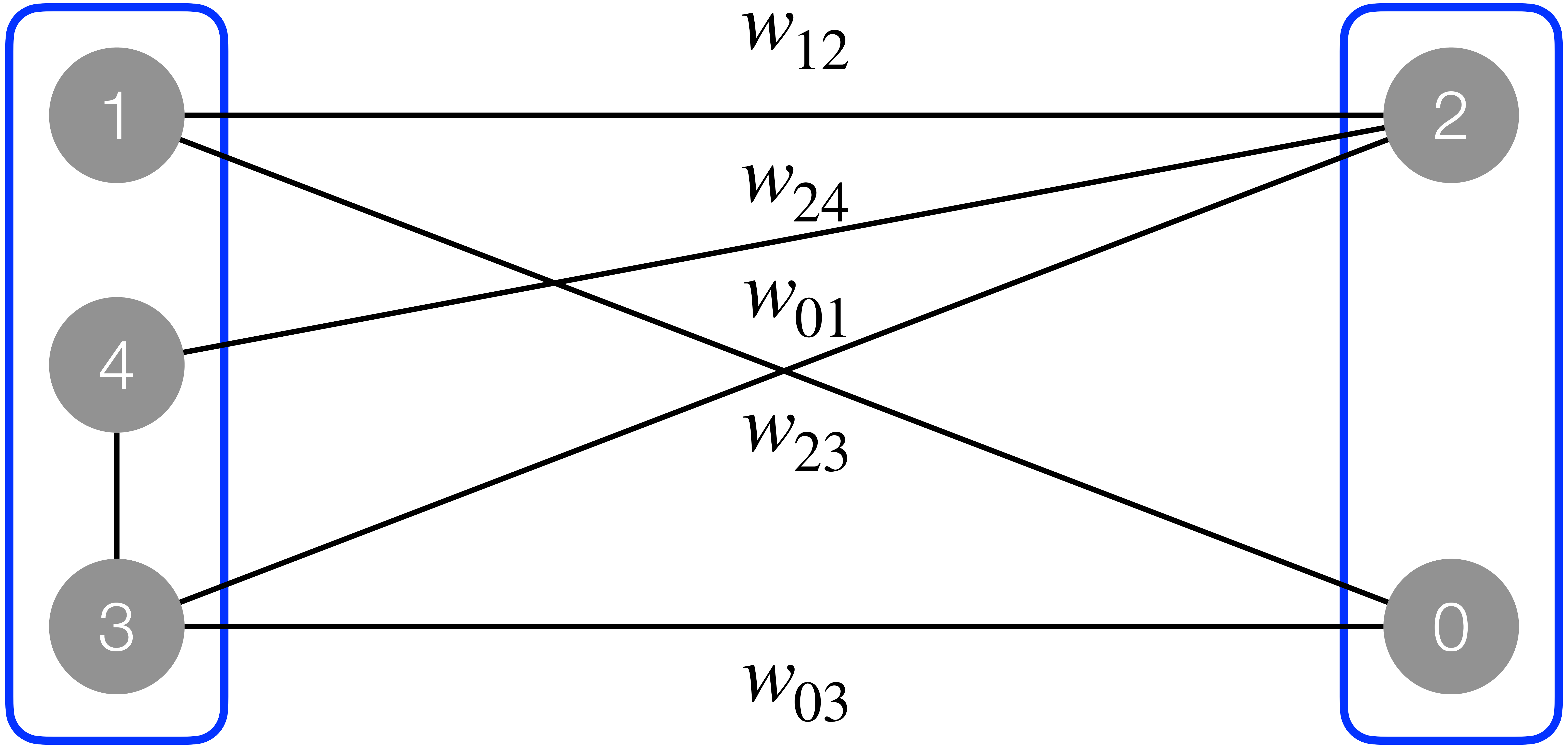


$$w_{ij} \geq 0$$

- Maximize the sum of weights that are cut by a given partition of the vertices into two sets
- If w_{ij} represents the distance between x_i and x_j, weighted max-cut is clustering
- This problem is equivalent to finding the ground state of $H = \sum_{i<j} w_{ij} Z_i Z_j$

QUBO Example

Weighted Max-cut



$w_{ij} \geq 0$

- Maximize the sum of weights that are cut by a given partition of the vertices into two sets
- If w_{ij} represents the distance between x_i and x_j, weighted max-cut is clustering
- This problem is equivalent to finding the ground state of $H = \sum_{i<j} w_{ij} Z_i Z_j$

🤫 A classical algorithm is likely to provide the best approximation ratio

Summary

Opportunities

- Noisy Intermediate-Scale Quantum era (soon)
    - Kernel method
    - Gradient-based optimization
    - Gradient-free optimization (e.g. quantum annealing)
- Full-fledge Quantum era (when?)
    - Basic linear algebra subprogram (such as matrix diagonalization and inversion)
- QML is the most natural for quantum data (e.g. from quantum sensing)

Summary

Challenges

- Variational quantum algorithms and non-adiabatic quantum annealing:
  - Practical quantum advantage without Haar measure?
  - Practical quantum advantage without adiabaticity?
- Classical data encoding
- Classical-Quantum hybrid model: Optimizer robust to quantum noise
- Dealing with quantum errors without relying on quantum error correction