

# 量子機械学習の可能性

慶應義塾大学 理工学部 / 量子コンピューティングセンター

山本直樹



# 目次

## 機械学習による分子構造デザイン

新奇量子状態を探索するための量子生成モデルの構築法

量子力学の特性を活かすカーネル設計法の開発

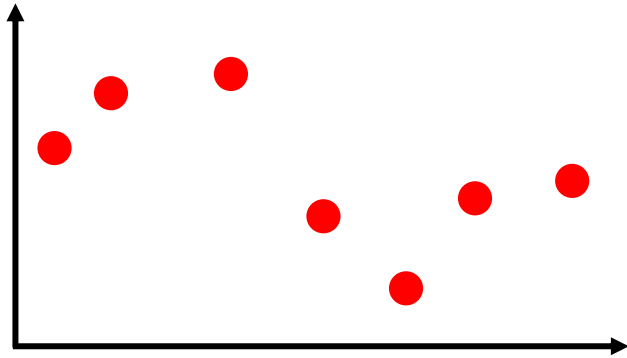
量子状態モデル選択のための情報量規準

## 機械学習による分子構造デザイン

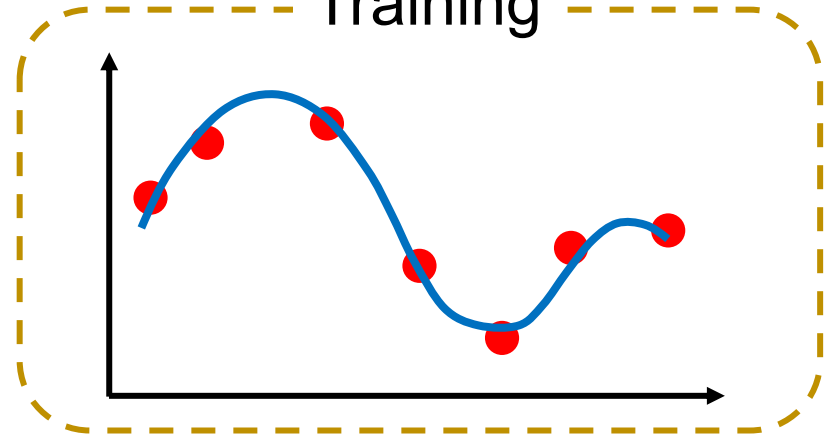
**Quantum-classical computational molecular design of deuterated high-efficiency OLED emitters, Gao, et al. Intelligent Computing (2023)**

まずはじめに、量子力学が登場する機械学習問題を紹介。  
このような問題には、やはり量子計算機が有効だろうか・・・？

Data given to us

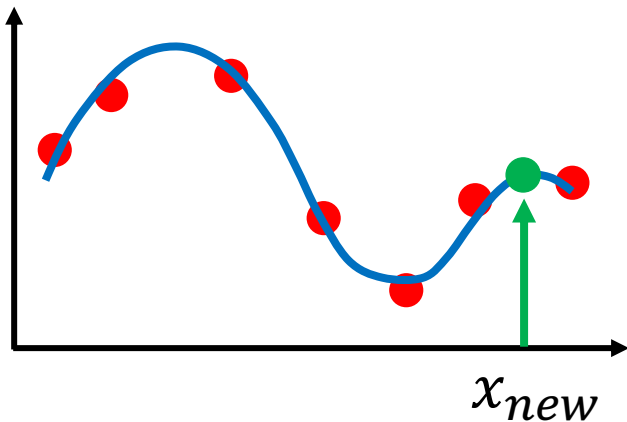


Training

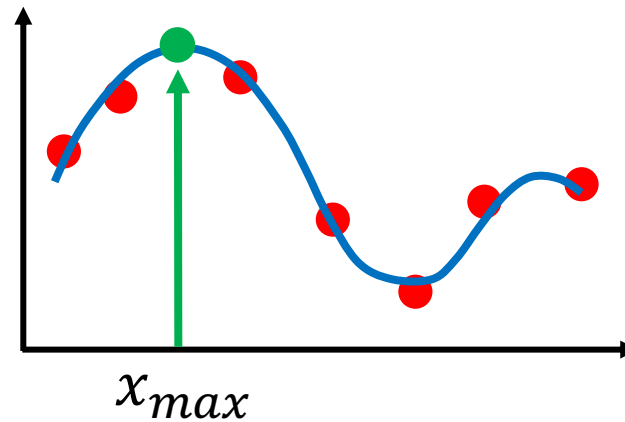


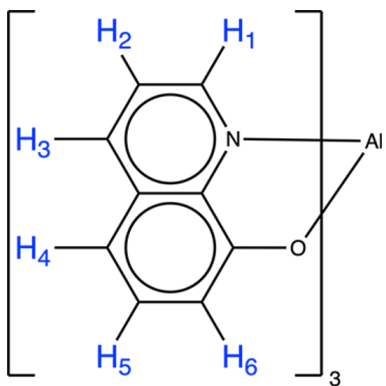
Prediction

See the property  
of a new data  $x_{new}$



Find  $x_{max}$  that maximizes  
the function





Replacing H via D improves the emission rate.  
Which H should be replaced?

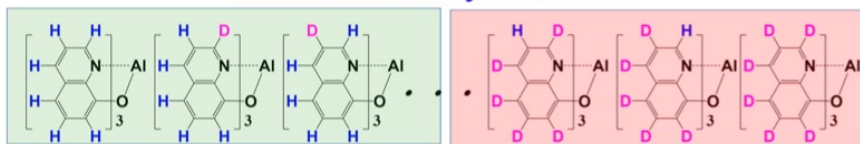
The data  $x$  is the alignment of H and D. The (hidden) function is the solver for Schrodinger Eq.

Using 13 training data, we determined the cost of emission (Franck-Condon factor):

$$C = \sum_{i=1}^6 h_i s_i + \sum_{i=1}^5 \sum_{j=i+1}^6 J_{ij} s_i s_j \quad (s=1: \text{H}, s=0: \text{D})$$

## Learning

### Quantum chemistry calculations



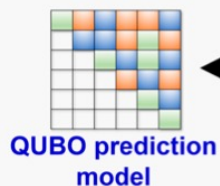
Training data

Test data

Repeat

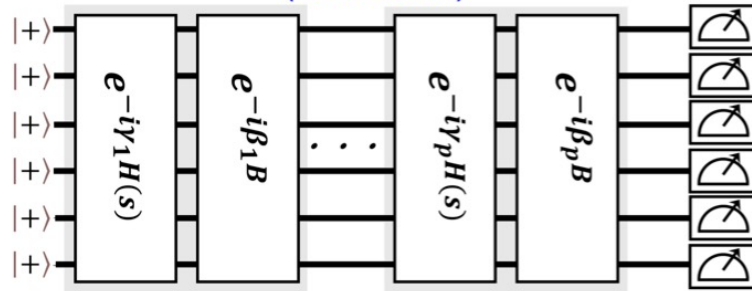
Train machine learning model

Validate machine learning model

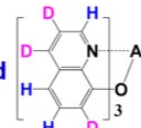


## Prediction

### Constrained and unconstrained optimization (VQE & QAOA)



Optimal deuterated molecule



Quantum computer

Constructed model:

$$C = \sum_{i=1}^6 h_i s_i + \sum_{i=1}^5 \sum_{j=i+1}^6 J_{ij} s_i s_j \quad (s=1: H, s=0: D)$$

Prediction: Use the model to compute the optimal set of D,  
under the constraint  $\#D \leq 3$ .

**QAOA:**

$$\text{Min. } \langle \psi(\beta, \gamma) | C | \psi(\beta, \gamma) \rangle$$

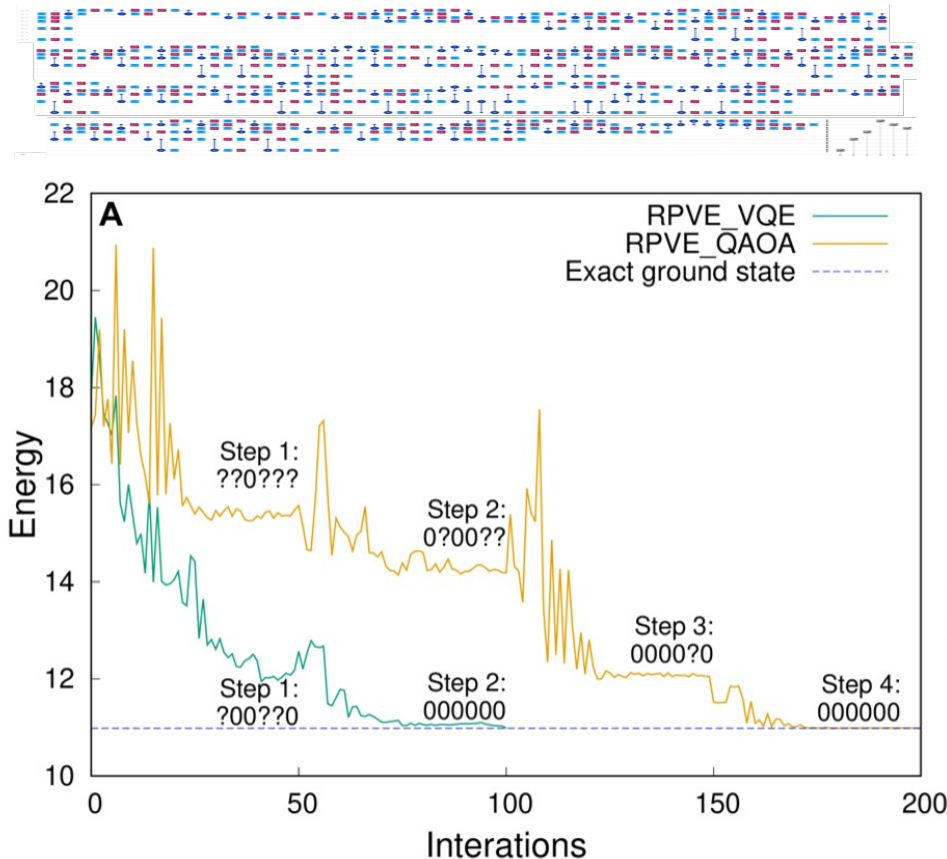
$$\text{where } |\psi(\beta, \gamma)\rangle = \prod_{a=1}^p e^{i\beta_a B} e^{i\gamma_a C} |+\rangle^n$$

**バイナリサーチ:**

計算基底で測定  $\rightarrow$  0 (or 1)の割合がある閾値以上の場合、その変数を0 (or 1)に固定。使用量子ビット数が徐々に減る。

データ(原子配置)から発光効率を計算する量子計算機をデザインできれば尚嬉しい

**Result:** (p=3、#CNOT = 156)



# 新奇量子状態を探索するための量子生成モデルの構築法

**Quantum Generative Model with Optimal Transport,  
Tezuka, Uno, Yamamoto, arXiv, 2022**

データ  $x$  が量子状態。温度や相互作用強度により、色々な量子状態が発現している。機械学習で、この発現プロセスを模倣する量子計算機をつくれませんか？

# 機械学習

教師あり学習

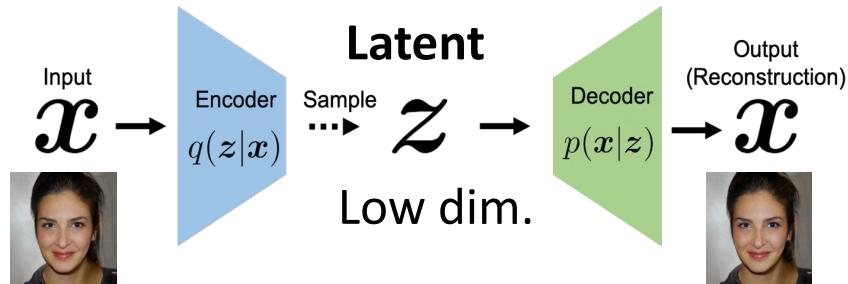
教師なし学習 — 生成モデル

強化学習

VAE (Variational AutoEncoder)

GAN (Generative Adversarial Network)

Flow-based

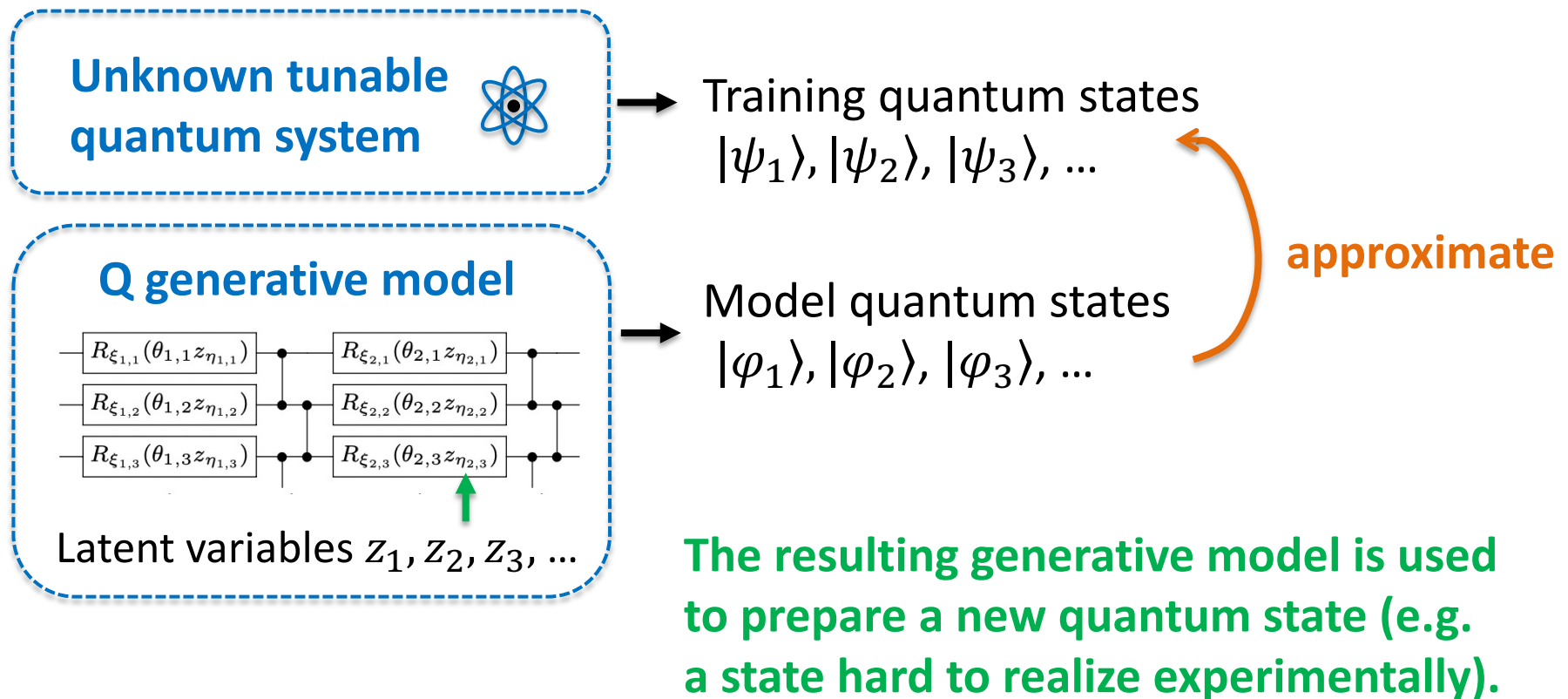


beta-VAE, ICLR2017



An unknown yet tunable quantum system prepares a set of training quantum states (e.g., temperature or the interaction strength is tunable).

The goal is to construct a quantum generative model that outputs a set of quantum states approximating the set of training quantum states.

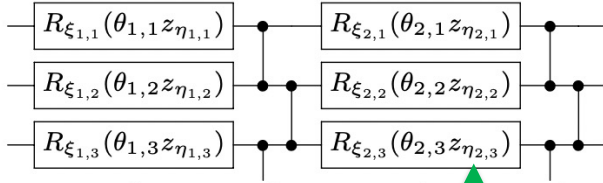


Unknown tunable quantum system



Training quantum states  $|\psi_1\rangle, |\psi_2\rangle, |\psi_3\rangle, \dots$

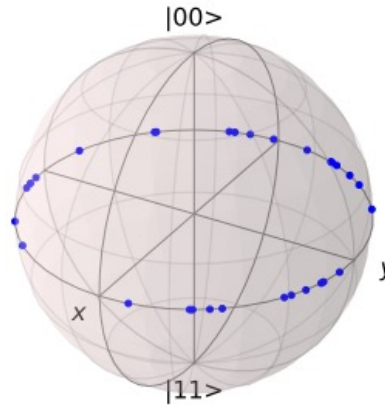
Q generative model



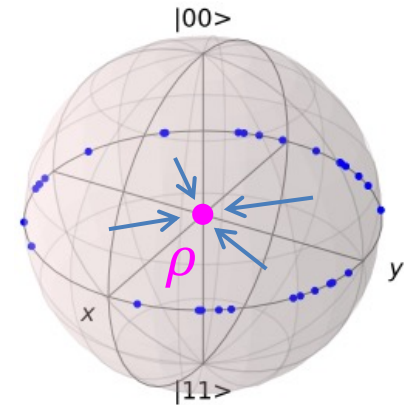
Latent variables  $z_1, z_2, z_3, \dots$

Model quantum states  $|\varphi_1\rangle, |\varphi_2\rangle, |\varphi_3\rangle, \dots$

approximate



$$\{|\psi_i\rangle\}^M$$



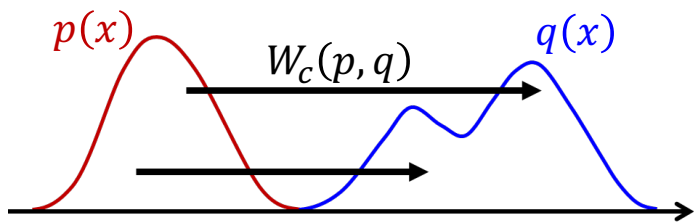
$$\rho = \sum_{i=1}^M p_i |\psi_i\rangle\langle\psi_i|$$

How to define the cost for measuring the gap between the two set?

Distance between  $\frac{1}{M} \sum_K |\psi_k\rangle\langle\psi_k|$  and  $\frac{1}{M} \sum_K |\varphi_k\rangle\langle\varphi_k|$  may not work.

Use the optimal transport distance (Wasserstein distance)

Wasserstein distance  $W_c$  is the cost for moving  $p(x)$  to  $q(x)$

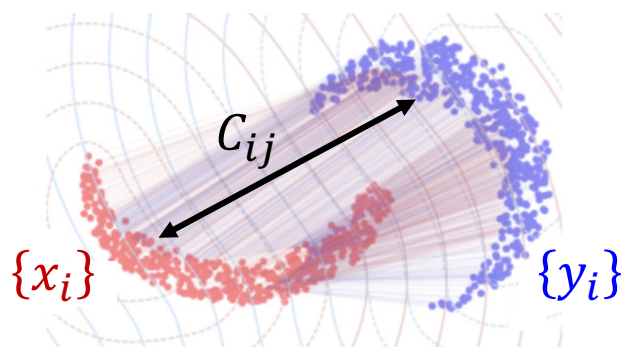


$$W_c(p, q) = \min_{\pi} \int c(x, y) d\pi(x, y)$$

Ground cost (given)   
 Transport plan

$$\int \pi(x, y) dx = q(y), \int \pi(x, y) dy = p(x), \pi(x, y) \geq 0$$

In the sample space:

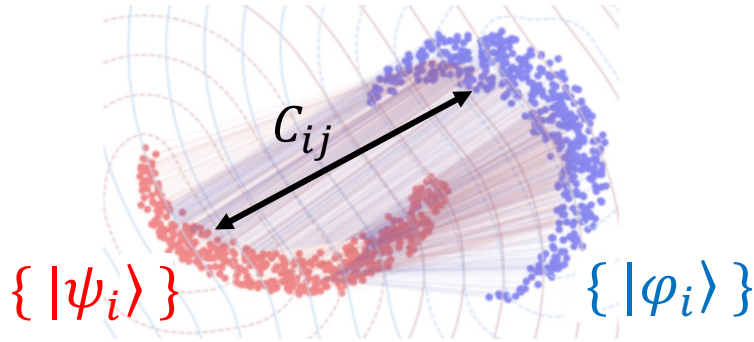


$$W_c(\{x_i\}, \{y_i\}) = \min_{P \in \mathbb{R}^{m \times n}} \sum_{i=1}^N \sum_{j=1}^M C_{ij} P_{ij}$$

$$\sum_{i=1}^N P_{ij} = \frac{1}{M} \quad \sum_{j=1}^M P_{ij} = \frac{1}{N} \quad P_{ij} \geq 0$$

--- can be efficiently solved via linear programming

Quantum Wasserstein distance  $W_q$  is the cost for moving  $\{|\psi_i\rangle\}$  to  $\{|\varphi_i\rangle\}$

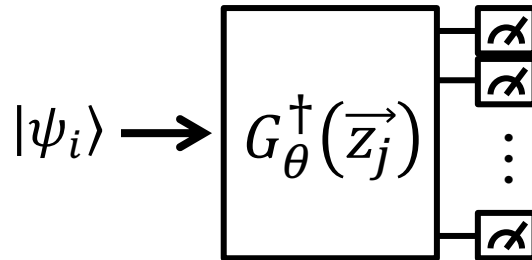


$$W_q(\{|\psi_i\rangle\}, \{|\varphi_i\rangle\}) = \min_{P \in \mathbb{R}^{m \times n}} \sum_{i=1}^N \sum_{j=1}^M C_{ij} P_{ij}$$

$$\sum_{i=1}^N P_{ij} = \frac{1}{M} \quad \sum_{j=1}^M P_{ij} = \frac{1}{N} \quad P_{ij} \geq 0$$

Ground cost  $C_{ij}$  may be set as the trace distance

$$D = \sqrt{1 - |\langle \psi_i | \phi_{\vec{\theta}}(\vec{z}_j) \rangle|^2} = \sqrt{1 - |\langle \psi_i | G_{\vec{\theta}}(\vec{z}_j) | 0 \rangle|^2}$$



Probability to have all zeros

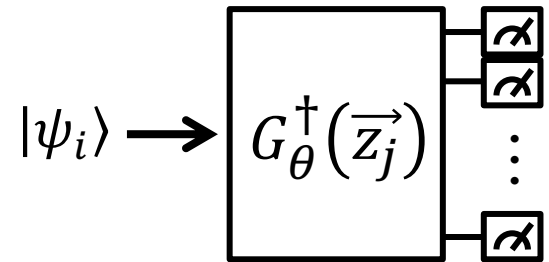
# Barren plateau issue

--- The gradient vectors exponentially decay as # qubit increases.

→ Localization of the cost mitigates the barren plateau issue.

$$c_{\text{global}}(|\psi\rangle, |\phi\rangle) = \sqrt{1 - |\langle\psi|G_{\theta}|0\rangle|^2}$$

Probability to have all zeros



$$c_{\text{local}}(|\psi\rangle, |\phi\rangle) = \frac{1}{n} \sqrt{\sum_{k=1}^n (1 - p^{(k)})}$$

Probability of the k-th bit to be zero

$$p^{(k)} = \text{Tr} \left[ P_0^{(k)} G_{\theta}^{\dagger} |\psi\rangle \langle\psi| G_{\theta} \right]$$

$$P_0^{(k)} = \mathbf{1}_1 \otimes \mathbf{1}_2 \otimes \cdots \otimes |0\rangle\langle 0| \otimes \cdots \otimes \mathbf{1}_n$$

$c_{\text{global}}(|\psi\rangle, |\phi\rangle)$ は距離の公理を満たすが、 $c_{\text{local}}(|\psi\rangle, |\phi\rangle)$ は満たさない。  
しかし、divergenceの公理を満たすため、コストとしては機能。

ただし、計算に必要なデータ数について注意が必要。

古典Wasserstein距離に関する、近似誤差の収束則(J.Weed et al 2019):

$$\mathbb{E} \left[ \left| \underbrace{\mathcal{W}_p(\alpha, \beta)}_{\text{真の距離}} - \underbrace{\mathcal{W}_p(\hat{\alpha}_M, \hat{\beta}_M)}_{\text{サンプリングに基づく推定値}} \right| \right] \lesssim O(M^{-1/N_z})$$

(注:  $M$ はサンプル数、 $N_z$ は潜在空間の次元)

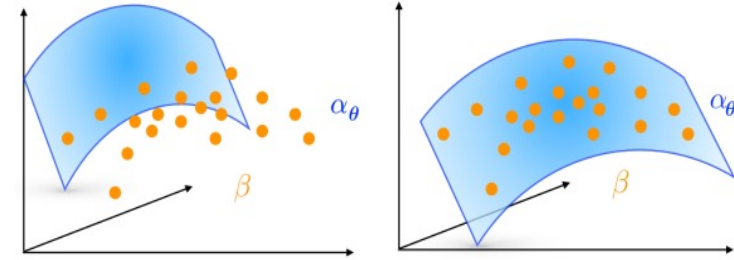
➡ 学習に必要なサンプルデータ数は、データ空間の次元に依存せず、潜在空間の次元 $N_z$ のみに依存する

localizeされた $W_q$ は距離ではなく divergence。学習コスト？

# これまでのまとめ

## 問題設定

量子データセット  $\{|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_M\rangle\}$  が与えられたとき、これらを生成する量子モデルを学習。

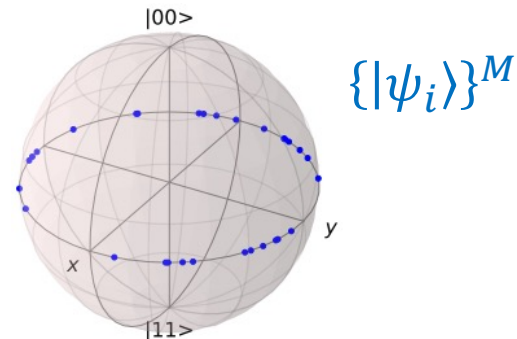
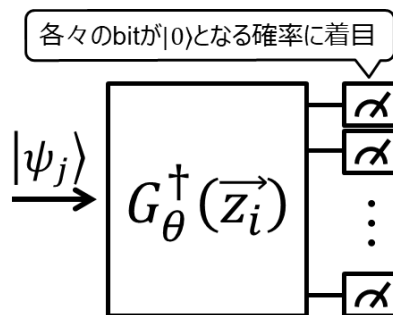


## 提案手法

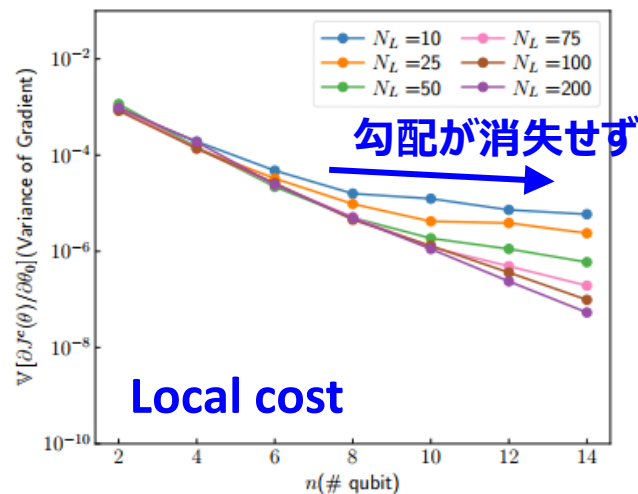
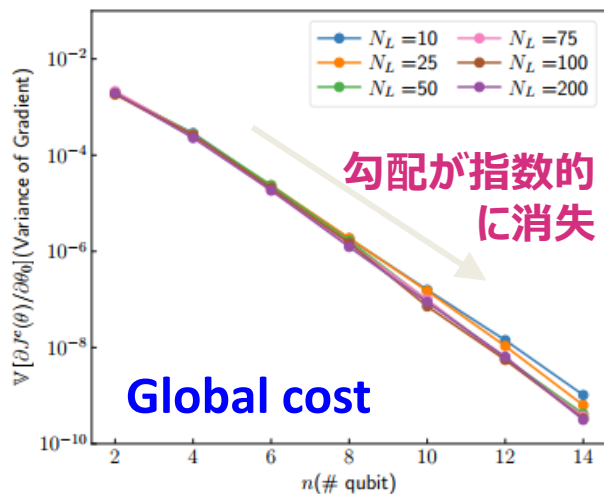
- 量子データセット  $\{|\psi_i\rangle\}^M$  と生成モデル  $G(\vec{z}, \vec{\theta})$  のinversion testで学習を行う。
- Optimal transport lossを導入し、量子状態の”分布”情報を残したまま学習。
- Local costを導入し、勾配消失の問題を緩和する。

## 懸念点

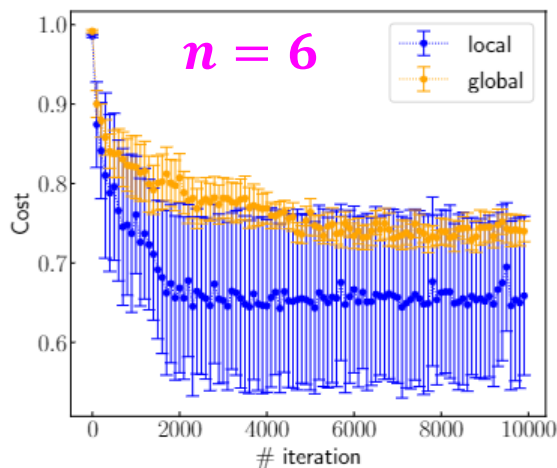
- Local cost化により、Costが距離の公理を満たさなくなる。学習に必要なサンプル数について、良い性質があるか？
- Local cost化によって、勾配消失問題はどれほど緩和されるか？



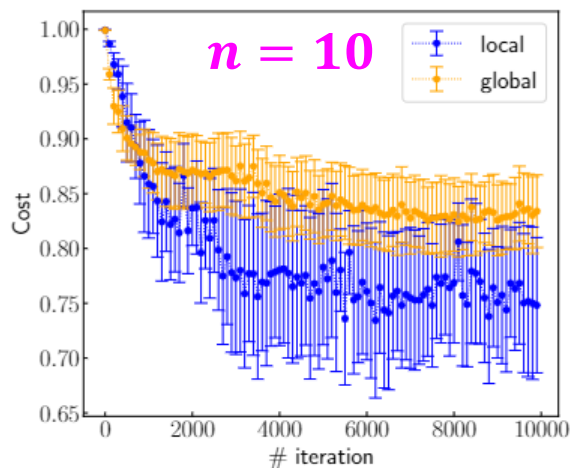
# Variance of the gradient (at the initial point)



# Change of the cost (dim(z)=2, n=6,10)



(a) Learning curve ( $n = 6$ )



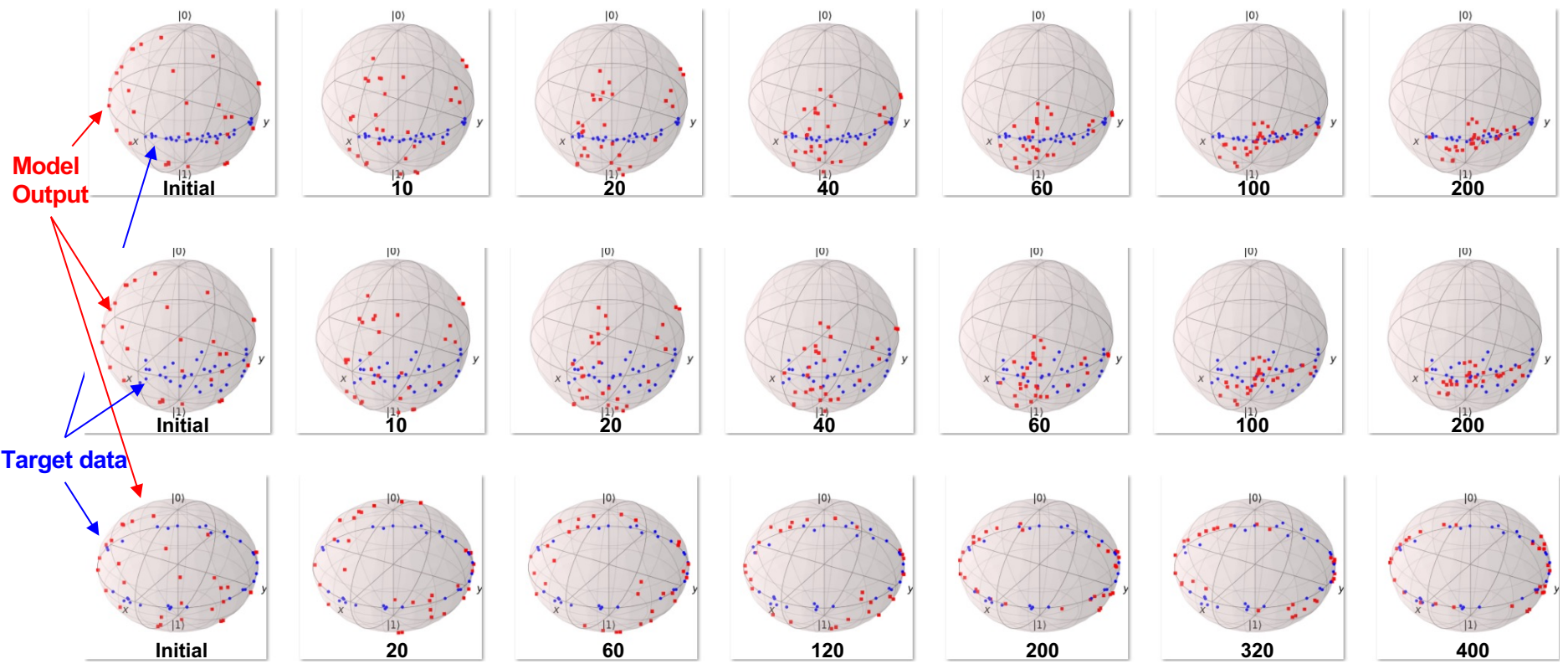
(b) Learning curve ( $n = 10$ )

Localization of the cost mitigates the barren plateau issue.

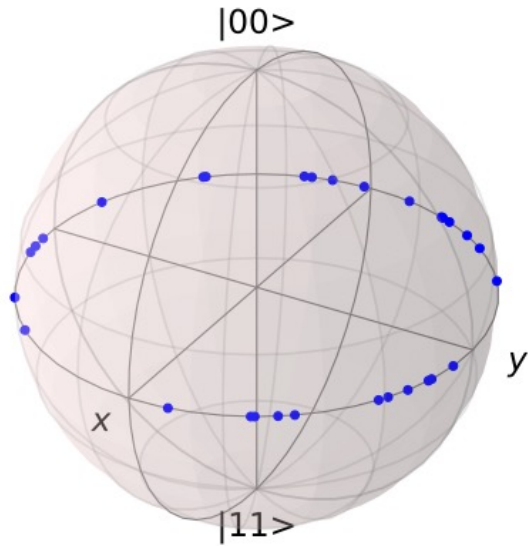


# Training process

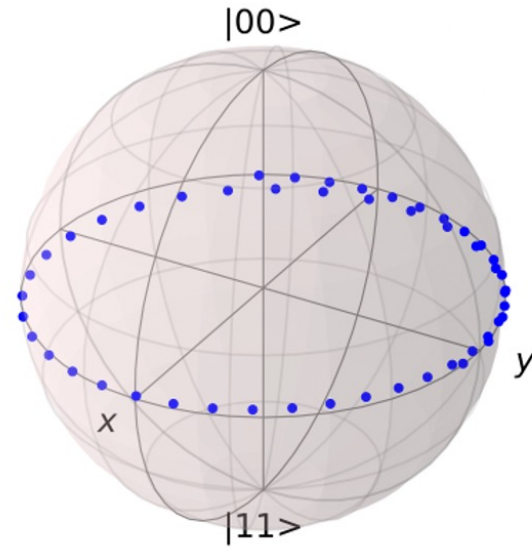
Iteration →



# Prediction process



Training quantum data



New quantum data produced from the generative model

“赤道情報”が正しく学習されている。

距離については次が成立:

サンプル数

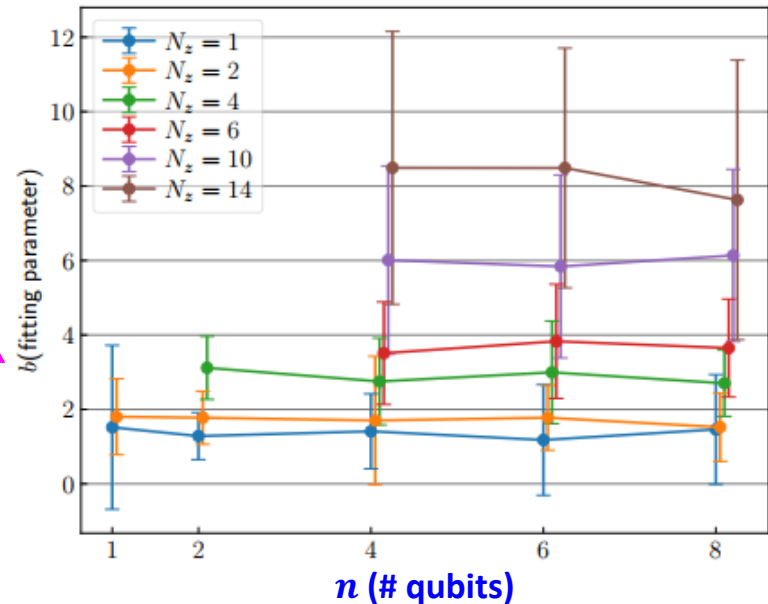
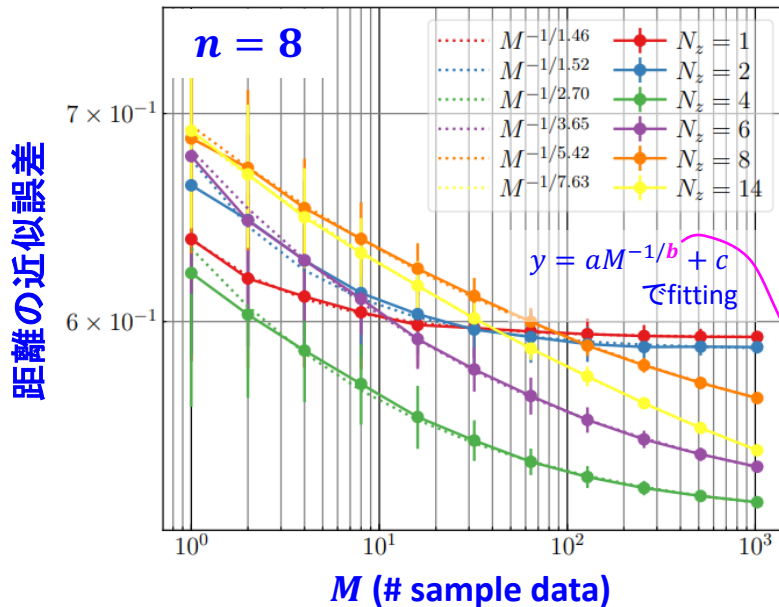
潜在空間の次元

$$\mathbb{E}[|\mathcal{W}_p(\alpha, \beta) - \mathcal{W}_p(\widehat{\alpha}_M, \widehat{\beta}_M)|] \lesssim O(M^{-1/N_z})$$

真の距離

サンプリングに基づく推定値

いま、localizeされた  $W_q$  は距離ではなく divergence。上記が成立するか？



コストの近似精度はHilbert空間の次元に依存せず、潜在空間の次元  $N_z$  に線型に依存する

## まとめ

- 量子データの“**分布**”を学習可能な量子生成モデルを提案した
- ポイントは
  - ◆ 最適輸送距離（Wasserstein距離）の導入
  - ◆ Local costの導入
- 学習に必要なデータ数が、対象そのものの次元でなく、潜在空間の次元 $N_z$ にのみ依存するため高次元系であっても、比較的少数のサンプル数で学習が可能

## 量子力学の特性を活かすカーネル設計法の開発

**Quantum Fisher kernel for mitigating the vanishing similarity issue,  
Suzuki, Kawaguchi, Yamamoto, arXiv:2210.16581, 2022**

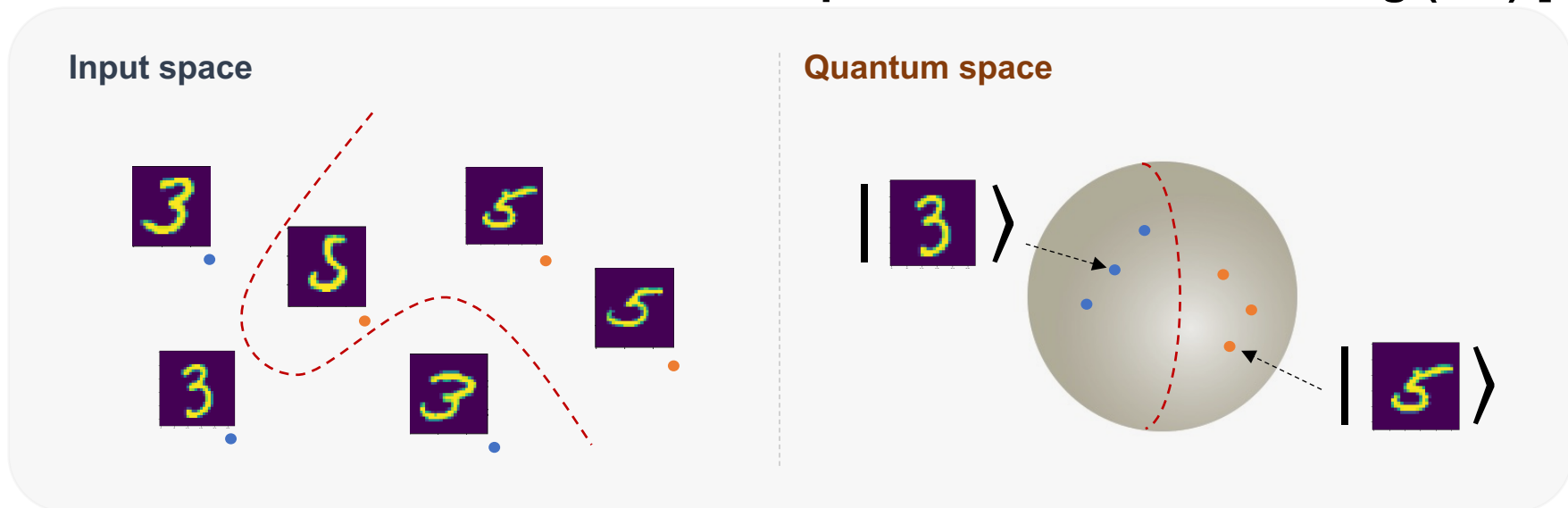
量子機械学習は、量子状態の空間で回帰や分類器を構成。良い分類器をつくるために、量子の特性を利用。そのような量子の特性を利用する量子機械学習器は、`量子ならではの`と期待。

# Quantum machine learning

## ■ Quantum machine learning (in NISQ\* era)

- Quantum computing possibly enhances the performance of machine learning (ML)

### Ex) : Quantum-enhanced feature space for machine learning (ML) [1]



**Large quantum space is exploited for feature extraction in ML tasks**

(e.g. 433 qubits  $\rightarrow 2^{433} \approx 10^{130}$  dim space)

[1] Havlíček, Vojtěch, et al. "Supervised learning with quantum-enhanced feature spaces." *Nature* 567, 7747 (2019).

# Background: Quantum kernel method

## ■ Quantum kernel methods [1]

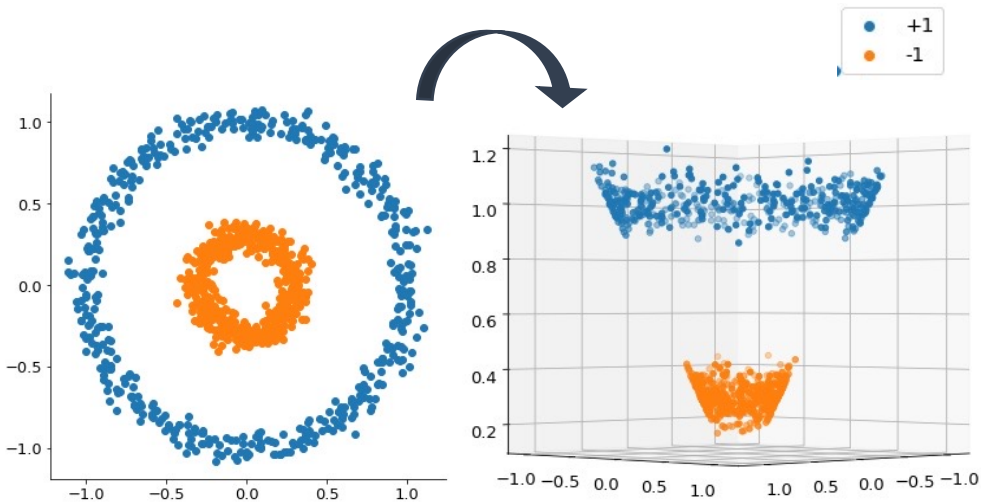
- Utilizing quantum computers for classical kernel method

Use Q computer

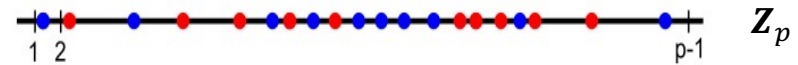
Kernel function:

$$k(x, x') = \phi(x)\phi(x')$$

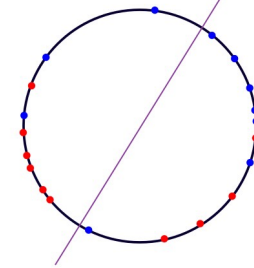
Feature map:  $\phi(x)$



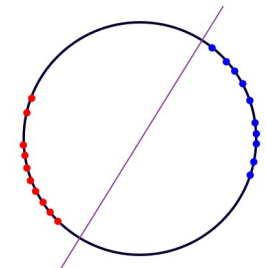
Ex) DLP-inspired datasets [2]



classical  $\phi(x)$



Quantum  $\phi(x)$



➔ Quantum kernel have potential to have **quantum advantage**

[1] Havlíček, Vojtěch, et al. "Supervised learning with quantum-enhanced feature spaces." *Nature* **567**, 7747 (2019).

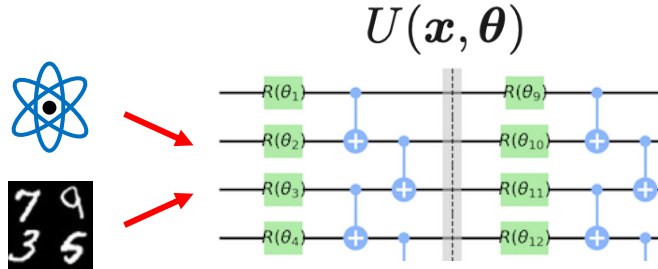
[2] Liu, Yunchao, Arunachalam, Temme. "A rigorous and robust quantum speed-up in supervised machine learning." *Nat. Phys.* **17**, 9 (2021).

# Details of quantum kernel method

## ■ Fidelity-based quantum kernel (QK)

- A function defined using the fidelity to measure the similarity between a pair of data  $\mathbf{x}, \mathbf{x}'$

Fidelity-based QK:



$$k_Q(\mathbf{x}, \mathbf{x}') = \text{Tr} [\rho_{\mathbf{x}, \theta} \rho_{\mathbf{x}', \theta}]$$

$$\rho_{\mathbf{x}, \theta} = U(\mathbf{x}, \theta) \rho_0 U^\dagger(\mathbf{x}, \theta)$$

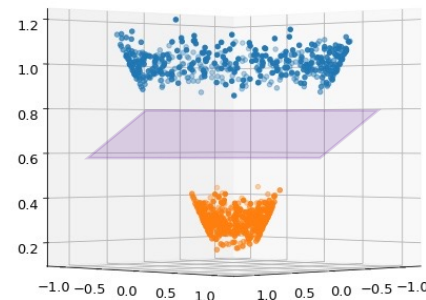
Gram matrix:

$$K = \begin{bmatrix} k_Q(\mathbf{x}_1, \mathbf{x}_1) & k_Q(\mathbf{x}_1, \mathbf{x}_2) & \cdots & k_Q(\mathbf{x}_1, \mathbf{x}_N) \\ k_Q(\mathbf{x}_2, \mathbf{x}_1) & & & \\ \vdots & \ddots & & \\ k_Q(\mathbf{x}_N, \mathbf{x}_1) & \cdots & & k_Q(\mathbf{x}_N, \mathbf{x}_N) \end{bmatrix}$$



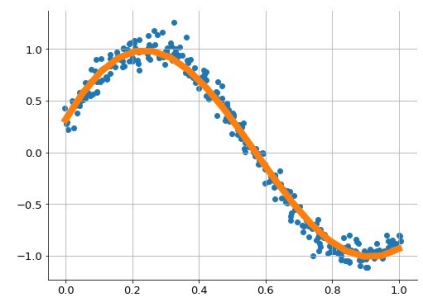
**Classification**

$$y(\mathbf{x}) = \sum_i \alpha_i t_i K(\mathbf{x}_i, \mathbf{x})$$



**Regression**

$$f(\mathbf{x}) = \sum_i \alpha_i K(\mathbf{x}_i, \mathbf{x})$$



[1] Havlíček, Vojtěch, et al. "Supervised learning with quantum-enhanced feature spaces." *Nature* **567**, 7747 (2019).

[2] Liu, Yunchao, Srinivasan Arunachalam, and Kristan Temme. "A rigorous and robust quantum speed-up in supervised machine learning." *Nat. Phys.* **17**, 9 (2021).



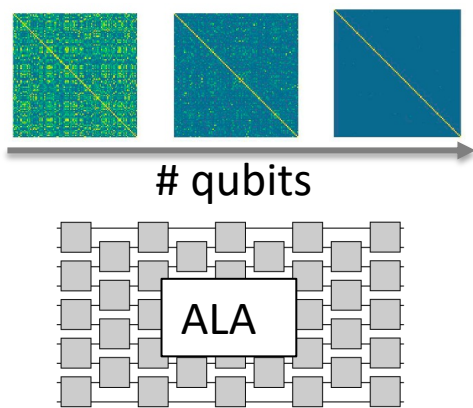
## <困難1>

(量子超越的)性能を出すために  $U(x, \theta)$  は帰納バイアス(データに対する事前知識など)をもつ必要があるが、構成法が全く非自明: **量子帰納バイアス問題**。

(例) Lie et al, Nat Phys. では離散対数データに対してショア回路を構成

## <困難2>

$U(x, \theta)$  に十分な表現力をもたせると、標準カーネル  $k_Q(x, x')$  が量子ビット数に対して指数的に小さくなる: **類似度消失問題** --- 経験的に知られていたが、EPFLのグループおよび本論文で証明された。



量子ニューラルネットワークにおける「勾配消失問題」は2018年に指摘され、以後対処法が多数提案されてきた。代表例: Alternating Layered Ansatz (ALA) を使うもの (Cerezo (LANL), Nat Comm 2021)。

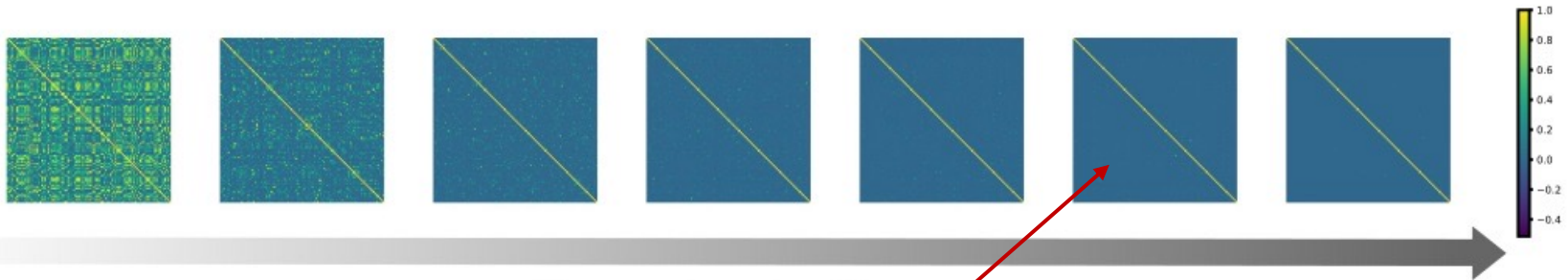
→ しかし本論文で、標準カーネル+ALAでも類似度消失することを証明

# Vanishing similarity issue

## ■ Vanishing similarity issue in the fidelity-based QK\*

- Exponential decrease of the expectation value and variance w.r.t. # qubits

Gram matrices  $K$  of the fidelity-based QK for different # qubit



$k_Q(x, x')$  decays exponentially fast for different inputs  $x \neq x'$

$$K = \begin{bmatrix} k_Q(x_1, x_1) & k_Q(x_1, x_2) & \cdots & k_Q(x_1, x_N) \\ k_Q(x_2, x_1) & \ddots & & \vdots \\ \vdots & & k_Q(x_i, x_j) & \vdots \\ k_Q(x_N, x_1) & \cdots & & k_Q(x_N, x_N) \end{bmatrix}$$

➡ An exponential number of measurement shots / overfitting & poor generalization performance

[3] Huang, Hsin-Yuan, et al. "Power of data in quantum machine learning." *Nat. Commun.* **12**, 2631 (2021).

\* This is empirically known [3], but analytically unexplored

# Vanishing similarity issue

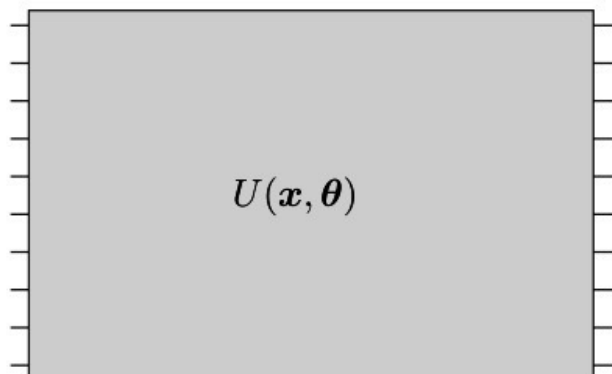
## ■ Analytical investigation (Setting)

Analytical results of the fidelity-based QK for two types of quantum circuits

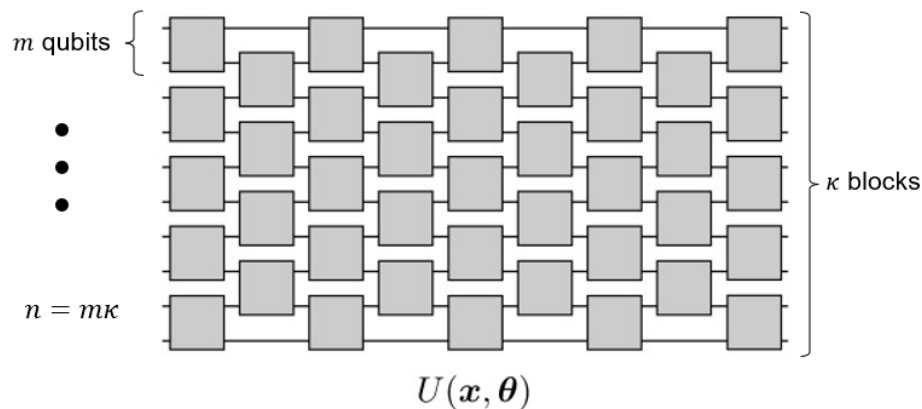
### Fidelity-based QK:

$$k_Q(\mathbf{x}, \mathbf{x}') = \text{Tr}[\rho_{\mathbf{x}, \theta} \rho_{\mathbf{x}', \theta}] \quad \text{with } \rho_{\mathbf{x}, \theta} = U(\mathbf{x}, \theta) \rho_0 U^\dagger(\mathbf{x}, \theta)$$

(1) Random quantum circuit acting on all  $n$  qubits



(2) Alternating layered ansatz (ALA)



■ : 2-design (unitary is sampled from an ensemble of unitary possessing the property of the Haar random unitary up to degree 2)

# Vanishing similarity issue

## ■ Analytical investigation (Result)

Analytical results of the fidelity-based QK for two types of quantum circuits

### Proposition 1 (Informal statement on the fidelity-based QK)

The expectation value and the variance of the fidelity-QK vanish exponentially fast w.r.t. #qubits for both types of quantum circuits.

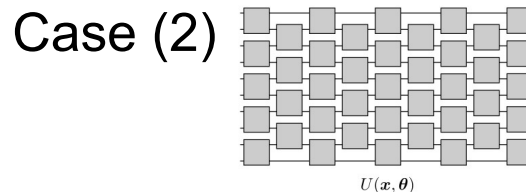
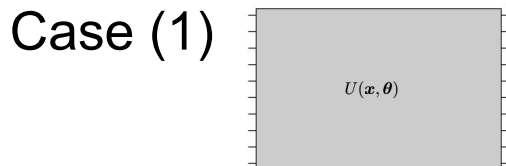
Case (1)

$$\langle k_Q \rangle = 1/2^n$$
$$\text{Var}[k_Q] = \frac{2^n - 1}{2^{2n}(2^n + 1)} \approx \frac{1}{2^{2n}}$$

Case (2)

$$\langle k_Q \rangle = 1/2^n$$
$$\text{Var}[k_Q] \leq \frac{2^\kappa}{(2^{2m} + 1)^\kappa} - \frac{1}{2^{2n}} \approx \frac{1}{2^{n(2 - \frac{1}{m})}}$$

### ➤ Fidelity-based QK cannot avoid the issue even with ALA



■ : 2-design

成果: 帰納バイアスを系統的に導入でき、類似度消失問題を緩和する量子カーネルを発見。

ポイント: フィッシャーカーネル。古典では

$$k_F(\mathbf{x}, \mathbf{x}') = \langle \mathbf{g}_{\mathbf{x}, \theta}, \mathbf{g}_{\mathbf{x}', \theta} \rangle_{\mathcal{J}^{-1}} = \mathbf{g}_{\mathbf{x}, \theta}^T \mathcal{J}^{-1} \mathbf{g}_{\mathbf{x}', \theta}$$

$\mathbf{g}_{\mathbf{x}, \theta} = \nabla_{\theta} \log P_{\mathbf{x}, \theta}$ : スコア関数

$P_{\mathbf{x}, \theta}$ : データ生成分布(実際は、データから生成モデルを構築)

量子版(量子フィッシャーカーネル)を提案

$$k_{QF}(\mathbf{x}, \mathbf{x}') = \langle \mathbf{L}_{\mathbf{x}, \theta}, \mathbf{L}_{\mathbf{x}', \theta} \rangle_{\mathcal{F}^{-1}} = \sum_{i,j} \mathcal{F}_{a,b}^{-1} \langle \mathbf{L}_{\mathbf{x}, \theta_a}, \mathbf{L}_{\mathbf{x}', \theta_b} \rangle_{\rho}$$

$\mathbf{L}_{\mathbf{x}, \theta}$ : スコア関数を量子化したもの

# Our proposal: Quantum Fisher kernel

## ■ Quantum Fisher kernel (QFK)

- A quantum extension of the classical Fisher kernel that incorporates the data structure into the kernel design [5]:

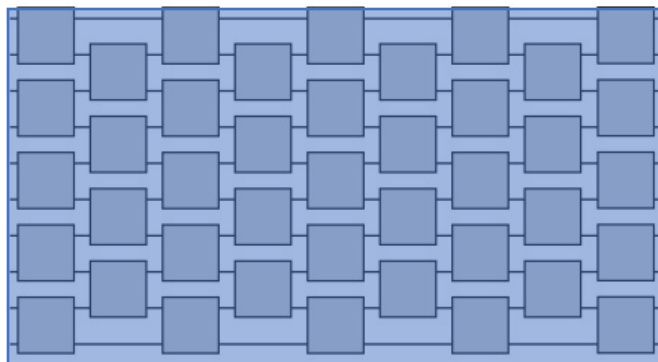
$$k_{QF}(\mathbf{x}, \mathbf{x}') = \frac{1}{2} \sum_i \text{Tr}[\rho_0 \{ \tilde{B}_{\mathbf{x}, \theta_i}, \tilde{B}_{\mathbf{x}', \theta_i} \}]$$

$\tilde{B}_{\mathbf{x}, \theta_i} = U_{1:i}^\dagger(\mathbf{x}, \boldsymbol{\theta}) B_{\theta_i} U_{1:i}(\mathbf{x}, \boldsymbol{\theta})$      $\rho_0$ : initial state     $\mathbf{x}$ : input data     $\boldsymbol{\theta}$ : parameters

$U_{1:i}(\mathbf{x}, \boldsymbol{\theta}) = U_i(\mathbf{x}, \theta_i) \dots U_2(\mathbf{x}, \theta_2) U_1(\mathbf{x}, \theta_1)$ : input- and parameter-dependent unitary

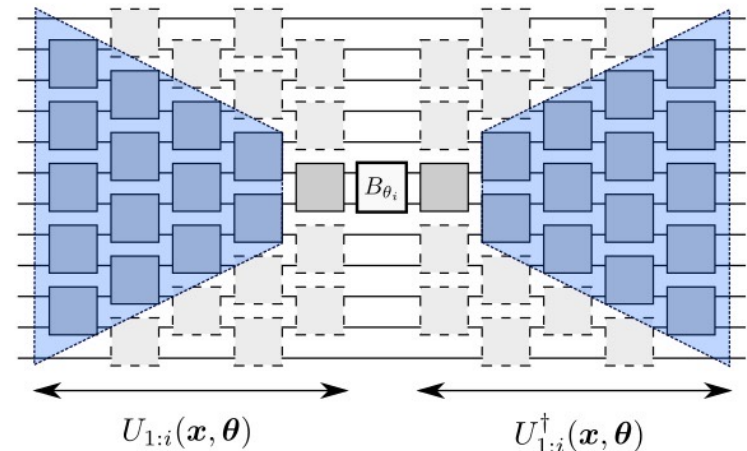
QFK measures local similarities using information geometric quantity

Fidelity-based QK : Global similarity



$U(\mathbf{x}, \boldsymbol{\theta})$

QFK : **Local similarity**



$U_{1:i}(\mathbf{x}, \boldsymbol{\theta})$

$U_{1:i}^\dagger(\mathbf{x}, \boldsymbol{\theta})$

# Vanishing similarity issue in QFK

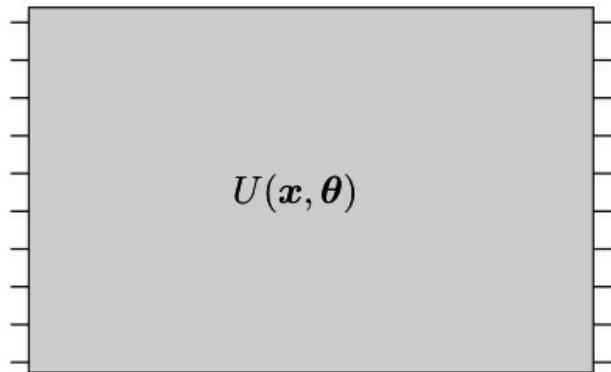
## ■ Analytical investigation (Setting)

- Analytical results of the QFK for two types of quantum circuits

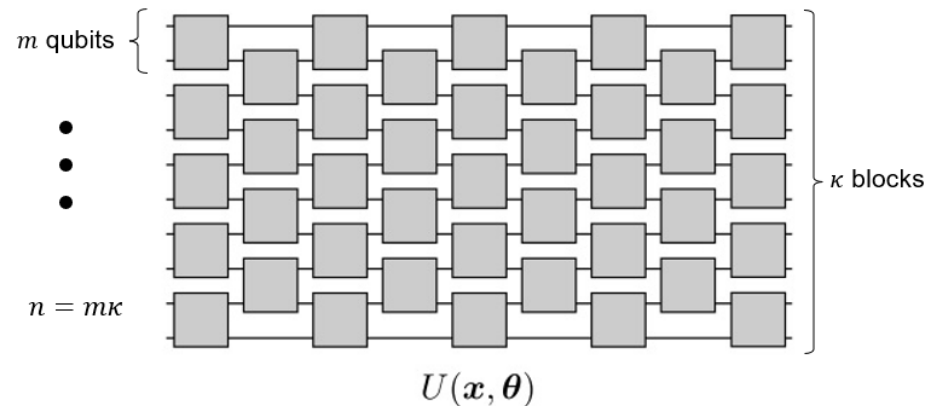
### The $i$ th component of QFK:

$$\tilde{k}_{QF} = \frac{1}{2} \text{Tr}[\rho_0 \{\tilde{B}_{x,\theta_i}, \tilde{B}_{x',\theta_i}\}] \quad \text{with} \quad \tilde{B}_{x,\theta_i} = U_{1:i}^\dagger(\mathbf{x}, \boldsymbol{\theta}) B_{\theta_i} U_{1:i}(\mathbf{x}, \boldsymbol{\theta})$$

(1) Random quantum circuit acting on all  $n$  qubits



(2) Alternating layered ansatz (ALA)

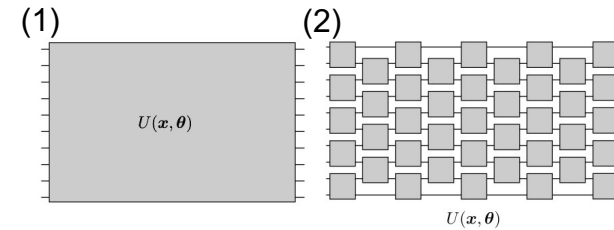


- : 2-design (unitary is sampled from an ensemble of unitary possessing the property of the Haar random unitary up to degree 2)

# Vanishing similarity issue in QFK

## ■ Analytical investigation (Result)

- Analytical results of the QFK for two types of quantum circuits



## Theorem 1 (Informal statement on the ALDQFK)

The expectation of the QFK is zero for both types of quantum circuits. The variance of the QFK does not depend on the number of qubits  $n$ , but on the size of the unitary blocks in the ALA,  $m$ , and the depth of the corresponding unitary block,  $d$  for the ALA, while the issue arises for the random quantum circuits.

Case (1)

$$\text{Var}[\tilde{k}_{QF}] \approx \frac{1}{2^{n+1}}$$

Case (2)

$$\text{Var}[\tilde{k}_{QF}] \geq \frac{2^{2md}(2^{md} - 1)}{2(2^{2m} - 1)^2(2^m + 1)^{4(d-1)}} \approx \frac{1}{2^{md}}$$

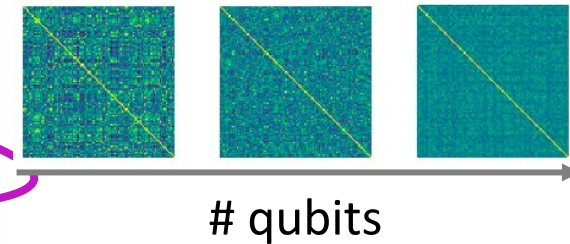
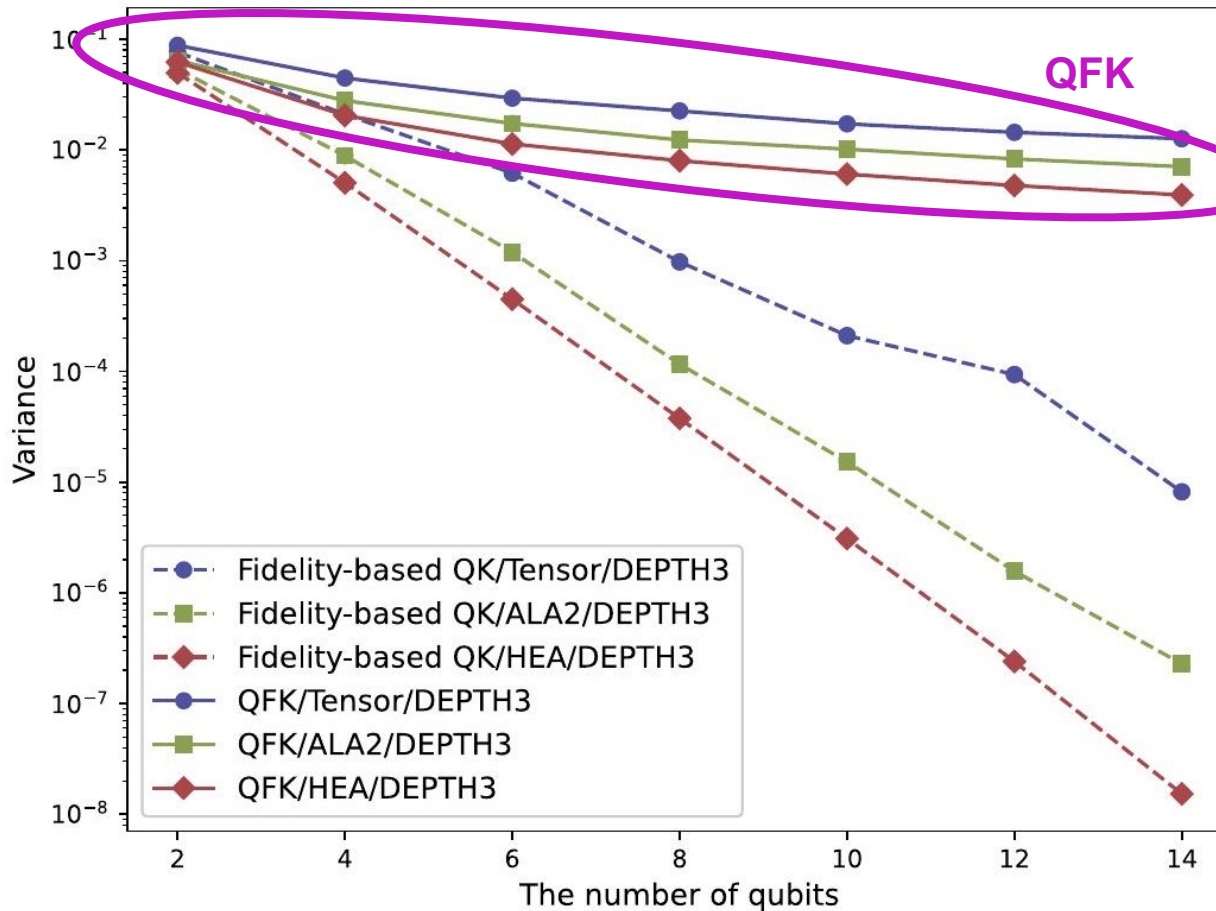
➤ QFK using ALA with shallow depth can avoid the vanishing similarity issue



# Numerical simulations

## ■ Numerical simulations to verify Proposition and Theorem

- Three types of quantum circuits with depth  $L = 3$  are used for both QKs



Quantum circuits:

Tensor-product  
(Tensor)

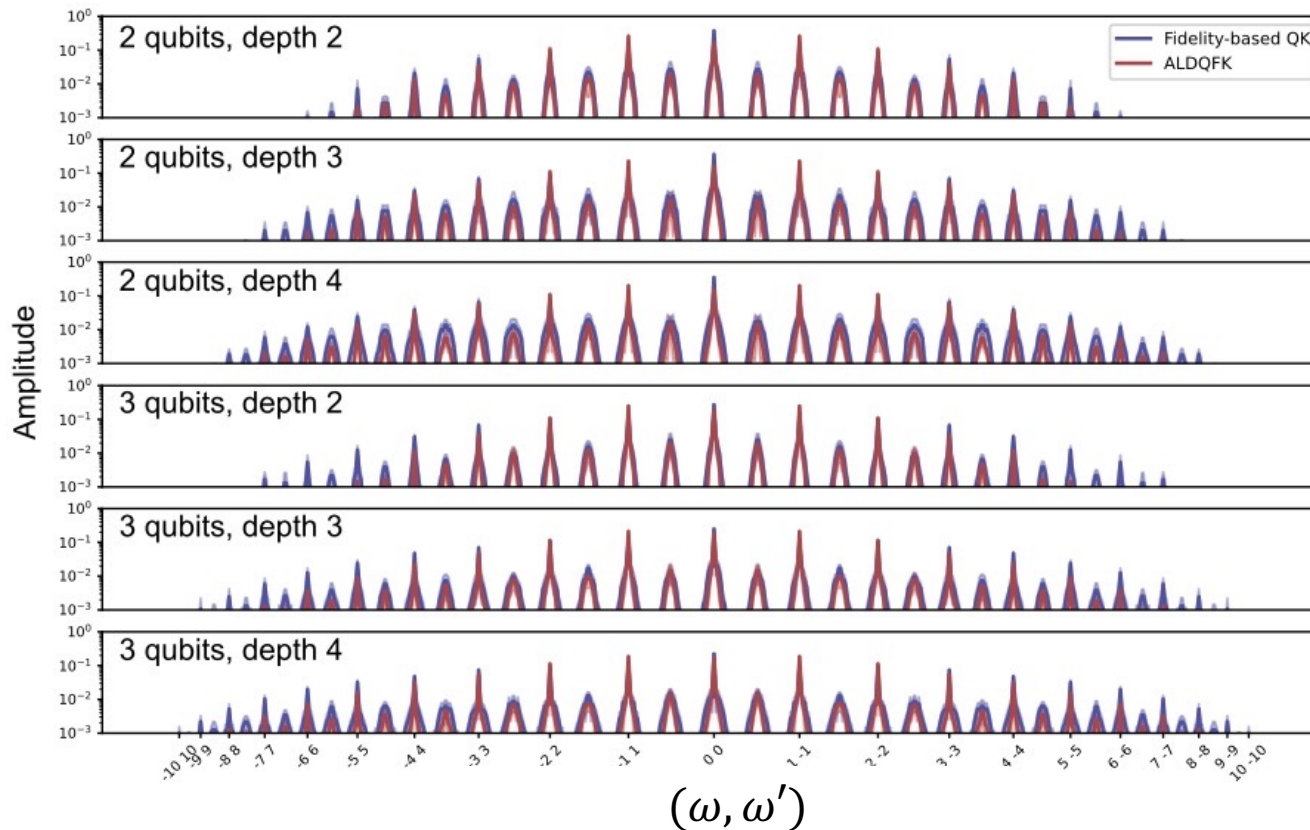
ALA with 2-qubit unitary  
blocks (ALA2)

Hardware efficient  
ansatz (HEA)

# Expressivity analysis of QFK

- Expressivity of the QFK in terms of Fourier analysis [6]
  - Compare the Fourier coefficients of the fidelity-based QK and the QFK

Fourier representation:  $k(x, x') = \sum_{\omega, \omega' \in \Omega} e^{i\omega x} e^{i\omega' x'} c_{\omega, \omega'}$



標準カーネル

フィッシャーカーネル

# Performance analysis of QFK

## ■ Performance analysis using a synthesized dataset

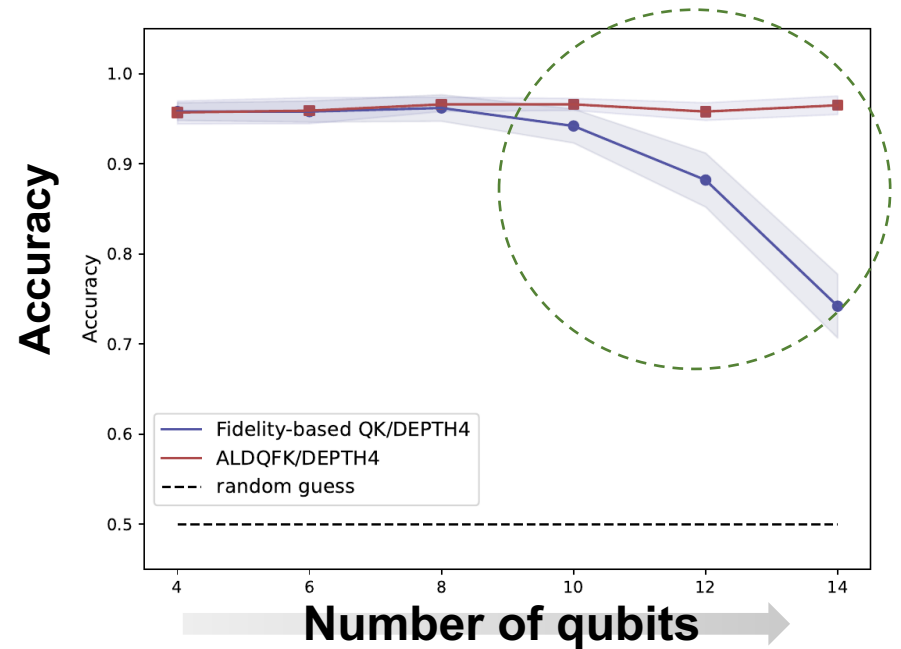
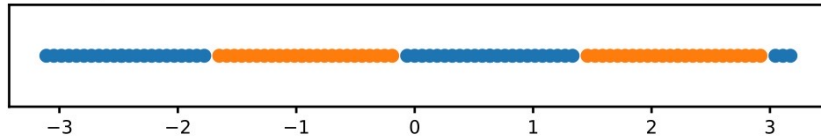
Compare the performance of QKs using a simple binary classification task

Synthesized dataset :  $\{x_i, y_i\}_i^{100}$

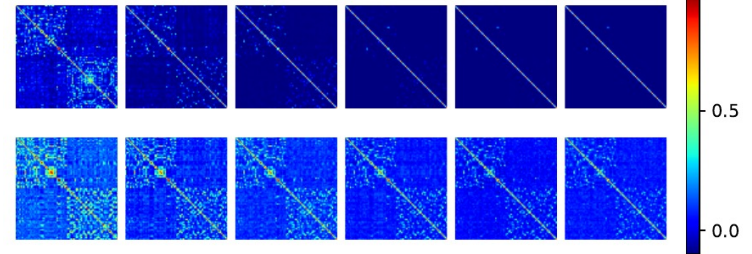
$$y_i = \text{sgn}(\sin(wx_i + 0.3))$$

$w$  : Frequency of the datasets

$$w = 2, \quad b = 0.3$$



Fidelity  
QK



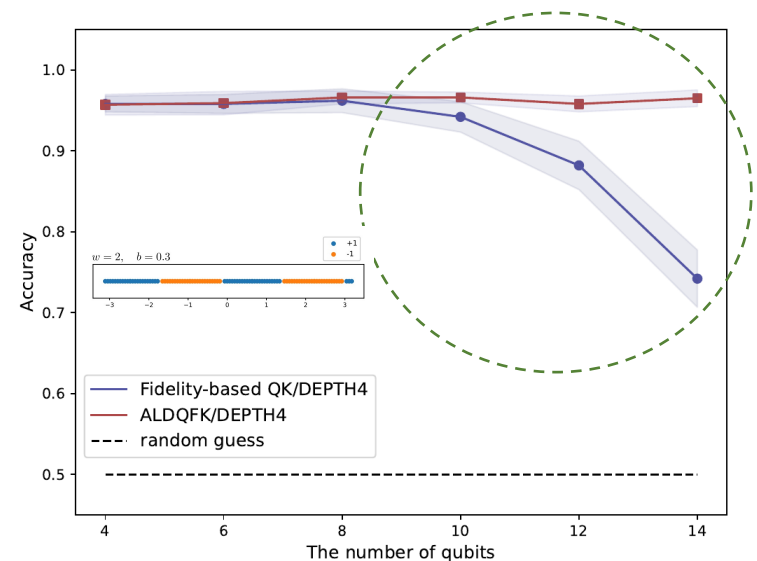
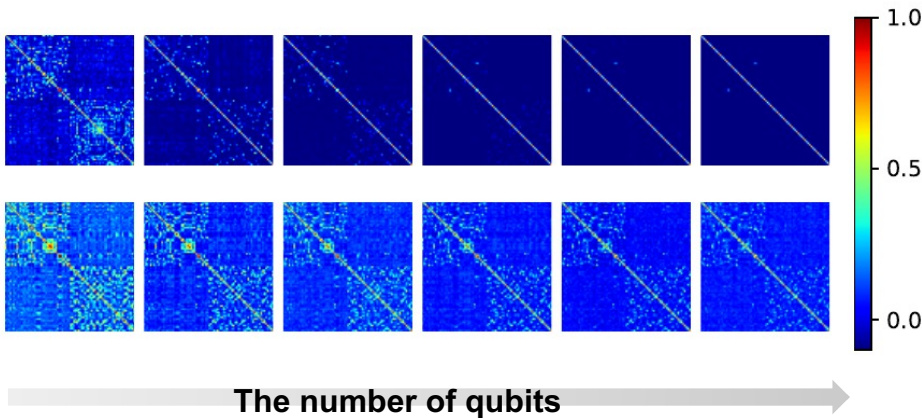
QFK

# Summary

## ■ Summary

- To analytically show the limitation of fidelity-based QK (the vanishing similarity issue)
- To propose a new class of quantum kernels, “*quantum Fisher kernel*” (QFK)
  - Analytically and numerically show the QFK can avoid the vanishing similarity issue
  - Show an example where the QFK can outperform the fidelity-based QK in performance

Fidelity-based QK (upper panel), QFK (lower panel)



## 量子状態モデル選択のための情報量規準

**Quantum information criteria for model selection  
in quantum state estimation  
Yano and Yamamoto, arXiv 2023**

量子機械学習は、量子状態の空間で回帰や分類を実行。  
モデルも、量子状態の空間に「住んでいる」。どんなモデルが  
良いか？量子状態の空間内部に分け入って考えてみよう。

**問題**: 未知のデータ生成源(確率分布)の数理モデルを定めたい。2つのモデル分布  $f_1(x|\theta)$ 、 $f_2(x|\theta)$  のどちらが良いか、データをもとに判定。

→ 生成源とモデルの距離(KL divergence)は赤池情報量規準(AIC)で評価される:

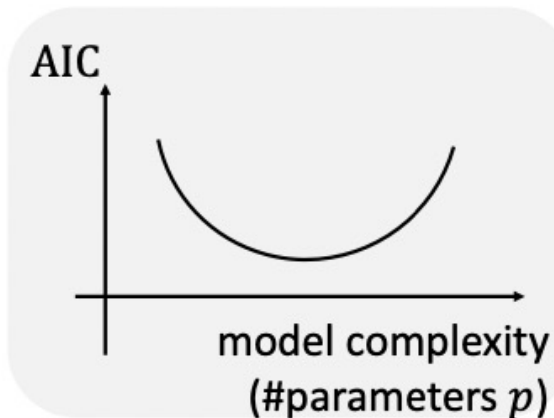
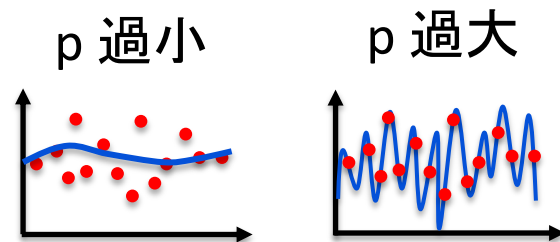
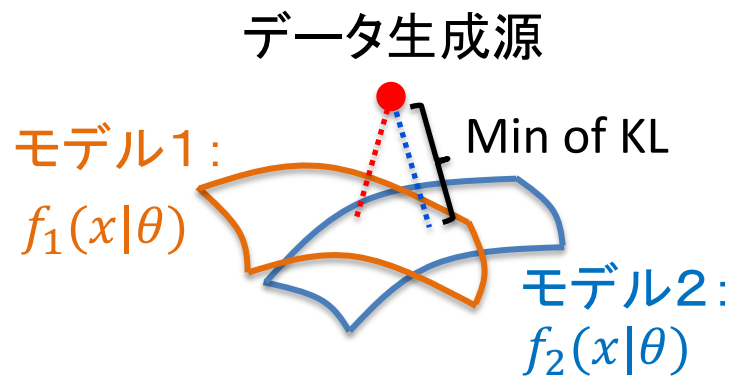
AIC = モデルのデータへの適合具合  
+ モデルの複雑さ(パラメータ数)

$$= -2\sum_k \log f(x_k|\hat{\theta}) + 2p$$

$\{x_k\}$ : データ    ↗    ↖    最尤推定パラメータ

モデルには適合具合と複雑さのトレードオフがある。このこと(オッカムの剃刀)を初めて数式化したのがAIC

モデル選択規準: AIC が小さくなるモデルを選ぶべし



## 注意: 情報量規準AICは確率変数

AIC は **期待対数尤度** に対して漸近的な不偏推定量

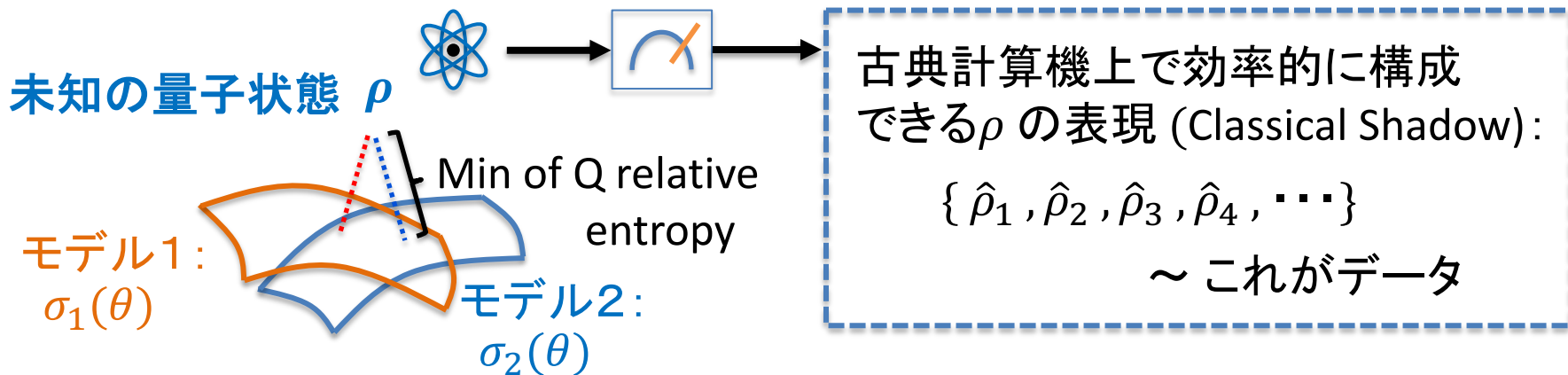
$$\mathbb{E}_{g(\mathbf{x}_n)} \left[ -\frac{1}{2n} \text{AIC} \right] = \mathbb{E}_{g(\mathbf{x}_n)} \left[ \mathbb{E}_{g(z)} \left[ \log f(Z|\hat{\theta}_C) \right] \right] + o(n^{-1})$$

AIC は有限のデータ  $\mathbf{x}_n = \{x_1, x_2, \dots, x_n\}$  によって揺らぐ

しかし比較するモデルが階層モデル  $M_1 \subset M_2 \subset \dots$  のとき  
この揺らぎはモデルに共通して現れるので、モデルの比較  
では揺らぎが相殺され、比較に影響を及ぼさない

今回考えた量子情報量規準も上記と同様の性質を持っている

**問題:** 未知の量子状態  $\rho$  の数理モデル  $\sigma(\theta)$  を定めたい。2つのモデル状態  $\sigma_1(\theta)$ 、 $\sigma_2(\theta)$  のどちらが良いか、“データ”をもとに判定。



生成源とモデルの距離(量子相対エントロピー)を評価する指標を導出:

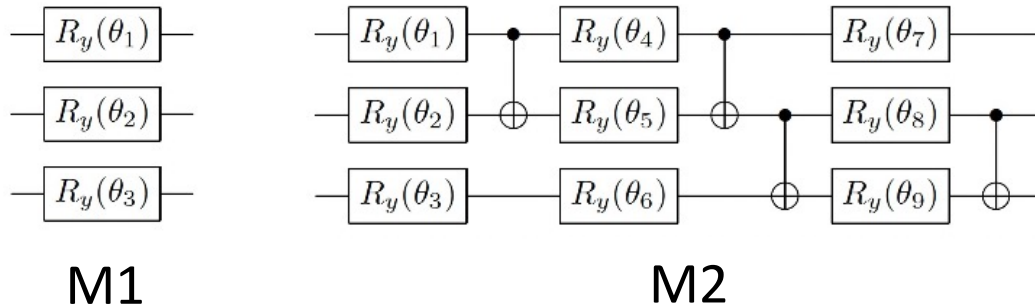
$$\text{QTIC} = -2\text{Tr}[(\sum_k \hat{\rho}_k) \log \sigma(\hat{\theta})] + 2\text{Tr}[F(\hat{\theta})J(\hat{\theta})^{-1}] \quad \text{--- これが小さくなるモデルを選ぶべし}$$

$$\hat{I}_{Q;ij}^{\text{emp}}(\hat{\theta}_Q) = \frac{1}{n} \sum_{\alpha=1}^n \text{Tr} \left( \hat{\rho}_\alpha \frac{\partial \log \sigma(\theta)}{\partial \theta_i} \Big|_{\theta=\hat{\theta}_Q} \right) \text{Tr} \left( \hat{\rho}_\alpha \frac{\partial \log \sigma(\theta)}{\partial \theta_j} \Big|_{\theta=\hat{\theta}_Q} \right)$$

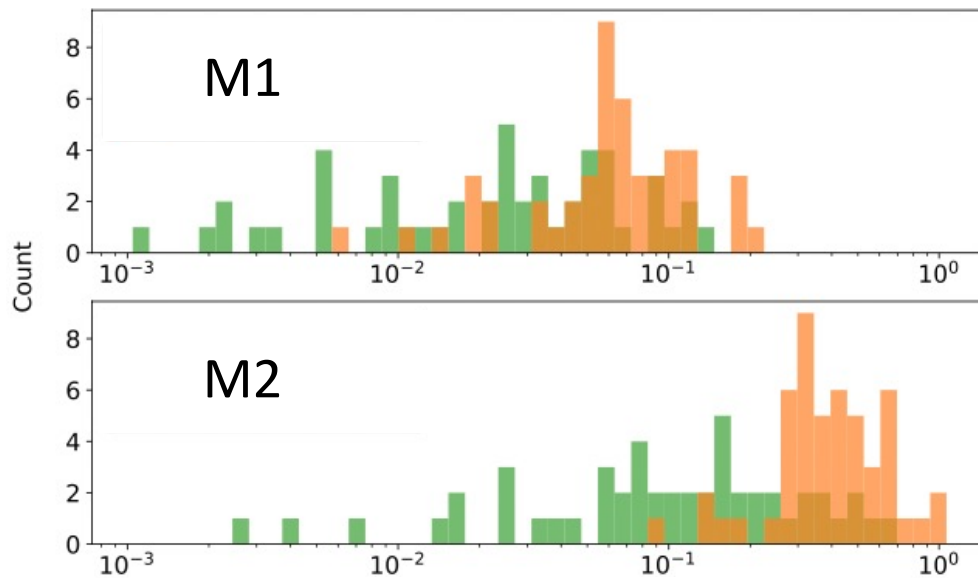
$$\hat{J}_{Q;ij}^{\text{emp}}(\hat{\theta}_Q) = -\text{Tr} \left( \left( \frac{1}{n} \sum_{\alpha=1}^n \hat{\rho}_\alpha \right) \frac{\partial^2 \log \sigma(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\hat{\theta}_Q} \right)$$



# 単純モデルM1 vs 複雑モデルM2 の例:



真の量子状態は  
モデルM1に  
含まれている



横軸: QAIC (を少し変えたもの)

AIC は日本発の統計・学習理論における金字塔。これの量子版構築は不変的な価値があると考えられる。課題は数多い。量子生成モデルの設計問題に応用したい。

# 目次

## 機械学習による分子構造デザイン

新奇量子状態を探索するための量子生成モデルの構築法

量子力学の特性を活かすカーネル設計法の開発

量子状態モデル選択のための情報量規準