Fermionic Born Machines: Classical training of quantum generative models based on Fermion Sampling

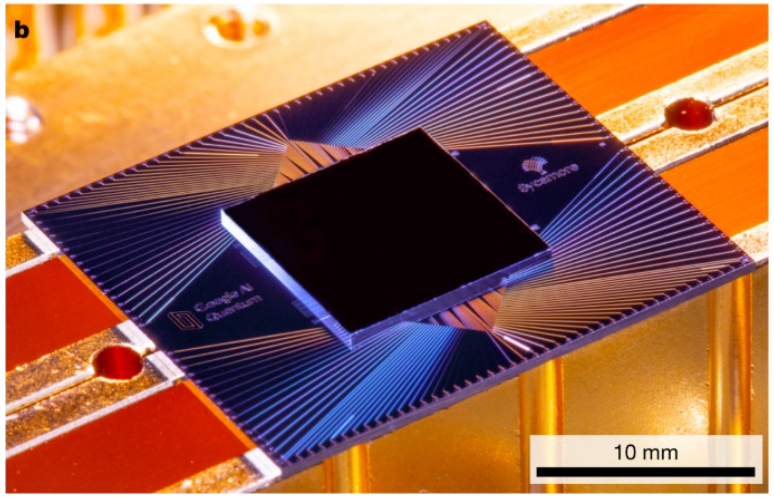
Bence Bakó, Zoltán Kolarovszki, Zoltán Zimborás

arXiv:2511.13844

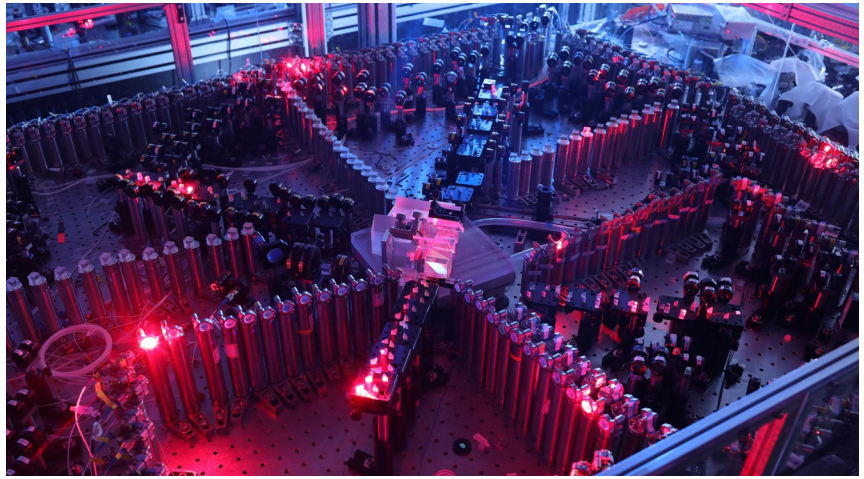
GPU Day 2026

Quantum advantage experiments

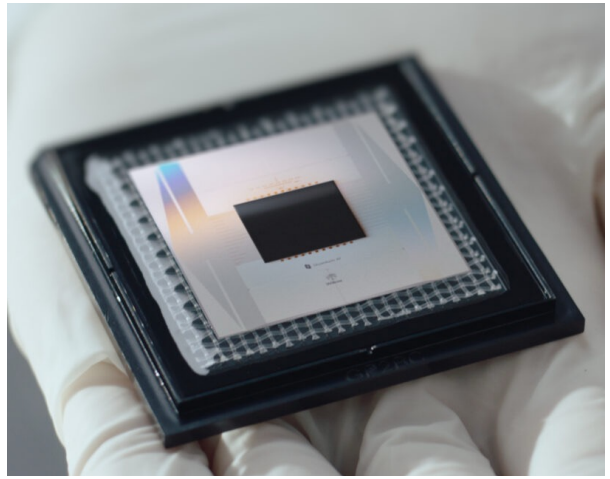
- Based on generating samples from classically hard probability distributions.



Google 2019



USTC 2019



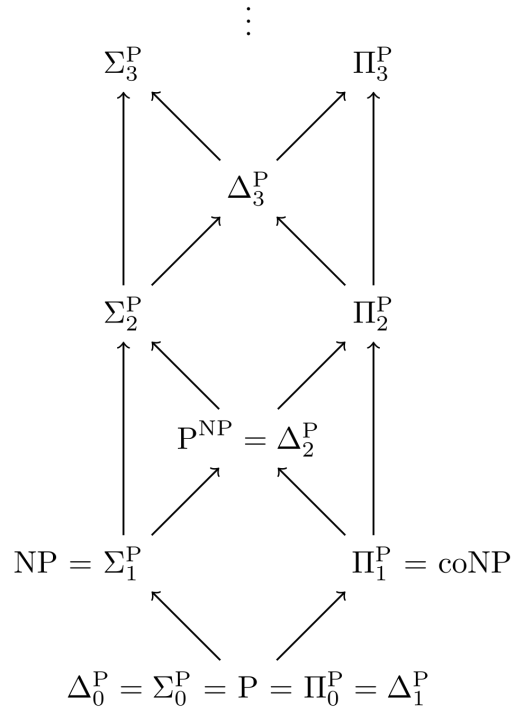
Google 2024

Motivation: Classical hardness guarantees

- Even certain very restricted quantum circuit classes are **on average hard to classically sample** from!

Quantum advantage scheme	Average-case hardness of near-exact computation of $p_x(V)$	Anticoncentration of $p_x(V)$	Certification	Experiments
Random Quantum Circuits [12]	✓	✓	✗ ✓	✓
Boson Sampling [9]	✓	✗	✗ ✓	✗ ✓
Gaussian Boson Sampling [17]	✓	✗	✗ ✓	✓
IQP [51]	✗	✓	✗	✗
Fermion Sampling **(this work)**	✓	✓	✗ ✓	-

M. Oszmaniec, et al. (2022), PRX Quantum, 3(2), 020328.



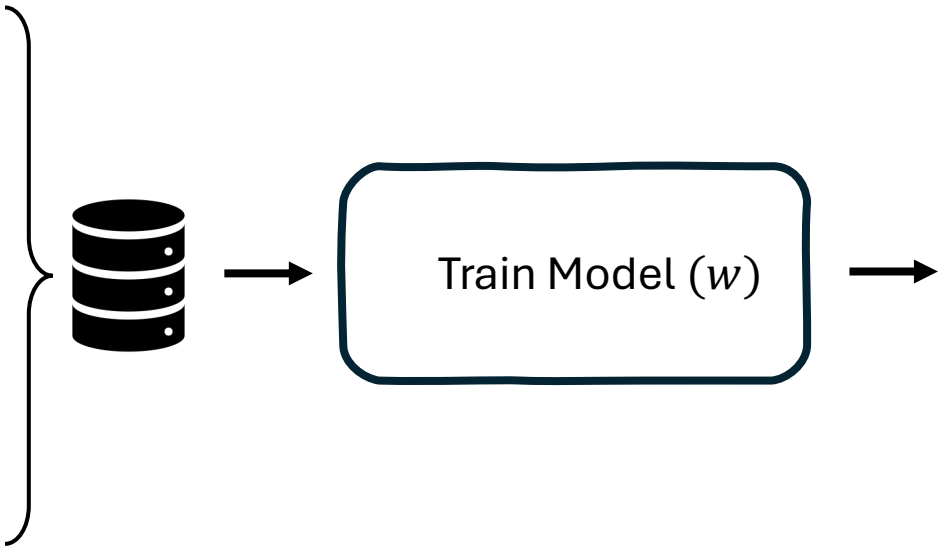
Wikipedia

Generative modeling

- Learn a representation of some **probability distribution** in order to **generate realistic samples**.

$x_i \sim p$

$y_i \sim q_w$



$d(p, q_w) \leq \epsilon$

QML: Potential and limitations

QML models have high expressivity

BUT

Training is challenging

QML: Potential and limitations

QML models have high expressivity

BUT

Training is challenging

- Barren plateaus
- Poor local minima
- COSTLY GRADIENTS!

QML: Potential and limitations

QML models have high expressivity

BUT

Training is challenging

- Barren plateaus
- Poor local minima
- COSTLY GRADIENTS!

Solution (?)

- Problem-specific models
- Incorporating symmetries
- Restricted circuit classes

Provably good trainability → effectively classical

- In many cases, **good trainability** implies classical **simulability of expectation values**.
- What is a bad news for supervised learning, can be **good news for generative learning!**

1. Efficient classical simulability of local expectation values

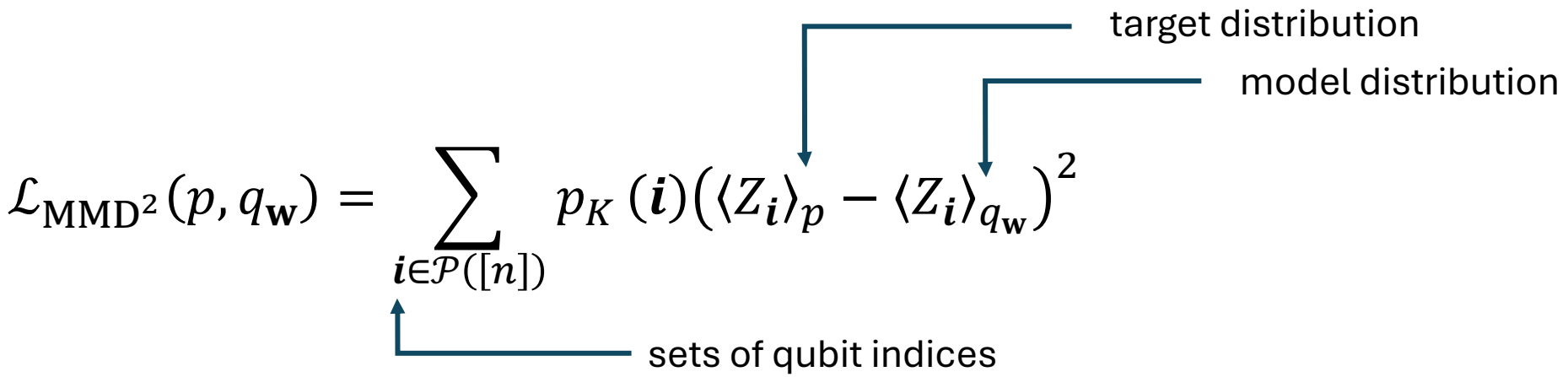
↔

2. Classical sampling hardness of the probability distribution

M. Cerezo, et al. *Nat Commun* **16**, 7907 (2025)

Classical training of quantum models

The **squared Maximum Mean Discrepancy (MMD)** can be expressed with expectation values.



Introduced for **IQP circuits**.

M. S. Rudolph, et al. npj QI 10, 116 (2024)

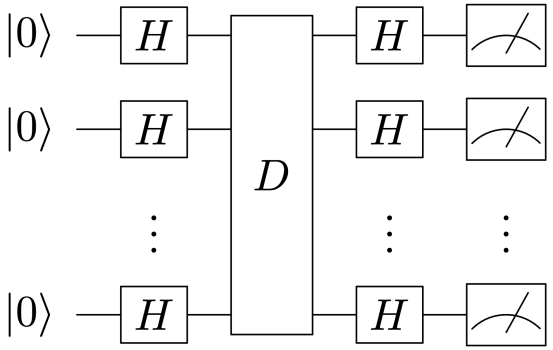
E. Recio-Armengol, S. Ahmed, and J. Bowles. arXiv:2503.02934 (2025).

Classical training of IQP circuits

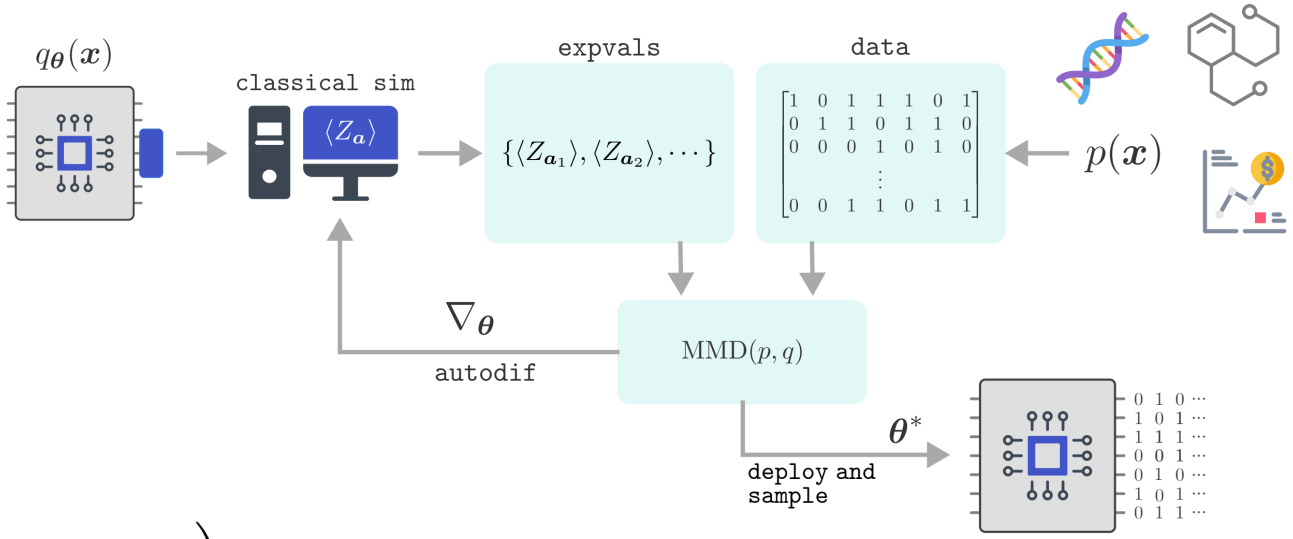
$$\langle \hat{Z}_a \rangle_{q_\theta} = \frac{1}{|\mathcal{Z}|} \sum_{i=1}^{|\mathcal{Z}|} \cos\left(\sum_j \theta_j \, (-1)^{g_j \cdot z_i} (1 - (-1)^{g_j \cdot a}) \right)$$

Unbiased estimator

$$\mathbb{E}_{\mathcal{Z}} \left[\langle \hat{Z}_a \rangle_{q_\theta} \right] = \langle Z_a \rangle_{q_\theta}$$



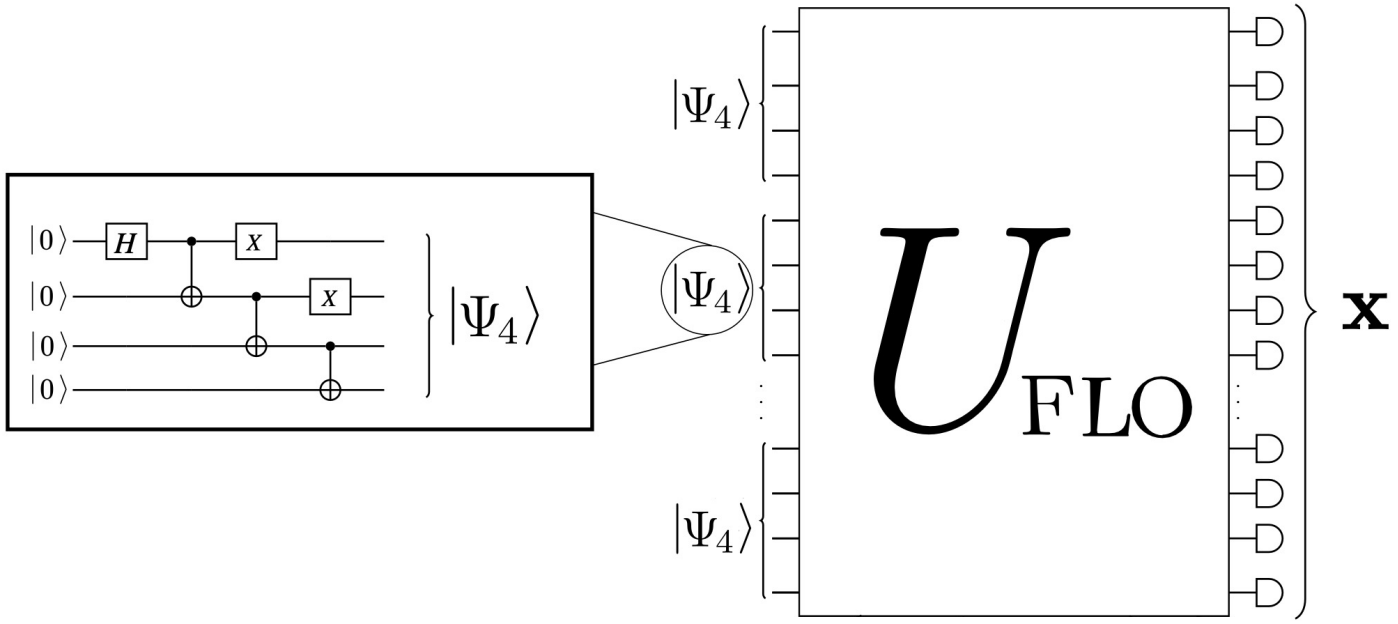
$$D(\boldsymbol{\theta}) = \exp\left(\sum_{\mathcal{A}} i\theta_{\mathcal{A}} \prod_{i \in \mathcal{A}} Z_i \right)$$

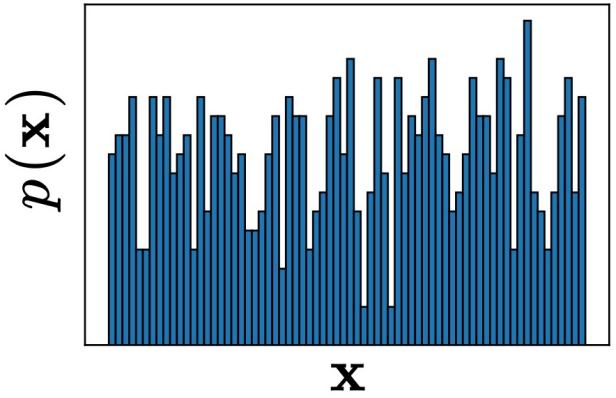


E. Recio-Armengol, et al., arXiv:2503.02934 (2025).

Fermion Sampling

magic state input + FLO circuit = sampling advantage





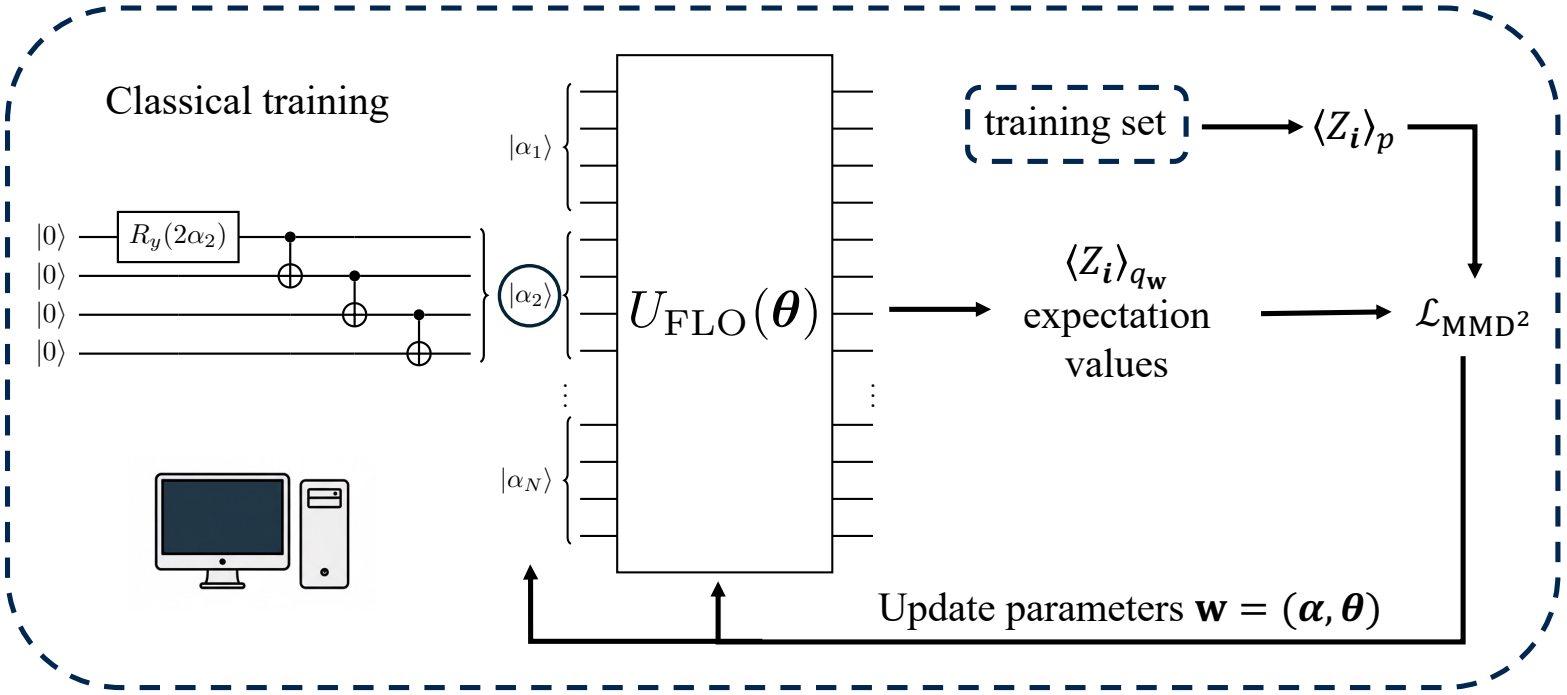
M. Oszmaniec, N. Dangniam, M. E. Morales, Z. Zimborás. *PRX Quantum*, *3*(2), 020328. (2022)

Fermionic Born Machines

Input states:

$$|\psi_j\rangle = \cos\alpha_j|0000\rangle + \sin\alpha_j|1111\rangle$$

- Can be transformed to any even fermionic state
- Only measure the first 3 modes in each 4-mode register



Estimating expectation values

1. Decompose input states into Gaussian operators with non-zero trace

$$|\psi_j\rangle\langle\psi_j| = |\Phi_0\rangle\langle\Phi_0| + |\Phi_0\rangle\langle\Phi_1| + |\Phi_1\rangle\langle\Phi_0| + |\Phi_1\rangle\langle\Phi_1|$$

where

$$|\Phi_0\rangle = \cos\alpha_j|0000\rangle + |1100\rangle \qquad |\Phi_1\rangle = \sin\alpha_j|1111\rangle - |1100\rangle$$

Estimating expectation values

For N registers: 4^N Gaussian constituents

1. Decompose input states into Gaussian operators with non-zero trace

$$|\psi_j\rangle\langle\psi_j| = |\Phi_0\rangle\langle\Phi_0| + |\Phi_0\rangle\langle\Phi_1| + |\Phi_1\rangle\langle\Phi_0| + |\Phi_1\rangle\langle\Phi_1|$$

where

$$|\Phi_0\rangle = \cos\alpha_j|0000\rangle + |1100\rangle \qquad |\Phi_1\rangle = \sin\alpha_j|1111\rangle - |1100\rangle$$

Estimating expectation values

For N registers: 4^N Gaussian constituents

1. Decompose input states into Gaussian operators with non-zero trace

$$|\psi_j\rangle\langle\psi_j| = |\Phi_0\rangle\langle\Phi_0| + |\Phi_0\rangle\langle\Phi_1| + |\Phi_1\rangle\langle\Phi_0| + |\Phi_1\rangle\langle\Phi_1|$$

where

$$|\Phi_0\rangle = \cos\alpha_j|0000\rangle + |1100\rangle \qquad |\Phi_1\rangle = \sin\alpha_j|1111\rangle - |1100\rangle$$

2. We can separate the Gaussian part of these states

$$\rho_{ij} = \frac{|\Phi_i\rangle\langle\Phi_j|}{\langle\Phi_i|\Phi_j\rangle}$$

$$|\psi_j\rangle\langle\psi_j| = \rho_{\text{Gauss}}(\alpha_j) + \sigma(\alpha_j)$$

$$\sigma(\alpha_j) = -\rho_{\text{Gauss}}(\alpha_j) + (\cos^2\alpha_j + 1)\rho_{00}(\alpha_j) + (\sin^2\alpha_j + 1)\rho_{11}(\alpha_j) - \rho_{01}(\alpha_j) - \rho_{10}(\alpha_j)$$

Estimating expectation values

- For 1-long Z-strings:
  Gaussian contributions
- For 2- and 3-long Z-strings:
  Gaussian contributions
  + a single register with the non-Gaussian part
- For 4- and 5-long Z-strings:
  Gaussian contributions
  + a single register with the non-Gaussian part
  + two registers with the non-Gaussian part

Non-Gaussian contribution

$$|\psi_j\rangle\langle\psi_j| = \rho_{\mathrm{Gauss}}(\alpha_j) + \sigma(\alpha_j)$$

Estimating expectation values

- For 1-long Z-strings:
  Gaussian contributions
- For 2- and 3-long Z-strings:
  Gaussian contributions
  + a single register with the non-Gaussian part
- For 4- and 5-long Z-strings:
  Gaussian contributions
  + a single register with the non-Gaussian part
  + two registers with the non-Gaussian part

Non-Gaussian contribution

$$|\psi_j\rangle\langle\psi_j| = \rho_{\mathrm{Gauss}}(\alpha_j) + \sigma(\alpha_j)$$

For l-long Z-strings → $\mathcal{O}(N^{\lfloor l/2 \rfloor})$

Estimating expectation values

- For 1-long Z-strings:
  Gaussian contributions
- For 2- and 3-long Z-strings:
  Gaussian contributions
  + a single register with the non-Gaussian part
- For 4- and 5-long Z-strings:
  Gaussian contributions
  + a single register with the non-Gaussian part
  + two registers with the non-Gaussian part

Non-Gaussian contribution

$$|\psi_j\rangle\langle\psi_j| = \rho_{\mathrm{Gauss}}(\alpha_j) + \sigma(\alpha_j)$$

Expectation values estimated exactly!

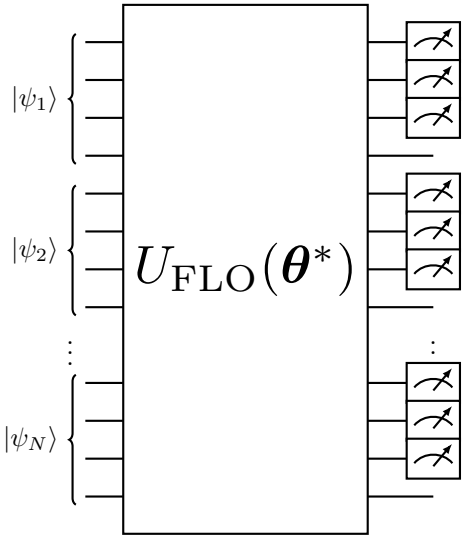
For l-long Z-strings → $\mathcal{O}(N^{\lfloor l/2 \rfloor})$

Inference (sampling)

After training:

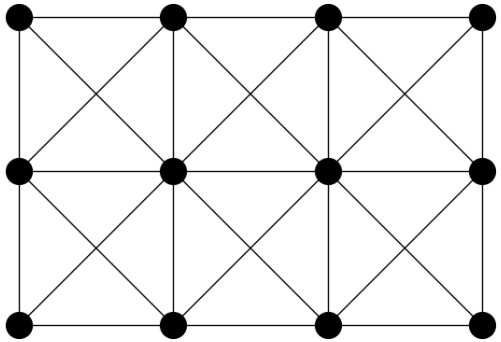
1. Fix the parameters
2. Decompose unitary
3. Use Jordan-Wigner transformation
4. Run circuit on qubit-based device

$$|\psi_j\rangle = \cos\alpha_j^*|0000\rangle + \sin\alpha_j^*|1111\rangle$$

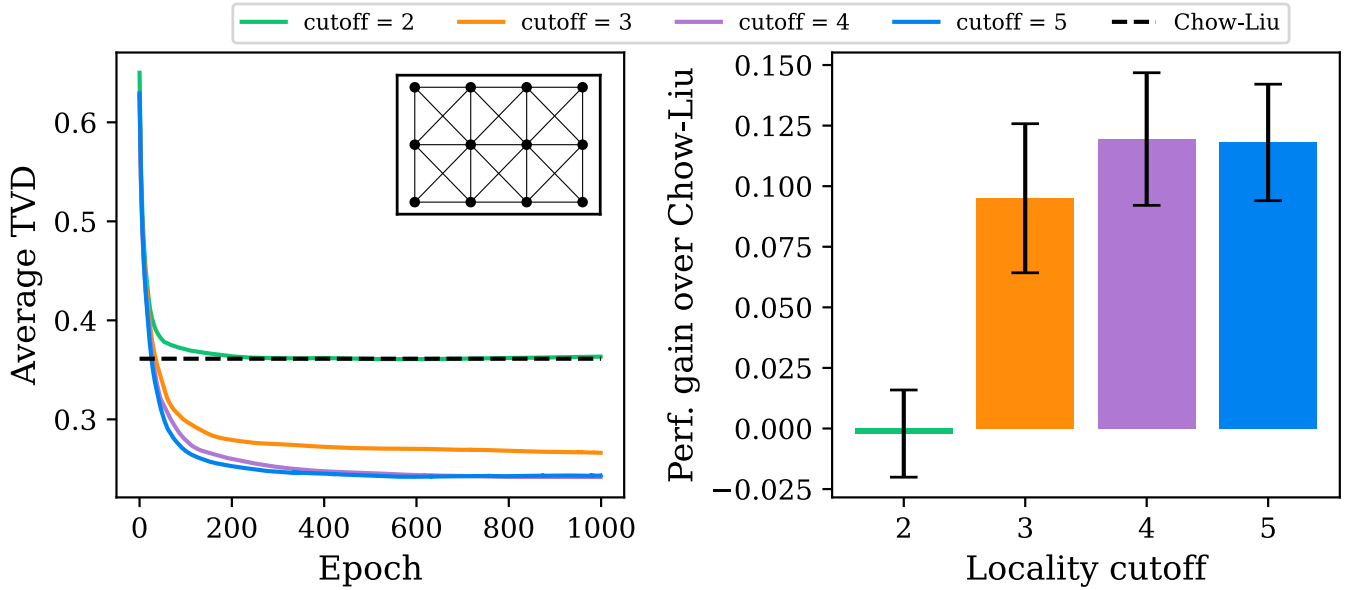


$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

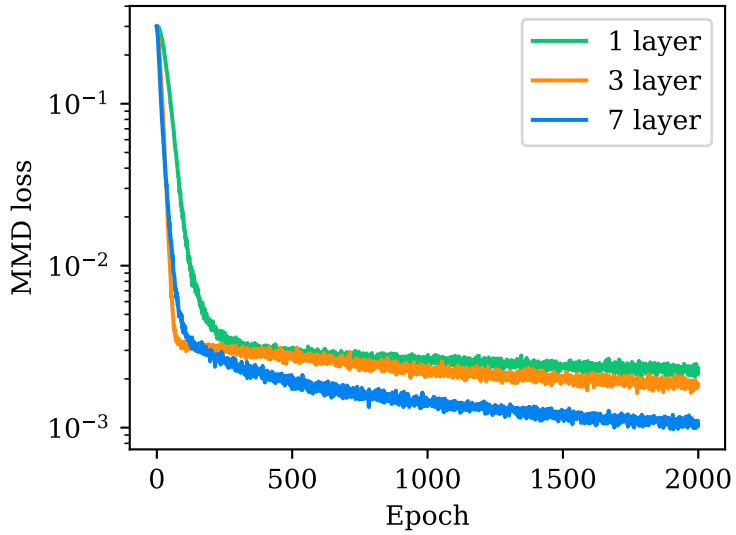
Small-scale experiment

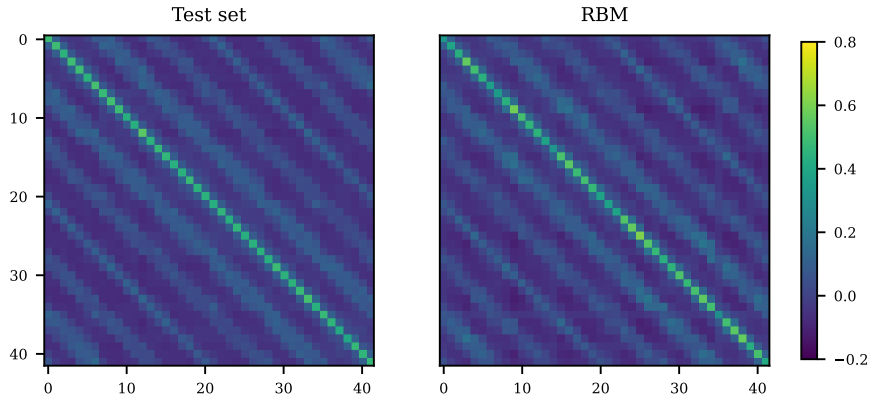


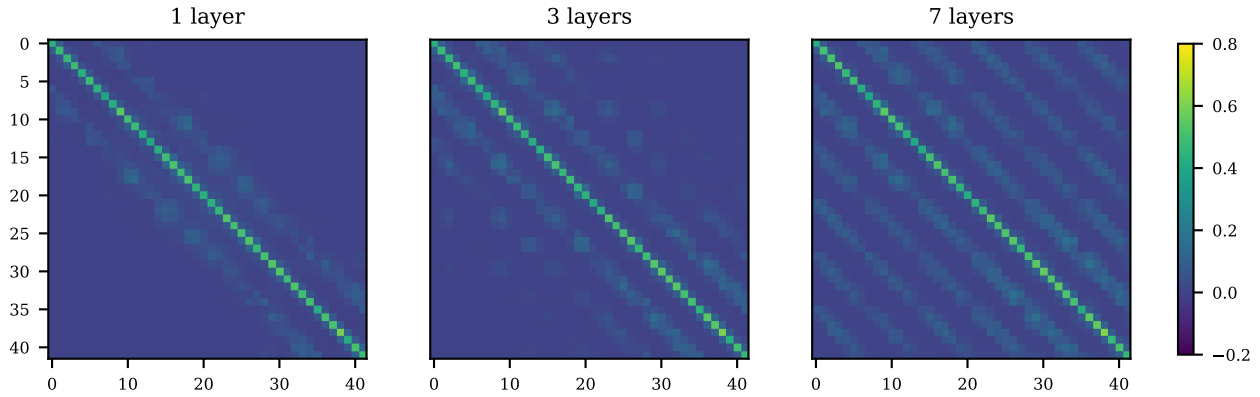
- 12 binary random variables with Markov network factorization
- 16 qubits
- 10 independent problems – average performance



Effect of overparametrization



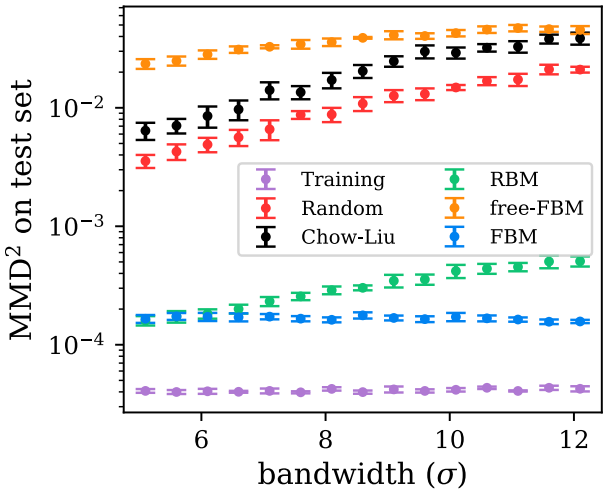


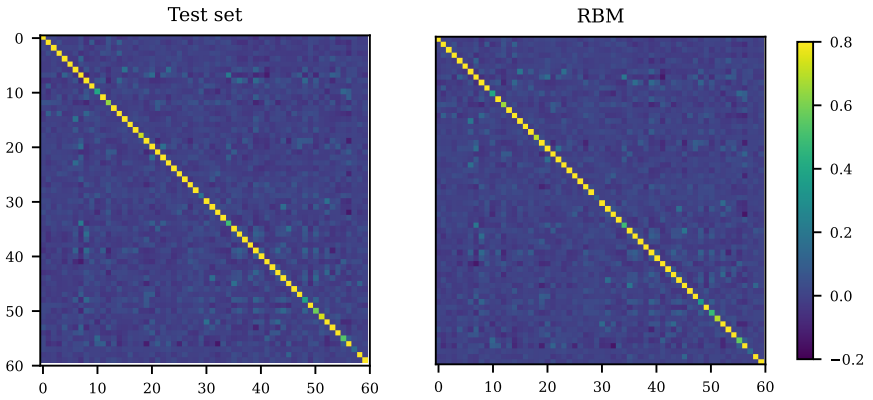


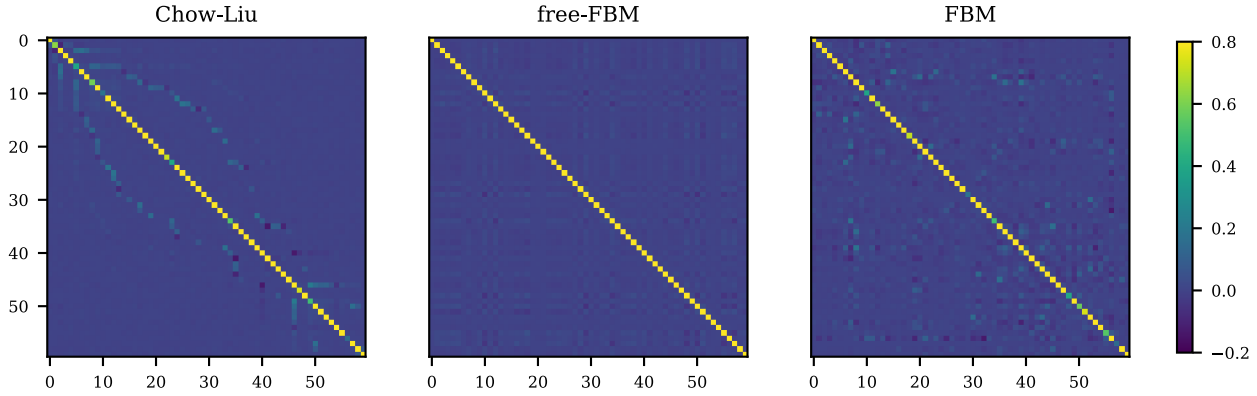
Molecular structure generation

- ZINC20 database
- Small molecules represented as SLIMES strings
- Converted to 60-bit long Morgan fingerprints (RDKit package)
- 60 binary random variables
- **80 qubits**
- 50k training and test samples

Molecular structure generation



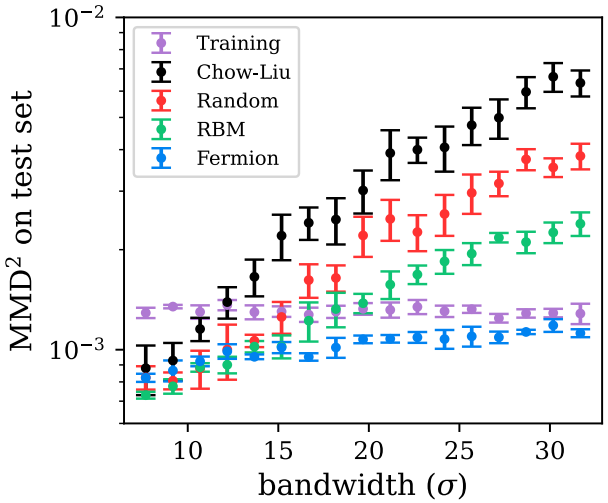


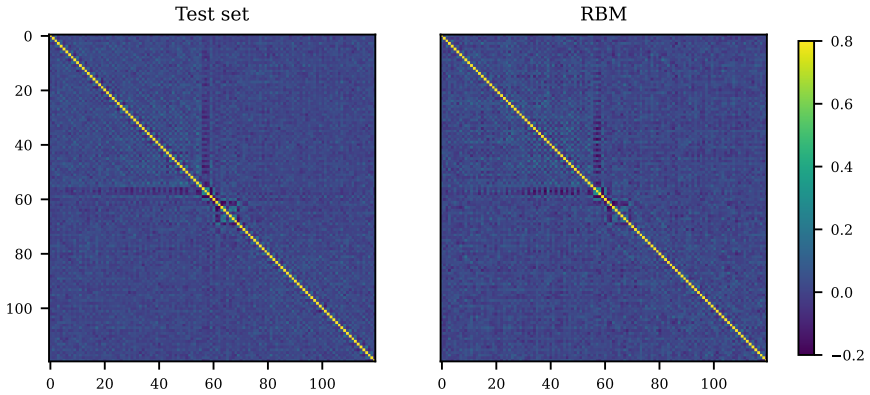


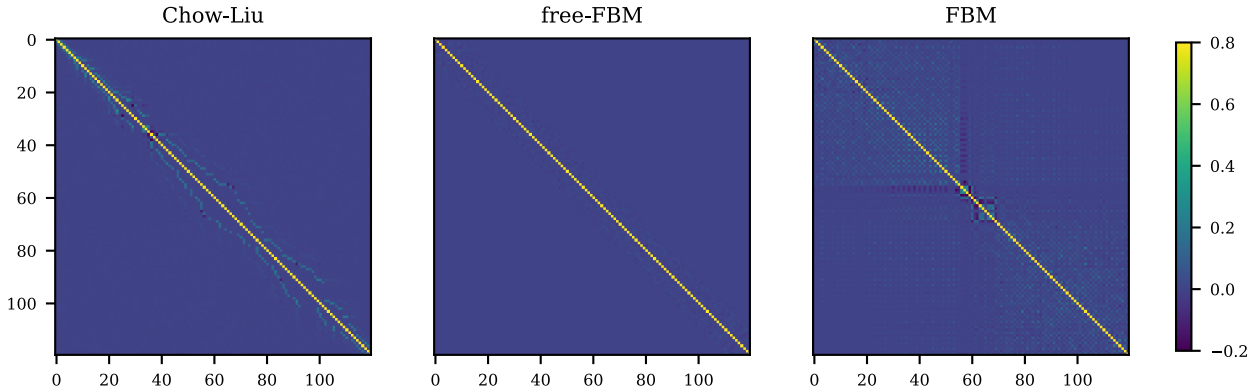
Gene sequences

- Splice-junction gene dataset
- 60-long gene sequence: A, G, C, T
- 3175 samples: 1270 for training, 1905 for test
- 120 binary random variables
- **160 qubits**
- 50-50 train-test split

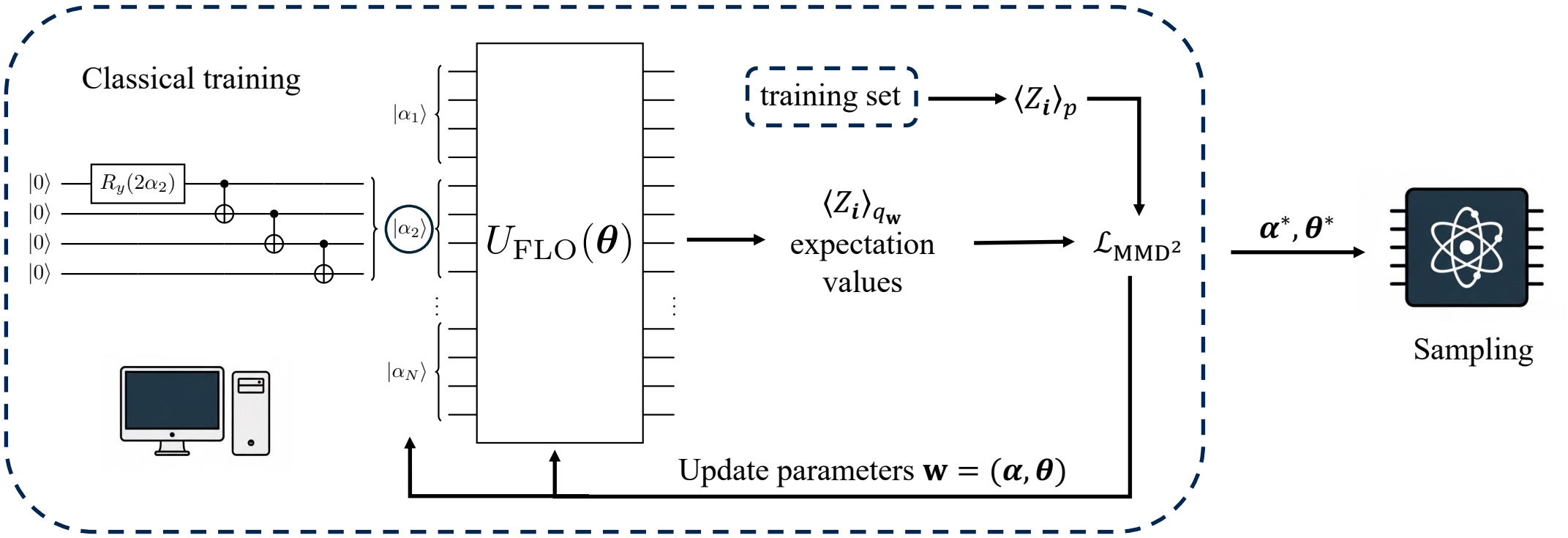
Gene sequences







Summary



- Good **trainability** properties!
- **Expectation values** estimated classically!
- **Sampling** requires a quantum computer!

Epilogue: Spectral methods are natural for quantum computers

- Fourier decomposition of a probability distribution

$$p(x) = \frac{1}{2\pi}\int_{\Omega} p\,(\omega) e^{-2\pi i x\omega} d\omega$$

Fourier coefficients: the MMD training fits these coefficients

- For binary random variables: Walsh-Hadamard transform ($\mathbb{Z}_2^n$)

$$p(x) = \frac{1}{2^n}\sum_{\omega\in\{0,1\}^n} p\,(\omega)(-1)^{x\cdot\omega}$$

Parity function

V. Belis, et al. "Spectral methods: crucial for machine learning, natural for quantum computers?." *arXiv:2603.24654.* (2026)