Introduction to Quantum Machine Learning (QML)

Bence Bakó

HUN-REN Wigner RCP

ELTE IK

Lectures on Modern Scientific Programming 2025

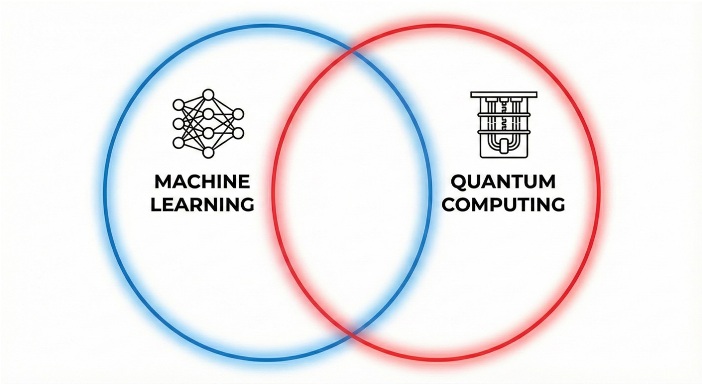
Outline

1. Why QML?
2. What is QML?
3. Basics ingredients of ML
4. Variational QML (quantum neural networks)
5. Quantum generative modelling
6. Recent advances in generative QML

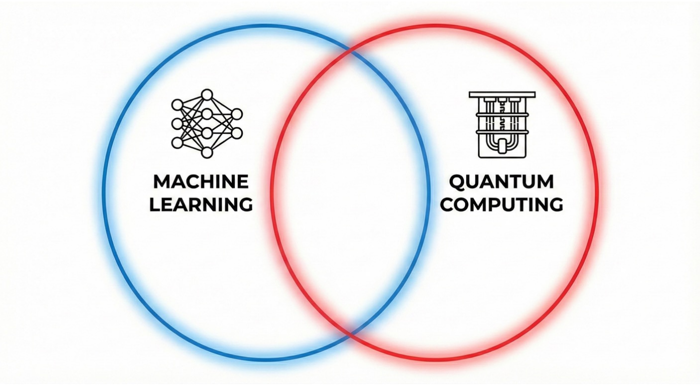
(Re)sources and further reading

- Book: Maria Schuld, Francesco Petruccione: **Machine Learning with Quantum Computers**
- Su Yeon Chang, Marco Cerezo: **A Primer on Quantum Machine Learning** (recent review)
- Pennylane demos, tutorials (**pennylane.ai**)
- Some images were generated by Gemini

1. Why QML?

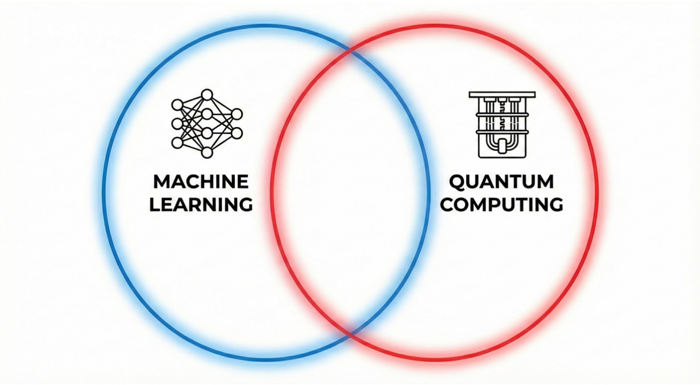


1. Why QML?



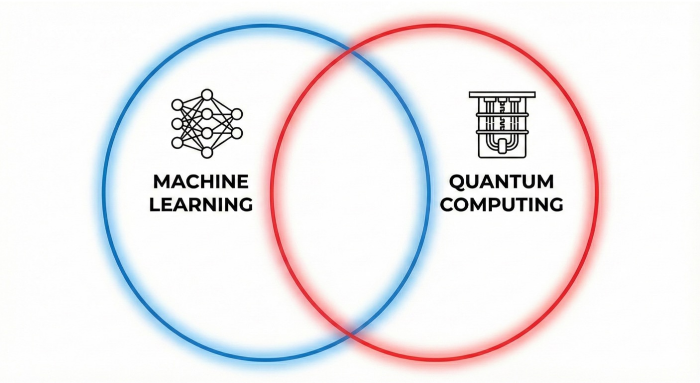
- Sounds cool (?)

1. Why QML?



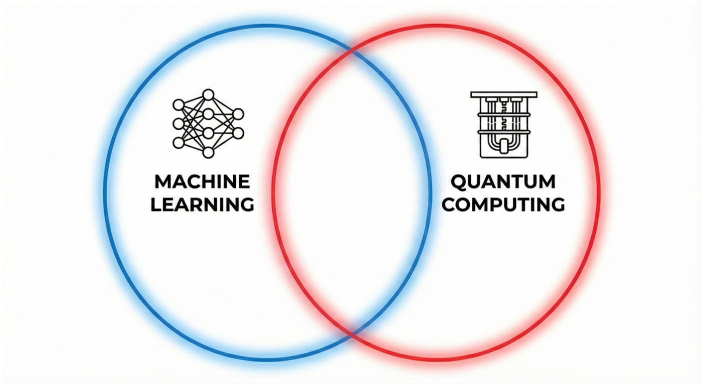
- Sounds cool (?)
- Algorithmic speedups for ML tasks

1. Why QML?



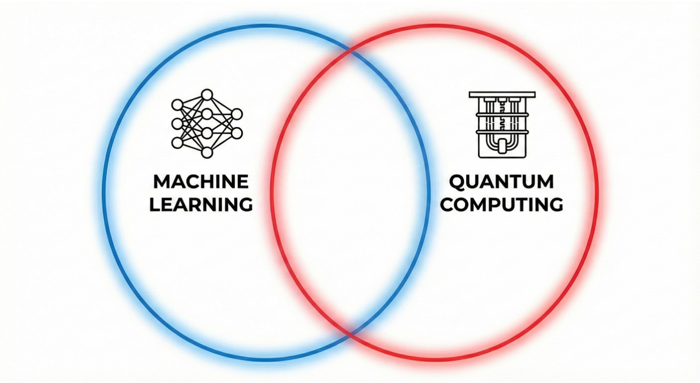
- Sounds cool (?)
- Algorithmic speedups for ML tasks
- ML on quantum data

1. Why QML?



- Sounds cool (?)
- Algorithmic speedups for ML tasks
- ML on quantum data
- Potential for better models:
  - Higher expressivity
  - Better generalization
  - ...

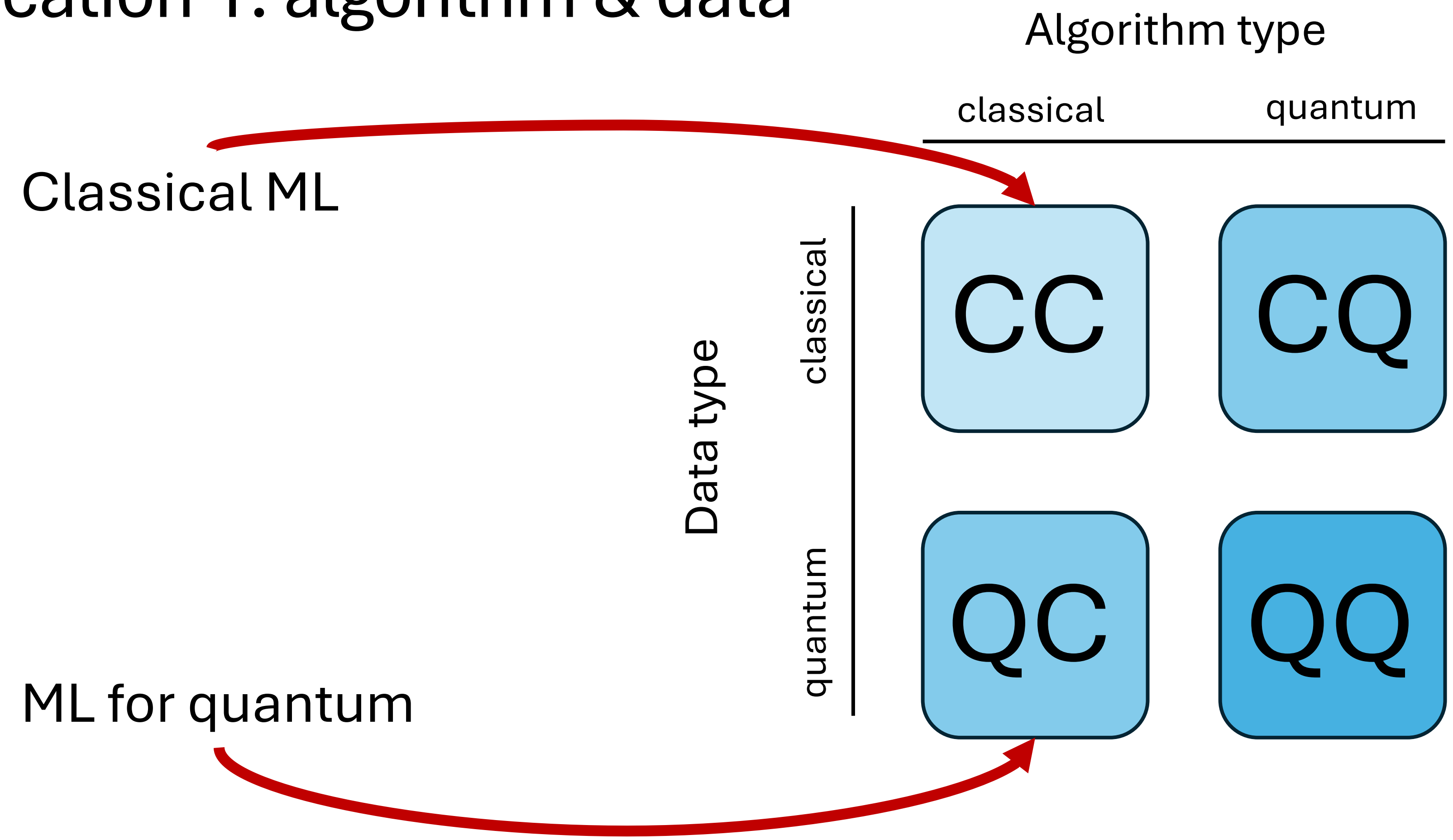
1. Why QML?



- Sounds cool (?)
- Algorithmic speedups for ML tasks
- ML on quantum data
- Potential for better models:
  - Higher expressivity
  - Better generalization
  - ...
- Sampling advantage for generative models

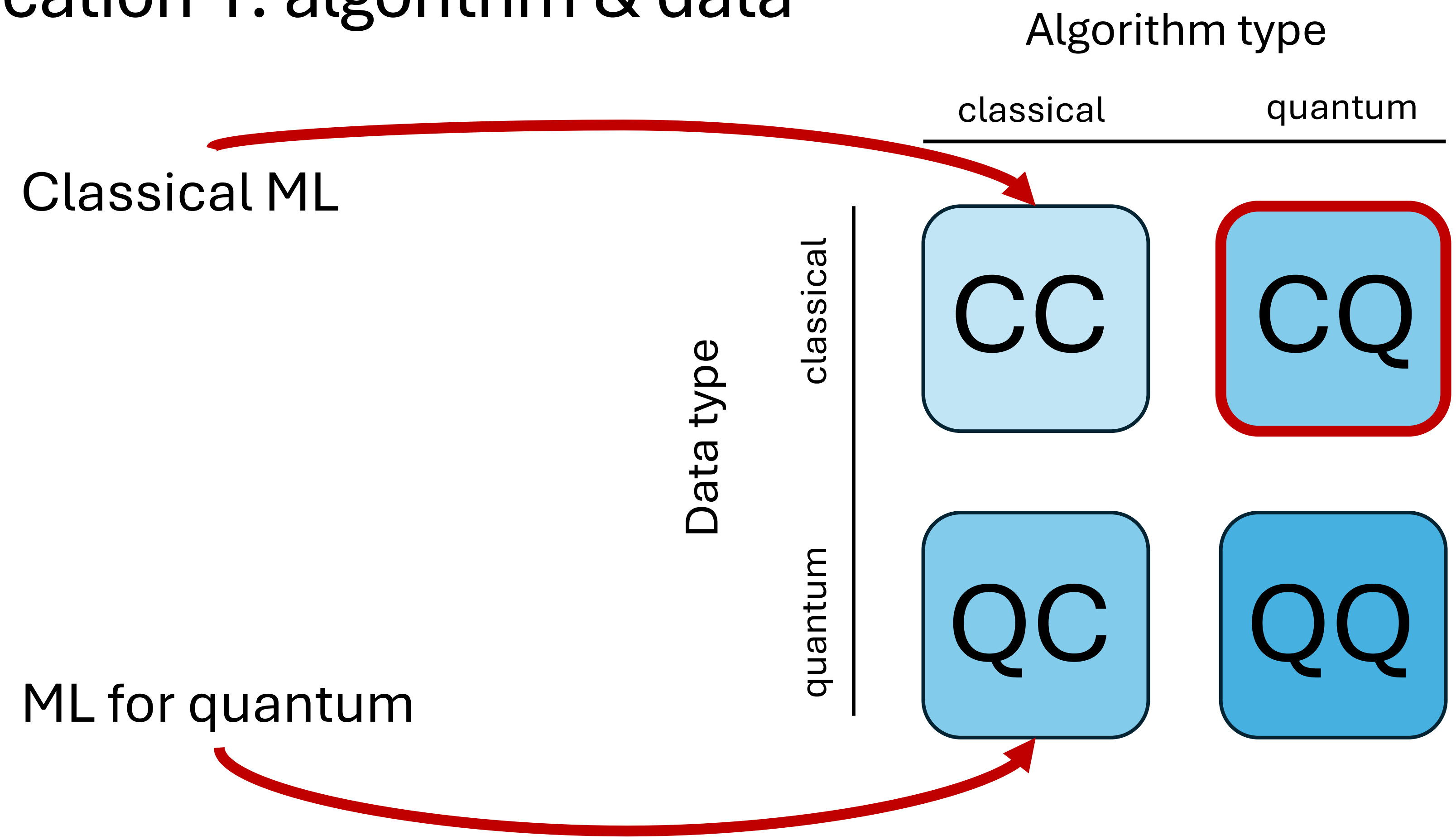
2. What is QML?

Classification 1: algorithm & data



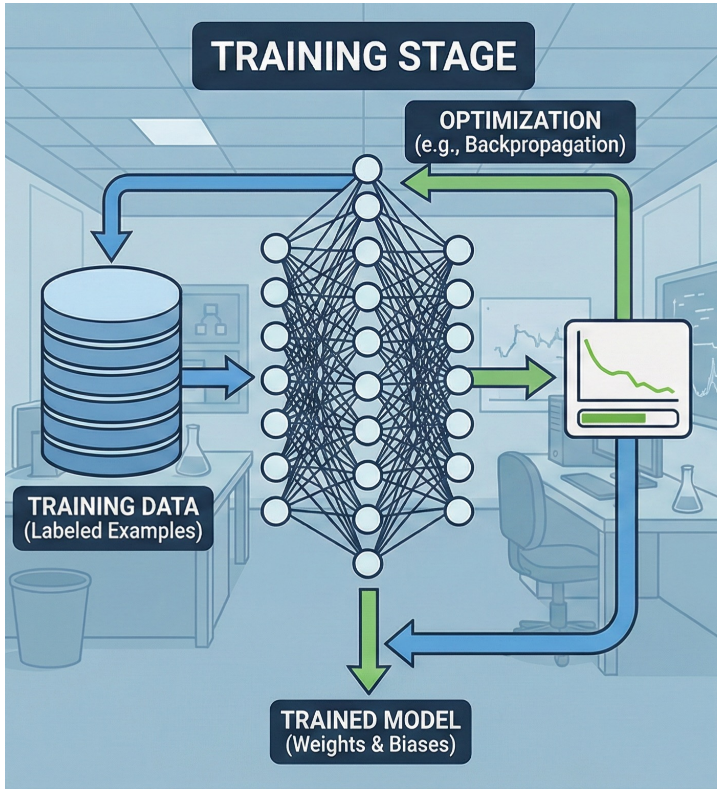
2. What is QML?

Classification 1: algorithm & data

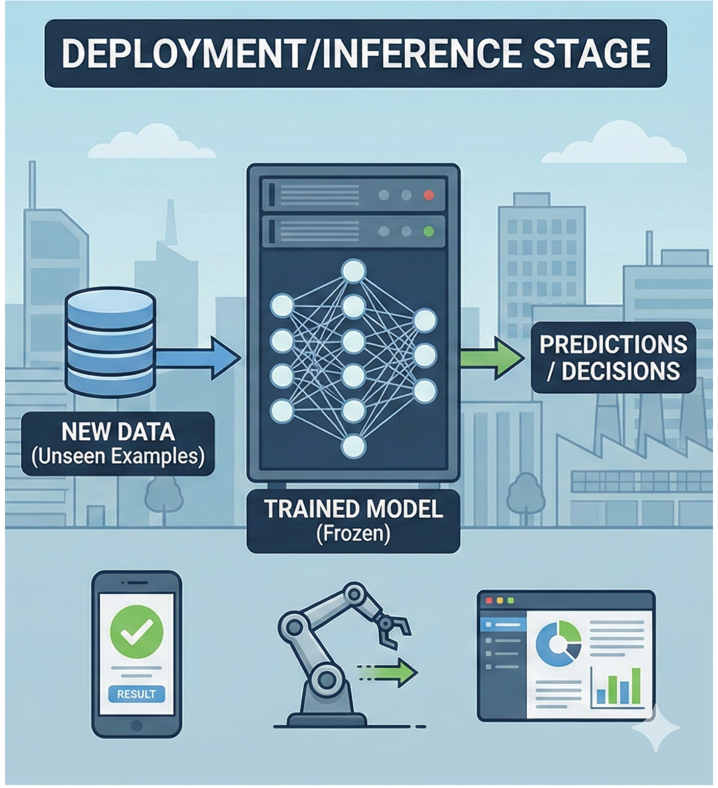


2. What is QML?

Classification 2: training & inference



Training algorithm \ Inference algorithm	classical	quantum
classical	CC	CQ
quantum	QC	QQ

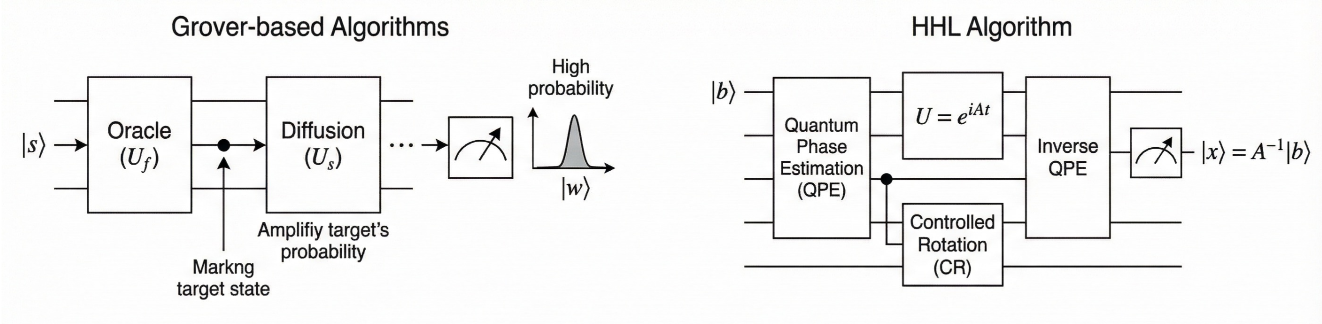


2. What is QML?

Classification 3: paradigm

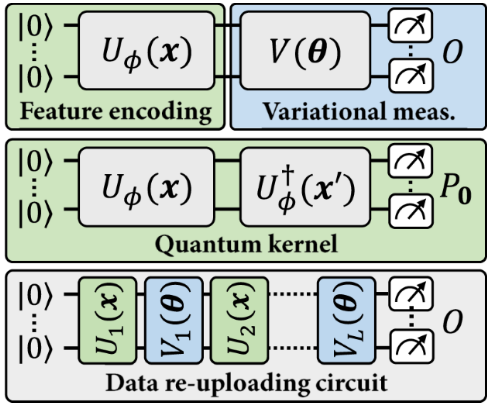
Pre-2018:

- Using established quantum algorithms to speed up ML pipelines
- **Grover**-based algorithms
- **HHL**-based algorithms
- Require large, **fault-tolerant** circuits
- Polinomial speedup guarantees



Post-2018:

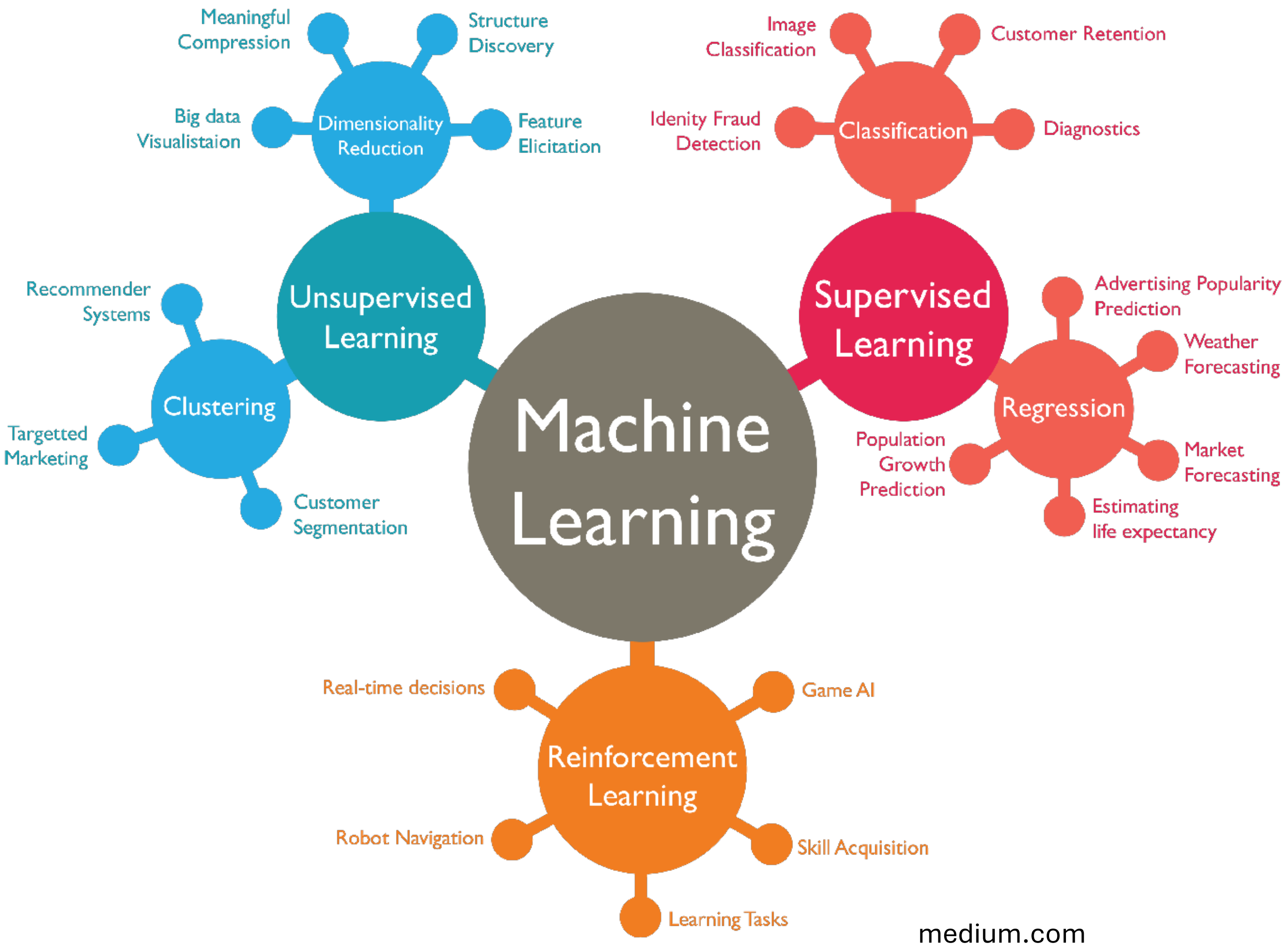
- Parametrized quantum circuits
- **Hybrid** algorithms
- Suitable for noisy devices (?)
- More **heuristic** approach



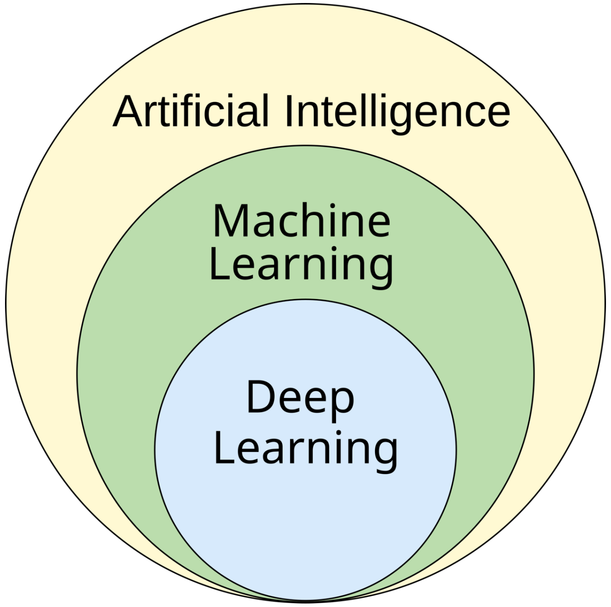
S. Jerbi et al. *Nature Communications*, *14*(1), 517.

3. Basic ingredients of ML

3. Basic ingredients of ML

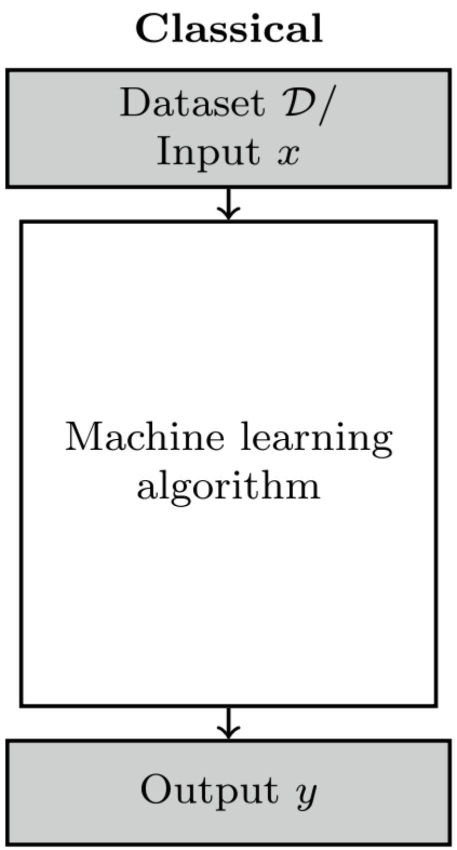


medium.com



wikipedia.org

3. Basic ingredients of ML



Example:

x – feature vector (real)
y – binary label (0/1)

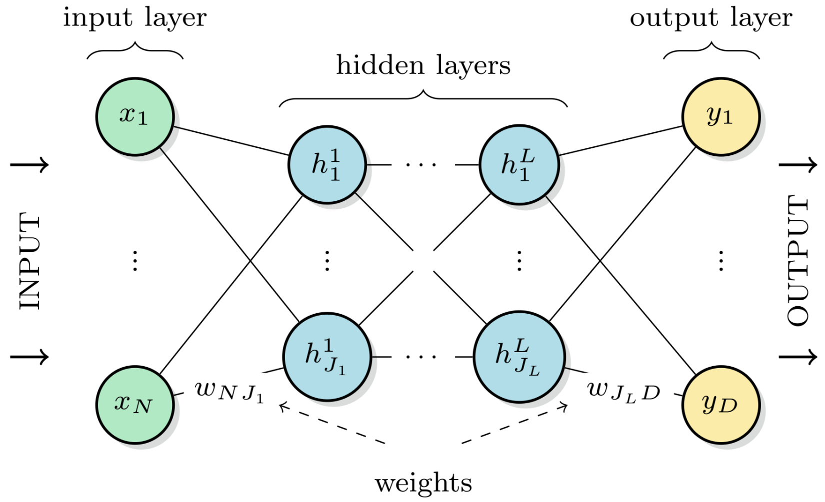
Feed-forward neural network

Predicted label: 0/1

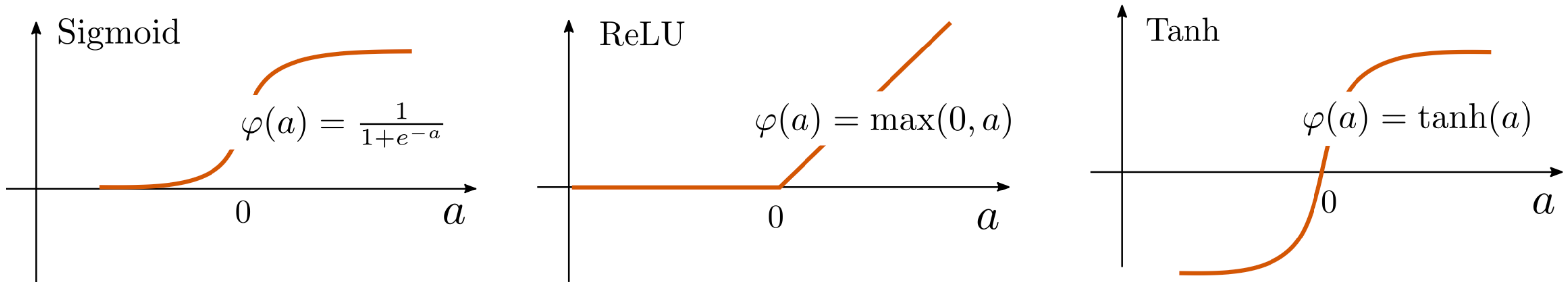
Feed-forward neural networks

- Parametrized hypothesis class
- Layers of linear transformations

$$\boxed{\boldsymbol{h}^i = \boldsymbol{W}^i \boldsymbol{h}^{i-1} + \boldsymbol{b}^i}$$



- Non-linear activation functions after each layer



Example

- Binary classification
- Samples **x**: feature vectors representing pictures
- Labels: 1/0 (dog / not dog)
- Train / test split

( ,1);( ,0);( ,1); ...

Loss function

- Quantifying empirical risk (based on training set)
- Squared error

$$L(f_\theta(x), y) = (f_\theta(x) - y)^2$$

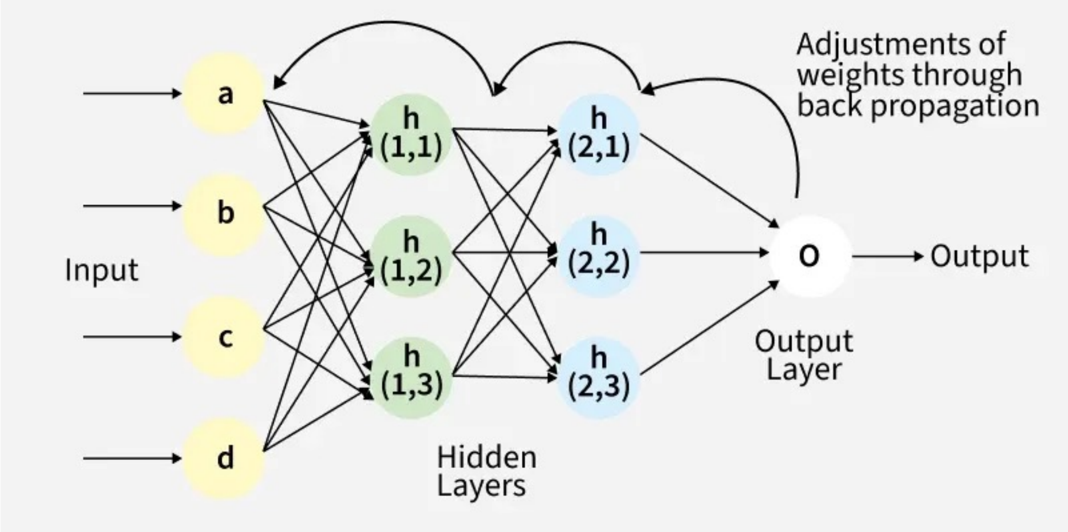
- Cross-entropy loss

$$L(f_\theta(x), y) = -\sum_{d=1}^{D} y_d \log p_d$$

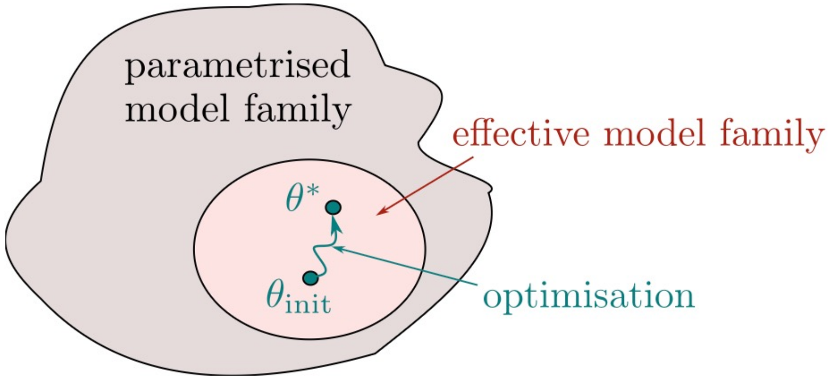
Optimization (training)

- Optimize parameters to minimize loss function

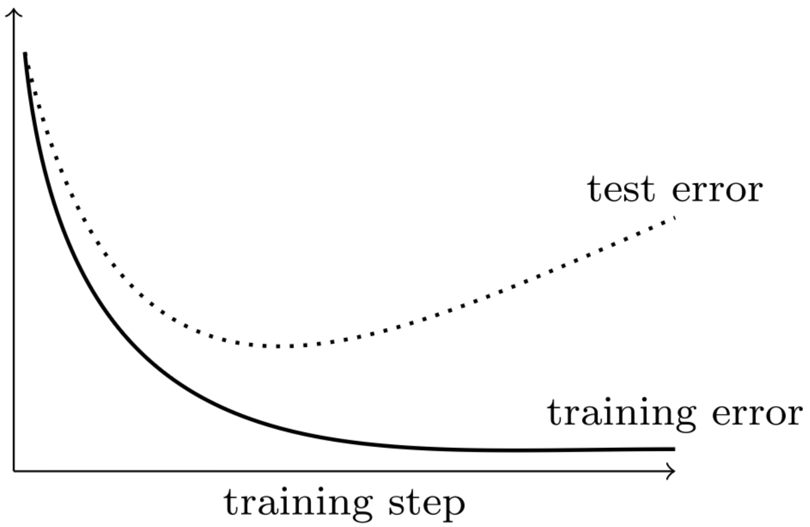
- Usually gradient descent:
  - Efficient algorithm: backpropagation

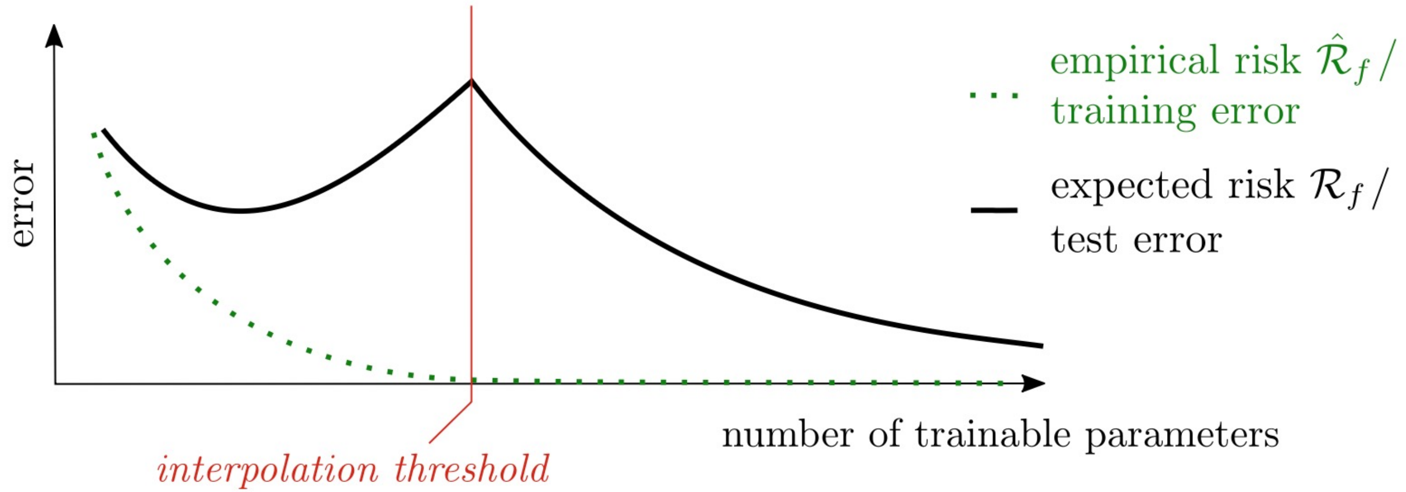


geeksforgeeks.org



Training vs test loss





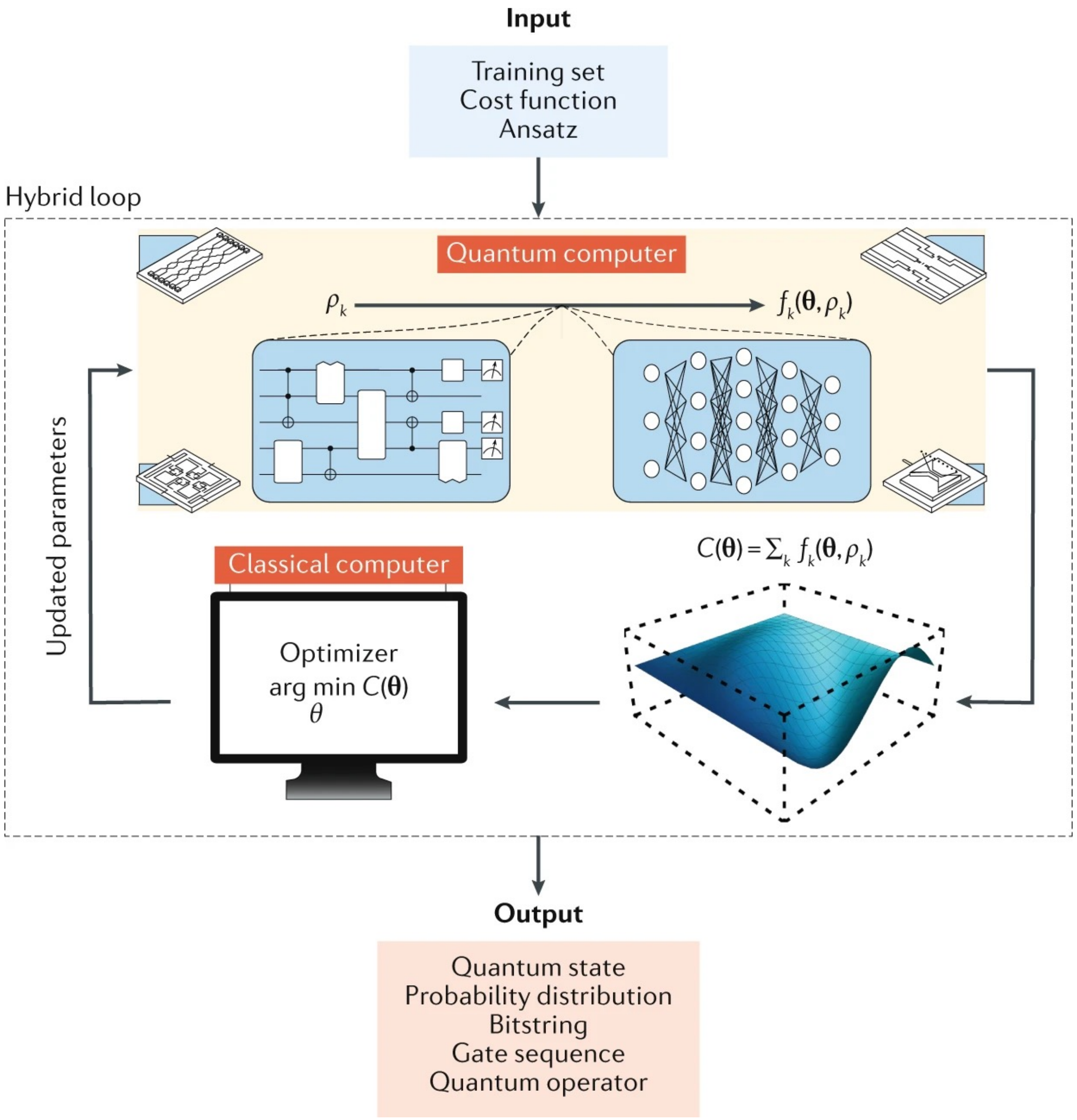
4. Variational QML

4. Variational QML

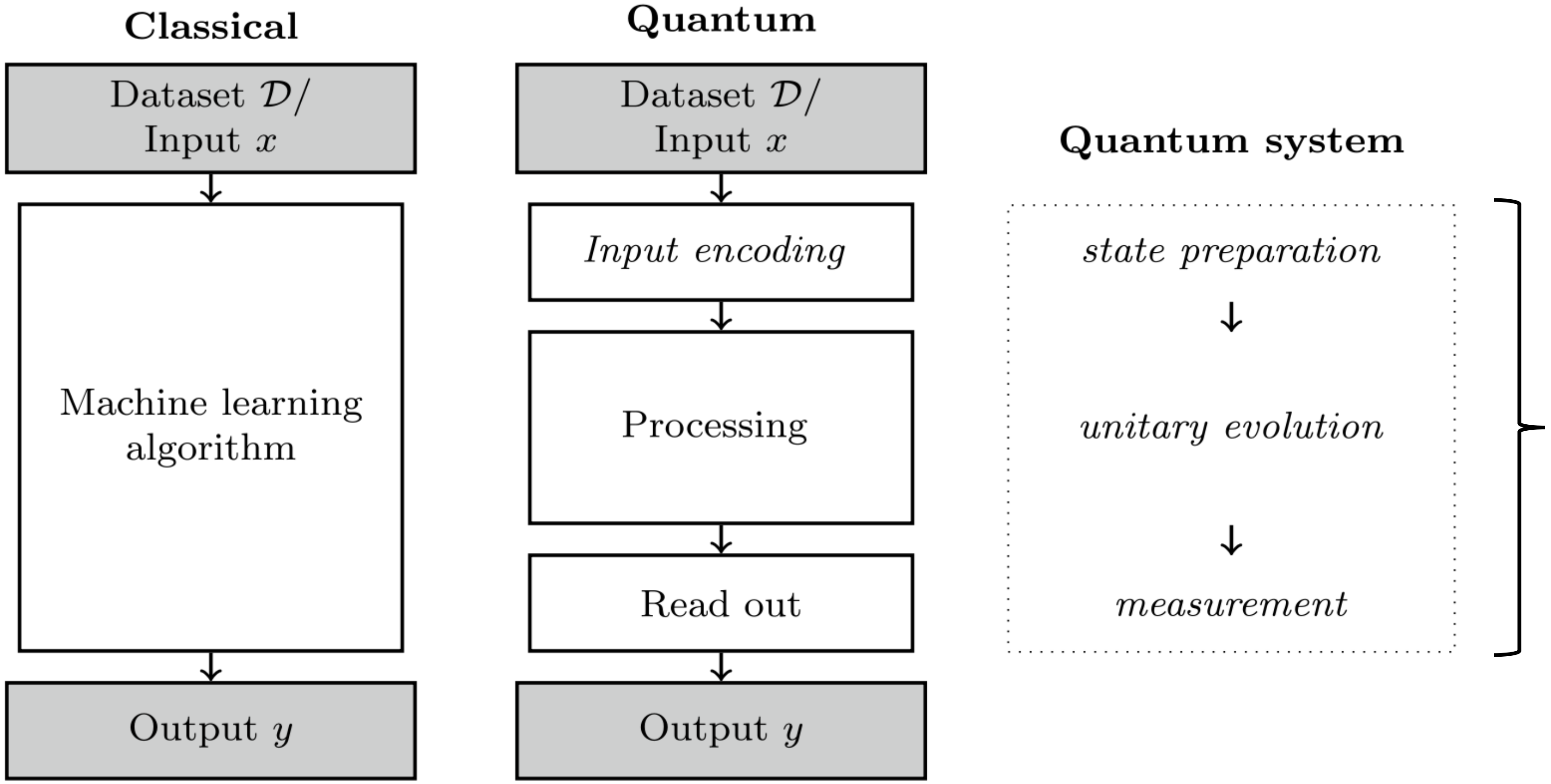
- Variational quantum circuits (VQC)
- Parametrized quantum circuits (PQC)
- Quantum neural networks (QNNs)

4. Variational QML

- Variational quantum circuits
- Parametrized quantum circuits
- Quantum neural networks

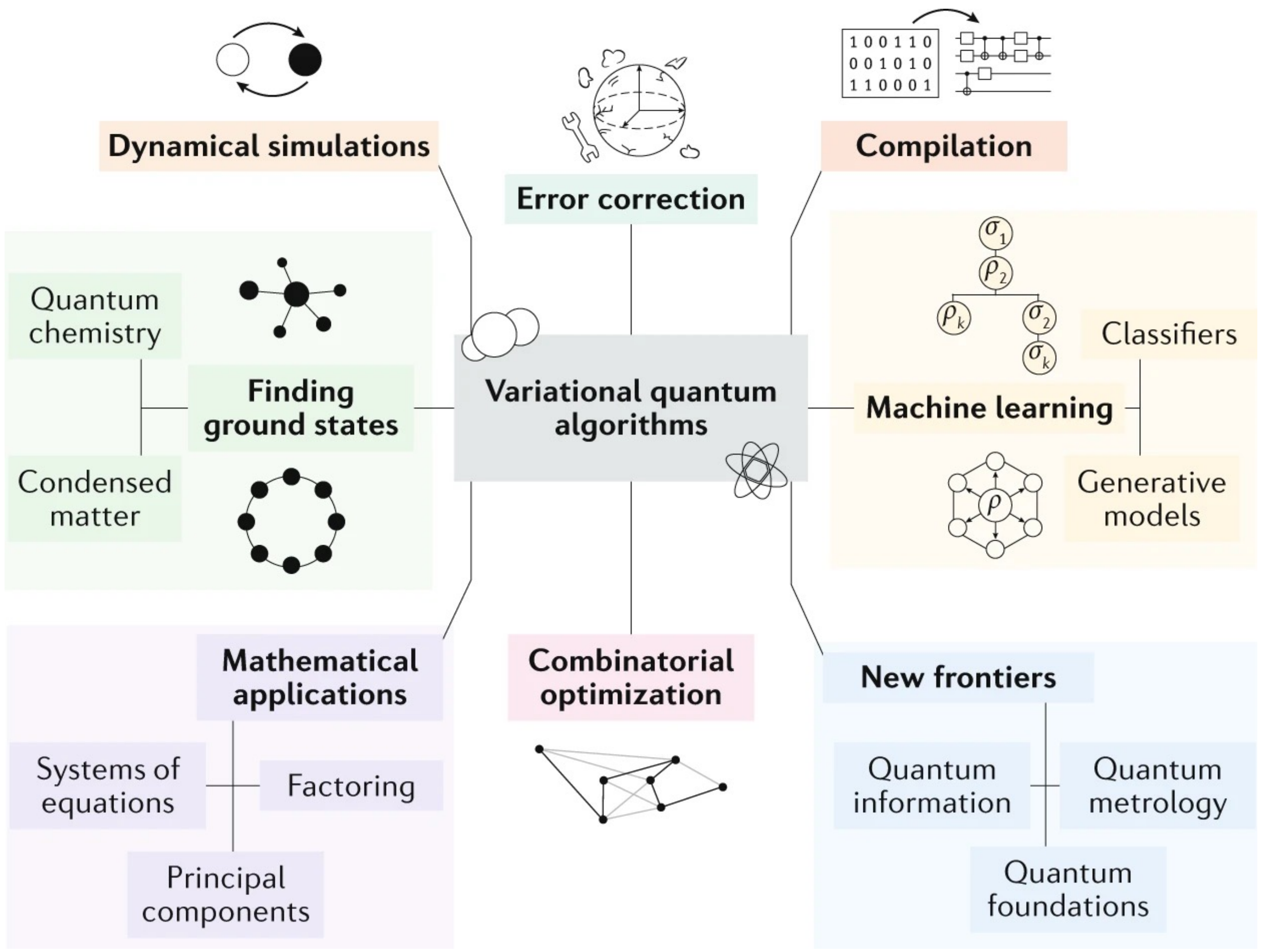


4. Variational QML



Ansatz
(parametrized hypothesis class)

Context: Variational Quantum Algorithms



M. Cerezo et al. (2021), Nat. Rev. Phys., 3(9), 625-644

Building blocks

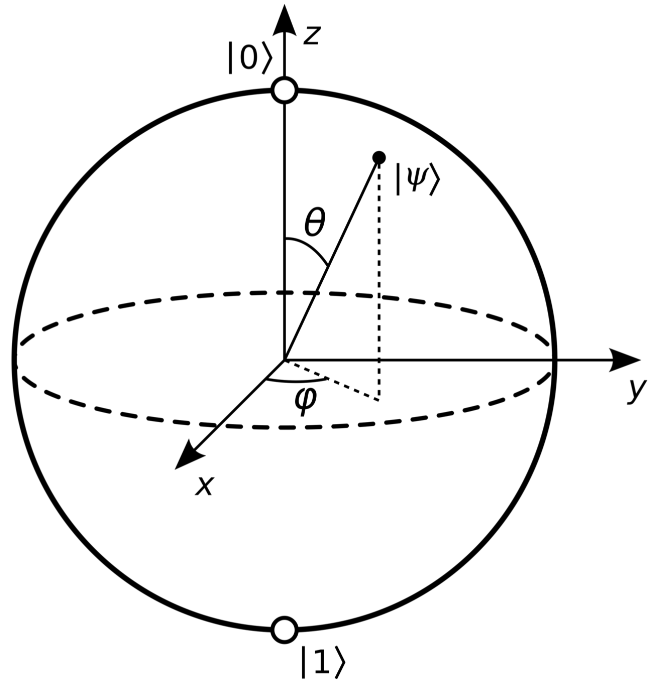
- Parametrized quantum logic gates

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

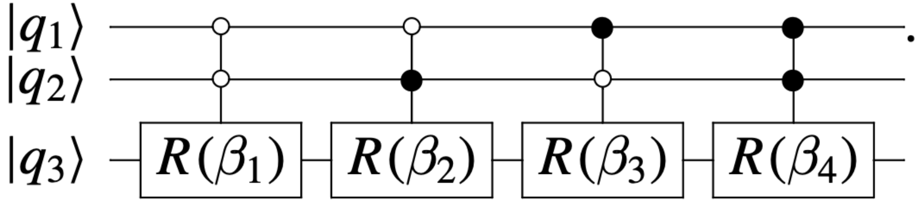
$$R_y(\theta) = e^{-i\frac{\theta}{2}\sigma_y} = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right) \\ \sin\left(\frac{\theta}{2}\right) & \cos\left(\frac{\theta}{2}\right) \end{pmatrix}$$

$$R_z(\theta) = e^{-i\frac{\theta}{2}\sigma_z} = \begin{pmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{pmatrix}.$$

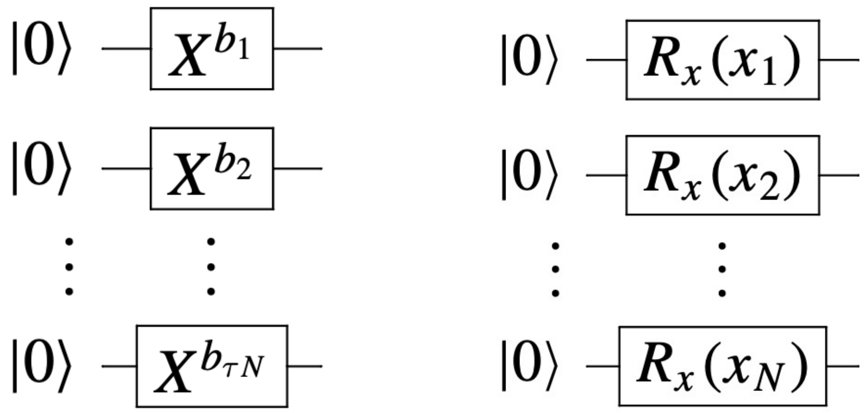
$$\mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$



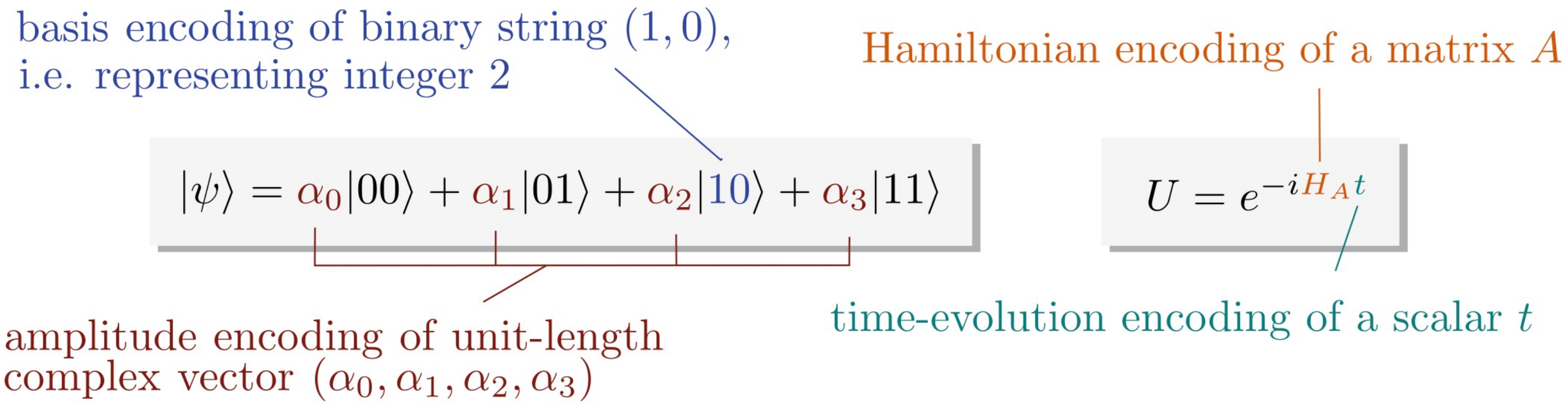
Bloch sphere



Encoding data

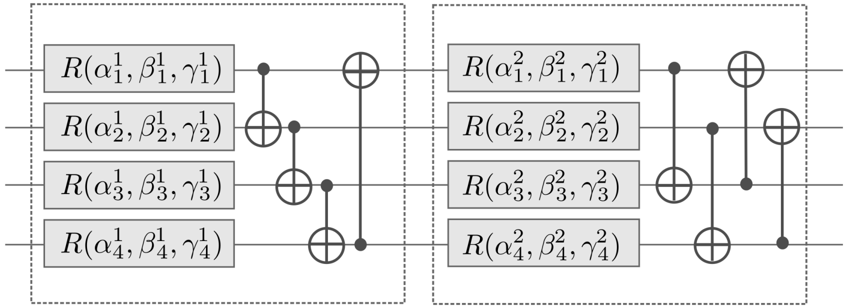


- Data embedding techniques:
  - Binary encoding for binary data
  - Amplitude encoding: normalized feature vector
  - Rotation encoding

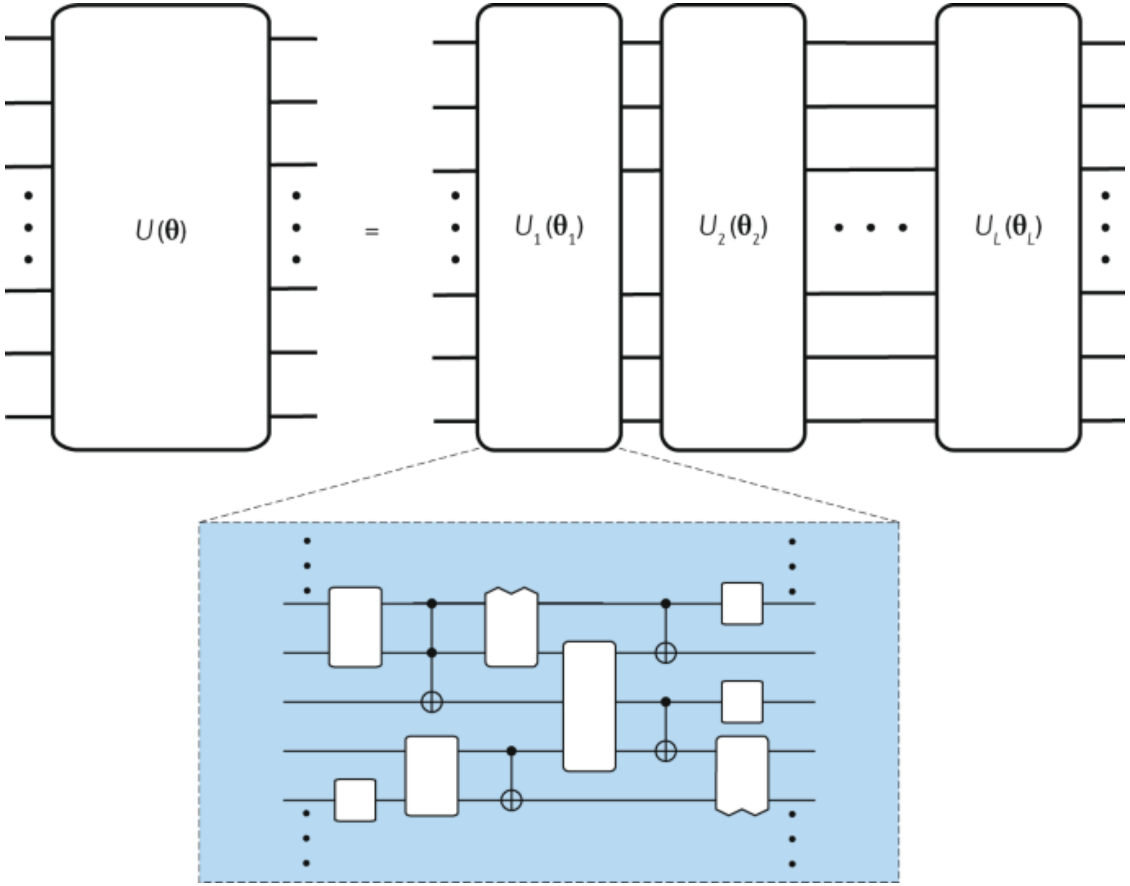


QNN layers

- Parametrized layers of operations repeated L times.
- Parametrized hypothesis class
- Different constructions based on:
  - Available hardware
  - Problem-specific information
  - …

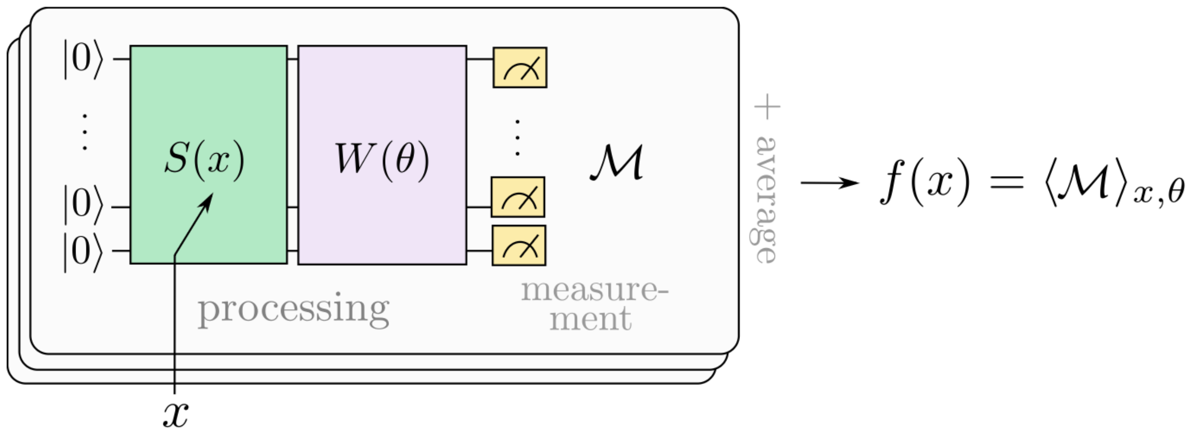


Strongly-entangling layers



M. Cerezo et al. (2021), Nat. Rev. Phys., 3(9), 625-644

Measurement

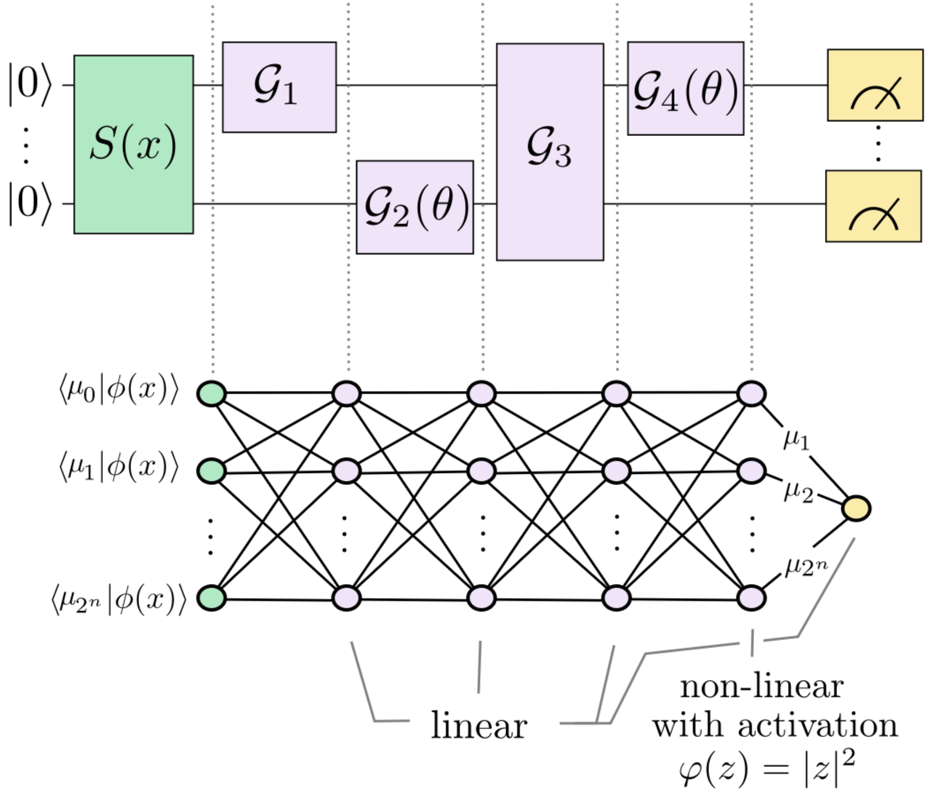


- Obtaining classical output

- To be fixed:
    - Output type: expectation value, (marginal) probability distribution, ...
    - Which qubits

- Estimating expectation values / marginals empirically

Source of nonlinearity



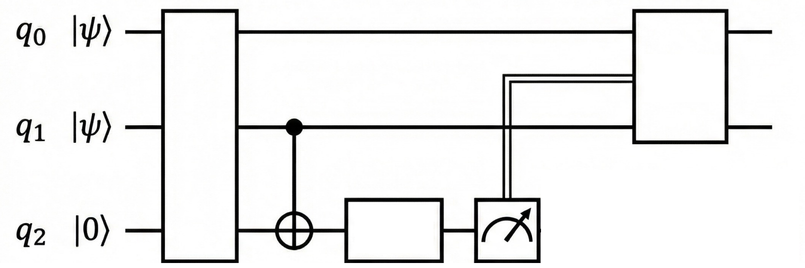
- Data encoding

$$|\psi(x)\rangle = S(x)|0\rangle$$

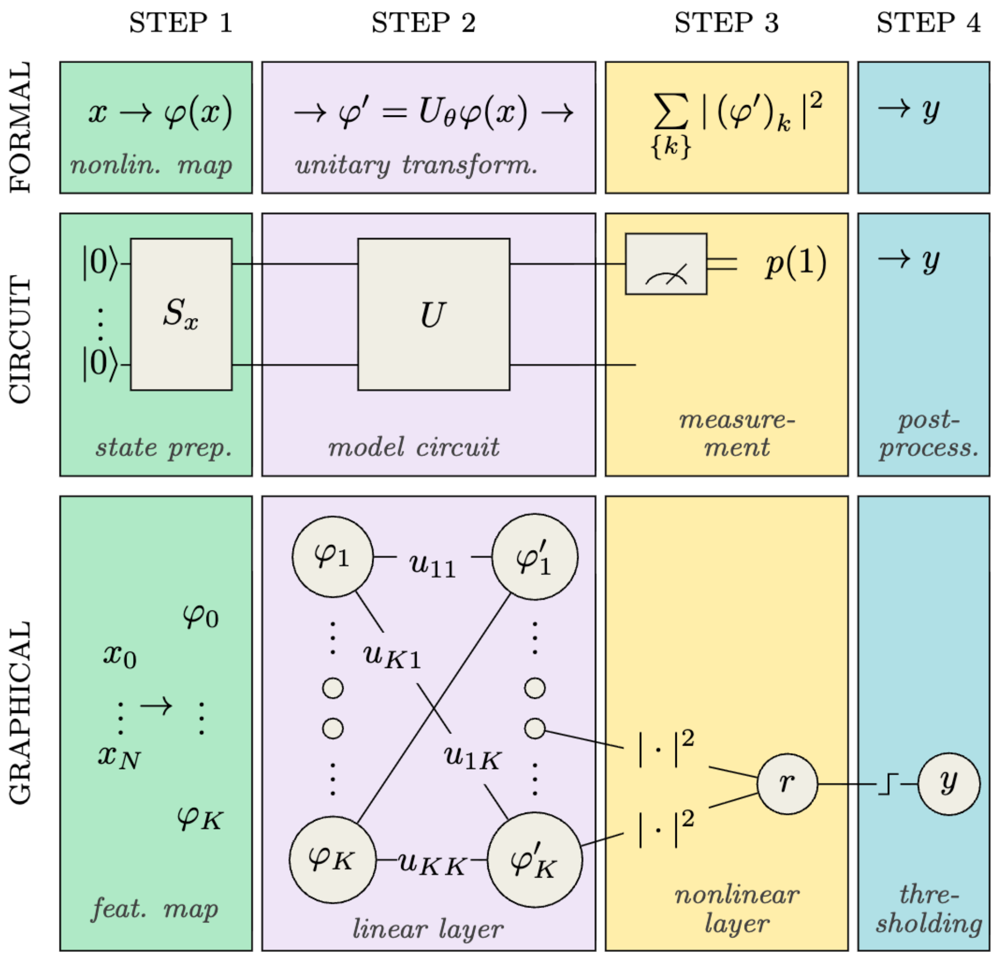
- Measurement

$$\langle\psi(x,\theta)|O|\psi(x,\theta)\rangle$$

- More advanced: measurement and feed-forward

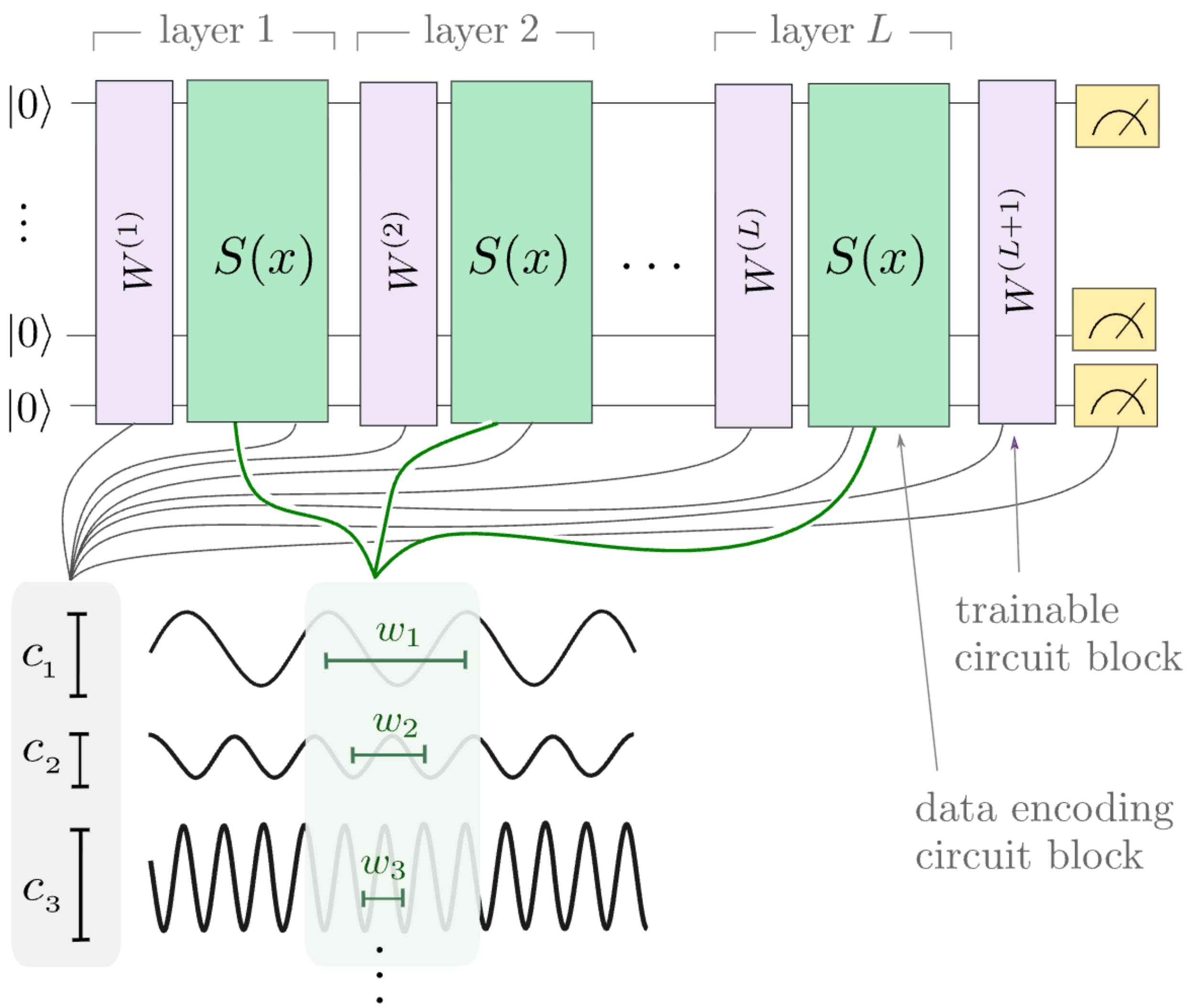


Example: circuit-centric quantum classifier



M. Schuld, et al. (2020). PRA, 101(3), 032308.

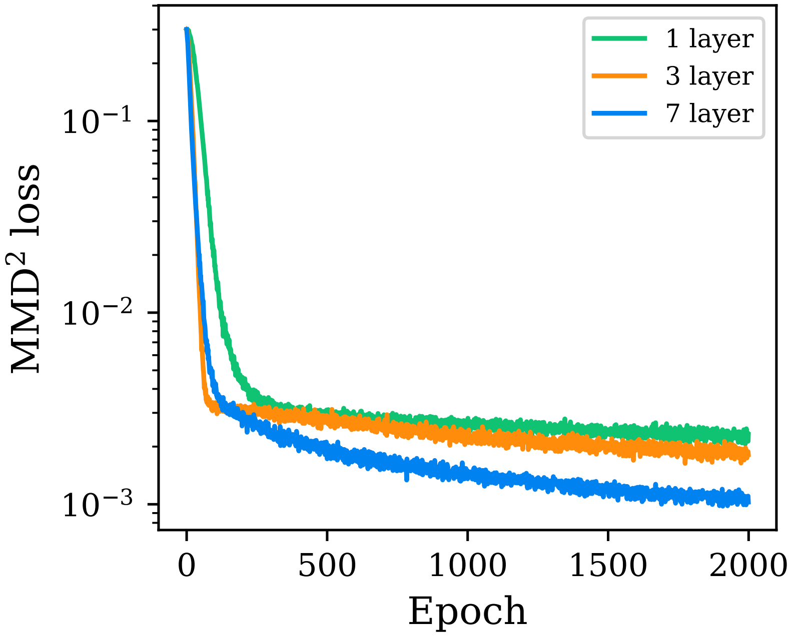
Are these models expressive?

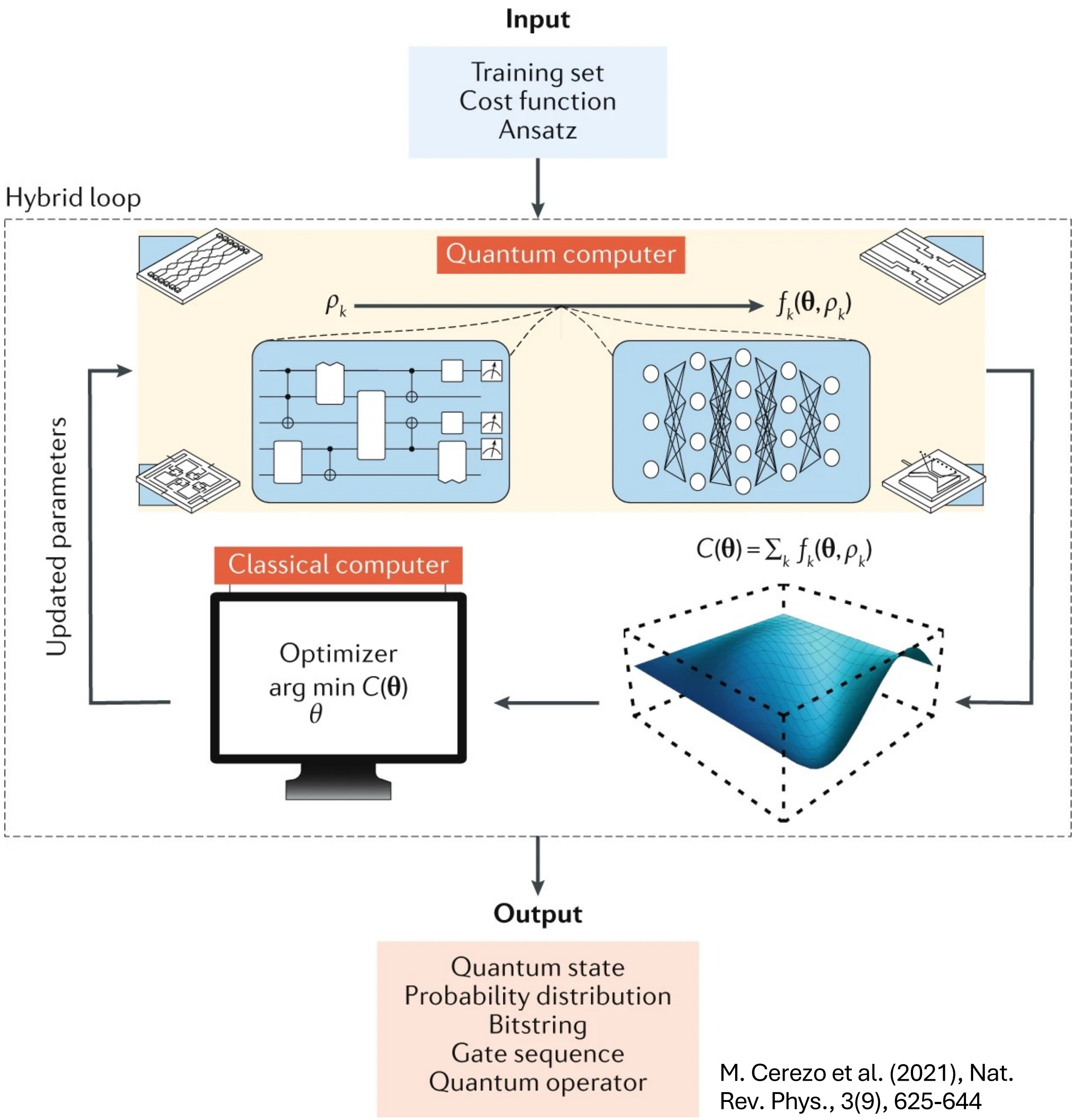


Data re-uploading

M. Schuld et al. (2021), PRA, 103(3), 032430

Hybrid training loop





M. Cerezo et al. (2021), Nat. Rev. Phys., 3(9), 625-644

Problems with this method

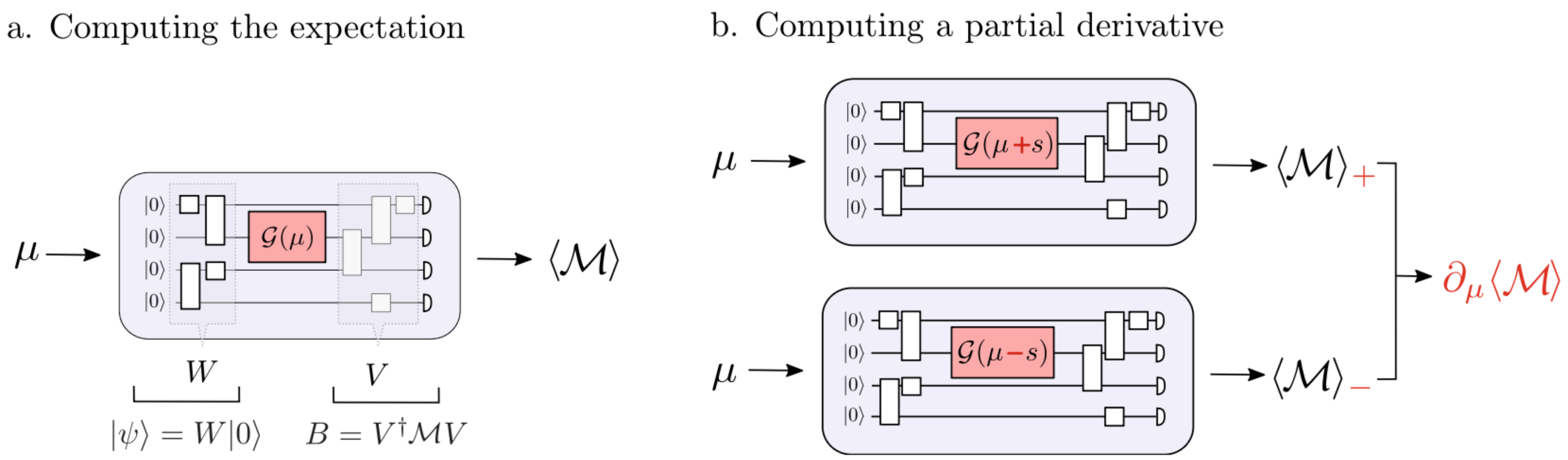
- There is no backpropagation algorithm; how to train these circuits efficiently?
- The high expressivity comes with a curse of dimensionality: vanishing gradients lead to exponential number of measurements.

Proposition 3 (Backpropagation scaling is impossible for quantum data using single copies)**.** *Given the quantum data setting where one seeks to train a variational model using copies of the unknown state ρ and the additional constraint of no quantum memory, then backpropagation scaling is not possible in the general case.*

A. Abbas, et al. (2023), NIPS '23, 1940.

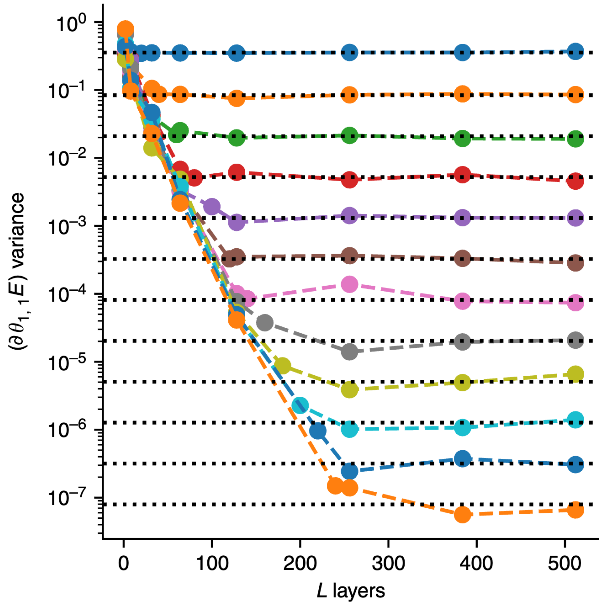
Parameter shift rules

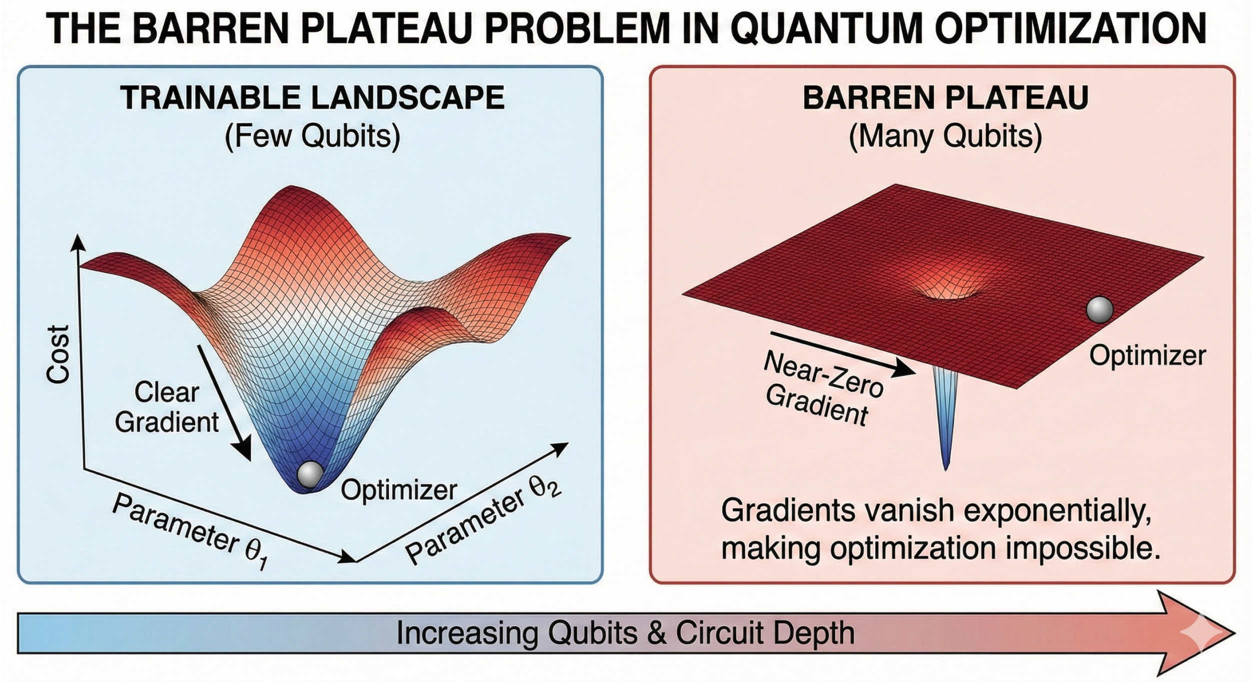
- Each partial derivative estimated separately using shifted quantum circuits.
- This cost scales linearly with the number of parameters.

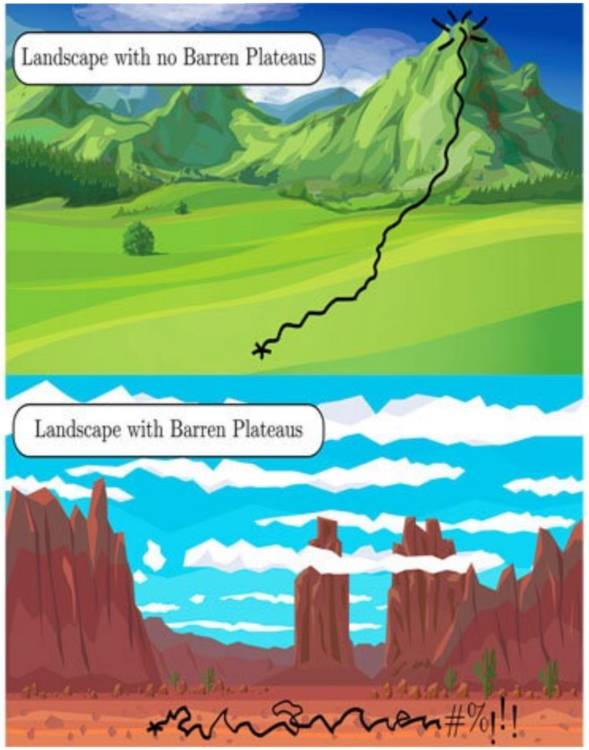


Barren plateaus

For very expressive models, the gradients vanish exponentially in the number of qubits.

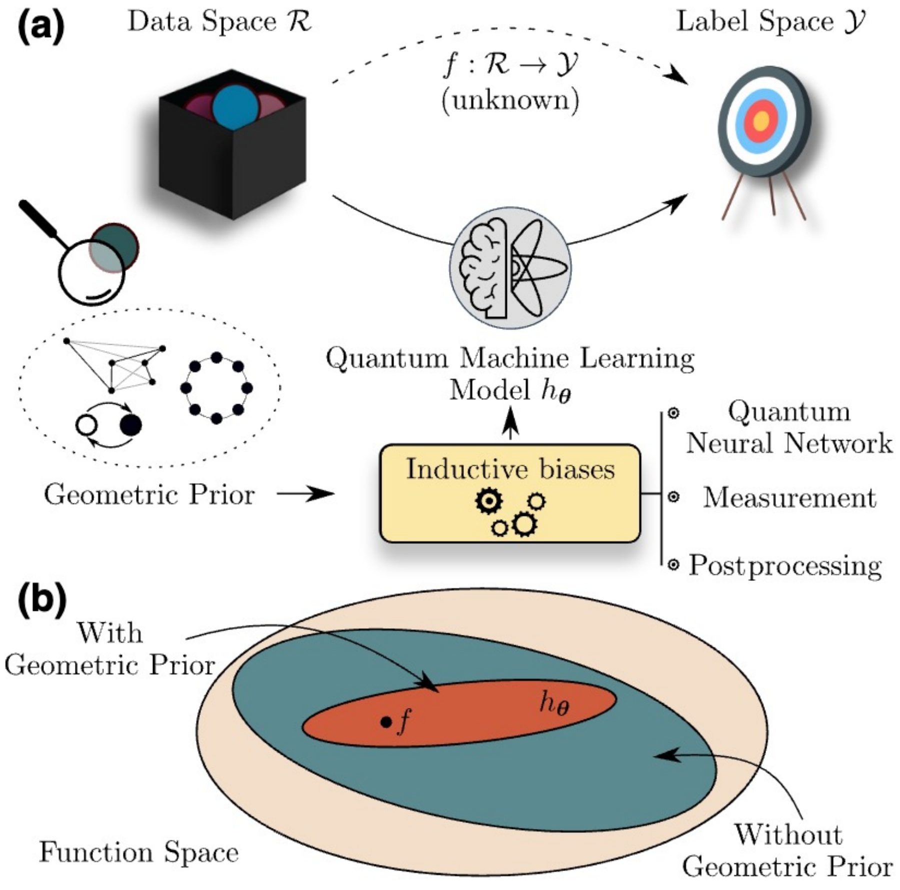




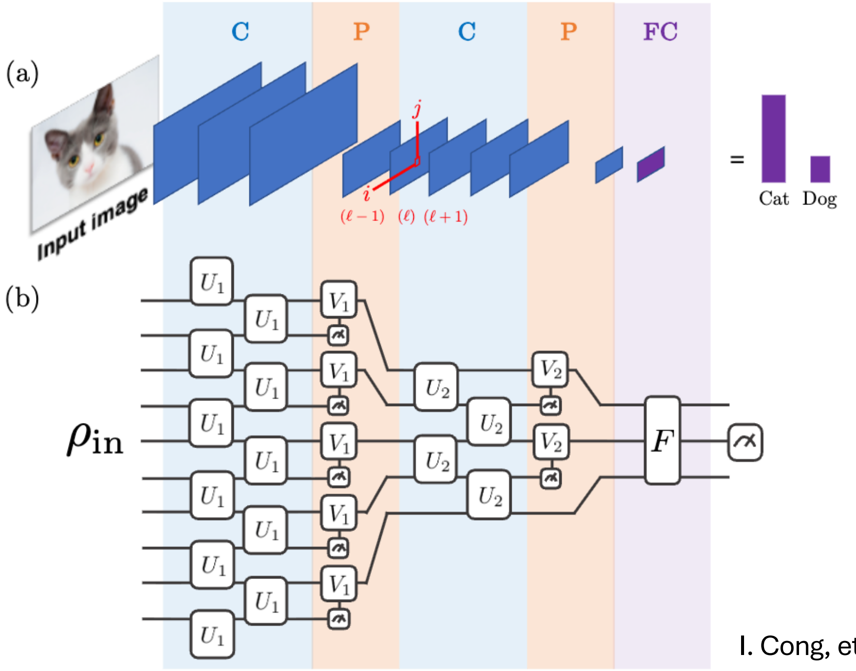


Structured models

- Provable absence of barren-plateaus for certain structured models:
  - Incorporating symmetries
  - Polynomial-size subspace (instead of exponential Hilbert space)



M. Larocca, et al. (2022), PRX Quantum, 3(3), 030341



I. Cong, et al. (2019), Nature Physics, 15(12), 1273-1278.

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!**

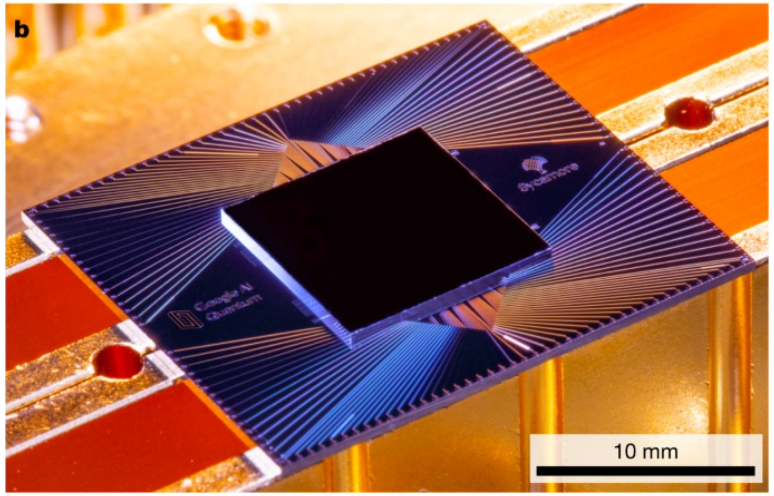
1. Efficient classical simulability of local expectation values	↔	2. Classical sampling hardness of the probability distribution

M. Cerezo, et al. (2025), Nature Communications, 16(1), 7907.

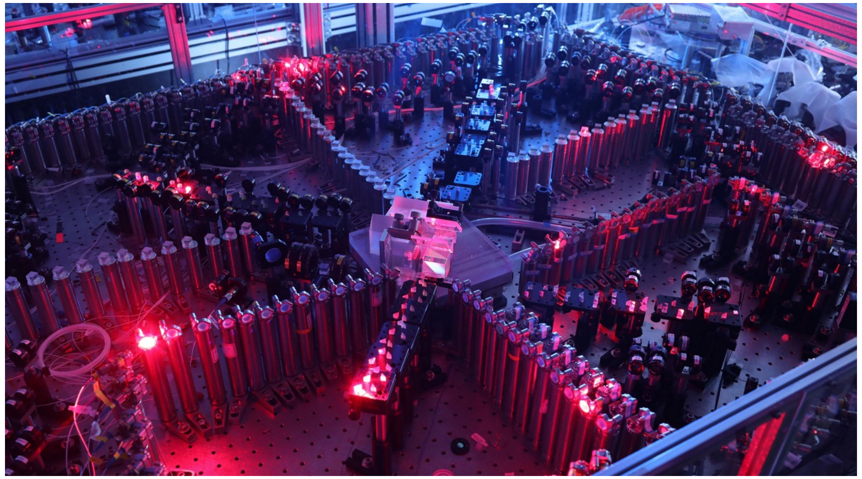
5. Quantum generative modelling

Motivation: Quantum advantage experiments

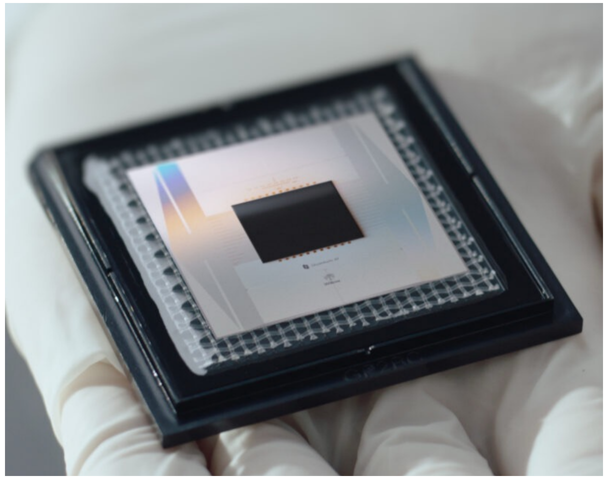
- Based on generating samples from classically hard probability distributions.



Google 2019



USTC 2019



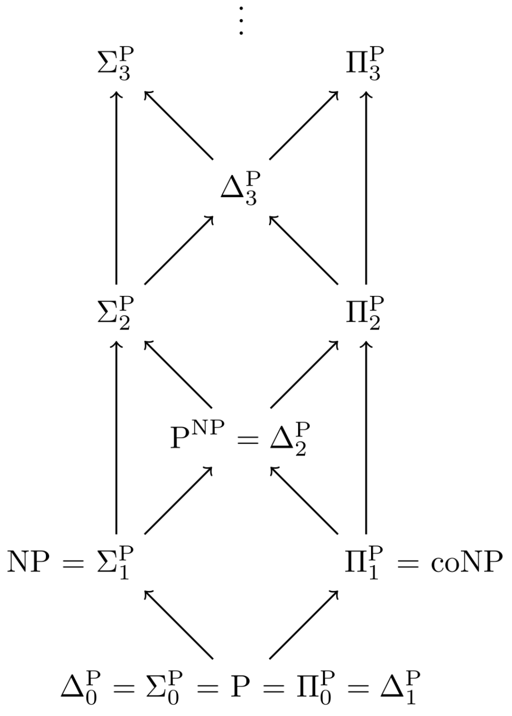
Google 2024

Motivation 2: 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)**	✓	✓	✗ ✓	–

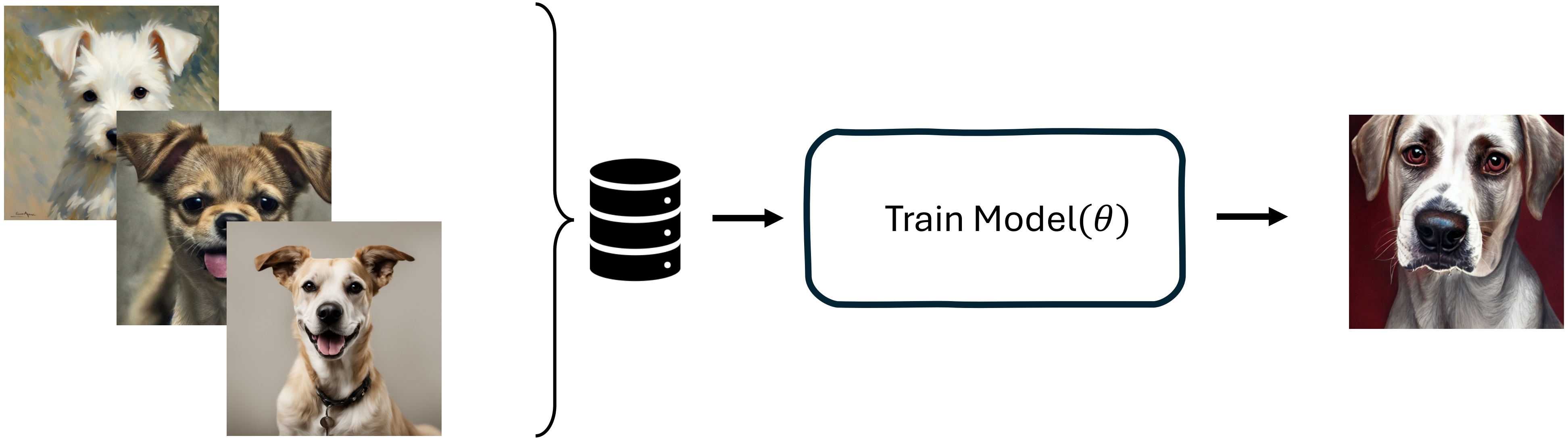
M. Oszmaniec, et al. (2022), PRX Quantum, 3(2), 020328.



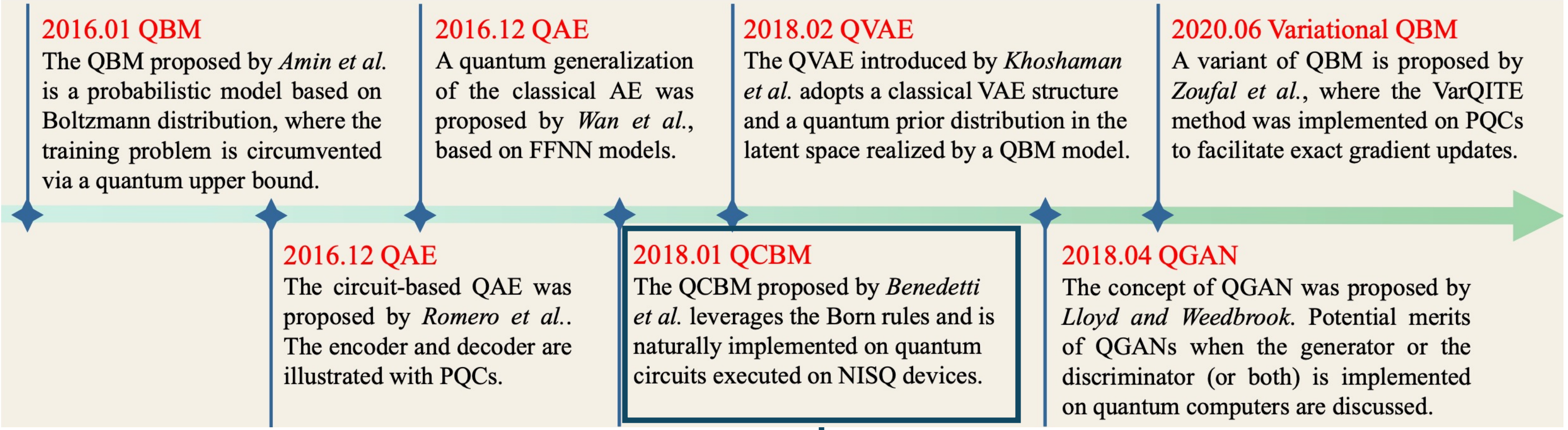
Wikipedia

Generative modeling

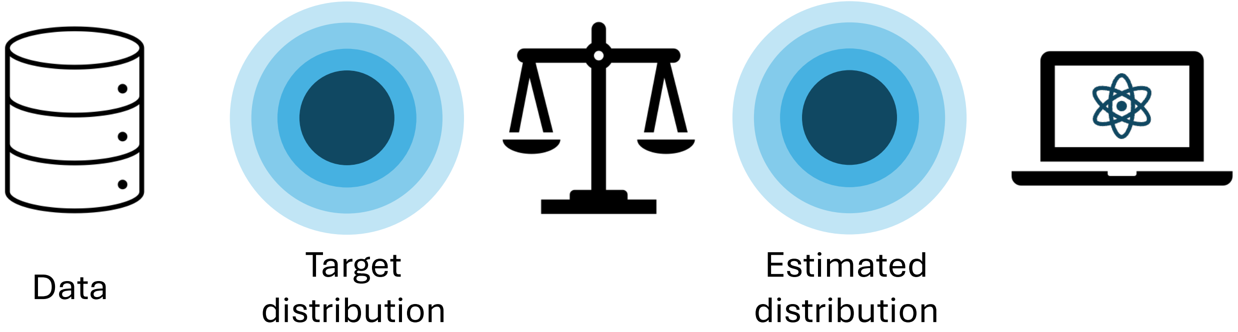
- Learn a representation of some **probability distribution** in order to **generate realistic samples**.



Quantum generative models

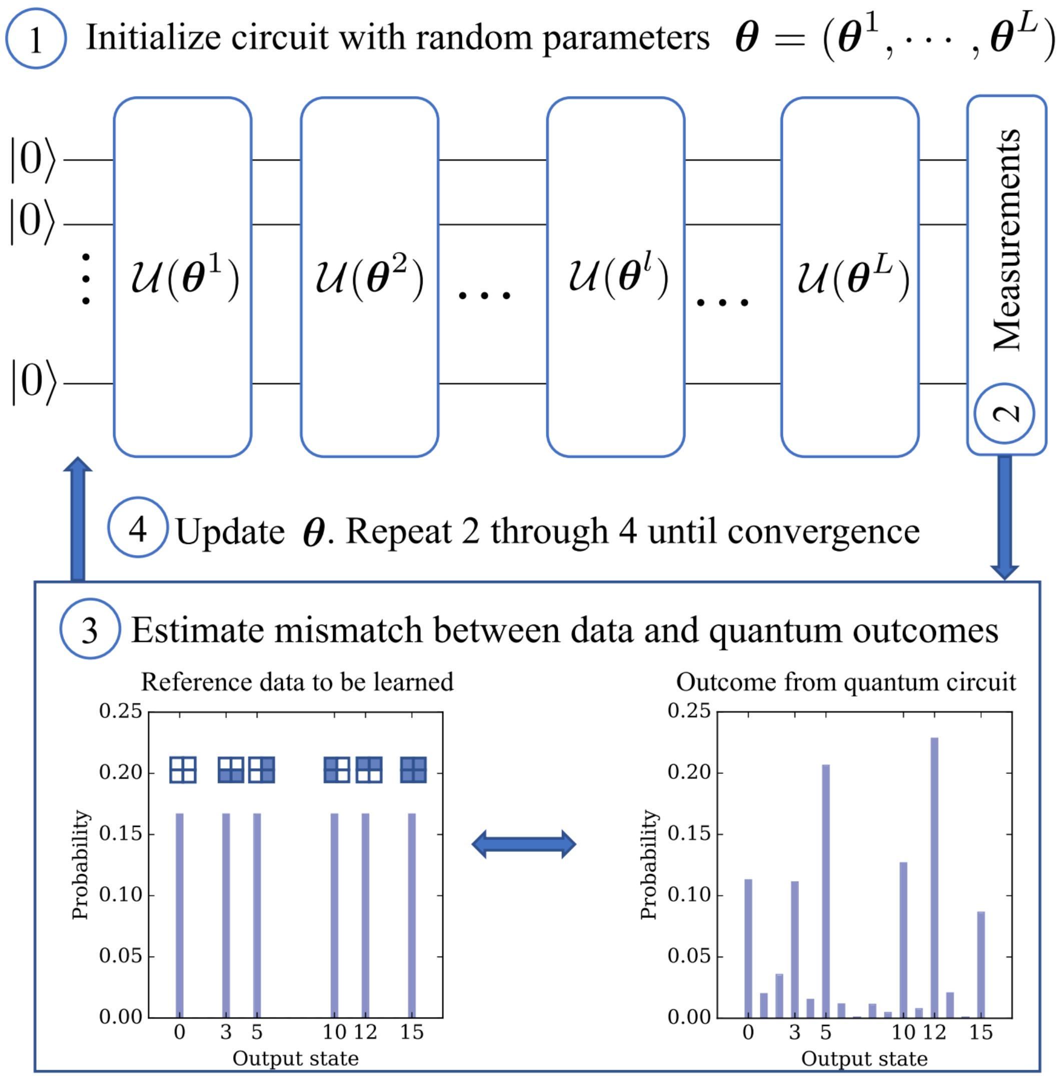


Source: J. Tian, et al., arXiv: 2206.03066, *2023.*



- Quantum counterpart of MLP (in generative setting)
- Inherits the Born rule

Quantum Circuit Born Machines (QCBMs)



Input: set of n-long bitstrings

QCBM task:

- Learn a quantum state via optimizing the parameters of a variational quantum circuit s. t.

$$|\psi\rangle = U(\boldsymbol{\theta})|\psi_0\rangle$$

$$P_{\boldsymbol{\theta}}(\boldsymbol{x}) = |\langle \boldsymbol{x}|\psi\rangle|^2$$

$$d(P_{\boldsymbol{\theta}}, P^*) \leq \varepsilon \qquad \text{where} \quad \varepsilon \in (0,1)$$

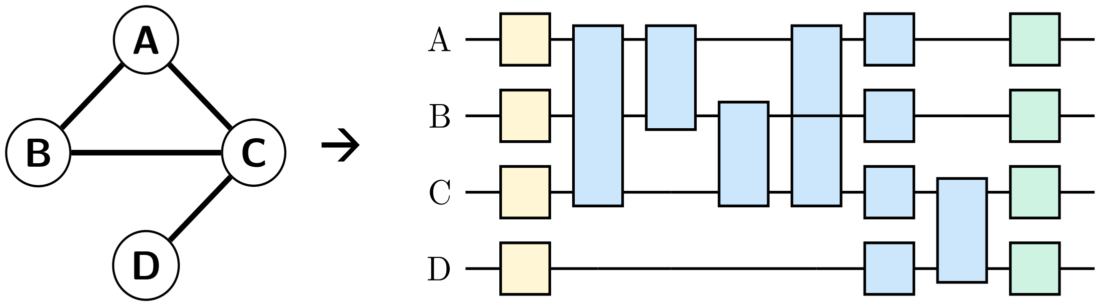
M. Benedetti, et al., npj QI, vol. 5, no. 1, p. 45, 2019.

Quantum Circuit Born Machines (QCBMs)

- Data encoding: None
- QNN layers: several possibilities
- Measurement: all qubits in the computational basis (one sample)
- Loss function:
  - Kulback-Leibler divergence
  - Maximum mean discrepancy
  - Etc.
- Training: e.g., gradient based optimizers

Example:
Quantum Circuit Markov Random Fields

Independencies: $(A \perp D \mid C)$ $(B \perp D \mid C)$



Problem-informed Ansatz

- Respects the Markov Network structure of the generative learning problem.
- Outperforms its problem-agnostic counterparts.

BB, D. Nagy, P. Hága, Z. Kallus, Z. Zimborás. Problem-informed Graphical Quantum Generative Learning. arXiv:2405.14072 (2024).

6. Recent advances in generative QML

Classical training, quantum inference

- Recent results show that specific QCBMs can be trained without using a quantum computer
- Final inference stage (sampling) still requires a quantum device (hardness guarantees)

Parametrized IQP circuits

Train on classical, deploy on quantum: scaling generative quantum machine learning to a thousand qubits

Erik Recio-Armengol*[1,2,3], Shahnawaz Ahmed[1], and Joseph Bowles†[1]

[1]Xanadu, Toronto, ON, M5G 2C8, Canada
[2]ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860, Castelldefels (Barcelona), Spain
[3]Eurecat, Centre Tecnològic de Catalunya, Multimedia Technologies, 08005 Barcelona, Spain

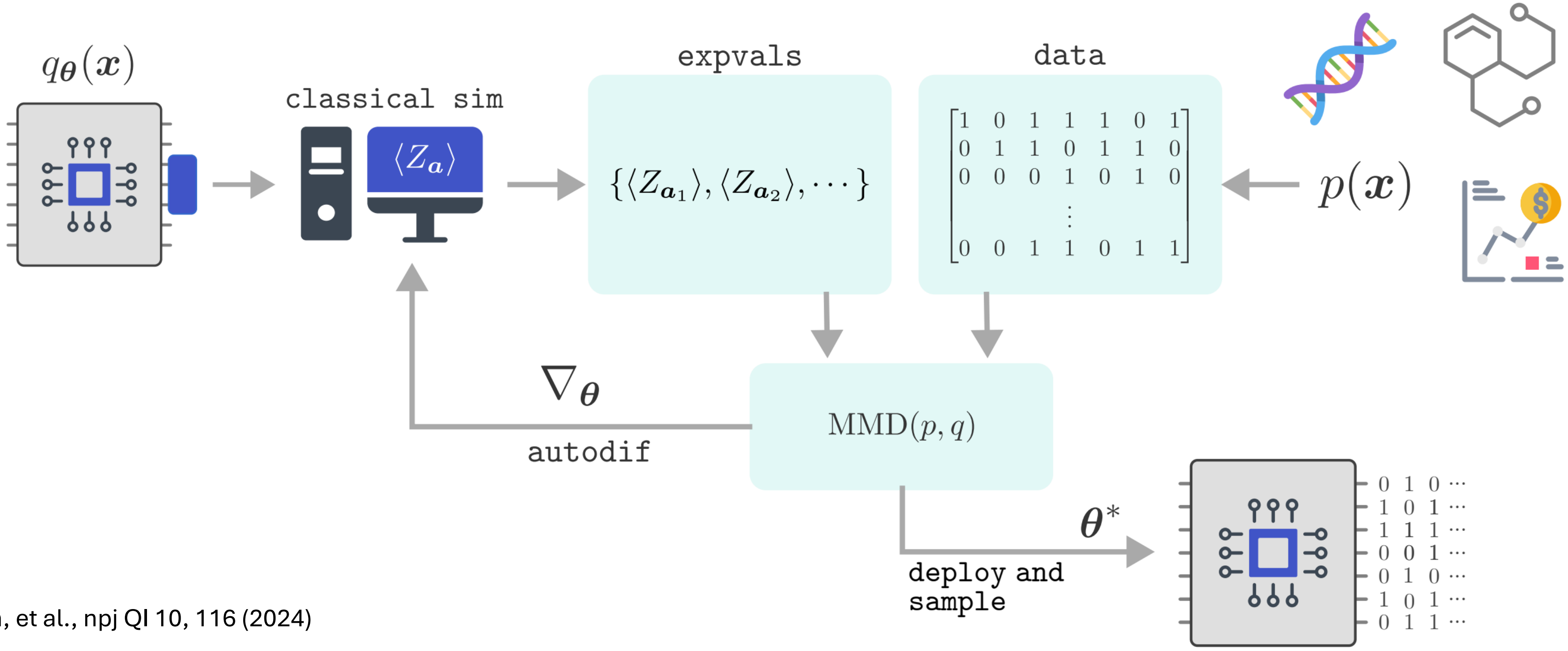
IQPopt: Fast optimization of instantaneous quantum polynomial circuits in JAX

Erik Recio-Armengol*[1,2] and Joseph Bowles†[3]

[1]ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860, Castelldefels (Barcelona), Spain
[2]Eurecat, Centre Tecnològic de Catalunya, Multimedia Technologies, 08005 Barcelona, Spain
[3]Xanadu, Toronto, ON, M5G 2C8, Canada

Classically training IQP circuits

<u>**Goal**</u>: learn joint probability distribution of binary random variables **p(x)**

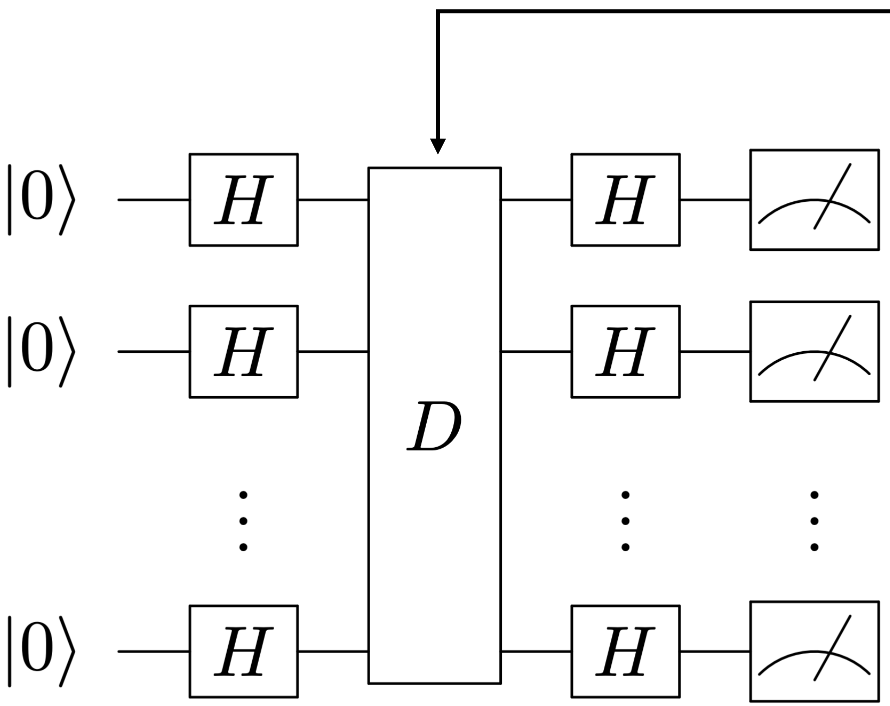


M. S. Rudolph, et al., npj QI 10, 116 (2024)

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

IQP circuits

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



- Post-IQP = PP
- Efficient sampling implies collapse of PH to the third level

Loss function formulation

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

model distribution

target distribution

$$\mathcal{L}_{MMD^2} = \sum_A C_\sigma(|A|)(\langle Z\rangle_A^p - \langle Z\rangle_A^q)^2$$

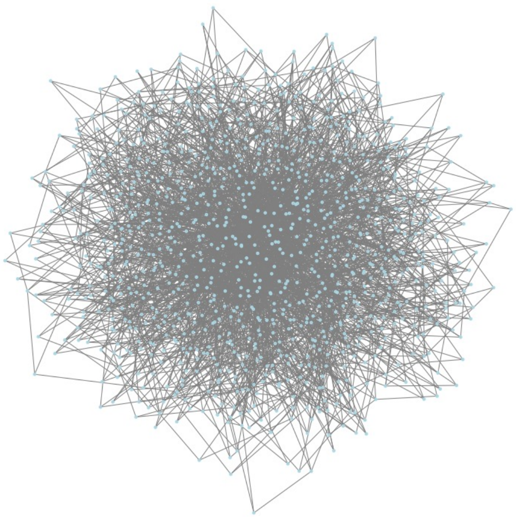
sets of qubit indices

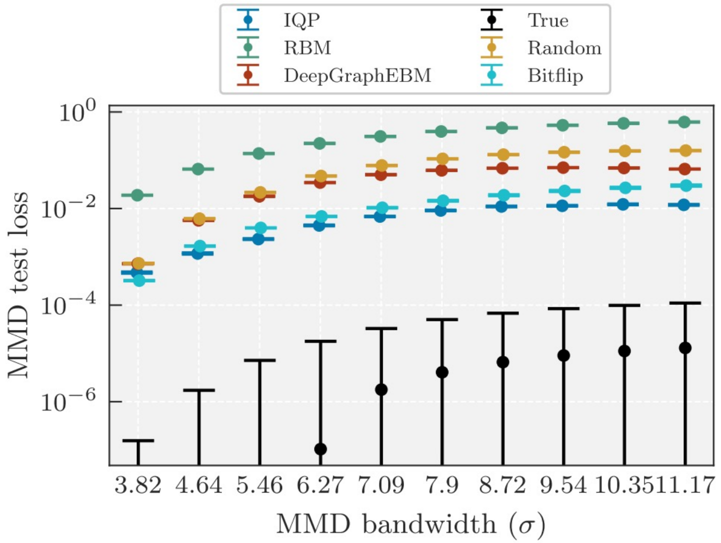
Recently implemented for **IQP circuits**.

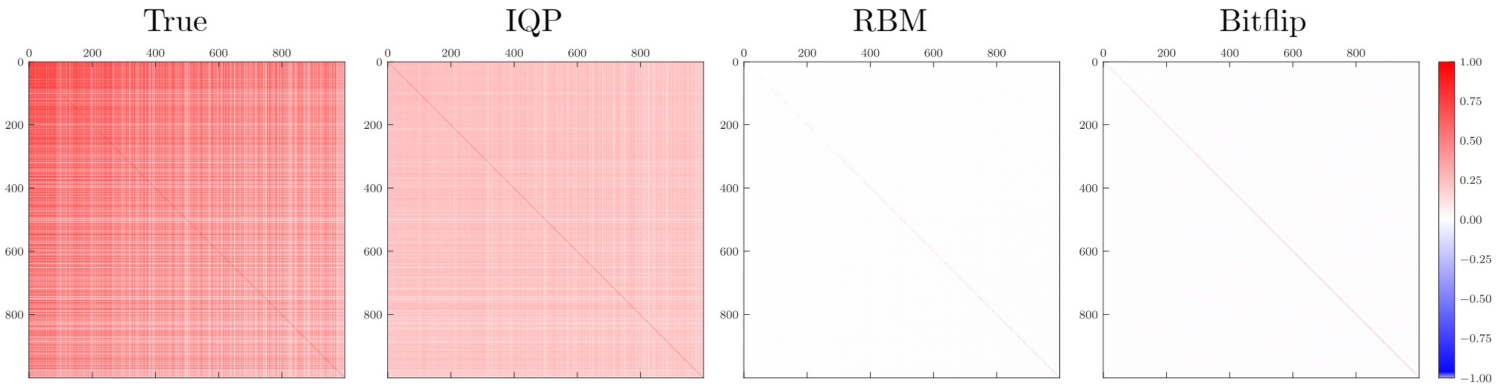
M. S. Rudolph, S. Lerch, S. Thanasilp, O. Kiss, S. Vallecorsa, M. Grossi, Z. Holmes. Trainability barriers and opportunities in quantum generative modeling. npj QI 10, 116 (2024)

E. Recio-Armengol, S. Ahmed, and J. Bowles. Training large-scale generative quantum machine learning models on classical hardware. arXiv:2503.02934 (2025).

Demonstration: Scale-free network







- 1000-node scale-free network (Barabasi-Albert)
- Ising model, with degree-biased local fields
- Dataset sampled using MCMC
- **1000-qubit model trained classically!**

Fermion Sampling circuits

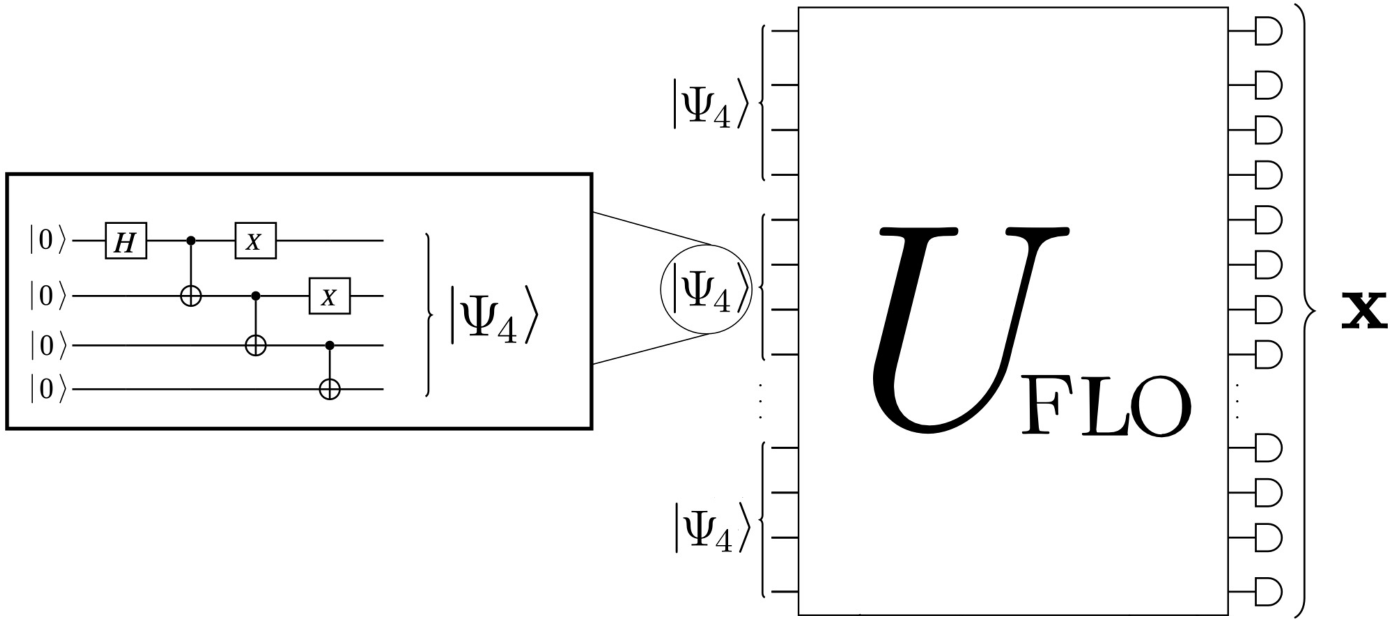
Fermionic Born Machines: Classical training of quantum generative models based on Fermion Sampling

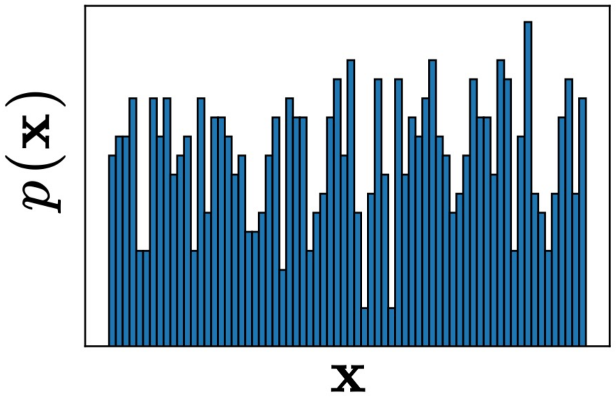
Bence Bakó[1,2], Zoltán Kolarovszki[1,2], and Zoltán Zimborás[1,3,4]

[1]HUN-REN Wigner Research Centre for Physics, Budapest, Hungary
[2]Eötvös Loránd University, Budapest, Hungary
[3]Algorithmiq Ltd, Kanavakatu 3C 00160 Helsinki, Finland
[4]University of Helsinki, Yliopistonkatu 4 00100 Helsinki, Finland

Fermion Sampling circuits

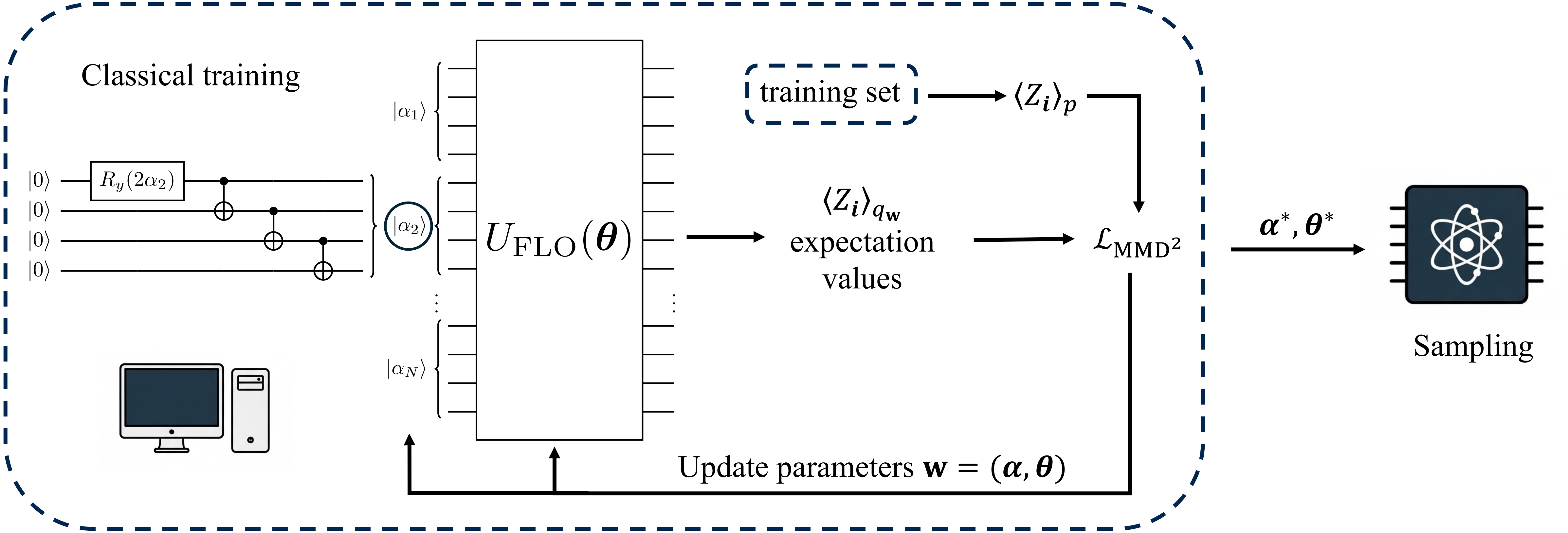
magic state input + FLO circuit = sampling advantage



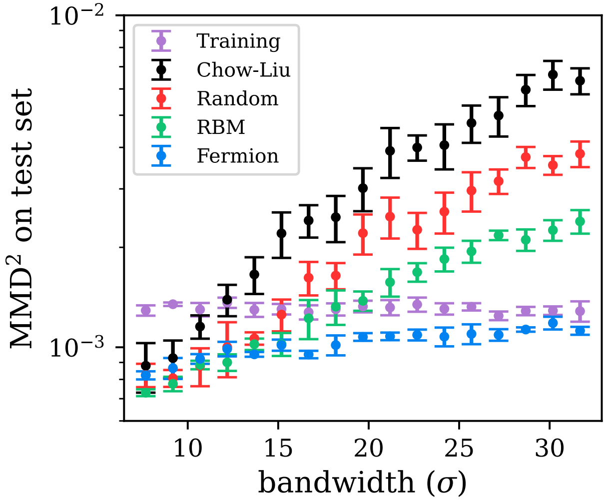


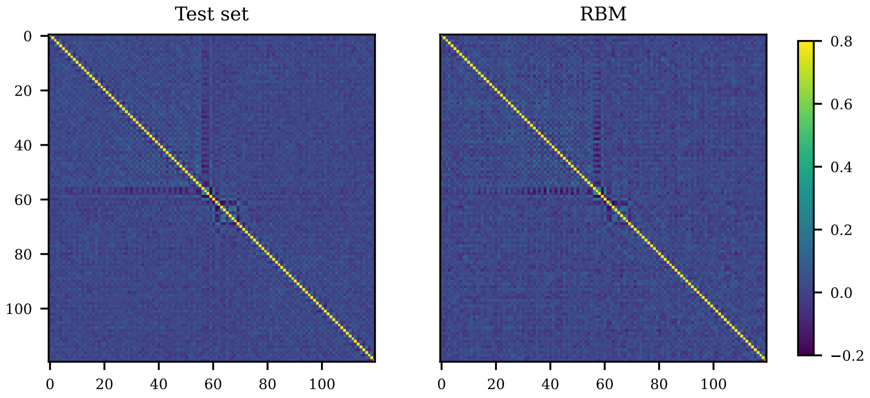
M. Oszmaniec, N. Dangniam, M. E. Morales, Z. Zimborás. Fermion sampling: a robust quantum computational advantage scheme using fermionic linear optics and magic input states. *PRX Quantum*, *3*(2), 020328. (2022)

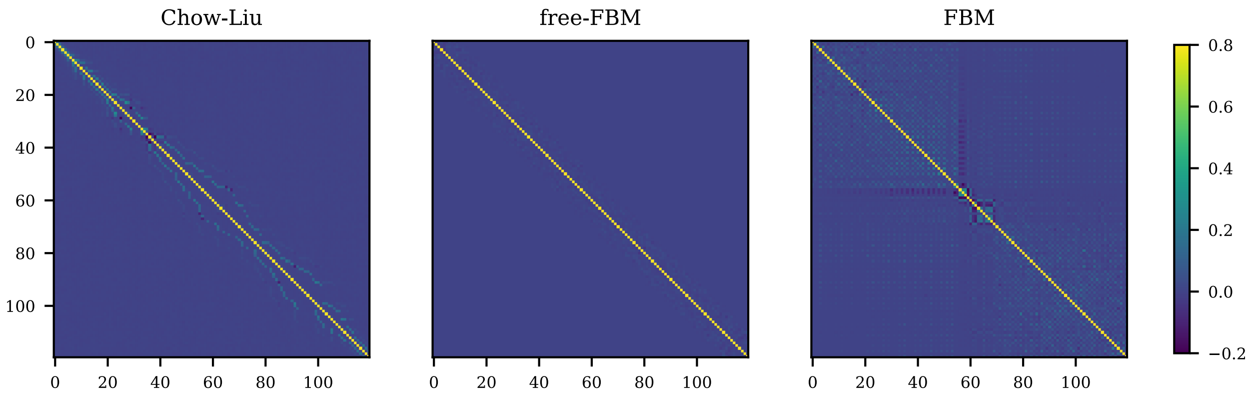
Fermionic Born Machines



FBM experiments (gene sequences)







160-qubit model trained classically!

Google experiment

Generative quantum advantage for classical and quantum problems

Hsin-Yuan Huang,[1,2,*] Michael Broughton,[1] Norhan Eassa,[1,3]
Hartmut Neven,[1] Ryan Babbush,[1] and Jarrod R. McClean[1,†]

[1] *Google Quantum AI, 340 Main Street, Venice, CA 90291, USA*
[2] *Department of Physics, California Institute of Technology, Pasadena 91125, CA, USA*
[3] *Department of Physics and Astronomy, Purdue University, West Lafayette, IN 47906, USA*

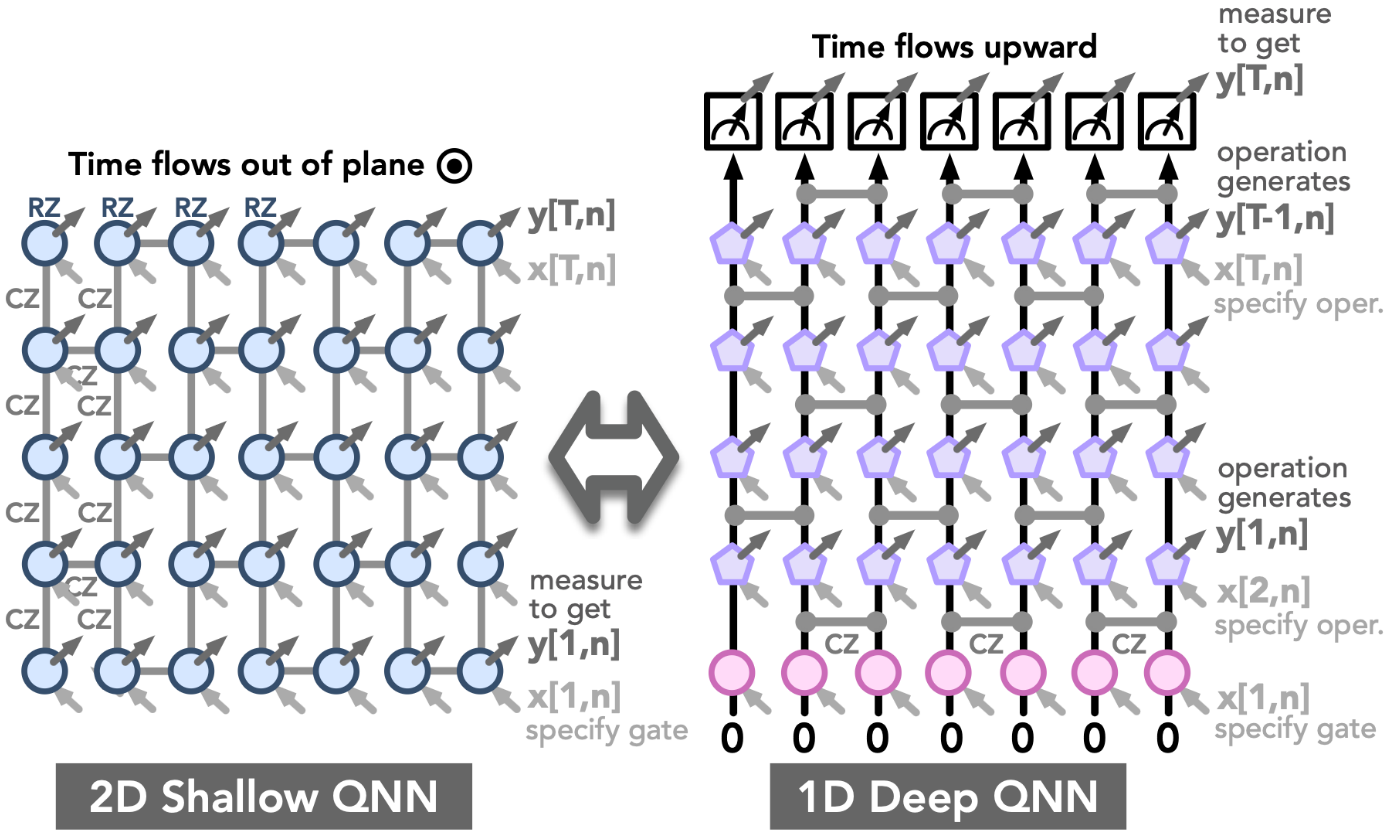
(Dated: September 12, 2025)

Learning tasks

1. **Generating bitstrings**
   - Input: pairs of bitstrings $\{\boldsymbol{x_i}, \boldsymbol{y_i}\}$
   - Goal: learn the conditional probability distribution $\boldsymbol{p}(\boldsymbol{y}|\boldsymbol{x})$
2. Generating compressed simulation circuits
   - Accessing the regime where classical compilers fail
3. Generating quantum states
   - Learning a unitary from local measurement data

Instantaneously deep QNNs (IDQNNs)

- Mapping between 2D shallow and 1D deep circuits:
  - Shallow circuits enable efficient training using **lightcone methods and local sampling** (polynomial overhead in the number of qubits)
  - Deep circuits enable experimental realisation of sampling



Shallow circuit architecture

- Given **n qubits** and **graph G**, using the **encoding** $\mathbf{0} \rightarrow |+\rangle, \mathbf{1} \rightarrow |\mathbf{0}\rangle$
  1. Prepare input state according to bitstring x
  2. Apply $R_Z(\theta_i)$ gates on each qubit
  3. Apply CZ gates according to G

- Varying the connectivity graph, these shallow circuits can implement deep QNNs

Learning-inference pipeline

1. Given a dataset with local decupling property
2. Take deep QNN, map it to its shallow counterpart by introducing additional qubits
3. Emulate training of model parameters

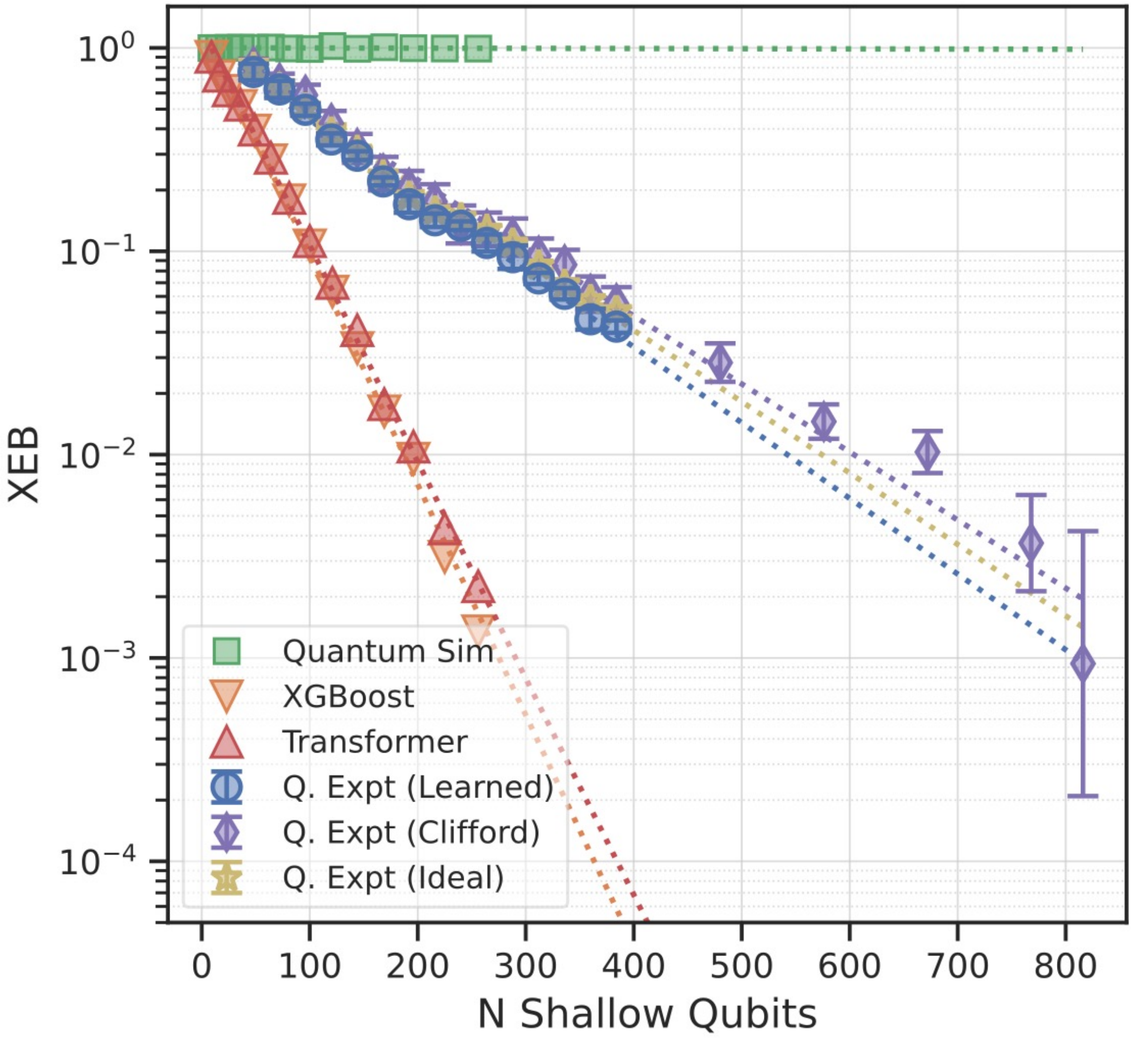
$$\hat{\beta}^j = \frac{1}{2} \arccos \left(1 - 2 \frac{1}{N_{\mathrm{sp}}} \sum_{t=1}^{N_{\mathrm{sp}}} y^j_{(t)} \right)$$

4. Reverse step 1 and perform inference using deep QNN

Experimental demonstration

- IDQNN in (2+1)D over up to 68 qubits
- Depth is constant t=12
- Mapped to 816 shallow qubits (depth 12)
- Training set: sampled from ideal IDQNN

$$\mathcal{F}_{\mathrm{XEB}}(E) = \frac{\langle p(s)\rangle_{s\in E} - \langle p(s)\rangle_{s\in\mathcal{U}}}{\langle p(s)\rangle_{s\in p} - \langle p(s)\rangle_{s\in\mathcal{U}}}.$$



Take home: potential and limitations

QML models have high expressivity

BUT

Training is challenging

Barren plateaus
- Barren plateaus
- Poor local minima
- COSTLY GRADIENTS!

Solution

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

Tutorial links

- Quantum classifier: https://colab.research.google.com/drive/1qOCjrdYSiUR8oVHYYRD8Vjphh35i7rJM?usp=copy
- Scalable generative learning: https://colab.research.google.com/drive/1aVkw6EDviO634ZRJV_a9D4mzBpELTjhN?usp=copy