INTRODUCTION TO QUANTUM MACHINE LEARNING

Oliver Sefrin, DLR-QT (Ulm)

WAW ML 10, 23.09.2024

Content

1. Basics of Quantum Computing
   - From Qubits to Quantum Circuits
   - Technological Realizations
   - Algorithms & Applications

2. Quantum Machine Learning
   - Basic Ideas
   - Data Encoding
   - Model *Ansätze*

3. Coding Demos
   - Supervised Learning
   - Generative Learning

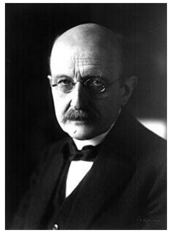
4. Wrap-Up
   - Challenges in QML
   - Conclusion & Outlook

Break of 10 mins at ~10:30

Quantum Computing Basics

DLR

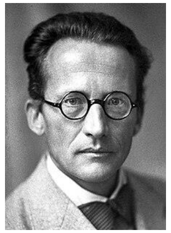
History of Quantum Physics and Quantum Computing

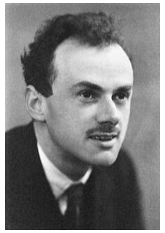



1900's:
- first ideas of quantized energy & light
- wave-particle duality

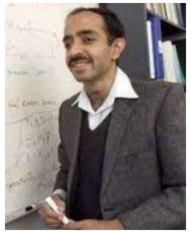

1980 & 1981:
proposal of quantum computers to simulate quantum physics

1920's:
development of quantum mechanics



1994 & 1996:
Shor algorithm (factorization), Grover algorithm (search)

Qubits & Quantum Superposition

The quantum bit or qubit has two basis states, e.g.,

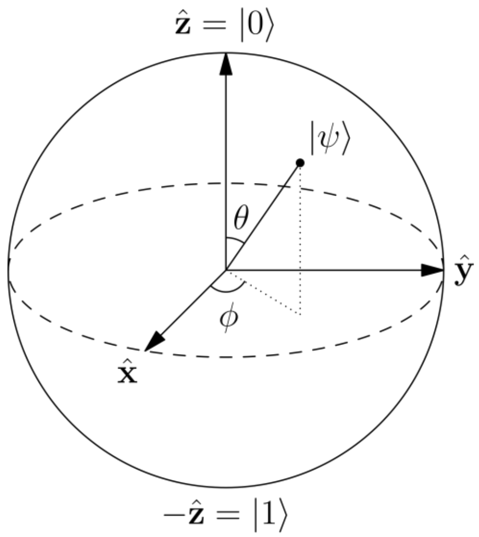
$$|0\rangle \equiv \begin{pmatrix}1\\0\end{pmatrix} \text{ and } |1\rangle \equiv \begin{pmatrix}0\\1\end{pmatrix} \in \mathbb{C}^2$$

In general, the qubit's state can be in any linear combination

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

with $\alpha, \beta \in \mathbb{C}$, $|\alpha|^2 + |\beta|^2 = 1$.

This is called superposition.



https://de.wikipedia.org/wiki/Bloch-Kugel

Only a measurement lets the wavefunction collapse to a basis state:

$$p(\text{"measure } |0\rangle\text{"}) = |\alpha|^2$$
$$p(\text{"measure } |1\rangle\text{"}) = |\beta|^2$$

Multiple Qubits

- Cf. *classical* bits:

$$n \text{ bits} \quad \rightarrow \quad 2^n \text{ combinations}$$
$$\text{e.g.}, n = 3: \quad \underbrace{000, 001, \dots, 110, 111}_{}$$

8 combinations

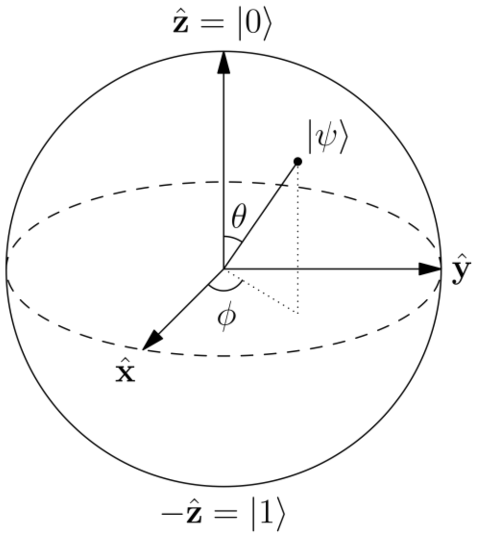
- Qubits:
  - Qubit A in state $|0\rangle$ and qubit B in state $|1\rangle$ → joint state: $|0\rangle_A |1\rangle_B \equiv |01\rangle_{AB}$
  - Qubit A in state $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and qubit B in state $\frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$
    → joint state: $\frac{1}{\sqrt{2}}(|0\rangle_A + |1\rangle_A) * \frac{1}{\sqrt{2}}(|0\rangle_B - |1\rangle_B) = \frac{1}{2}(|00\rangle - |01\rangle + |10\rangle - |11\rangle)$

→ superposition of 2^n states possible

Operations on Qubits

- Gate logic as in classical computer science
- Operations have to be unitary (i.e., norm-preserving)

- Important single-qubit gates:
  - NOT gate $|0\rangle \mapsto |1\rangle, \quad |1\rangle \mapsto |0\rangle$
  - Hadamard gate $|0\rangle \mapsto \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad |1\rangle \mapsto \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$
  - Pauli rotations



https://de.wikipedia.org/wiki/Bloch-Kugel

- Two-qubit gates:
  - Controlled-NOT gate

$$
\begin{aligned}
|0\rangle_{\text{control}}|0\rangle_{\text{target}} &\mapsto |0\rangle_{\text{control}}|0\rangle_{\text{target}} \\
|0\rangle_{\text{control}}|1\rangle_{\text{target}} &\mapsto |0\rangle_{\text{control}}|1\rangle_{\text{target}} \\
|1\rangle_{\text{control}}|0\rangle_{\text{target}} &\mapsto |1\rangle_{\text{control}}|1\rangle_{\text{target}} \\
|1\rangle_{\text{control}}|1\rangle_{\text{target}} &\mapsto |1\rangle_{\text{control}}|0\rangle_{\text{target}}
\end{aligned}
$$

Universal Gate Sets

- We can approximate any unitary operation with finite sequence of operations from universal gate set

- Ex.: {Hadamard-gate, Pauli-Rotations, CNOT-gate}

→ any *n*-qubit operation can be decomposed into one- and two-qubit gates

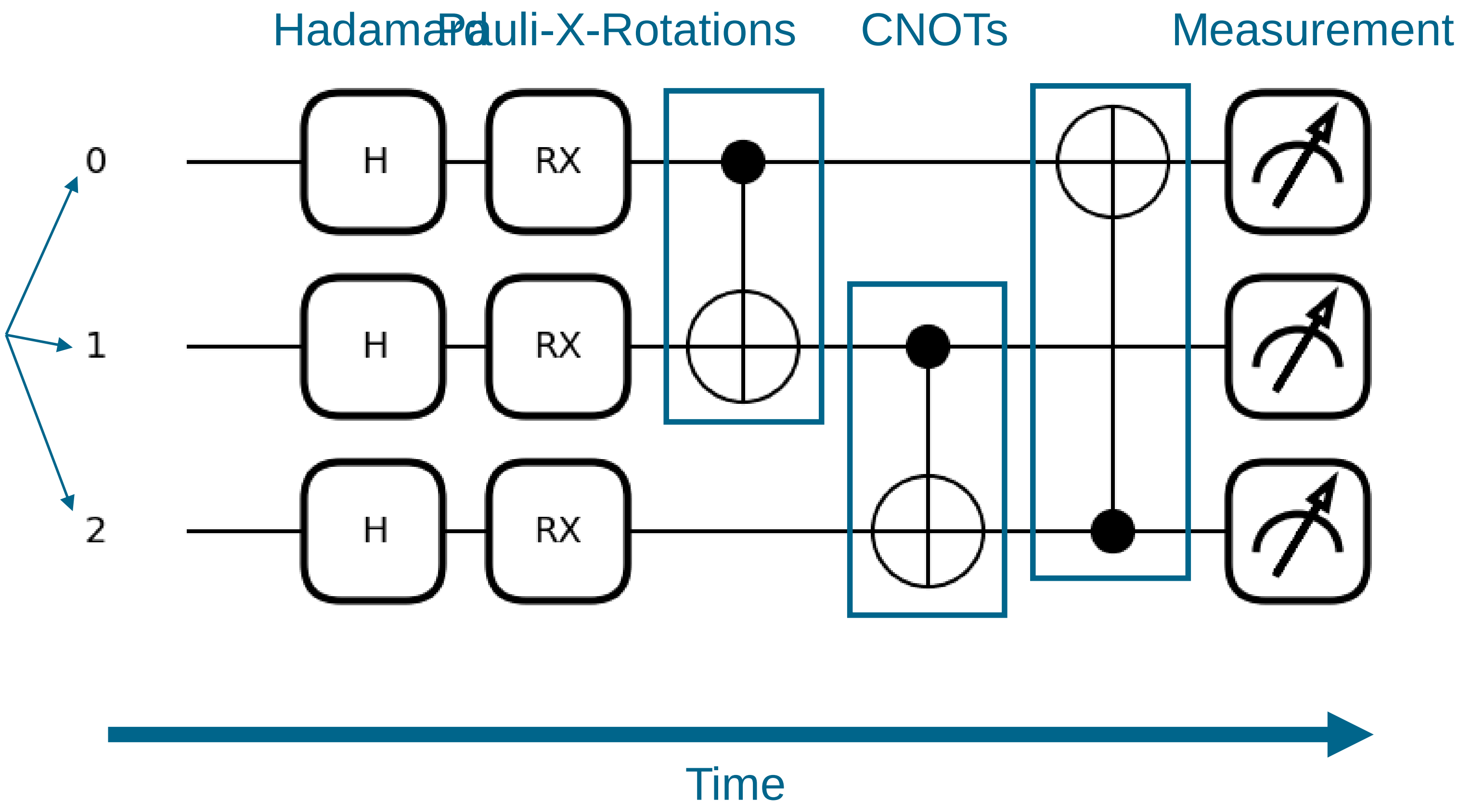
Entanglement

- "quantum correlation"
- Multi-qubit state $|\psi\rangle_{\mathrm{AB}}$ cannot be expressed as product of subsystems $|\varphi\rangle_A$ and $|\gamma\rangle_B$
- Ex.:
  - $\frac{1}{\sqrt{2}}(|00\rangle_{\mathrm{AB}} + |11\rangle_{\mathrm{AB}})$ is entangled
  - $\frac{1}{\sqrt{2}}(|01\rangle_{\mathrm{AB}} + |11\rangle_{\mathrm{AB}}) = \frac{1}{\sqrt{2}}(|0\rangle_{\mathrm{A}} + |1\rangle_{\mathrm{A}}) \otimes |1\rangle_{\mathrm{B}}$ is not entangled

- Capability to create entanglement is a crucial resource in QC

Circuit Diagrams

Implicit
initialization
of qubits as $|0\rangle$



Technology of Quantum Computers

Universal Quantum Computer

DiVincenzo Criteria

1. Well-characterized and scalable qubits
2. Qubit initialization
3. Long coherence times
4. Universal set of gates
5. Measurement of individual qubits

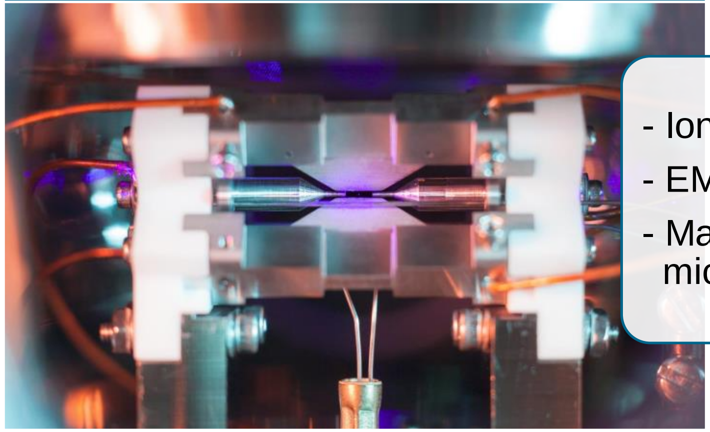
Experimental Realizations

Superconducting QC

- Different realizations:
  - Flux-qubit
  - Phase-qubit
  - Charge-qubit
- Manipulation via microwave pulses, coupling circuits, …

https://www.newscientist.com/article/2372828-superconducting-qubits-have-passed-a-key-quantum-test/

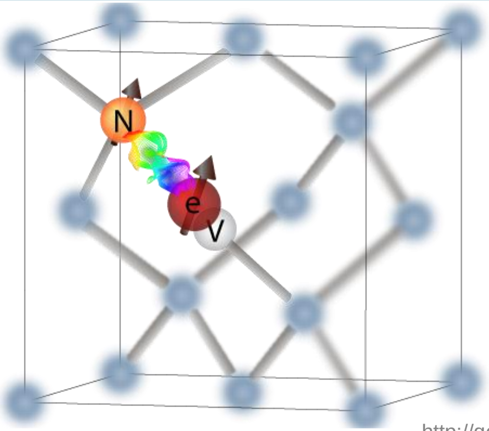
Ion Trap QC



- Ion energy levels as qubits
- EM fields trap chain of ions
- Manipulation via microwave pulses

https://www.physics.ox.ac.uk/research/group/ion-trap-quantum-computing

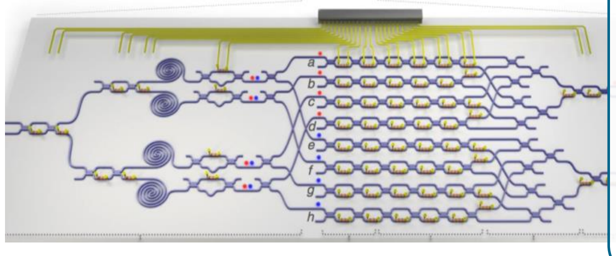
NV Center QC



- **N**itrogen & **V**acancy in diamond lattice
- Nitrogen nuclear spin as qubit
- Readout & coupling from NV electron spin

http://qeg.mit.edu/research.php

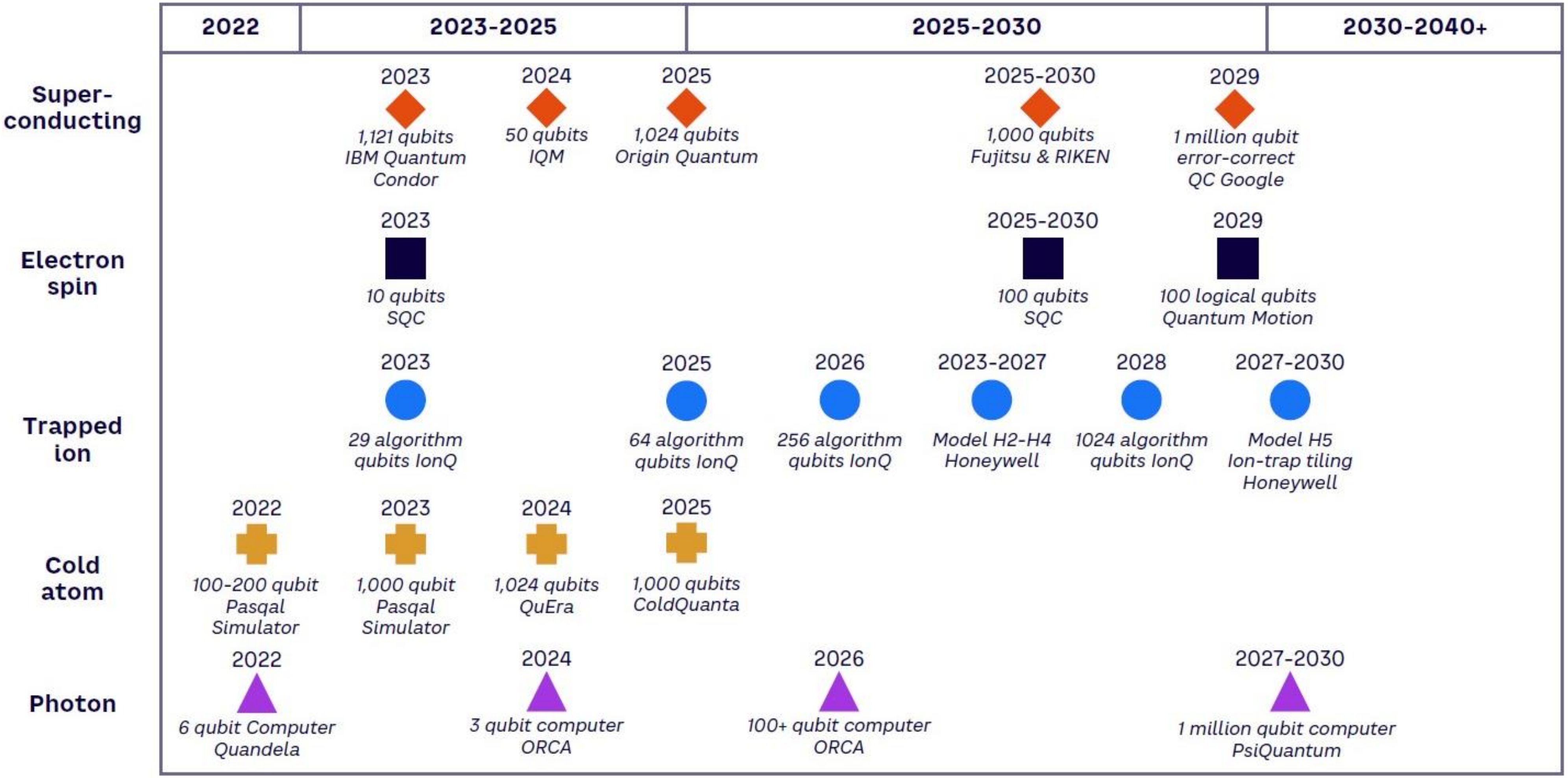
Photonic QC



- encode special states as Qubits (GKP states)
- Operations via beam splitters, phase shifters, …
- Readout via homodyne measurements

https://arstechnica.com/science/2018/09/engineering-tour-de-force-births-programmable-optical-quantum-computer/

Experimental Realizations: Roadmaps



Source: Arthur D. Little, Olivier Ezratty
https://www.adlittle.com/en/insights/viewpoints/quantum-computing

The NISQ Era & beyond

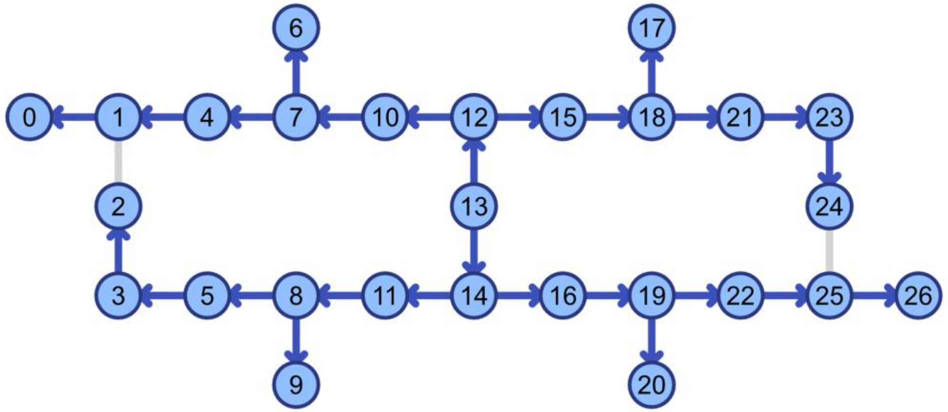
NISQ = “noisy, intermediate-scale quantum”:

- Large error rates ($\sim 10^{-3}$ for two-qubit gates)
- Relatively low qubit numbers
- Limited connectivity

Quantum Error Correction

- Encode several physical qubits into logical qubits
- Requires lower error rates to be useful
- Overhead: 100-1000 physical qubits per logical qubit

https://www.marketing-boerse.de/news/details/2124-quantensprung-in-ehningen-bei-stuttgart/177821



https://www.ibm.com/quantum/blog/whole-device-entanglement

Quantum Algorithms

What Quantum Computing is not

- “Quantum Computing = massive parallel computing” ?

- Idea:
  1. Prepare superposition: $|\psi\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} |i\rangle$
  2. Apply unitary U (to each element): $U|\psi\rangle = \frac{1}{\sqrt{2^n}} \sum_{i=0}^{2^n-1} U|i\rangle$
  3. …
  4. Success?

- Caveat: measurement → wavefunction collapse

Some Quantum Algorithms

	Shor	Grover	HHL
Use Case	Prime factorization	Search in unstructured database	Solving linear equation systems $A\vec{x} = \vec{b}$
Quantum Complexity	$O((\log N)^2 (\log\log N))$	$O(\sqrt{N})$	$O(\log(N)\,\kappa^2)$
Best Classical Complexity	$O(e^{1.9(\log N)^{1/3}(\log\log N)^{2/3}})$	$O(N)$	$O(N\kappa)$
Speed Up	Exponential	Quadratic	Exponential

Common caveat: circuit complexity → require fault-tolerant QC

Potential Applications

- Quantum chemistry: ground state estimation, molecule simulation
- Optimization:
  - quantitative finance (e.g., with reinforcement learning)
  - pharma/healthcare (drug discovery)
- Material science (solving PDEs, simulation)
- Any classical classification or regression task

Summary

Basics

- Qubits are the basic building block of quantum computers
- Superposition and entanglement of multiple qubits are essential resources
- Universal gate sets of one- and two-qubit gates allow arbitrary operations

Technology

- Different technologies are examined for QC; no clear-cut favorite
- Currently: noisy and intermediate scale (NISQ) era

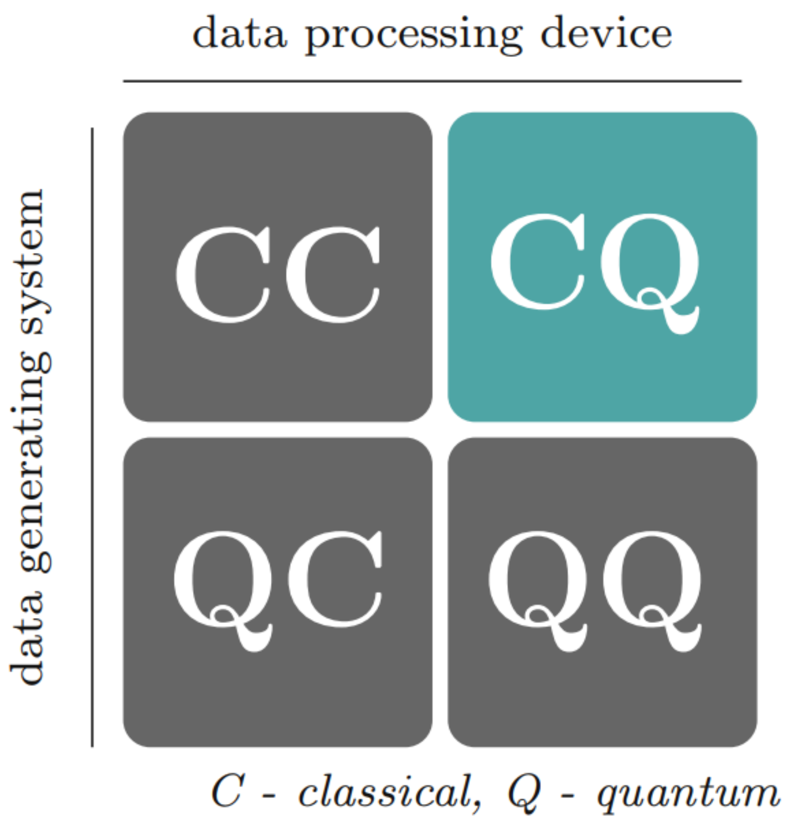
Algorithms

- QC ≠ free parallel computing
- Algorithms with proven speedups exist for some problem classes, but require better quantum computers (& quantum error correction)

Quantum Machine Learning

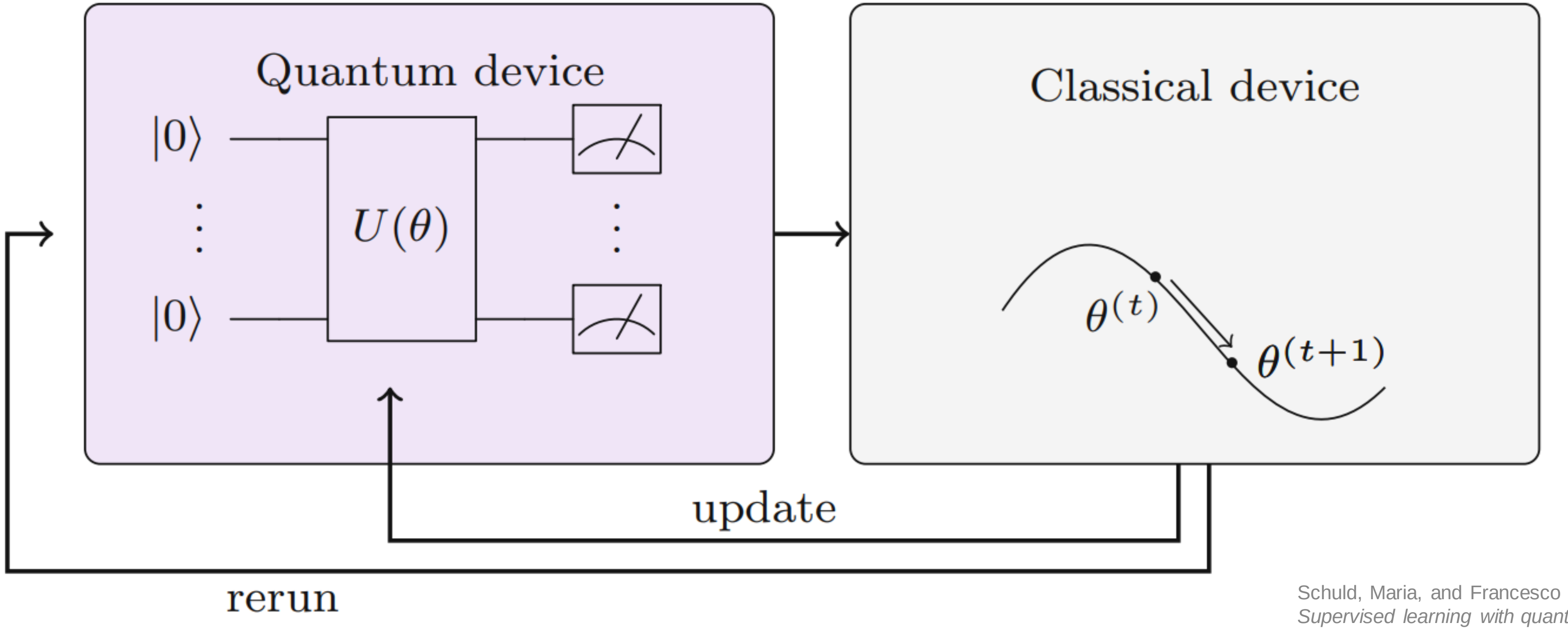
Overview

- Term “Quantum Machine Learning” coined ca. 2013
- Initially mainly kernel methods, later variational QML
- Mostly: processing of classical data with quantum computers



Schuld, Maria, and Francesco Petruccione.
Supervised learning with quantum computers. Vol. 17. Berlin: Springer, 2018.

Variational QML

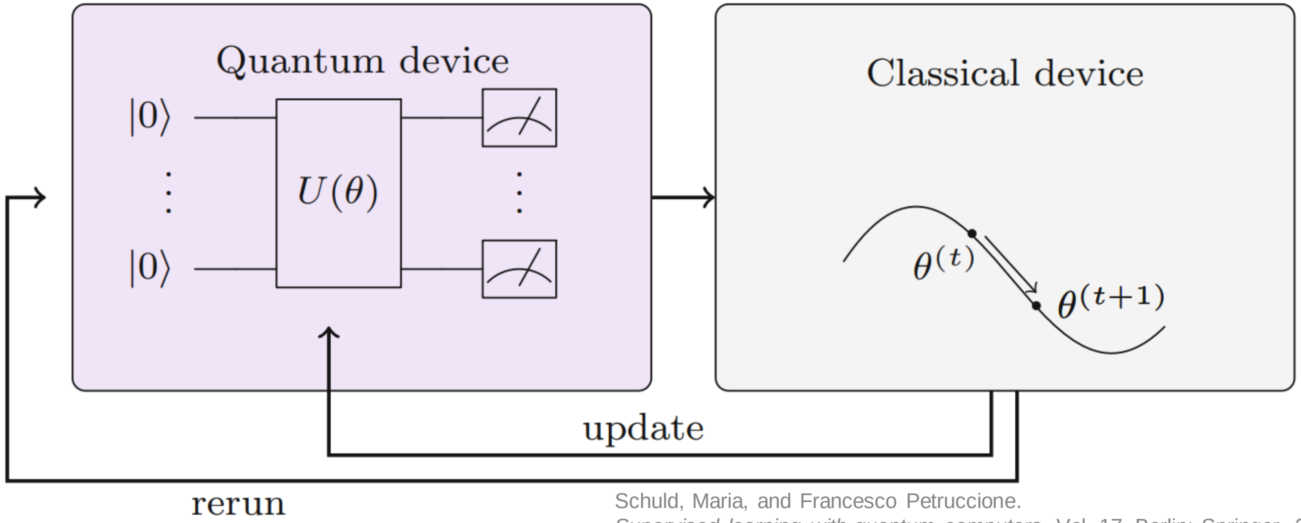


Schuld, Maria, and Francesco Petruccione.
Supervised learning with quantum computers. Vol. 17. Berlin: Springer, 2018.

Hybrid algorithm:
Classical optimization of a **parametrized quantum circuit**

Ingredients of a QML Algorithm

1. Preparation of input data on QC → Encoding
2. Choice of parametrized quantum circuit → Model ansatz
3. Output interpretation
4. Definition of loss function
5. Choice of optimization method



Schuld, Maria, and Francesco Petruccione.
Supervised learning with quantum computers. Vol. 17. Berlin: Springer, 2018.

Data Encodings

Basis Encoding

- Encode classical bitstrings $(b_1, b_2, \ldots, b_n) \in \{0,1\}^n$ as quantum state $|b_1, b_2, \ldots, b_n\rangle$
- *n* bit bitstring requires *n* qubits

Amplitude Encoding

- Classical feature vector $\boldsymbol{x} \in \mathbb{R}^{2^n}$
- Encode entries x_i of (normalized) $\boldsymbol{x}$ as *amplitudes* of quantum state

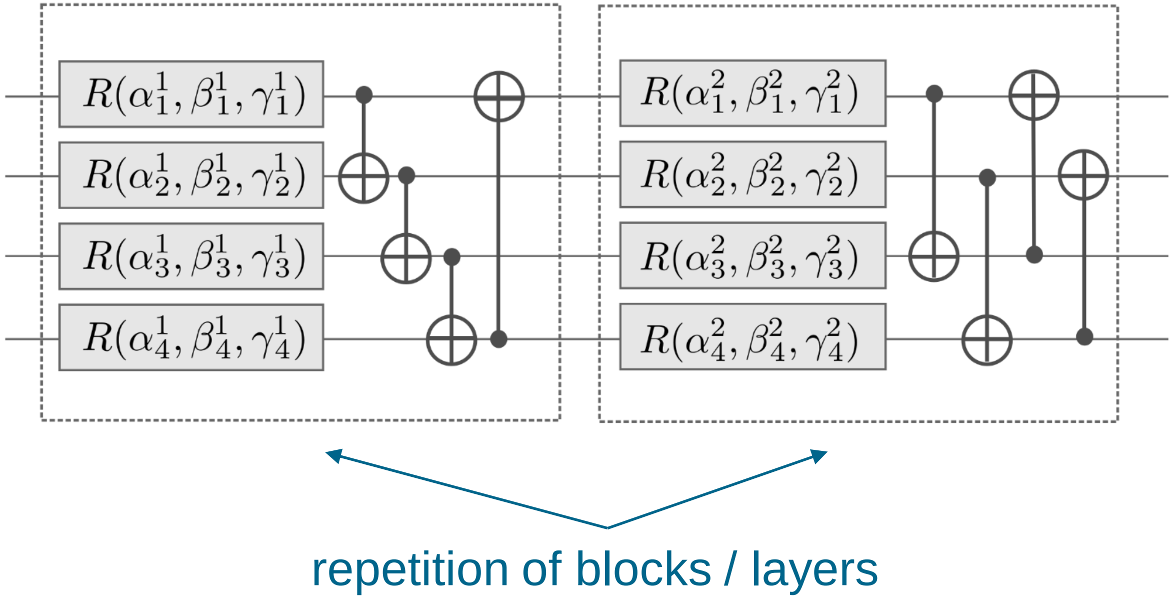
$$|\psi_x\rangle = \frac{1}{|\boldsymbol{x}|} \sum_{i=0}^{2^n - 1} x_i |i\rangle$$

- Feature vector of length 2^n requires *n* qubits → logarithmic scaling

Variational Model *Ansatz*

- Problem-tailored approaches
  - Unitary coupled cluster ansatz (UCC, quantum chemistry)
  - Quantum alternating operator ansatz (QAOA, optimization)

- Hardware efficient ansatz
  - Heuristic approach
  - Device-native gates & low circuit depth



https://docs.pennylane.ai/en/stable/code/api/pennylane.StronglyEntanglingLayers.html

Optimization

- Analogous to classical ML:
  - Start with randomly initialized parameters
  - Optimize each parameter θ_i w.r.t. the loss function $\mathcal{L}$

- Gradient-free methods:
  - Nelder-Mead, particle swarm opt., genetic algorithms

- Gradient methods:
  - Simultaneous perturbation stochastic approximation (SPSA)
  - Stochastic gradient descent, Adam, …
  - Parameter-shift rule: $\nabla_\theta \mathcal{L}(x;\theta) = \frac{1}{2}\left[\mathcal{L}\left(x;\theta+\frac{\pi}{2}\right) - \mathcal{L}\left(x;\theta-\frac{\pi}{2}\right)\right]$, if θ belongs to a Pauli-rotation

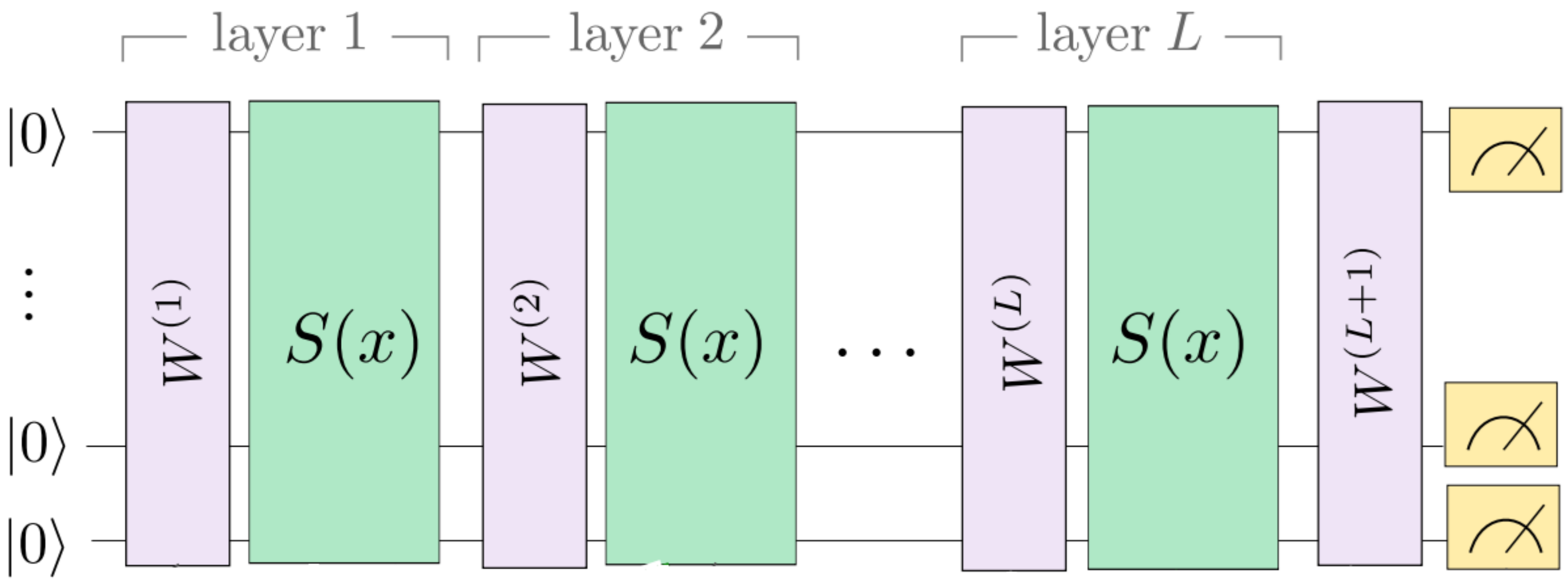
Where does non-linearity come in?

- Some encodings, e.g. angle encoding

- The measurement!
  - Maps amplitudes α_i to the absolute squared probabilities $|\alpha_i|^2$
  - Randomly samples an output state according to the probabilities

- In classification tasks: threshold functions, e.g. $\mathrm{sgn}(\cdot)$

Data Re-Uploading



Schuld, Maria, Ryan Sweke, and Johannes Jakob Meyer.
"Effect of data encoding on the expressive power of variational quantum-machine-learning models." Physical Review A 103.3 (2021): 032430.

Repeated application of encoding unitary in **alternating fashion** with parametrized blocks can enhance the **circuit expressivity**

Challenges in QML & Conclusion

Barren Plateaus

= "vanishing gradients problem in QML"

- For large classes of variational quantum circuits with random initialization:
  - Average value of gradient $\frac{\partial \mathcal{L}}{\partial \theta}$ of loss function $\mathcal{L}$ is zero
  - Variance of $\frac{\partial \mathcal{L}}{\partial \theta}$ decreases exponentially with the qubit number

→ Large "plateaus" in cost landscape which hinder learning

McClean, Jarrod R., et al. *"Barren plateaus in quantum neural network training landscapes."* Nature communications 9.1 (2018): 4812.

Classical Simulability

- Approaches to prevent / mitigate barren plateaus:
  - Local loss functions
  - Special parameter initializations
  - Circuit architectures which exploit problem symmetries
  - …

→ These approaches make the problem also classically efficiently simulable

Cerezo, Marco, et al. *"Does provable absence of barren plateaus imply classical simulability? Or, why we need to rethink variational quantum computing."* arXiv preprint arXiv:2312.09121 (2023).

Advantages of QML?

Property	Problems studied in quantum computing	Problems solved by machine learning
classical performance	**low** – problems are carefully selected to be provably difficult for classical computers	**high** – machine learning is applied on an industrial scale and many algorithms run in linear time in practice
size of inputs	**small** – near-term algorithms are limited by small qubit numbers, while fault-tolerant algorithms usually take short bit strings	**very large** – may be millions of tensors with millions of entries each
problem structure	**very structured** – often exhibiting a periodic structure that can be exploited by interference	**"messy"** – problems are derived from the human or "real-world" domain and naturally complex to state and analyse
theoretical accessibility	**high** – there is a large bias towards problems about which we can theoretically reason	**shifting** – theory is currently been re-built around the empirical success of deep learning
evaluating performance	**computational complexity** – the dominant measure to assess the performance of an algorithm is asymptotic runtime scaling	**practical benchmarks** – machine learning research puts a strong emphasis on empirical comparisons between methods

Schuld, Maria, and Nathan Killoran. *"Is quantum advantage the right goal for quantum machine learning?"*. Prx Quantum 3.3 (2022): 030101.

Conclusion

- QML is promising in some aspects, but:
  - No actual quantum advantage is in sight
  - Classical hardware is far advanced & especially developed for ML
  - Questions of trainability (barren plateaus) and classical simulability are still unclear

- However:
  - Big players are involved (Google, IBM, Intel)
  - Can bridge the gap from the NISQ era to fault tolerant era
  - Still ongoing research