Quantum Machine Learning: An Introduction

Osvaldo Simeone

King's College London

ELLIIT Workshop, 19/10/2022

This Talk

- Context
- Quantum machine learning: a taxonomy
- Unsupervised learning via Born machines
- Supervised learning: probabilistic vs. deterministic models
- Research directions
- Conclusions

Context

Promises and Hype

The Tell

Quantum computing will be the smartphone of the 2020s, says Bank of America strategist

Published: Dec. 12, 2019 at 2:40 p.m. ET

By Chris Matthews

12

Exponentially more computing power may revolutionize health care and cybersecurity

FAST COMPANY

09-28-21 | FAST COMPANY INNOVATION FESTIVAL

IBM CEO: Quantum computing will take off 'like a rocket ship' this decade

But Arvind Krishna says that some hard quantum physics problems await as the market pushes for larger and larger quantum systems.

CLASSIQ Q.A.D PLATFORM INSIGHTS

‹ Back to Insights

PRESS RELEASES

CLASSIQ RESEARCH REVEALS THAT QUANTUM COMPUTING IS THE FUTURE – AND IT'S COMING SOON

22 September, 2021

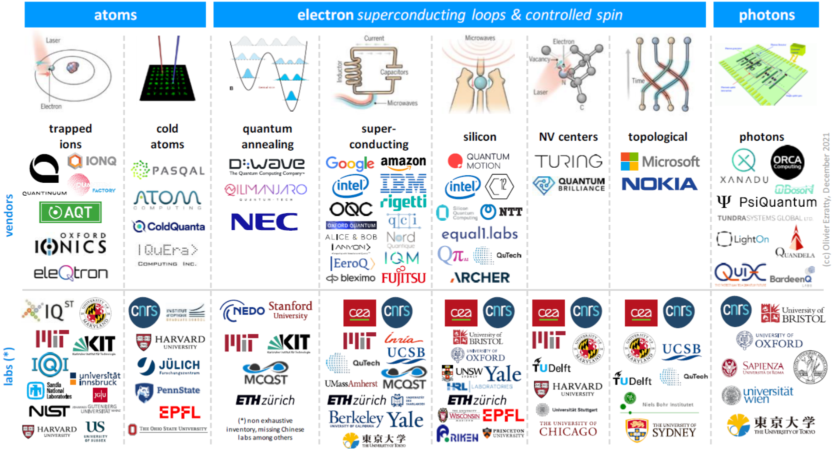
Quantum Computing Paranoia Creates a New Industry

Even though quantum computers don't exist yet, security companies are preparing to protect against them.

by Tom Simonite January 30, 2017

MIT Technology Review

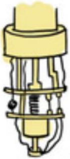
Players



[O Ezratty '22]

Quantum Algorithms and Today's Technology

- The traditional design of quantum algorithms assumes **large and reliable** quantum computers.
- Quantum machine learning is emerging as a programming paradigm suited for current **noisy intermediate-scale quantum (NISQ) computers**.[1]



	Fault-tolerant	Near-term
# qubits	millions	10-1000
errors	corrected	mitigated
use	Shor, Grover, HHL...	variational circuits
research	computational complexity	run it and see
available	in 5-30 years?	now

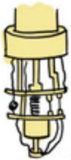
[M. Schuld '21]

[1] M. Schuld and F. Petruccione, Machine Learning with Quantum Computers, Springer, 2021.

Quantum Algorithms and Today's Technology

- The traditional design of quantum algorithms assumes **large and reliable** quantum computers.
- Quantum machine learning is emerging as a programming paradigm suited for current **noisy intermediate-scale quantum (NISQ) computers**.[1]

	Fault-tolerant	Near-term
# qubits	millions	10-1000
errors	corrected	mitigated
use	Shor, Grover, HHL...	variational circuits
research	computational complexity	run it and see
available	in 5-30 years?	now

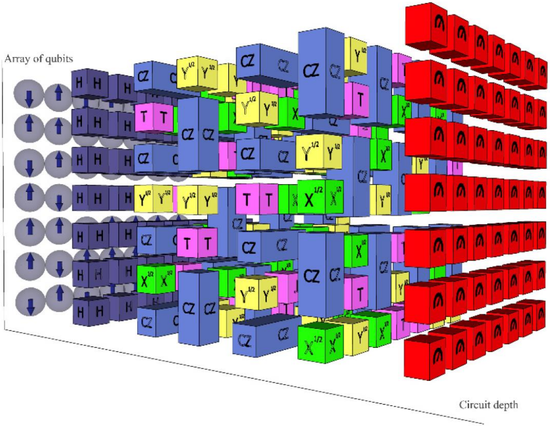


[M. Schuld '21]

[1] M. Schuld and F. Petruccione, Machine Learning with Quantum Computers, Springer, 2021.

Quantum Circuit

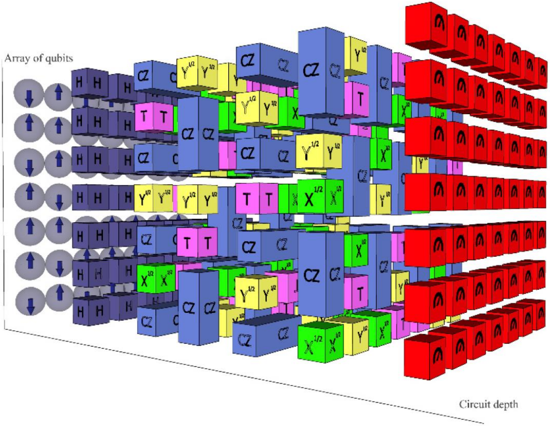
- A **quantum algorithm** is specified by a quantum circuit operating on a set of n qubits.
- A **quantum circuit** consists of a sequence of **quantum gates** that are applied sequentially and **in place** to the n qubits...



[Hidary '19]

Quantum Circuit

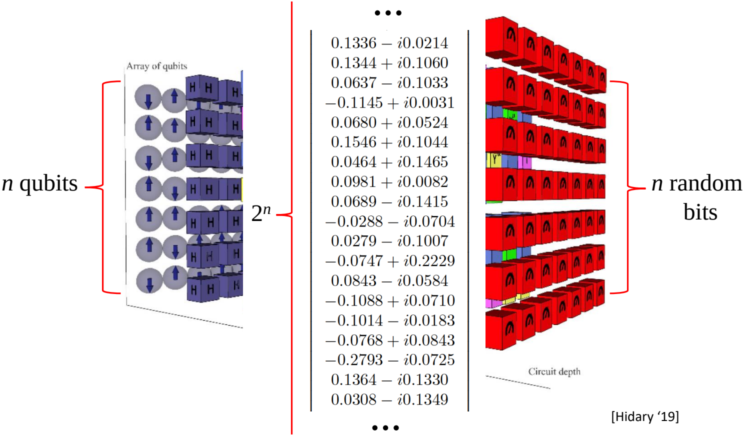
- … followed by **measurements** that convert the state of the *n* qubits into *n* classical bits.



[Hidary '19]

Quantum Circuit

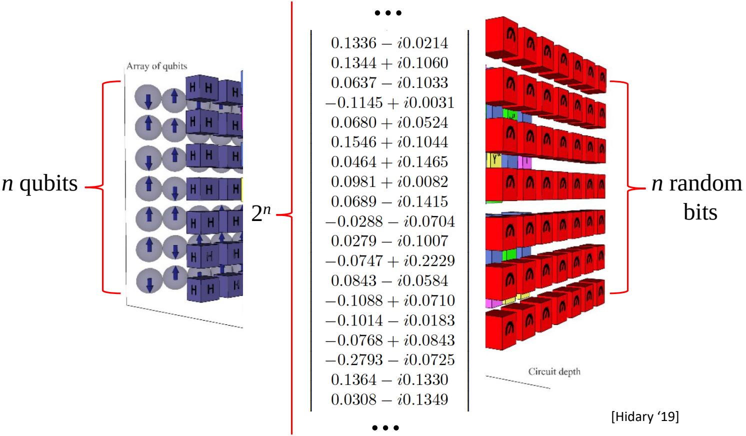
- The state of n qubits is described by a 2^n-**dimensional complex (amplitude) vector**.
- Quantum measurements are inherently **random**: **"collapse"** of the waveform.



[Hidary '19]

Quantum Circuit

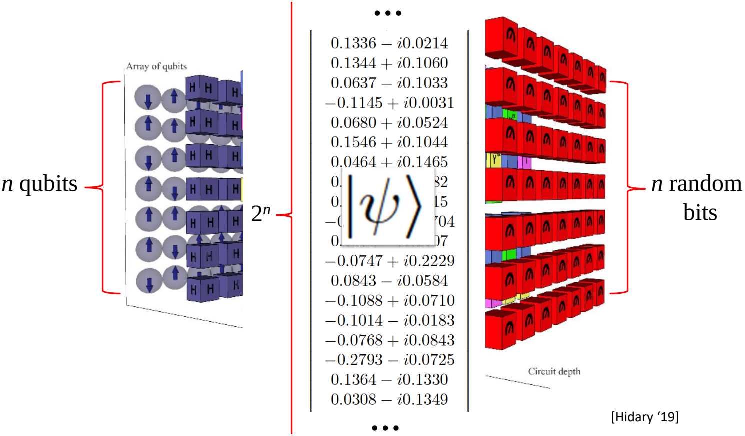
- The state of n qubits is described by a 2^n-**dimensional complex (amplitude) vector**.
- Quantum measurements are inherently **random**: **"collapse"** of the waveform.



[Hidary '19]

Quantum Circuit

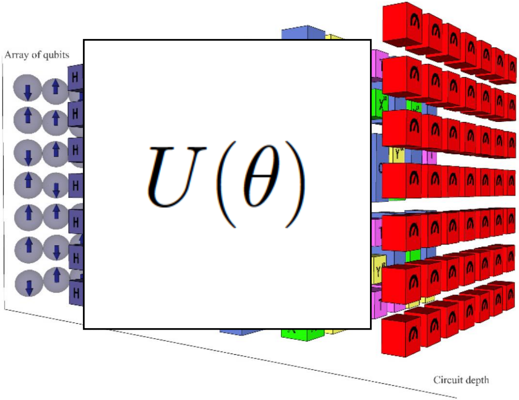
- State vectors are represented using **Dirac's ket notation** $|\psi\rangle$.



[Hidary '19]

Parameterized Quantum Circuit

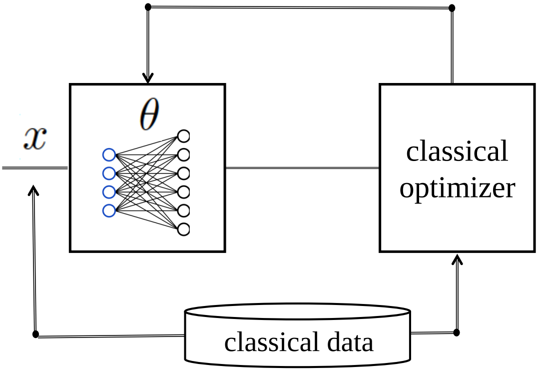
- A **parameterized quantum circuit** (PQC) is defined by a fixed sequence of quantum gates whose operation depends on a vector of **classical parameters** θ.
- PQCs are also known as **quantum neural networks**.



[Hidary '19]

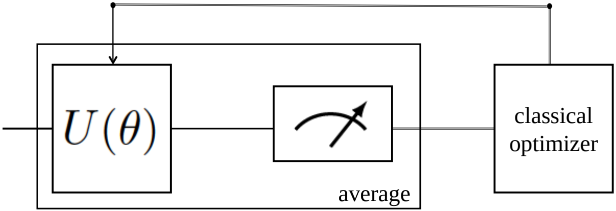
Classical Machine Learning

- Classical machine learning relies on **parameterized functions** $f(x|\theta)$, e.g., neural networks.
- The parameters θ are optimized by comparing the model output with classical data.



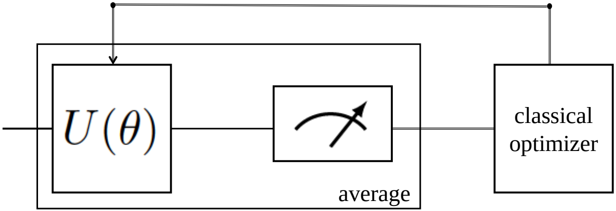
Quantum Machine Learning

- In **quantum machine learning**, the parameters of the PQC $U(\theta)$ are designed using classical optimization based on measurements of the output of the PQC and (possibly) data.
- By keeping the quantum computer in the loop, the classical optimizer can account for the **non-idealities and limitations** of quantum operations.

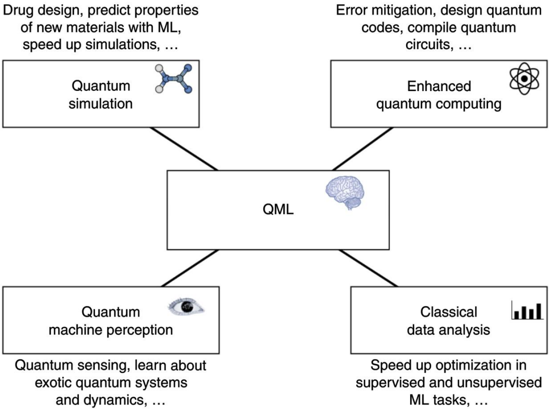


Quantum Machine Learning

- In **quantum machine learning**, the parameters of the PQC $U(\theta)$ are designed using classical optimization based on measurements of the output of the PQC and (possibly) data.
- By keeping the quantum computer in the loop, the classical optimizer can account for the **non-idealities and limitations** of quantum operations.

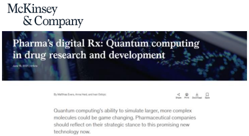


Quantum Machine Learning: Functionalities



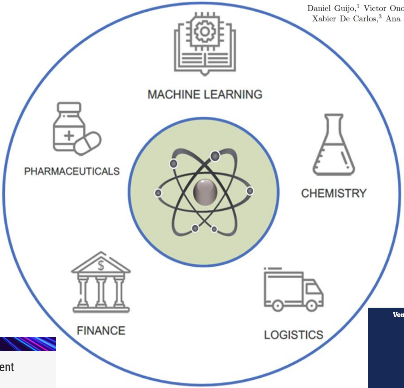
[Cerezo et al '22]

Quantum Machine Learning: Applications?



Quantum artificial vision for defect detection in manufacturing

Daniel Guijo,[1] Victor Onofre,[1] Gianni Del Bimbo,[1] Samuel Mugel,[2] Daniel Estepa,[3] Xabier De Carlos,[3] Ana Adell,[3] Aizea Lojo,[3] Josu Bilbao,[3] and Román Orús[1,4,5]

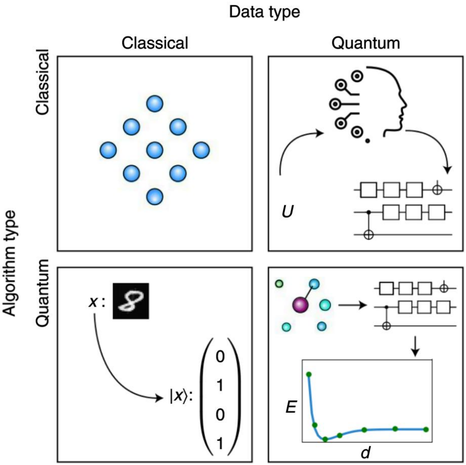


https://quantumcomputingreport.com/

Quantum Machine Learning:
A Taxonomy

Quantum Machine Learning

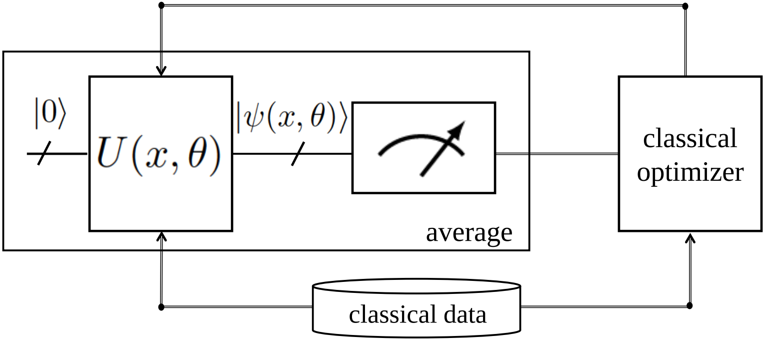
- Generalizing classical machine learning, in quantum machine learning **data and/or processing** are quantum.



[Cerezo et al '22]

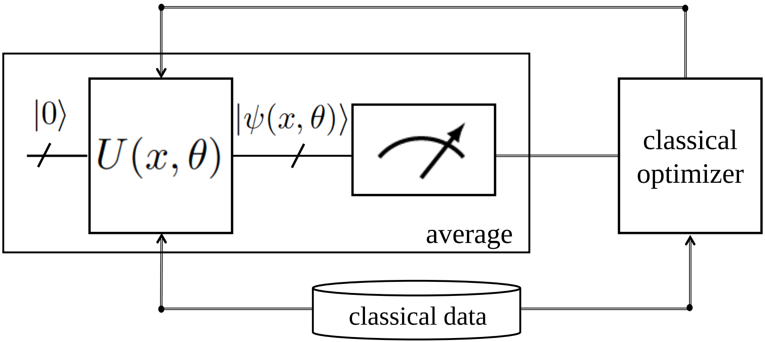
Quantum Machine Learning: CQ

- Currently, the most common quantum machine learning case is "CQ": data are classical, while processing is quantum.
- The measurement outputs are compared to classical data to optimize parameters θ.
- Can implement (more efficiently?) classical machine learning tasks.



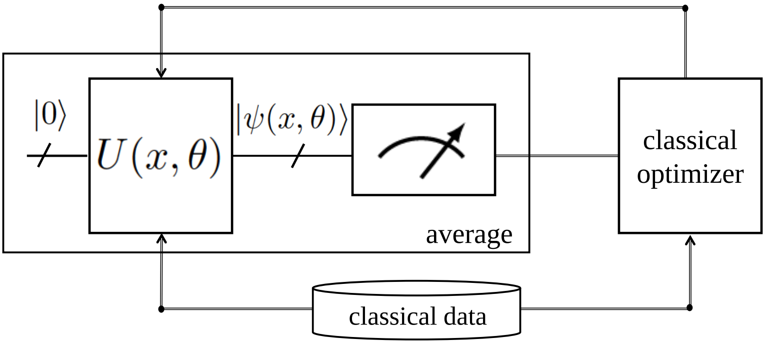
Quantum Machine Learning: CQ

- Currently, the most common quantum machine learning case is "CQ": data are classical, while processing is quantum.
- The measurement outputs are compared to classical data to optimize parameters θ.
- Can implement (more efficiently?) classical machine learning tasks.



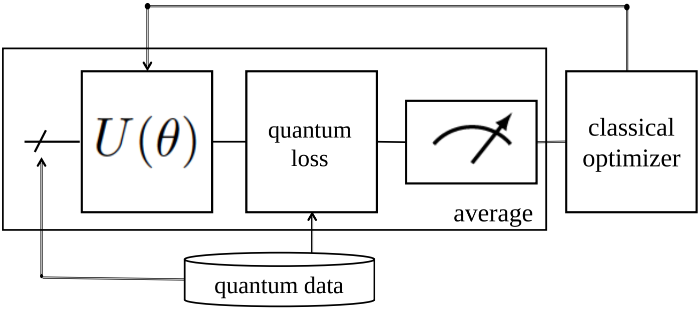
Quantum Machine Learning: CQ

- Currently, the most common quantum machine learning case is "CQ": data are classical, while processing is quantum.
- The measurement outputs are compared to classical data to optimize parameters θ.
- Can implement (more efficiently?) classical machine learning tasks.



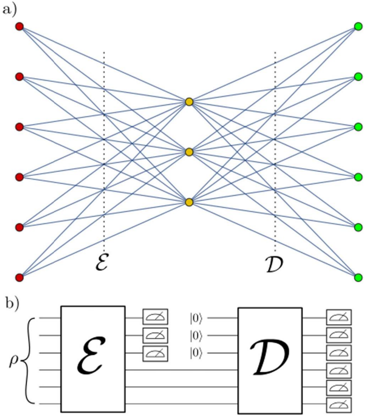
Quantum Machine Learning: QQ

- In the "QQ" case, the quantum state produced by the PQC is compared with quantum data to optimize θ.



Quantum Machine Learning: QQ

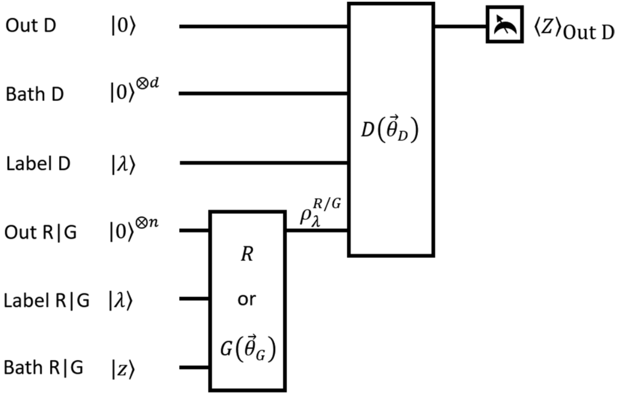
- **Quantum autoencoders** for compression[2]



[2] J. Romero, et al, "Quantum autoencoders for efficient compression of quantum data," Quantum Science and Technology, 2017.

Quantum Machine Learning: QQ

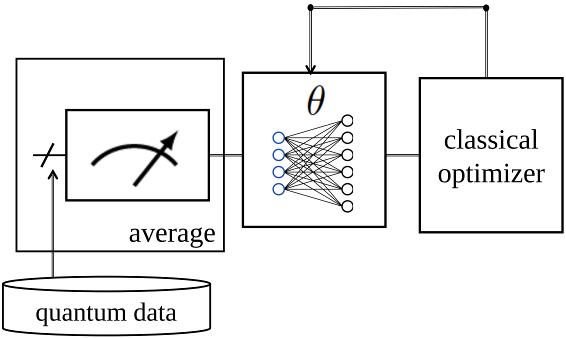
- **Quantum generative adversarial networks**[3]



[3] P. Dallaire-Demers and N. Killoran, "Quantum generative adversarial networks," Physical Review A, 2018.

Quantum Machine Learning: QC

- In the "QC" case, there is no PQC, and the outputs of measurements of a quantum state are processed by a classical machine learning model.
- Example: quantum tomography[4].



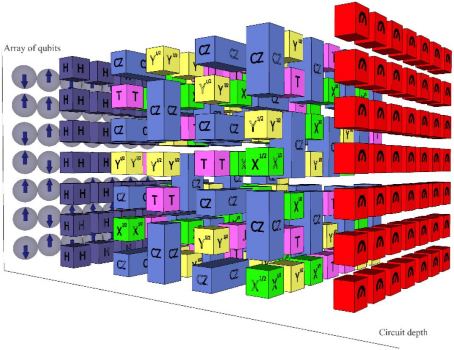
[4] V. Gebhart, "Learning quantum systems," arXiv:2207.00298 , 2022.

Parameterized Quantum Circuits

Quantum Circuit

- Quantum gates implement multiplications by **unitary matrices (reversible norm-preserving linear transformations)**.
- Measurements convert quantum information into n classical, random, bits by following **Born's rule**:
  - Given the $2^n \times 1$ state vector $|\psi\rangle = [\alpha_x]$ with $x \in \{0,1\}^n$

$$\Pr[\text{bit string } x] = p(x) = |\alpha_x|^2$$

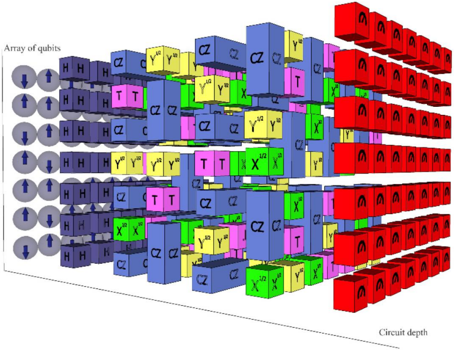


[Hidary '19]

Quantum Circuit

- Quantum gates implement multiplications by **unitary matrices (reversible norm-preserving linear transformations)**.
- Measurements convert quantum information into n classical, random, bits by following **Born's rule**:
  - Given the $2^n \times 1$ state vector $|\psi\rangle = [\alpha_x]$ with $x \in \{0,1\}^n$

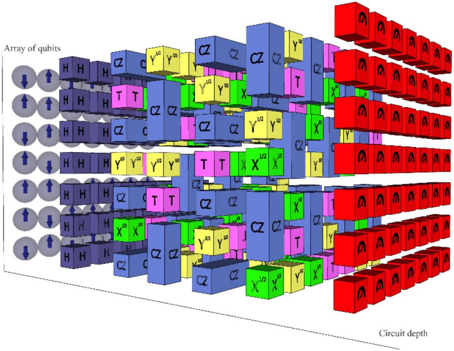
$$\Pr[\text{bit string } x] = p(x) = |\alpha_x|^2$$



[Hidary '19]

Quantum Circuit

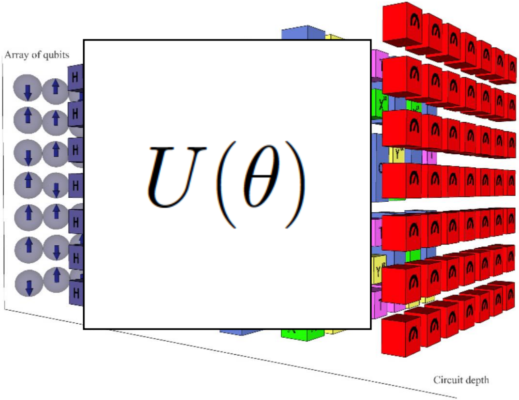
- Measurement outputs may be averaged to mitigate **shot noise**:
  - $x \rightarrow$ real number o_x = realization of an **observable**
  - $\langle O \rangle$ = **expected values of the observable**



[Hidary '19]

Parameterized Quantum Circuit

- A PQC is defined by a fixed sequence of quantum gates that can depend on a vector of classical parameters θ:
  - Parameterized **unitary** matrix $U(\theta)$



[Hidary '19]

Ansatz

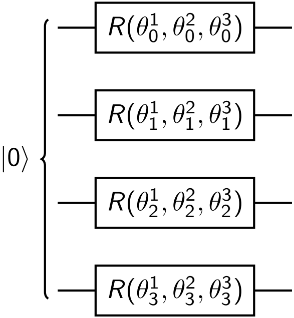
- The choice of the architecture of the PQC is akin to the choice of the **model class** in classical machine learning (e.g., neural network architecture).
- In quantum machine learning, we refer to the architecture of the PQC $U(\theta)$ as the **ansatz** (from the German term for "approach" or "attempt").
- As for the model class in machine learning, one should choose the ansatz, if possible, based on domain knowledge (e.g., in quantum chemistry).

Ansatz

- The choice of the architecture of the PQC is akin to the choice of the **model class** in classical machine learning (e.g., neural network architecture).
- In quantum machine learning, we refer to the architecture of the PQC $U(\theta)$ as the **ansatz** (from the German term for "approach" or "attempt").
- As for the model class in machine learning, one should choose the ansatz, if possible, based on domain knowledge (e.g., in quantum chemistry).

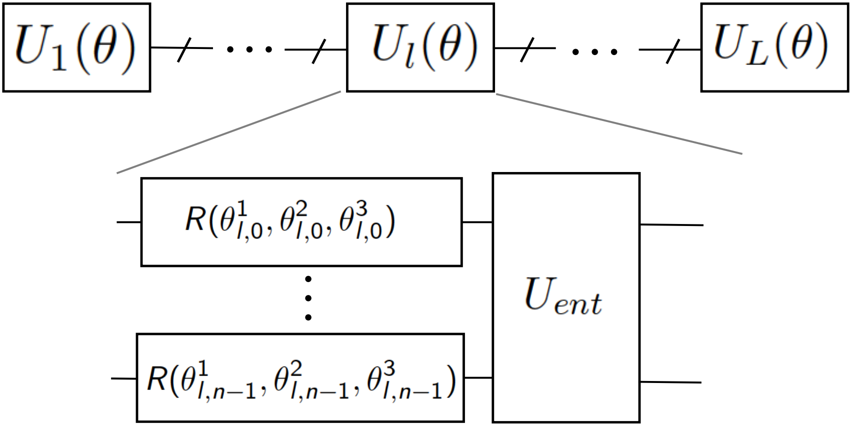
Mean-Field Ansatz

- The mean-field ansatz applies separate parameterized rotations on the qubits.



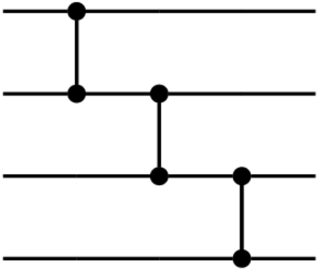
Hardware-Efficient Ansatz

- The hardware-efficient ansatz applies layers of separate rotations and **fixed multi-qubit, entangling, gates**.

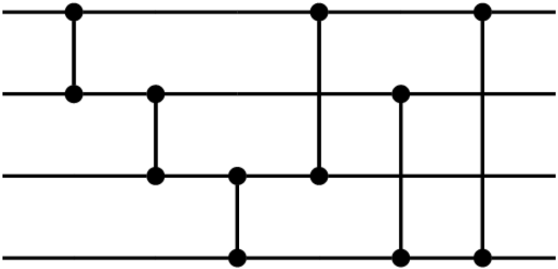


Hardware-Efficient Ansatz

- The multi-qubit entangling gate U_{ent} consists of a fixed cascade of **two-qubit gates**.

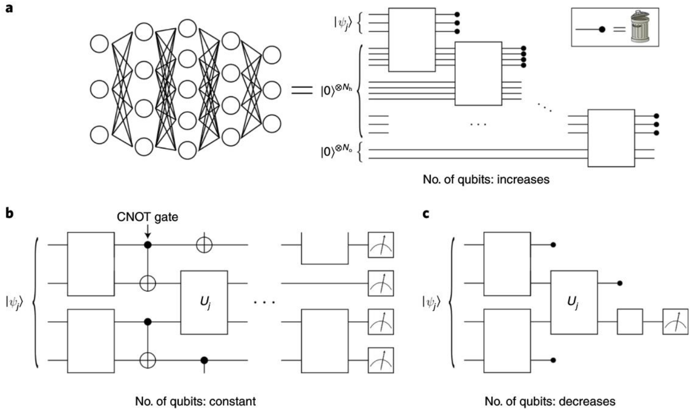


linear entangling gate
(based on CZ)



full entangling gate
(based on CZ)

Ansatzes with Increasing/ Decreasing Number of Qubits

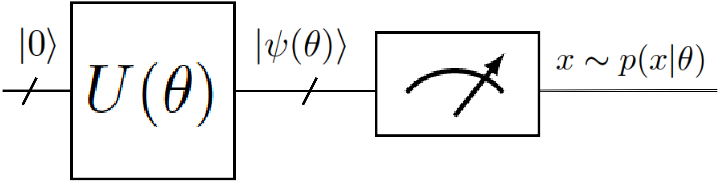


[Cerezo et al '22]

Unsupervised Learning via Born Machines

Unsupervised Generative Learning

- The measurement of the output of a PQC on n qubits produces a **random n-bit string** $x \sim p(x|\theta)$ by Born's rule.
- **Born machine**: Generative model for binary strings x implemented via a PQC[5]



[5] B. Coyle B, et al, "The Born supremacy: Quantum advantage and training of an Ising Born machine," Quantum Information, 2020.

Unsupervised Generative Learning

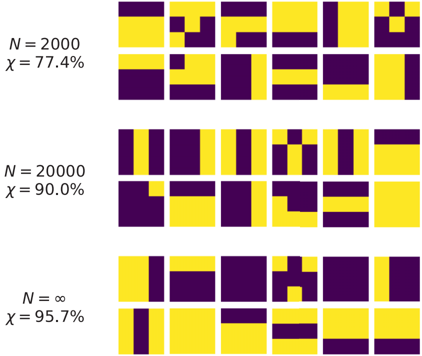
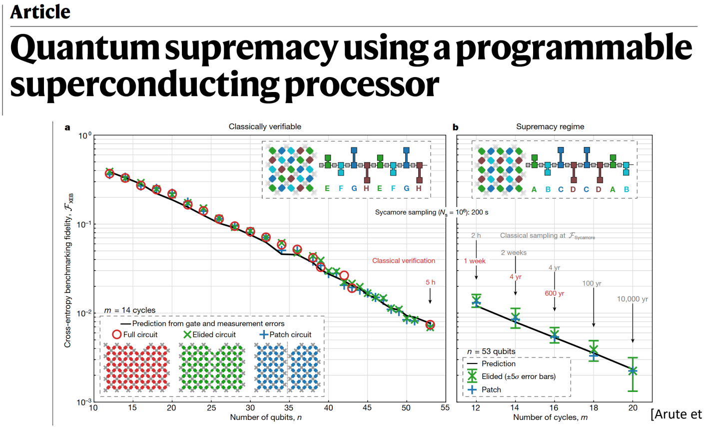


FIG. 4. 3×3 BARS-AND-STRIPES samples generated from the QCBMs. The circuit parameters used here are from the final stages of Adam training with different batch sizes N in Fig. 3(a). χ is the rate of generating valid samples in the training dataset. For illustrative purposes, we only show 12 samples for each situation with batch size N.

[Liu and Wang '18]

Quantum Circuits as Samplers

- Current claims of **quantum supremacy/ advantage** rest on the capacity of quantum circuits to generate samples from **joint discrete distributions** in a more efficient manner than classical devices[6].

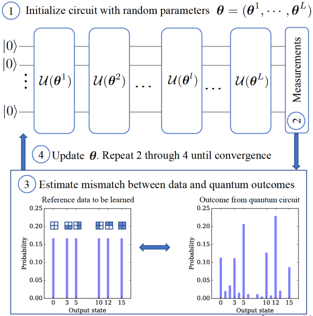


[Arute et al '19]

[6] X. Gao et al, "Enhancing generative models via quantum correlations," arXiv 2021.

Training Born Machines

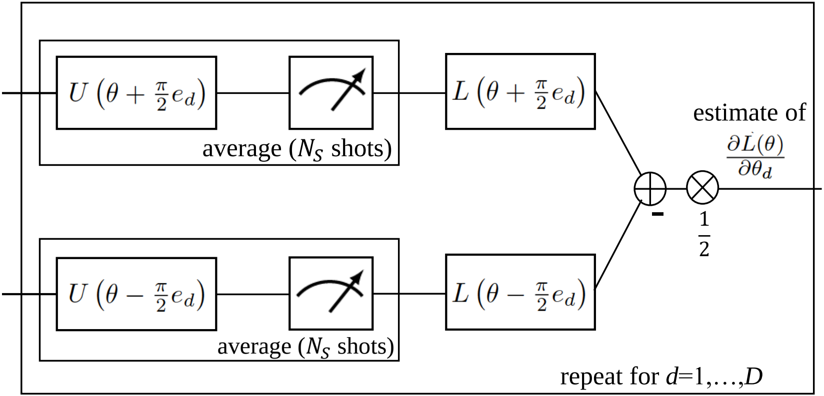
- Training is based on comparing samples $x \sim p(x|\theta)$ with training data.
- This is done by minimizing a measure of **divergence** between data distribution and $p(x|\theta)$.
- This is expressed in the form of the **expected value of a cost observable**.



[Tian et al '22]

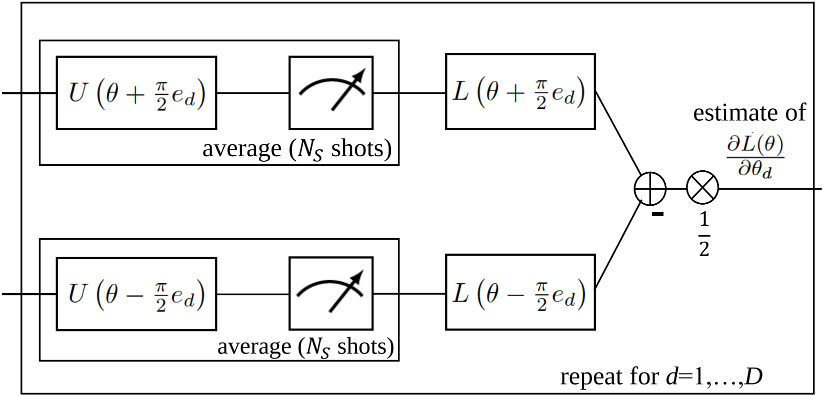
Gradient Descent

- Optimization is often carried out via **gradient descent**.
- Unlike classical machine learning models, **backprop** is **not** applicable, since we do not have access to internal workings of the PQC.
- The gradient is instead estimated via a perturbation-based method known as **parameter shift rule**.



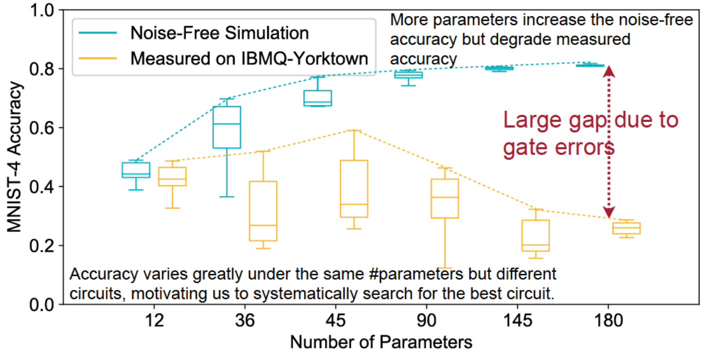
Gradient Descent

- Optimization is often carried out via **gradient descent**.
- Unlike classical machine learning models, **backprop** is **not** applicable, since we do not have access to internal workings of the PQC.
- The gradient is instead estimated via a perturbation-based method known as **parameter shift rule**.



Barren Plateaus

- Implementing gradient descent is practically made complicated by the fact that the **loss landscape** with generic, **unstructured**, ansatzes is not well behaved as the number of qubits increases.[7]



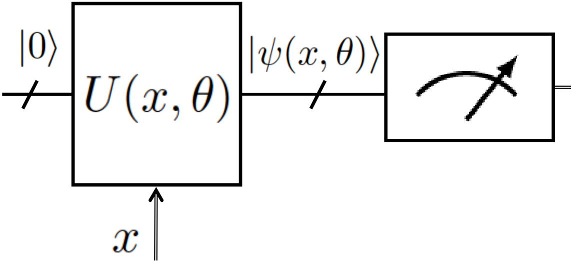
[Wang et al '22]

[7] J. McClean, et al, "Barren plateaus in quantum neural network training landscapes," Nature communications, 2018.

Supervised Learning:
Probabilistic vs. Deterministic Models

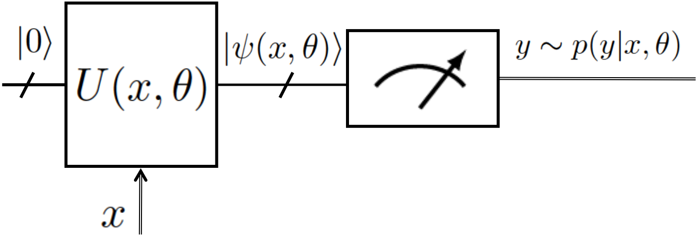
Supervised Learning

- **Input** x is typically encoded in the operation of the PQC $U(x, \theta)$ in a manner similar to the model parameters θ (**angle encoding**).
- (There are other ways to embed classical information into a quantum state.)



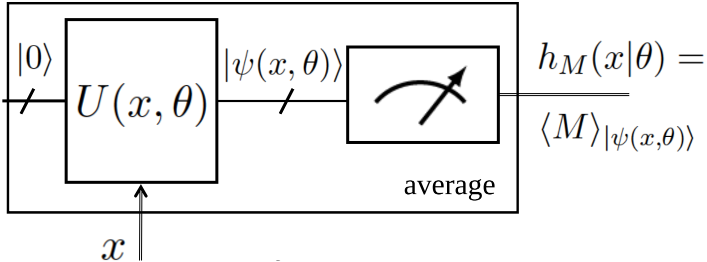
Supervised Learning: Probabilistic Models

- For **classification**, the target variable is a binary string $y \in \{0,1\}^m$, with $m \leq n$.
- Probabilistic models obtain a **randomized decision** y through a measurement of the qubits.



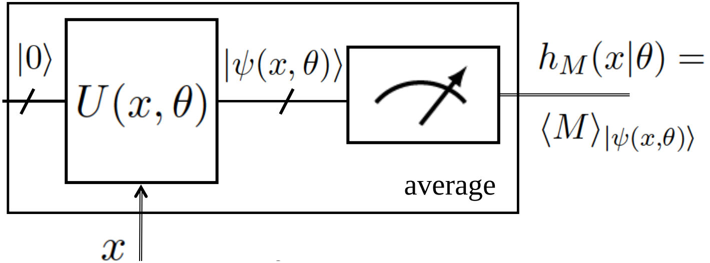
Supervised Learning: Deterministic Models

- **Deterministic models** implement a parameteric function $h_M(x|\theta)$ of the input x, which may be used for regression or classification.
- Function $h_M(x|\theta)$ is evaluated by estimating expectations of one or more observables.
- Unlike probabilistic modes, shot noise averaging is required also for inference (and not only for learning).



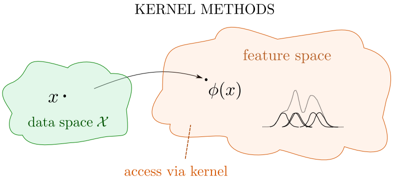
Supervised Learning: Deterministic Models

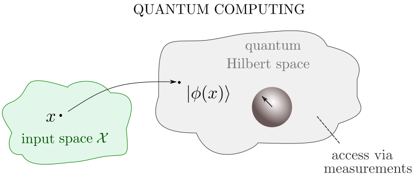
- **Deterministic models** implement a parameteric function $h_M(x|\theta)$ of the input x, which may be used for regression or classification.
- Function $h_M(x|\theta)$ is evaluated by estimating expectations of one or more observables.
- Unlike probabilistic modes, shot noise averaging is required also for inference (and not only for learning).



Supervised Learning: Deterministic Models

- Deterministic quantum models are akin to classical **kernel methods** in that they operate over a large feature space – the Hilbert space of dimension 2^n – via linear operations.[8]





[8] M. Schuld, "Supervised quantum machine learning models are kernel methods," arXiv, 2021.

Research Directions

Challenges

- **Architecture:**
  - What are the "right" building blocks for quantum machine learning models?
  - How to scale up classical input and/or output data?
  - How to integrate classical and quantum machine learning models?
- **Optimization:**
  - How to improve the performance of gradient descent in the presence of barren plateaus?
  - How to account for "quantum noise"?
- **Theory:**
  - What are the data requirements for quantum machine learning, particularly for generative modeling?

Challenges

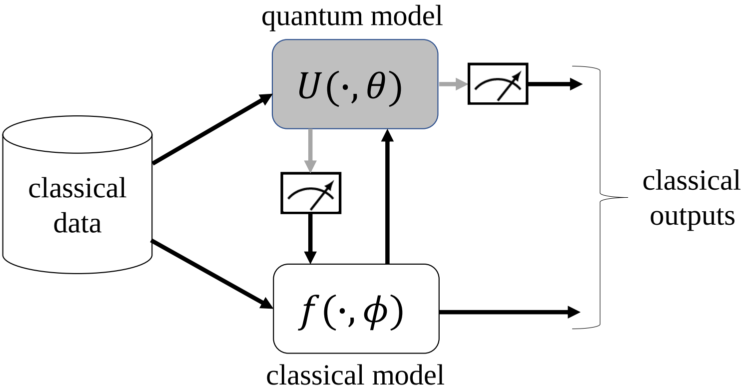
- **Architecture:**
  - What are the "right" building blocks for quantum machine learning models?
  - How to scale up classical input and/or output data?
  - How to integrate classical and quantum machine learning models?
- **Optimization:**
  - How to improve the performance of gradient descent in the presence of barren plateaus?
  - How to account for "quantum noise"?
- **Theory:**
  - What are the data requirements for quantum machine learning, particularly for generative modeling?

Challenges

- **Architecture:**
  - What are the "right" building blocks for quantum machine learning models?
  - How to scale up classical input and/or output data?
  - How to integrate classical and quantum machine learning models?
- **Optimization:**
  - How to improve the performance of gradient descent in the presence of barren plateaus?
  - How to account for "quantum noise"?
- **Theory:**
  - What are the data requirements for quantum machine learning, particularly for generative modeling?

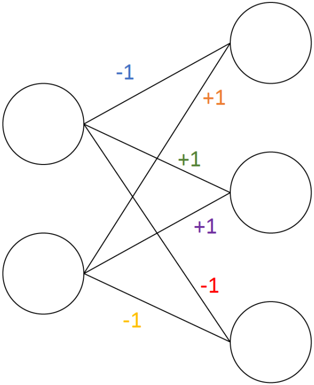
Hybrid Classical-Quantum Models

- **C-QC** = classical data - quantum-classical processing
- Possible two-way interaction between classical and quantum models
- Data generally processed and output by both models



Binary Neural Networks

- **Binary** neural network: ± 1 weights[9]
- We model the distribution $q(\theta)$ implicitly via a **Born machine**[10]

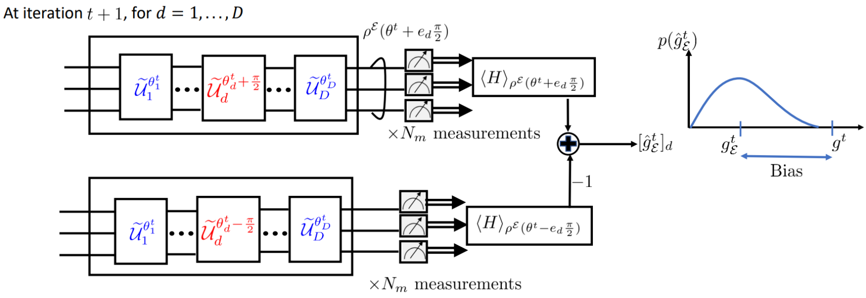


[9] W. Tang, et al, "How to train a compact binary neural network with high accuracy?," AAAI, 2017.

[10] I. Nikoloska and O. Simeone, "Quantum-Aided Meta-Learning for Bayesian Binary Neural Networks with Born Machines," IEEE MLSP, 2022.

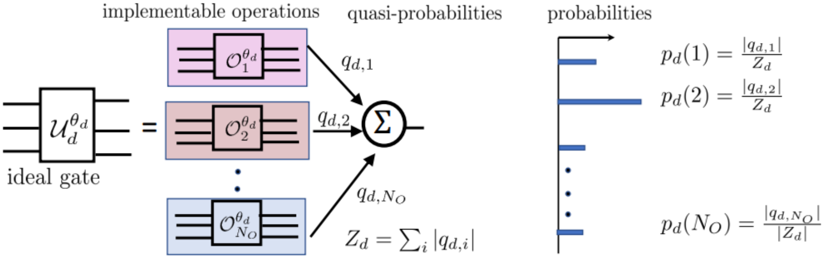
Training on Noisy Quantum Computers

- Gradients are to be estimated using an actual NISQ computer.
- Quantum noise causes the estimate of the gradient to be **biased**.



Quantum Error Mitigation

- **Quantum error correction** to fully compensate for quantum noise requires increasing the number of qubits beyond the current reach of quantum technology.
- **Quantum error mitigation** trades space (qubits) with time, running multiple noisy circuits to emulate a noiseless one.[11]

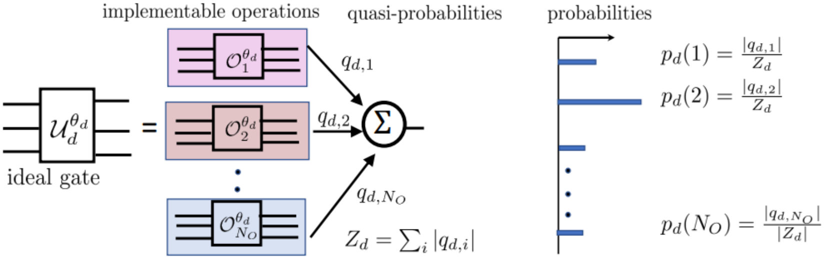


Quasi-probabilistic decomposition of ideal gate

[11] K. Temme, et al, "Error mitigation for short- depth quantum circuits", Physical review letters, 2017.

Quantum Error Mitigation

- **Quantum error correction** to fully compensate for quantum noise requires increasing the number of qubits beyond the current reach of quantum technology.
- **Quantum error mitigation** trades space (qubits) with time, running multiple noisy circuits to emulate a noiseless one.[11]

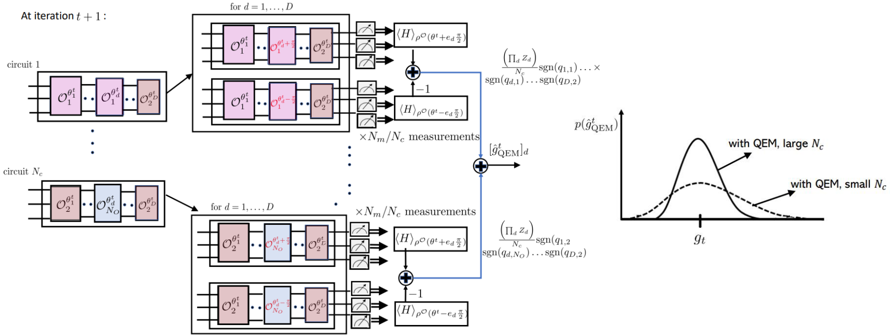


Quasi-probabilistic decomposition of ideal gate

[11] K. Temme, et al, "Error mitigation for short- depth quantum circuits", Physical review letters, 2017.

Training with Quantum Error Mitigation

- Quantum error mitigation removes the bias, but increases the variance.



Training with Quantum Error Mitigation

- Number of iterations required to ensure an error floor δ for a fixed number of measurements N_m per iteration (with respect to the solution for the noiseless circuit):[12]

Schemes	**Iteration Complexity**
shot-noise only	$\tilde{\mathcal{O}}\left(\log\frac{1}{\delta}+\frac{V}{\mu\delta}\right)$
shot and gate noise	$\tilde{\mathcal{O}}\left(\log\frac{1}{\delta-B^{\mathcal{E}}\mu}+\frac{V^{\mathcal{E}}}{\mu\delta}\right)$
shot and gate noise with QEM	$\tilde{\mathcal{O}}\left(\log\frac{1}{\delta}+\frac{V^{\mathrm{QEM}}}{\mu\delta}\right)$

Parameters	**Scaling**
variance V	$\mathcal{O}(D/N_m)$
bias $B^{\mathcal{E}}$	$\mathcal{O}(D\gamma)$
variance $V^{\mathcal{E}}$	$\mathcal{O}(Dc(\gamma)/N_m)$
variance V^{QEM}	$\mathcal{O}(c_1(\gamma)D/N_m)+\mathcal{O}(c_2(\gamma)D/N_c)$

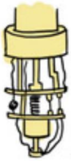
[12] S. T. Jose and O. Simeone, "Error Mitigation-Aided Optimization of Parameterized Quantum Circuits," in preparation.

Conclusions

Conclusions

- Quantum machine learning is an emerging paradigm suited for NISQ computers.
- **Classical-quantum machine learning** may be of more immediate relevance for engineering applications, particularly when implemented using hybrid quantum-classical models...

	Fault-tolerant	Near-term
# qubits	millions	10-1000
errors	corrected	mitigated
use	Shor, Grover, HHL...	variational circuits
research	computational complexity	run it and see
available	in 5-30 years?	now



[M. Schuld '21]

Conclusions

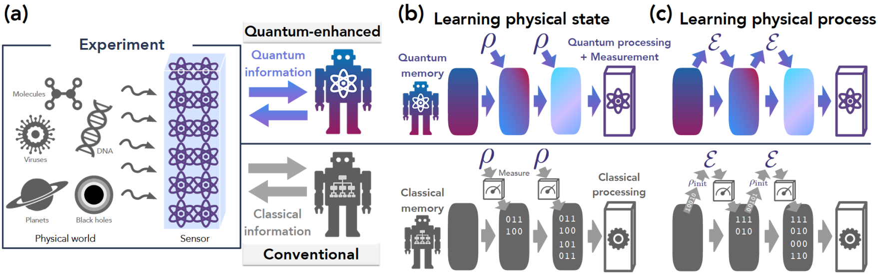
- ...although one should be aware of the **differences** between tasks suitable for classical and quantum machine learning.

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 and Killoran '22]

Conclusions

- In the long run, **quantum-quantum machine learning** applications to science and engineering may prove more impactful.



[Huang et al '22]

For More...

- O. Simeone, An Introduction to Quantum Machine Learning for Engineers, Foundations and Trends in Signal Processing, 2022, https://arxiv.org/abs/2205.09510