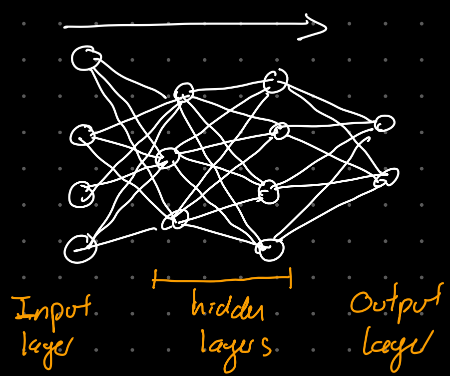
Class 5: PWF 2021 - QML 12/03/2021

QNN = parametrized quantum circuit

As such, it needs an ansatz?

★ Dissipative QNN   Beer Nat. Comm 11, 808 (2020) arxiv 1902.10445

It naturally generalizes the famous, and widely used feed-forward neural network.



O = neuron

— = perception

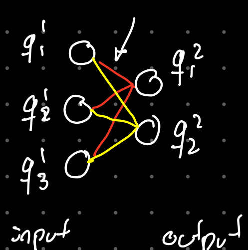
layer-to-layer information transmission.

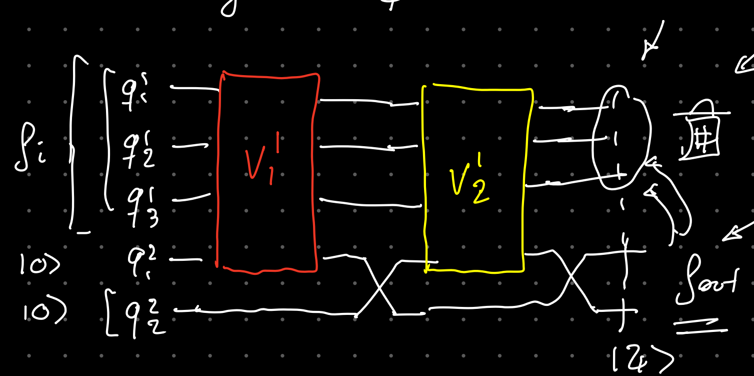
We are going to generalize to QML

O = neuron → qubit

— = perception → unitaries acting on the qubits

Sharma arxiv 2005.12458 always has barren plateaus. maybe they can be avoided





$$\rho_{out} = \mathrm{Tr}_{in,hid}\left[U(\vec{\theta}) \left(\rho^{in} \otimes |\vec{0}\rangle\langle\vec{0}|_{hid,out} \right) U^{\dagger}(\vec{\theta}) \right]$$

PRO: Can carry out universal QC

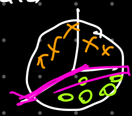
CON: creates big entangled states & information is stored in non-local quantum correlation. Requires a lot of qubits.

☆ Quantum Convolutional Neural Network

Cong Nature Phys 15, 1273 (2019)
arxiv 1810.03787

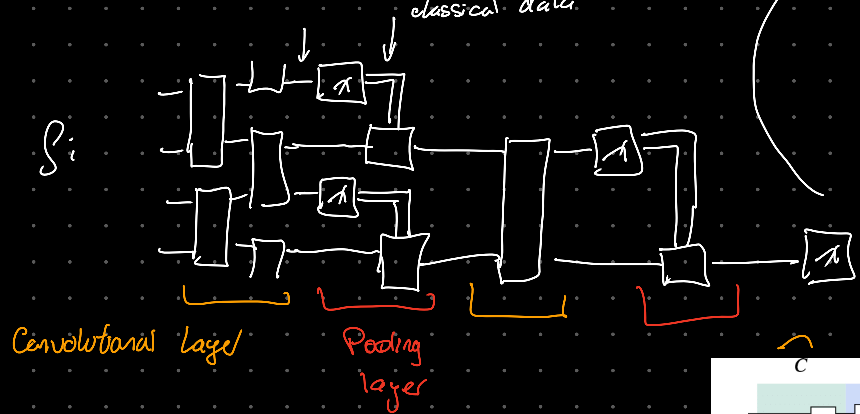
The idea is to reduce the dimension of the data & keep the relevant features.

→ $\langle z \rangle$



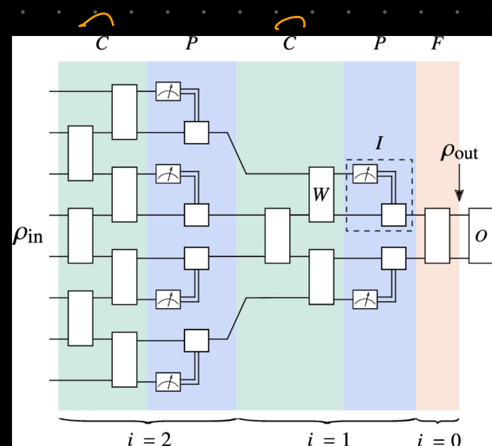
given an n-qubit quantum state

classical data



Convolutional layers: parametrized gates acting on the system

Pooling: reduce the dimension of dat



PRO: • shown to work for classification tasks →

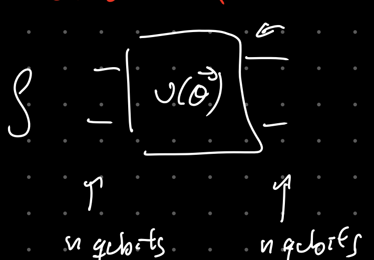
CON • shallow $O(\log(n))$

not obvious how general its applicability is.

Big PRO → NEVER HAS A BARREN PLATEAU

Pesah PRX 11, 041011 (2021)
arxiv 2011.02966

☆ Parametrized Quantum Circuit QNN



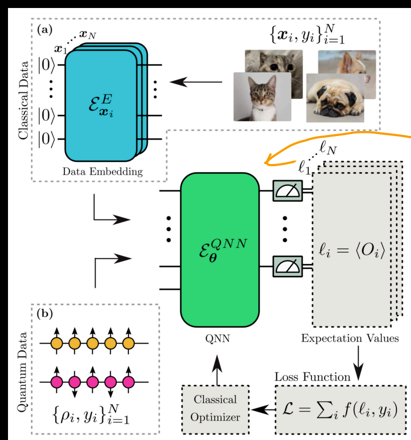
PRO: maybe more general than QCNN

CON: → Not obvious what the ansatz should be

– hardware efficient ansatz

Benchmarking a QML model

Ansats →



Ansatz

← choice

↑ choice

Benchmarking should be done with Quantum Data.

Quantum Data:

- outputs of some physical process
- synthetic quantum data: well characterized, easy to prepare, Good quality

Phase I   Phase II

$\xrightarrow[g_c]{}$ $H(g)$

1) VQE   randomly pick a field g

$|0\rangle \ldots |0\rangle$ — $U(\vec{\theta})$

$\langle H(g) \rangle = C(\vec{\theta})$

$|\psi(\vec{\theta})\rangle$ = ground state of $H(g)$

if $g > g_c$ we assign label $1 = g_i$

" $g < g_c$ " " " $-1 = g_i$

once we finish VQE $\{|\psi_i\rangle, g_i\}$

for a dataset of N points we need to run N VQE experiments.

2) We train a generative model to create the data set

we take the data preparation task and we apply the QML learning methodology

Train a QML model $\longrightarrow$ generates data on demand

Schatzki arxiv 2109.03400

we measure a quantity on $|\psi^{out}\rangle$ that characterizes the dataset

- level of entanglement in $|\psi^{out}\rangle$
- Order parameter (superconducting gap)

Framework Let $\mathcal{P}$ be a probability distribution of Quantum States

sample $|\psi^{in}\rangle \sim \mathcal{P}$

$$\mathop{E}_{|\psi^{in}\rangle \sim \mathcal{P}} \left[\mathcal{M}[|\psi^{out}\rangle] \right] = \eta$$

we can use a loss: $\mathcal{L}_{MSE}(\vec{\theta}) = \frac{1}{|S|} \sum_{i=1}^{|S|} \left(\mathcal{M}[|\psi^{out}\rangle] - \eta \right)^2$

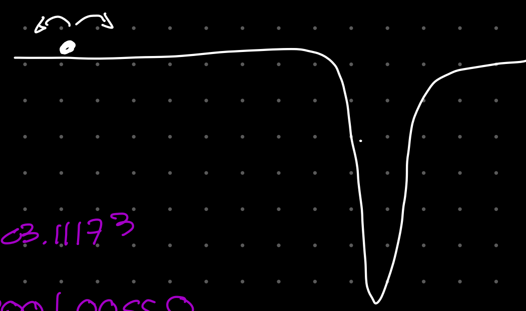
success if $\mathcal{M}[|\psi^{out}\rangle] \in [\eta - \delta, \eta + \delta]$

For binary classification: we have 2 optimization.

Challenges for near-term Quantum Computing

Barren Plateau:

- The cost or loss function landscape becomes, in average, exponentially flat with the problem size.



- The partial derivatives of the cost/loss vanish, in average, exponentially with the problem size.

McClean Nat Com 9, 4812 (2018) arxiv 1803.11173

Cerezo Nat Com 12, 1791 (2021) arxiv 2001.00550

Arrasmith arxiv 2104.05868

$$\rightarrow \underbrace{\Delta C}_{C(\theta_A)-C(\theta_B)} \in O\left(\tfrac{1}{2^n}\right) \qquad \frac{\partial C}{\partial \theta_a} \in \left(\frac{1}{2^n}\right)$$

Statistical Error $O\left(\frac{1}{\sqrt{N}}\right)$ ← shots

$n = 4$ $\qquad \Delta C \sim 0{,}0625 \rightarrow N = 256$ shots

$n = 20$ $\qquad \Delta C \sim 9{,}5 \cdot 10^{-7} \rightarrow N = 1{,}1 \cdot 10^{12}$ shots

A TRILLION

$$\frac{1}{\sqrt{N}} \sim \frac{1}{2^n} \rightarrow \boxed{N = 2^{2n}}$$

if we have a Barren Plateau, we have no hope of a Quantum Advantage

Mathematical definition of Barren Plateau

$C(\vec{\theta})$ has a Barren Plateau in parameter θ_a if

$$\mathrm{Var}_{\vec{\theta}}\left[\frac{\partial C(\vec{\theta})}{\partial \theta_a}\right] \leq F(n) \qquad F(n) \in O\left(\frac{1}{b^n}\right) \quad b > 1$$

Chebychev's Inequality

$$P[|X - \langle X \rangle| \geq c] \leq \frac{Var[X]}{c^2} \qquad \left\langle \frac{\partial C(\vec{\theta})}{\partial \theta_\mu} \right\rangle = 0$$

$$P\left(\left| \frac{\partial C(\vec{\theta})}{\partial \theta_\mu} \right| \geq c \right) \leq \frac{Var\left[\frac{\partial C(\vec{\theta})}{\partial \theta_\mu} \right]}{c^2} \leq \frac{F(n)}{c^2}$$

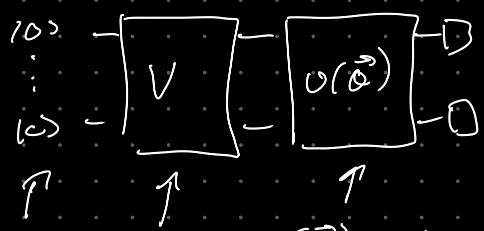
Alternative definition $\quad Var_{\theta_A}[C(\vec{\theta}_B) - C(\vec{\theta}_A)] \leq F(n)$

Narrow gorges: valley that contains the minimum

The volume of parameters that lead to the valley, vanishes exponentially.

Example of a cost that has a Barren Plateau:

Consider a problem of fixed state compilation:



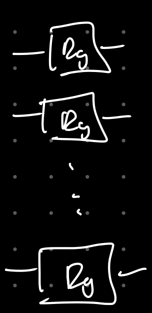
we want $U(\vec{\theta})|\vec{0}\rangle$ to be the same as $V|\vec{0}\rangle$

$$C(\vec{\theta}) = 1 - Tr[|\vec{0}\rangle\langle\vec{0}| \, U(\theta) V |\vec{0}\rangle\langle\vec{0}| V^\dagger U^\dagger(\vec{\theta})]$$

$$= 1 - p(0 \dots 0)$$

$V = \mathbb{1}$

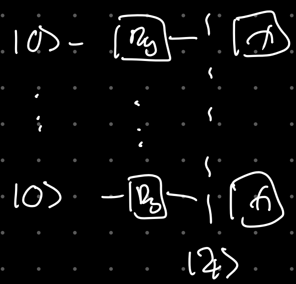
Toy model $\quad U(\vec{\theta})$



we know that

$\theta_i = 0 \ \forall i$

$C(\vec{\theta}) = 0$



$|0\rangle_i - \boxed{R_y} - |\psi_i\rangle = \cos\left(\frac{\theta_i}{2}\right)|0\rangle + \sin\left(\frac{\theta_i}{2}\right)|1\rangle$

$|\psi\rangle = \bigotimes_{i=1}^{n} |\psi_i\rangle = \bigotimes_{i=1}^{n}\left(\underline{\cos\left(\frac{\theta_i}{2}\right)}|0_i\rangle + \sin\left(\frac{\theta_i}{2}\right)|1_i\rangle\right)$

$\Rightarrow \; C(\vec{\theta}) = 1 - |\langle \vec{0}|\psi\rangle|^2$ → compares 2 objects living in an exponentially large Hilbert-Space

$$\boxed{C(\vec{\theta}) = 1 - \prod_{j=1}^{n} \cos^2\left(\frac{\theta_i}{2}\right)}$$

for visualizato purpose $\theta_i = \theta_1$ if $i \leq \frac{n}{2}$

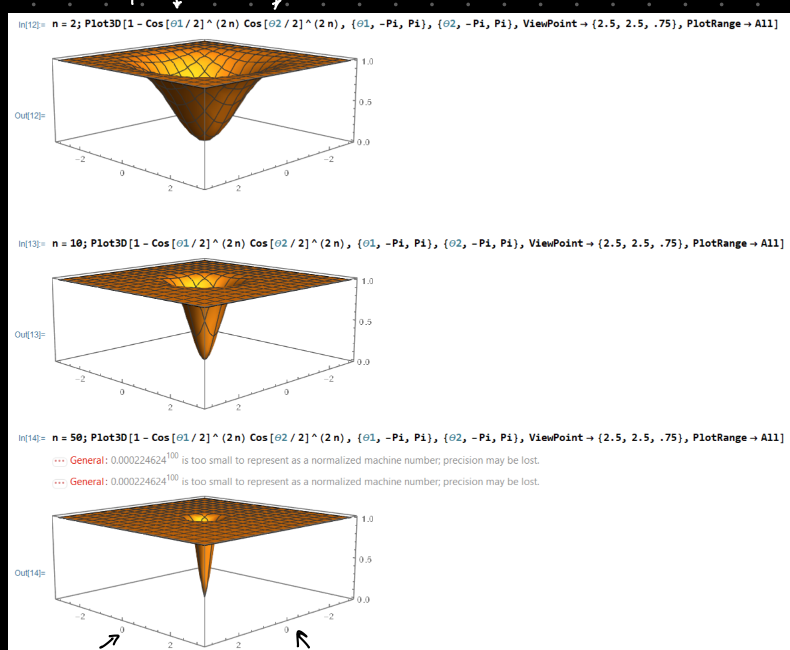
$\theta_i = \theta_2$ if $i > \frac{n}{2}$

1- ↓ ↓

4

20

100



$$\frac{\partial C(\vec{\theta})}{\partial \theta_k} = -\cos\left(\frac{\theta_k}{2}\right)\sin\left(\frac{\theta_k}{2}\right)\prod_{j \neq k} \underline{\underline{\cos^2\left(\frac{\theta_j}{2}\right)}}$$

$$\left\langle \frac{\partial C(\vec{\theta})}{\partial \theta_k} \right\rangle_{\vec{\theta}} = 0$$

$$\langle \; \rangle_{\vec{\theta}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta_1 \; \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta_2 \ldots \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta_n$$

$$\left\langle \frac{\partial C(\vec{\theta})}{\partial \theta_k} \right\rangle = \left(\prod_{j \neq k} \frac{1}{2\pi} \int d\theta_j \cos^2\left(\frac{\theta_j}{2}\right) \right) \left(\frac{-1}{2\pi} \int \cos\left(\frac{\theta_k}{2}\right) \sin\left(\frac{\theta_k}{2}\right) d\theta_k \right)$$

$=0$

$$\mathrm{Var}[X] = \langle X^2 \rangle - \langle X \rangle^2$$

$$\mathrm{Var}\left[\frac{\partial C(\vec{\theta})}{\partial \theta_k}\right] = \left\langle \left(\frac{\partial C(\vec{\theta})}{\partial \theta_k}\right)^2 \right\rangle - \underbrace{\left\langle \frac{\partial C(\vec{\theta})}{\partial \theta_k} \right\rangle^2}_{=0}$$

$$\mathrm{Var}\left[\frac{\partial C(\vec{\theta})}{\partial \theta_k}\right] = \left(\prod_{j \neq k} \underbrace{\frac{1}{2\pi} \int d\theta_j \cos^4\left(\frac{\theta_j}{2}\right)}_{\frac{3}{8}} \right) \left(\underbrace{\frac{1}{2\pi} \int \cos^2\left(\frac{\theta_k}{2}\right) \sin^2\left(\frac{\theta_k}{2}\right) d\theta_k}_{1/8} \right)$$

$$= \left(\frac{3}{8}\right)^{n-1} \frac{1}{8}$$

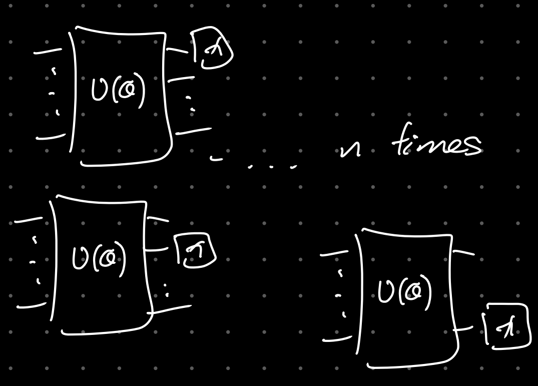
Barren Plateaus depend on the locality

Local cost function compare objects at the small problem size or even qubit size

$$C_L = 1 - \mathrm{Tr}[U(\vec{\theta})|\vec{0}\rangle\langle\vec{0}|U^\dagger(\theta)\, O_L]$$

$$O_G = |\vec{0}\rangle\langle\vec{0}| \qquad = 1 - \frac{1}{n}\sum_{j=1}^{n} P_j(0)$$

$$O_L = \frac{1}{n}\sum_{j=1}^{n} |0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}}$$



$$\mathrm{Var}\left[\frac{\partial C_L(\vec{\theta})}{\partial \theta_k}\right] = \frac{1}{8n^2}$$

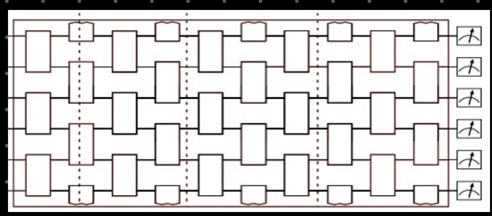
Not all cost functions are the same

Global ✗

Local ✓

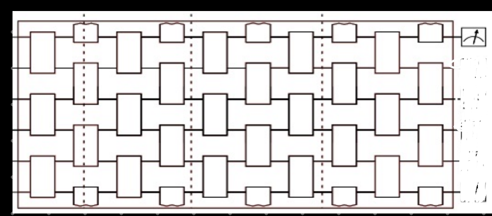
$A(U) \nearrow$ $\mathrm{Var}[\frac{\partial C}{\partial \theta}] \searrow$

Global



Barren Plateau

Local

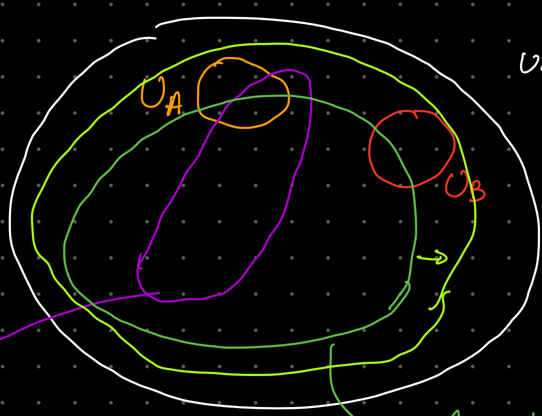


NO BP — transition — BP

log(n)

There multiple sources of Barren Plateaus

- Locallity
- Expressible



Unitary group space of degree d

$U(d)$

Higher volume in the group

= high expressivity

Accesible space by $U(\vec{\theta})$ by Problem

Accesible space by a problem-agnostic

Hardware efficient, have the property that their expressibility increases with the deph

Expressibility [non-mathematical definition] : a measure of how uniformly the circuit creates unitaries over the unitary group

A circuit with expressibility $A(U)$ is small

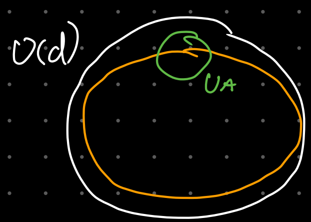
$A(U)$ = distance to the uniform distribution of unitaries (Haar measure)

Holmes arxiv 2101.02138 (2021)

$$\mathrm{Var}\left[\frac{\partial C(\vec{\theta})}{\partial \theta_u}\right] \leq A(U) \cdot c$$

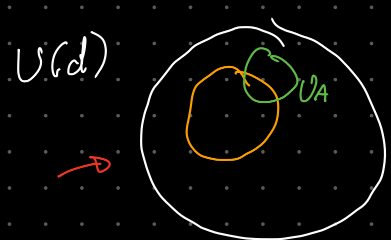
The more expressible the circuit, the smaller the gradients, and thus the more shots we need to train.

Highly expressible circuits



will have Barren plateaus

Low expressibility circuits



might not have a Barren Plateau

does not guarantee absence of Barren plateau (e.g. global cost function)

Larocca arxiv 2105.14377

↳ Problem inspired ansatz (QAOA, UCC, HVA, QOC...) are not immune to barren plateaus.

The way to analyze if an architecture has barren plateaus is by a principled analysis of computing

$$\mathrm{Var}\left[\frac{\partial C(\theta)}{\partial \theta_n}\right]$$

Other sources: Entanglement

Marrero arXiv 2010.15968

Patti: PRR 3,033090 (2021) arXiv 2012.12658

If $U(\vec{\theta})$ generates too much entanglement between "visible" qubits (that we measure) and "hidden" qubits, the due to the quantum correlations, the reduced state of the visible qubits is exponentially close to being maximally mixed

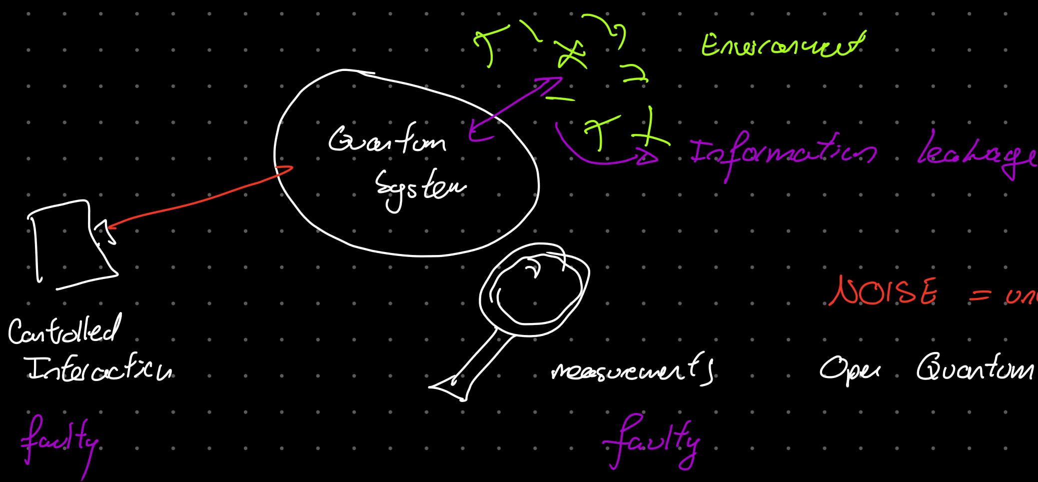
$$\rho_{visible} \sim \frac{\mathbb{1}}{2^{n_{visible}}}$$

Causes of Barren Plateaus:

- Globality
- Expressiveness
- Entangling

Quantum Noise

So far we have considered near-term QC for closed quantum systems that is, system that suffer no unwanted interactions with the environment.
In reality, there is no way to perfectly isolate quantum systems (HARD)
→ unwanted interactions of the environments



NOISE = unwanted interactions

Open Quantum System

$$|\psi\rangle \longrightarrow \rho$$

state are pure in a closed system — state is now mixed

Quantum Operator Formalism — Nielsen & Chuang

Describes transformations between quantum states

$$\rho \to \rho' = \mathcal{E}(\rho)$$

$\mathcal{E}$ = quantum operation
quantum channel

$\mathcal{E} : \mathcal{H}_A \longmapsto \mathcal{H}_B$

$$\mathcal{E}_U(\rho) = U\rho U^\dagger$$

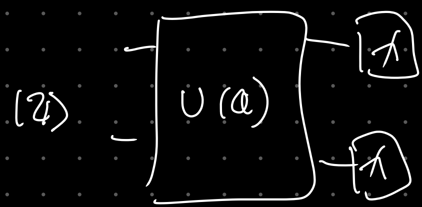
$$\mathcal{E}_m(\rho) = M_m \rho M_m$$

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger$$

E_k = Krauss operation

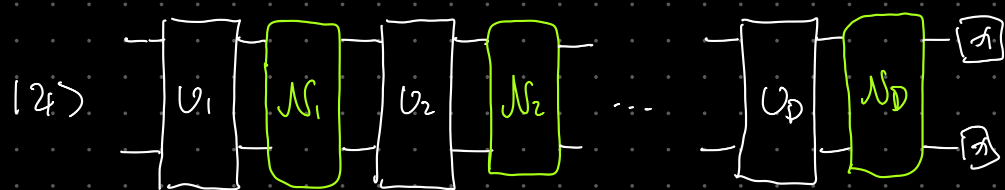
Global depolarizing channel

$$\mathcal{E}(\rho) = p\rho + (1-p)\frac{\mathbb{1}}{2^n}$$



we model noise as a noisy channel acting after every set of parallel gates.

$$U(\theta) = \prod_{i=1}^{D} U_i(\theta_i)$$



we can model the action of noise in the circuit

$\mathcal{N}$ = global depolarizing

$$\rho^{out} = U_D \cdots U_2 U_1 |\psi\rangle\langle\psi| U_1^\dagger U_2^\dagger \cdots U_D^\dagger$$

$$\tilde{\rho}^{out} = \mathcal{N}[U_D \mathcal{N}[\cdots \mathcal{N}[U_2 \mathcal{N}[U_1 \rho U_1^\dagger] U_2^\dagger] \cdots] U_D^\dagger]$$

$$\boxed{\tilde{\rho}^{out} = p^D \rho^{out} + (1-p^D)\frac{\mathbb{1}}{2^n}}$$

$$C(\vec{\theta}) = \mathrm{Tr}[U|\psi\rangle\langle\psi|U^\dagger O]$$

the noisy cost $\quad \tilde{C}(\theta) = \mathrm{Tr}[\tilde{\rho}^{out} O]$

$$\tilde{C}(\vec{\theta}) = p^D \, \mathrm{Tr}[\underline{\rho^{out}} O] + (1-p^D)\, \mathrm{Tr}\left[\frac{O\mathbb{1}}{2^n}\right]$$

$$= \underline{\underline{p^D}}\, C(\vec{\theta}) + \underline{\frac{(1-p^D)}{2^n}\mathrm{Tr}[O]} \quad (*)$$

Descale factor

constant and independent of D

p^D is a number that decreases exponentially with the depth.

In the limit $D \to \infty$

$$\tilde{\rho}^{out} = \frac{\mathbb{1}}{2^n}, \quad \tilde{C}(\vec{\theta}) = \text{constant} \quad (*)$$

Wang Nat Com 12, 6961 (2021)

arXiv 2007.14384

Noise-induced Barren Plateau

Solutions:
- Build better devices with smaller noise-probability ($p \sim 1$)
- Use shallow circuits

$\underline{D \sim O(\log(n))}$ $\to$ the supression is p^D

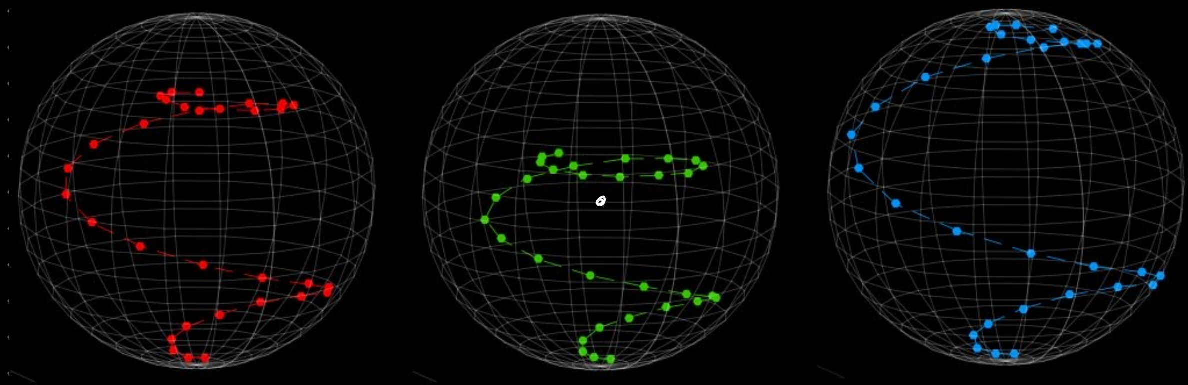
$p^{\log(n)} \sim \mathrm{poly}(n)$

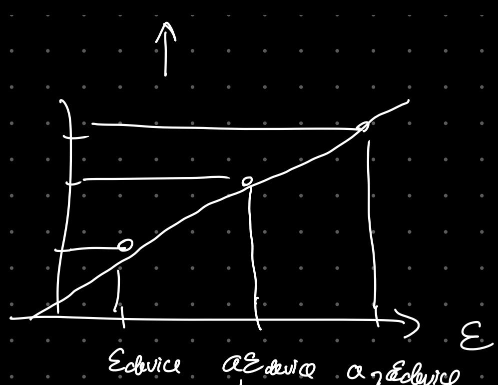
Error mitigation

How do we deal with noise, to recover noiseless expectation values.

Hope: we can train a VQA, QML with noise, and use error mitigation to recover correct and noiseless information

Noiseless $C(\vec{\theta})$ $\xrightarrow{\text{Noise}}$ Noisy Cost $\tilde{C}(\vec{\theta})$ $\xrightarrow[\text{mitigation}]{\text{error}}$ Recover $C_{EM}(\vec{\theta}) \sim C(\vec{\theta})$





Temme PRL 119, 180509 (2017) arxiv 1612.02058

Wang arxiv 2109.01051 (2021)

What are the Hopes for a Quantum Advantage

Huang PRL 126, 190505 (2021)

Huang Nat Com 12, 1 (2021)

Abbas Nat Comp Sci 1, 403 (2021)

Kubler arxiv 2106.03747

Liu Nat Phys 17, 1013 (2021)

Caro arxiv 2111.05292 (2021)

Huang arxiv 2112.00778