Approximate Gibbs state preparation with dynamic parameterized quantum circuits

Christa Zoufal
IBM Quantum

WQBM Seminar Series
December 8th, 2025

Dynamic parameterized quantum circuits: expressive and barren-plateau free

Abhinav Deshpande, Marcel Hinsche, Sona Najafi, Kunal Sharma, Ryan Sweke, Christa Zoufal

Classical optimization of parameterized quantum circuits is a widely studied methodology for the preparation of complex quantum states, as well as the solution of machine learning and optimization problems. However, it is well known that many proposed parameterized quantum circuit architectures suffer from drawbacks which limit their utility, such as their classical simulability or the hardness of optimization due to a problem known as "barren plateaus". We propose and study a class of dynamic parameterized quantum circuit architectures. These are parameterized circuits containing intermediate measurements and feedforward operations. In particular, we show that these architectures: 1. Provably do not suffer from barren plateaus. 2. Are expressive enough to describe arbitrarily deep unitary quantum circuits. 3. Are competitive with state of the art methods for the preparation of ground states and facilitate the representation of nontrivial thermal states. These features make the proposed architectures promising candidates for a variety of applications.

IBM

Collaborators

Abhinav Desphande

Marcel Hinsche

Sona Najafi

Kunal Sharma

Ryan Sweke

IBM Quantum

Christa Zoufal – ouf@zurich.ibm.com

Outlook

- Let’s explore dynamical ansatz classes
- Gibbs state preparation application
  - Infidelity matching
  - McLachlan’s variational principle

Barren Plateaus (BPs)

IBM **Quantum**

Exponentially vanishing gradients ↔ exponentially flat loss landscape

$$\mathbb{E}_\omega \left[\partial_\omega L\left(\omega\right)\right] = 0$$

$$\mathrm{Var}_\omega \left[\partial_\omega L\left(\omega\right)\right] \in \mathcal{O}\left(\frac{1}{b^n}\right),\ b > 1$$

n representing the number of qubits

Known causes

- Ansatz close to a t-design [1-3]
- Global observable [4-5]
- Extensive entanglement paired with partial traces [6]
- Particular noise models [7]

[1] J. McClean, et al., Nature Communications 9 (2018).
[2] Z. Holmes, et al., PRX Quantum 3 (2022).
[3] M. Larocca, et al., Quantum 6, 824 (2022).
[4] M. Cerezo, et al., Nature Communications 12 (2021).
[5] S. Thanasilp, et al., Quantum Machine Intelligence 5, 21 (2023).
[6] C. Ortiz Marrero, et al., PRX Quantum 2, 040316 (2021).
[7] S. Wang, et al., Nature Communications 12 (2021).

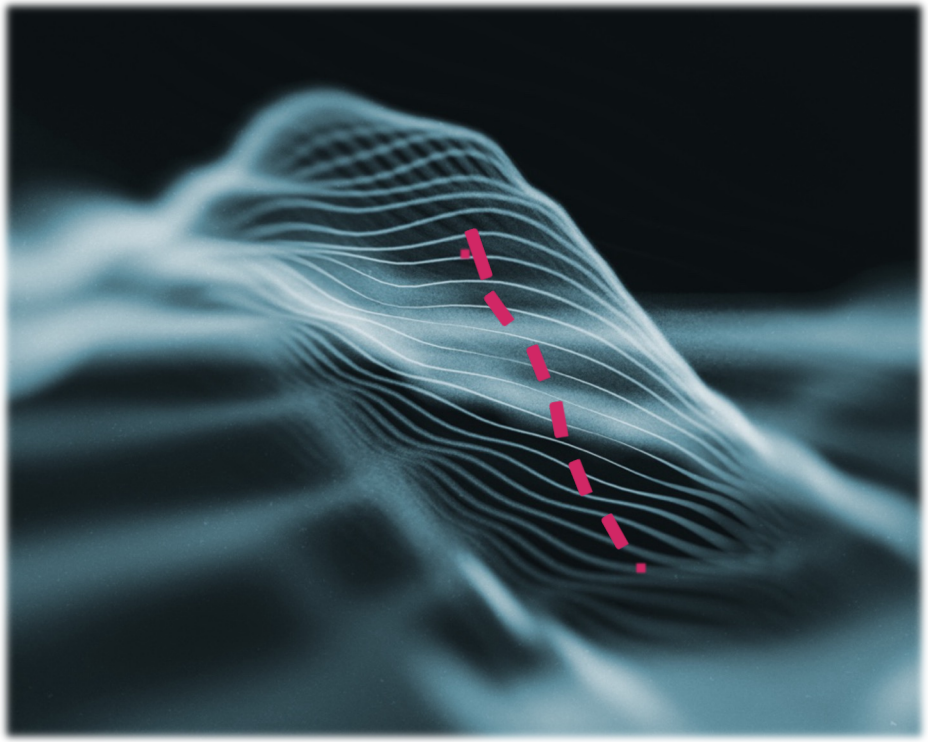
What are we looking for?

IBM **Quantum**

Trainable model with efficient access to non-exponentially vanishing gradients

Ansatz that is sufficiently **expressive**, i.e., capable of representing the target state

A system that is **not** trivially classically **simulable**



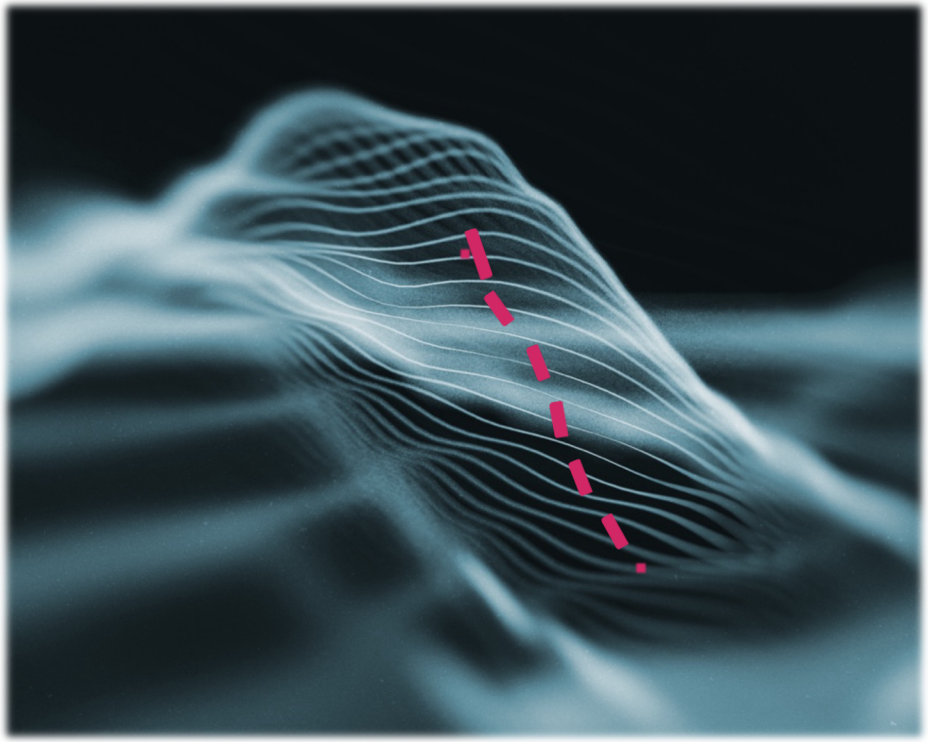
Our contribution

IBM **Quantum**

Our model provably does not suffer from barren plateaus given random initialisation

Numerical evidence that our model can reliably represent interesting states

Proven worst-case hardness but (full disclosure) the model is likely to be average-case easy



Note: We can smoothly transition between provably barren plateau-free and provably hard settings

Dynamic parameterized quantum circuit ansatz

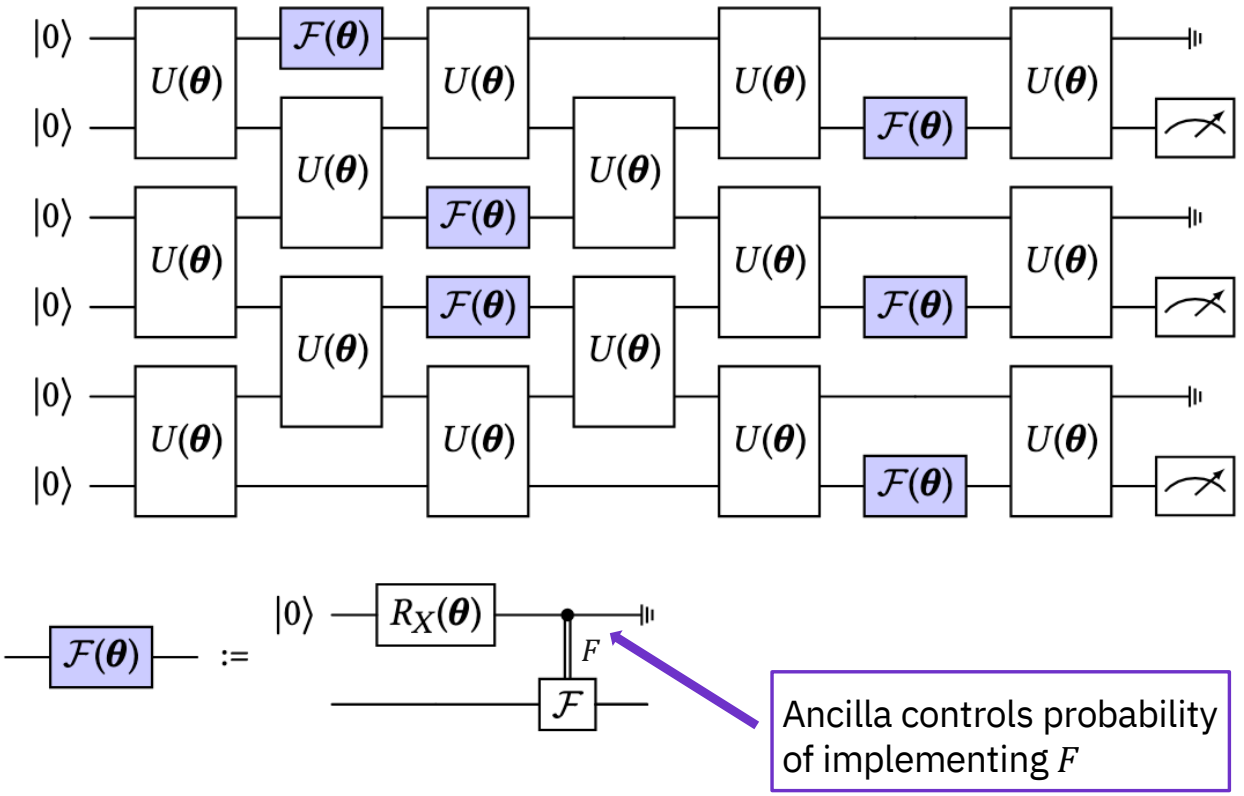
IBM **Quantum**

DPQC components

- Parameterized (or unparameterized) two-qubit unitary gates $U(\theta)$
- Parameterized no-nunitary single-qubit dynamic operations $\mathcal{F}(\theta)$

Probabilistic feedforward operation F
➔ Implemented via parameter θ controlled non-unitary action such as a reset

$$\mathcal{F}(\theta)(\cdot) = \cos^2\left(\frac{\theta}{2}\right)\mathbb{I}(\cdot) + \sin^2\left(\frac{\theta}{2}\right)F\,(\cdot)$$



Potential Applications

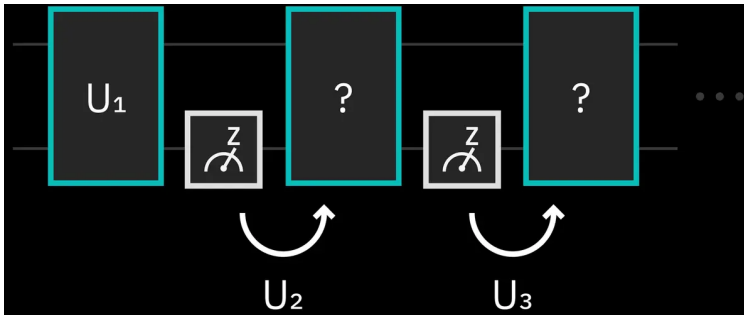
Ground state search:
Given that $\mathcal{S}_n$ represent the set of n −qubit quantum states and H an n −qubit Hamiltonian.

$$\min_{\rho \in \mathcal{S}_n} Tr[H\rho]$$

Thermal state preparation:
Given an n-qubit Hamiltonian H and an inverse temperature β. Find the state that minimizes the free energy, i.e., $\rho_{Gibbs} = \frac{e^{-\beta H}}{Tr[e^{-\beta H}]}$.

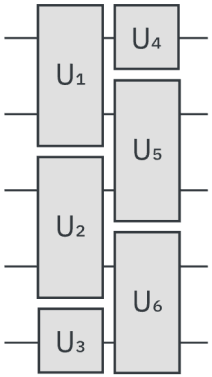
Quantum process learning:
Let $\mathfrak{I}_n$ be the set of all n-qubit channels and $d_\diamondsuit$ a distance induced by the diamond norm. Then, we are looking for an accessible channel $T' \in \mathfrak{I}_n$ which minimizes $d_\diamondsuit(T, T')$ the to a target channel $T \in \mathfrak{I}_n$.

What are Dynamic Circuits?



Regular quantum circuit

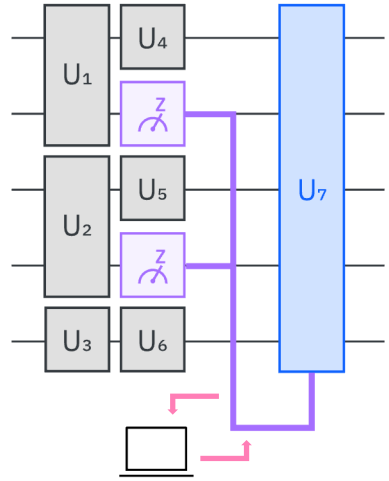
- Unitary gates
- Final measurements



Dynamic circuit

Additional features:

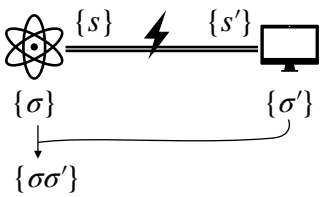
- Mid-circuit measurements
- Real-time classical computing
- Feedforward



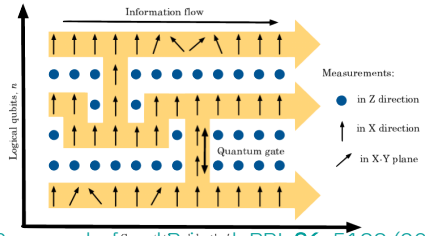
What can we do with dynamic circuits?

Dynamic circuits can be vastly more powerful than regular circuits.
Applications that require or can benefit from dynamic circuits:

Quantum Error Correction

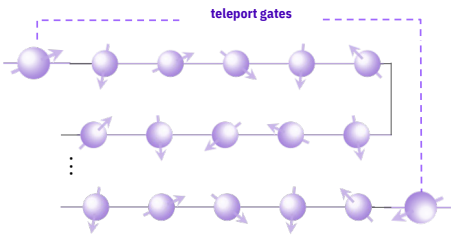


Measurement-based Quantum Computing

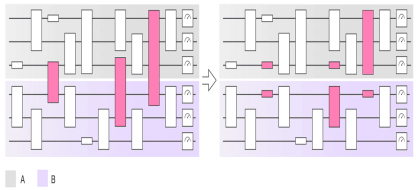


[Raussendorf and Briegel, PRL **86**, 5188 (2001)]

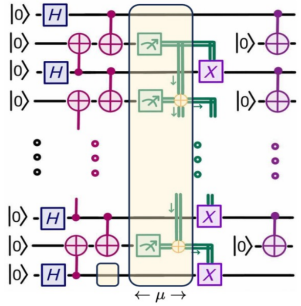
Entanglement Routing



Circuit Knitting

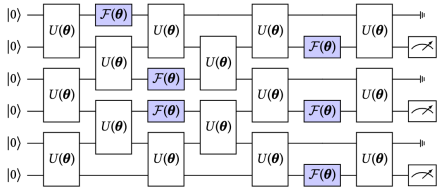


Preparation of GHZ and other topologically interesting states



[Bäumer et al., PRX Quantum **5**, 030339 (2024)]

Variational Ansatz



Relation to existing non-unitary architectures

IBM **Quantum**

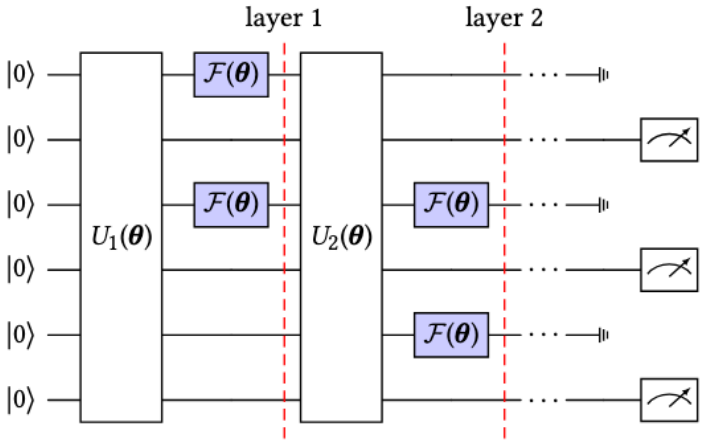
Quantum convolutional neural networks (QCNNs)[1]:
Provable absence of barren plateaus[2]
Based on specific non-unitary operation ➔ a particular subset of the DPQC architectures (classically simulable)

Dissipative quantum neural networks[3]:
Special cases of this form coincide with special cases of DPQC architectures (control over dissipative dynamics?)

PQCs subject to non-unital noise[4]:
Non-unital noise channels can be viewed as specific instances of the non-unitary operations allowed for in DPQC architectures (lacking control over density, type, and application probability)

➔ *Interpolation* between *expressive* and BP-free case non-trivial or even impossible

Note: Similar DPQC architectures involving mid-circuit measurements explored in terms of noise resilience (numerically)[5] and in terms of their ability to use mid-circuit measurements for shallow state preparation[6]



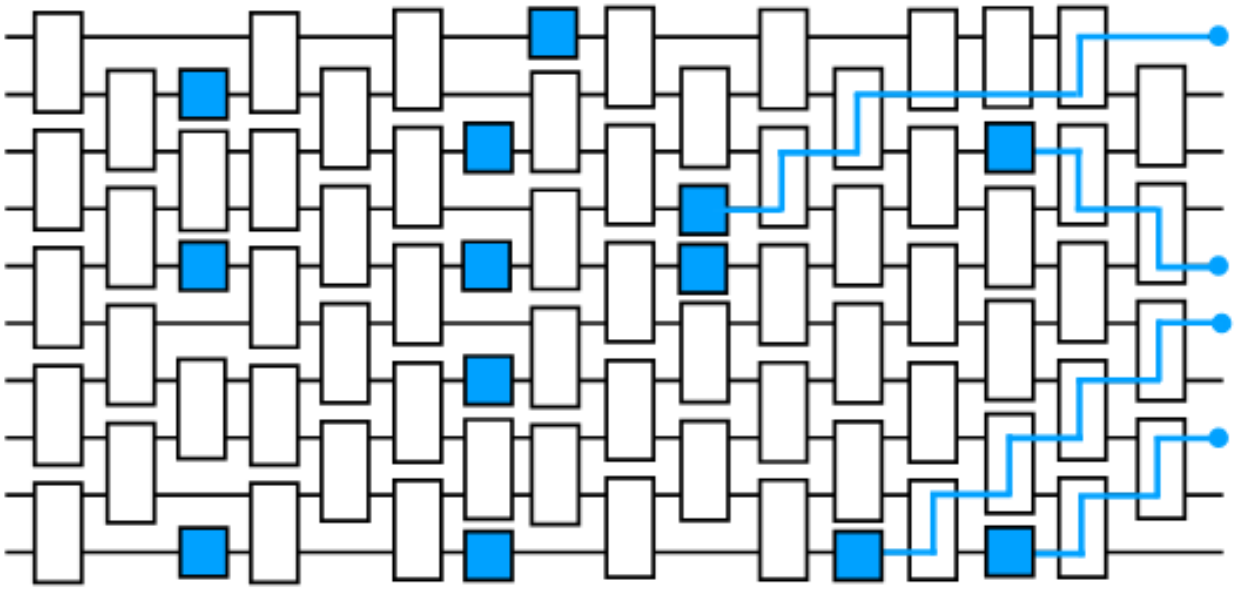
[1] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. "Quantum Convolutional Neural Networks". Nature Physics 15.12 (2019)
[2] Arthur Pesah,et al. "Absence of Barren Plateaus in Quantum Convolutional Neural Networks". Physical Review X 11.4 (2021)
[3] Kerstin Beer, et al. "Training Deep Quantum Neural Networks". Nature Communications 11.1 (2020)
[4] Antonio Anna Mele, et al. "Noise-Induced Shallow Circuits and Absence of Barren Plateaus". arXiv:2403.13927 (2024)
[5] Yigal Ilin and Itai Arad. "Dissipative Variational Quantum Algorithms for Gibbs State Preparation". arXiv:2407.09635 (2024)
[6] Yuxuan Yan, et al. "Variational LOCC-assisted Quantum Circuits for Long-Range Entangled States". arXiv:2409.07281 (2024

Gradient Variance

Feedforward distance

Definition:
Shortest paths from a qubit measurement to a feedforward operation through the backwards light cone of the measurement

The feedforward distance f describes the maximum length of paths, over all qubits on which the observable is supported.



Gradient variance

Ansatz: DPQC $C(\boldsymbol{\theta})$ with random initialization, i.e., the parameters of the two-qubit gates $U(\theta)$ are drawn from a locally scrambling ensemble (also assuming Haar-random two-qubit gates)

Initial state: $|0\rangle\langle 0|^{\otimes n}$

Loss function: $\mathrm{L}(\boldsymbol{\theta}) = \mathrm{Tr}\left[HC(\boldsymbol{\theta})(|0\rangle\langle 0|^{\otimes n})\right]$ with H corresponding to a k-local Hamiltonian

Theorem 1 (Absence of barren plateaus in DPQCs for k-local Hamiltonians – informal)

$$Var_{\boldsymbol{\theta}}\, L(\boldsymbol{\theta}) \geq \left(\frac{1}{5}\right)^{k(f+1)} \|H\|_{HS}^2$$

with $\|.\|_{HS}^2$ representing the Hilbert-Schmidt norm and f the feedforward distance explained before.

Noise-robustness of gradient variance

IBM **Quantum**

Ansatz: DPQC $C(\boldsymbol{\theta})$ with random initialization, i.e., the parameters of the two-qubit gates $U(\theta)$ are drawn from a locally scrambling ensemble (also assuming Haar-random two-qubit gates)

Initial state: $|0\rangle\langle 0|^{\otimes n}$

Noise: assuming a unital single-qubit noise channel after every operation introducing an average infidelity of $\frac{\gamma}{2}$ $\rightarrow \tilde{C}(\boldsymbol{\theta})(\,.\,)$

Loss function: $\mathrm{L}(\boldsymbol{\theta}) = \mathrm{Tr}\left[HC(\boldsymbol{\theta})(|0\rangle\langle 0|^{\otimes n})\right]$ with H corresponding to a k-local Hamiltonian

Theorem 2 (Noise robustness of Theorem 1 – informal)

$$Var_{\boldsymbol{\theta}}\, L(\boldsymbol{\theta}) \geq \left(\frac{(1-\gamma)^2}{5}\right)^{k(f+1)} \|H\|_{HS}^2$$

with $\|.\|_{HS}^2$ representing the Hilbert-Schmidt norm and f the feedforward distance explained before.

Notes

The absence of BPs does not imply trainability!

→ It is not directly given that the optimal or even a good solution may be found, if gradients do not vanish exponentially.

The bounds are **independent** of the number of qubits $\boldsymbol{n}$ and the depth $\boldsymbol{d}$ of the circuit. This fact is what enables us to consider deep circuits of this form, restoring expressivity.

Proof idea of this result relies on the **stat-mech model**: Analysis of a biased random walk over a configuration space of identity and swap operators, interpreted as bitstrings $\{0, 1\}^n$.

→ Variance of the loss function is related to the probability of obtaining a significant number of 1s, or large Hamming weight in the output of the random walk.

Expressivity

Expressivity

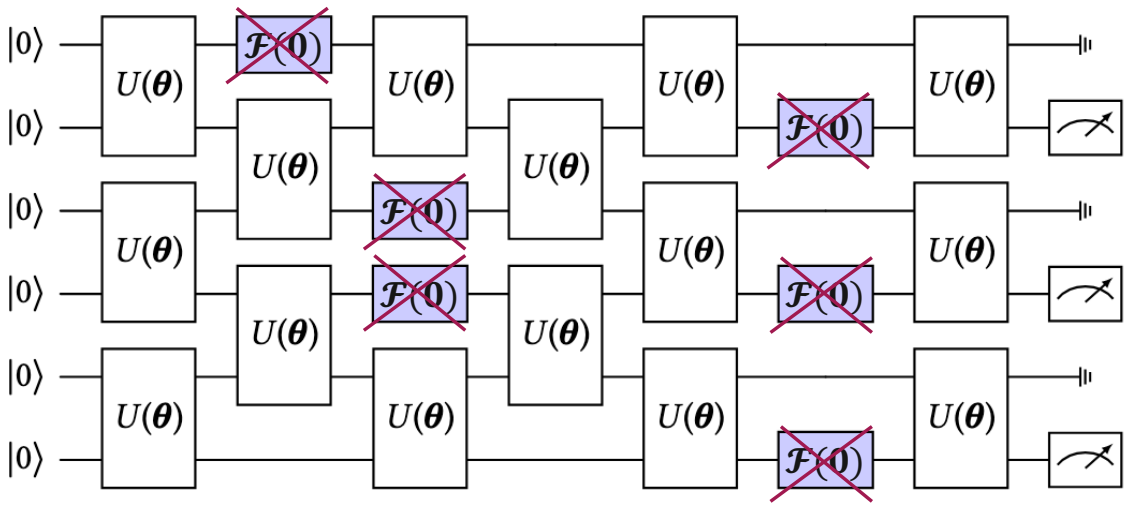
IBM **Quantum**

Observation (Expressivity of DPQC architectures with probabilistic feedforward—informal).

Setting all the circuit parameters that control the probability of implementing an F gate to 0 ➔ unitary ansatz

DPQC architecture $C(\boldsymbol{\theta})$ with connectivity graph G and depth d with all feedforward probabilities set to 0 ➔ architecture can realize all unitary operations of depth d on G.

It follows that one can in principle *prepare interesting pure states* using such an architecture.



Connecting expressivity and absence of BPs

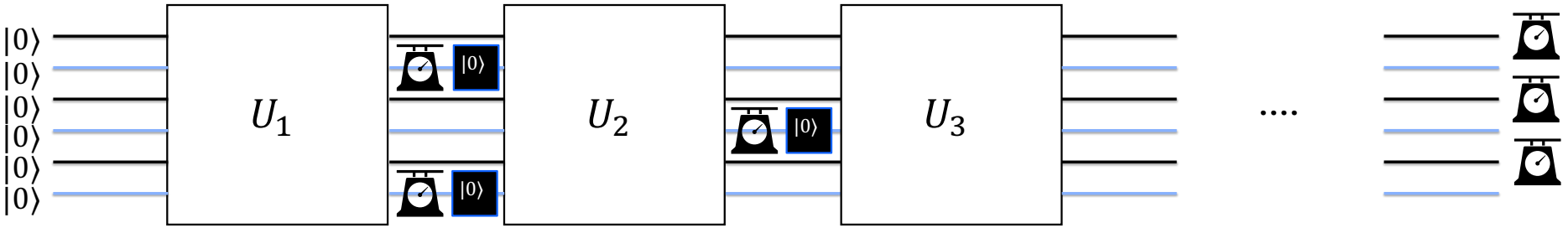
IBM **Quantum**

Observation (DPQC architectures: connecting expressivity and absence of BPs).

DPQC architectures allow one to interpolate smoothly between highly expressive unitary architectures and BP-free nonunitary architectures with a constant feedforward depth!

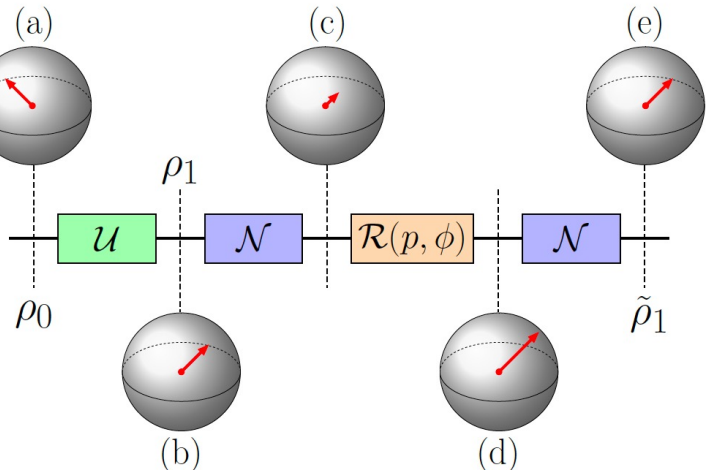
Hardness

Worst-case hardness

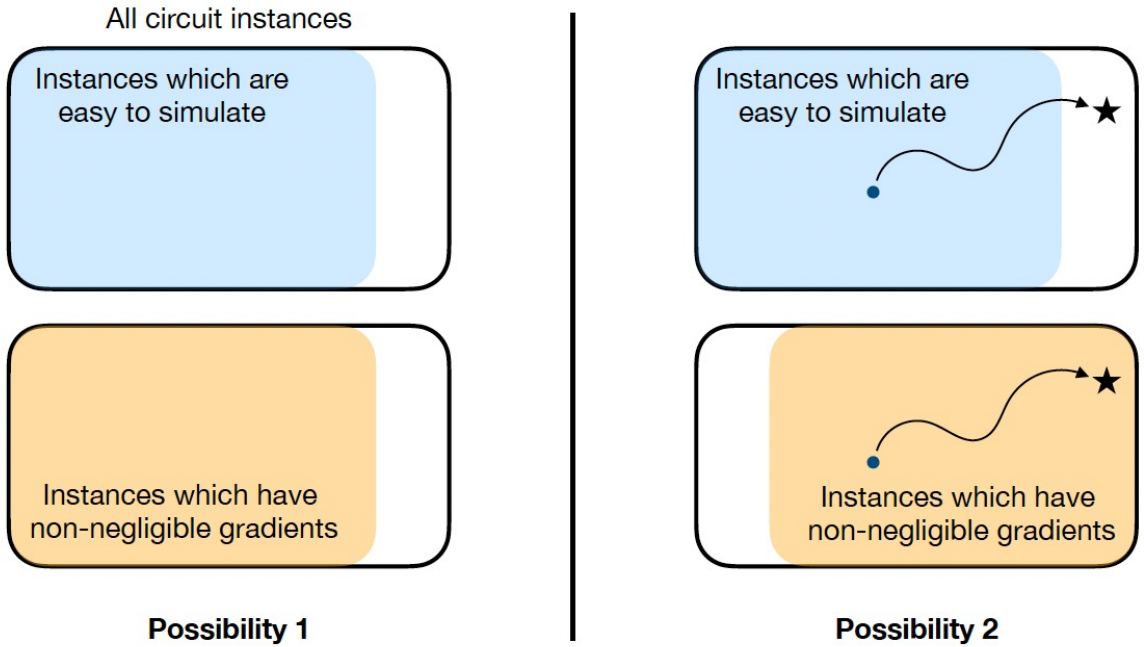


Avoid resets to have a deep, purely unitary circuit.

We can tune the parameters to increase the *number of acting resets* and, hence, *control the information flow*.



What about average-case hardness?



Possibility 1: classical computers are good enough to replicate all trainable variational quantum algorithms.

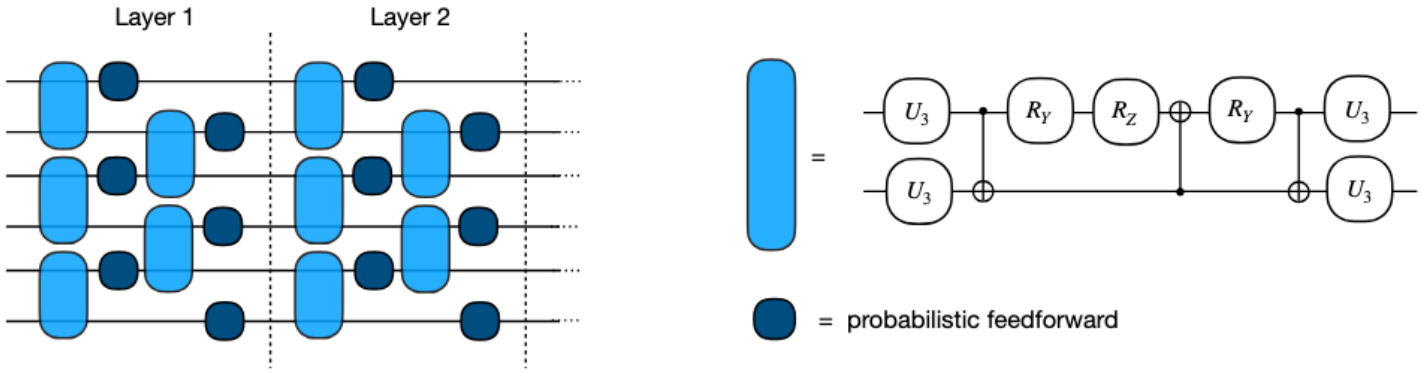
Possibility 2: quantum computers can genuinely help to reach parts of parameter or solution space that classical algorithms cannot access by themselves.

Numerical Simulation Results

Thermal state preparation

XY-Model

$$H_{TFI} = -\sum_{j=1}^{n} \frac{3}{4} X_j X_{j+1} + \frac{1}{4} Y_j Y_{j+1} - \frac{1}{2} \sum_{j=1}^{n} Z_j$$

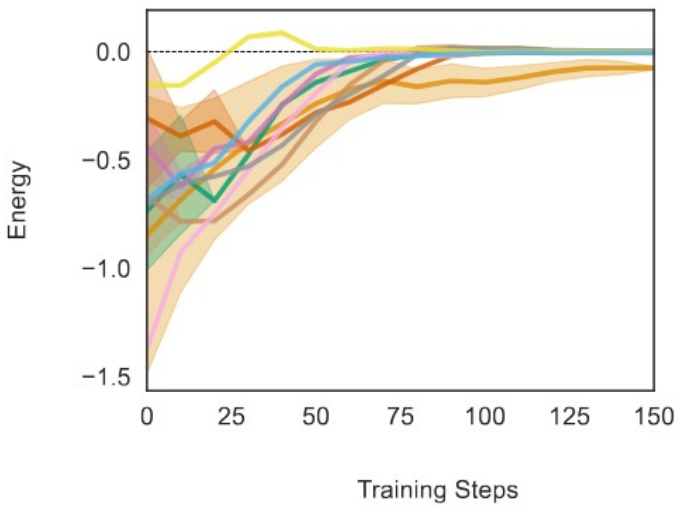


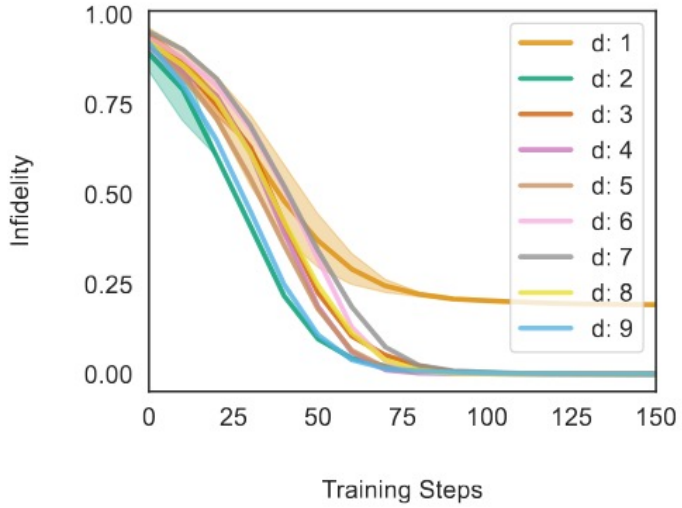
Model considered for a periodic 1D chain for up to 10 qubits and $\beta = 2$

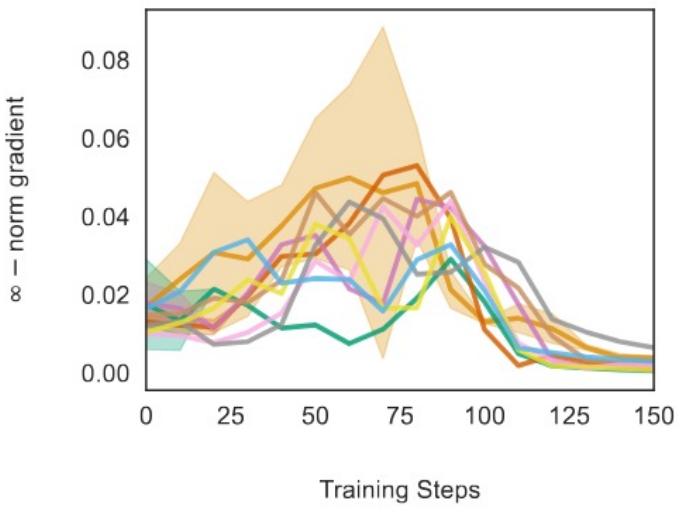
Probabilistic feedforward operations $\mathcal{F}(p)(\cdot) = \cos^2\left(\frac{p}{2}\right)(\cdot) + \sin^2\left(\frac{p}{2}\right) F\,(\cdot)$

Considered preparation protocols: fitting of infidelity to exact Gibbs state (not scalable), variational quantum imaginary time evolution (prone to instabilities)

Infidelity Training





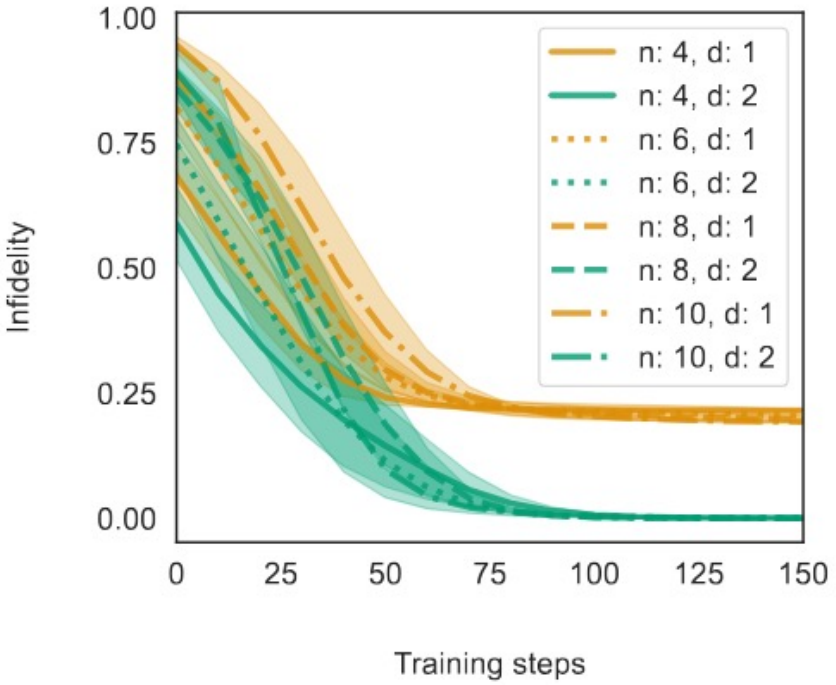


The setting does not suffice the assumptions necessary to guarantee our theoretical results!

→ Energy converges
→ Infidelities approach the order of 10^{-5}
→ Gradient norm does not decrease significantly for larger d

We know from the infidelity fitting that our ansatz is **sufficiently expressive** to represent the underlying thermal states!

What depth do we need?



➔ $d = 2$ seems to provide sufficient expressivity

Variational Quantum Time Evolution

$$H = \sum_i \theta_i h_i$$

Variational Ansatz

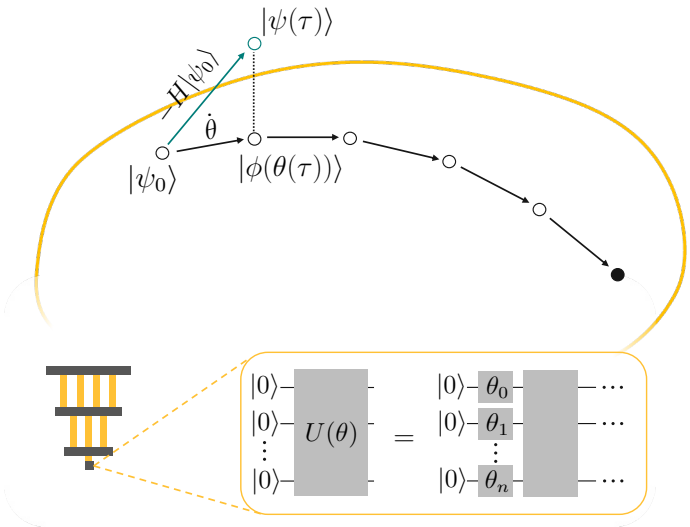
$$|\psi_{\boldsymbol{\omega}}(t)\rangle = U\big(\omega(t)\big)|0\rangle$$

Approach:

- State evolution → Parameter evolution with **McLachlan [1]**
- Minimize the error between the **variational trajectory** and the actual gradient using a **constant depth** Ansatz

Properties:

- Ensures that $\omega \in \mathbb{R}$
- In the case of real time evolution not necessarily energy preserving
- Unlike PFs: circuit depth does not (necessarily) increase with the number time-steps and the locality of the system



[1] A variational solution of the time-dependent Schrödinger equation, A. McLachlan

McLachlan's Variational Principle

$$H = \sum_i \theta_i h_i$$

Variational Ansatz

$$|\psi_{\boldsymbol{\omega}}(t)\rangle = U\big(\omega(t)\big)|0\rangle$$

Quantum Imaginary Time Evolution

➔ VarQITE

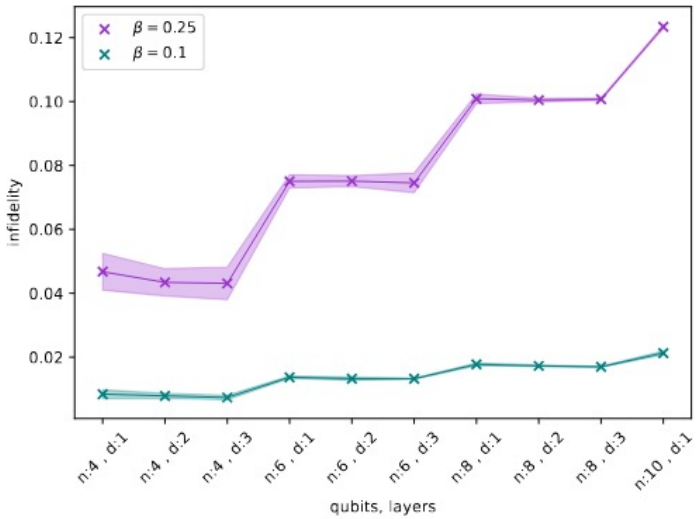
$$\delta \left\| \left(\frac{\partial}{\partial t} + H - E_t \right) |\psi_{\boldsymbol{\omega}}(t)\rangle \right\|_2 = 0$$

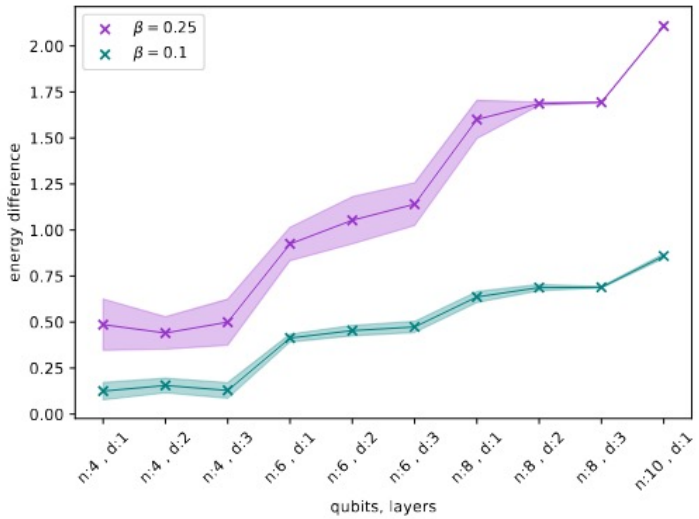
$$\mathrm{Re}\left(\frac{\partial \langle \psi_{\boldsymbol{\omega}}(t)|}{\partial \omega_i} \frac{\partial |\psi_{\boldsymbol{\omega}}(t)\rangle}{\partial \omega_j} - \frac{\partial \langle \psi_{\boldsymbol{\omega}}(t)|}{\partial \omega_i} |\psi_{\boldsymbol{\omega}}(t)\rangle\langle\psi(t)| \frac{\partial |\psi_{\boldsymbol{\omega}}(t)\rangle}{\partial \omega_j} \right) \dot{\omega}_j = -\mathrm{Re}\left(\frac{\partial \langle \psi_{\boldsymbol{\omega}}(t)|}{\partial \omega_i} H |\psi_{\boldsymbol{\omega}}(t)\rangle \right)$$

Quantum Geometric Tensor (QGT)
prop. to the Quantum Fisher Information (QFI)

$$\longrightarrow \quad F^Q_{ij} \dot{\omega}_j = -\mathrm{Re}(C_i) \quad \propto \frac{\partial \langle E_{\boldsymbol{\omega}}(t) \rangle}{\partial \omega_i}$$

Variational Quantum Imaginary Time Evolution





n ∈ {4, 6, 8, 10} and 10 random seeds per setting

→ Infidelities between the order of 10^{-2} and 10^{-1}
→ Energy match is not particularly good
→ We observe better training behavior for larger temperatures. This is aligned with the expectation that Gibbs states are more difficult to prepare for lower temperatures

Caveats and criticisms

1. **Small-scale experiments:**

   Number of system qubits at most 12 ➔ exponentially small gradient might not be very small

2. **No shot noise**

   We use tensor network simulator to simplify the evaluation of expectation values ➔ no shot noise in the numerics

3. **No hardware noise:**

   Numerical simulations are not affected by the noise typically present in actual quantum hardware. Robustness against hardware-induced noise unclear

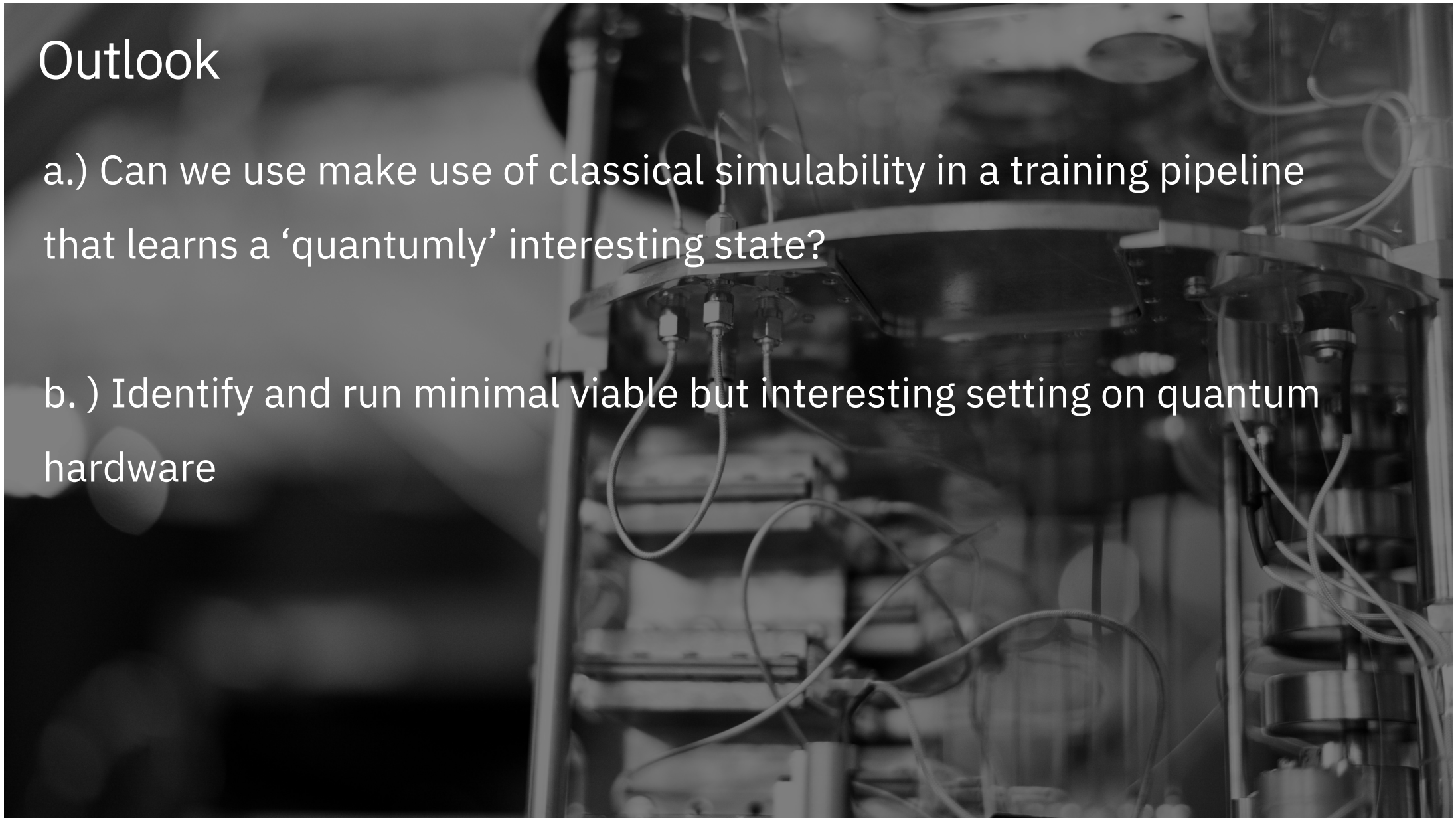
4. **Scalability vs. performance trade-off in thermal state preparation:**

   Thermal state preparations are either based on a cost function that requires the evaluation of the fidelity (which does not scale) or an ODE-based approach (which does not perform as well as the infidelity matching --> can further engineering help?)

Conclusion

IBM **Quantum**

Trainable model with efficient access to non-exponentially vanishing gradients	→	**BP-free** for randomly initialized parameter setting
Ansatz that is sufficiently **expressive**, i.e., capable of representing the target state	→	Numerical indication of **capability** to represent interesting ground and thermal states
A system that is **not** trivially classically **simulable**	→	We know that the model can represent **non-trivial pure** quantum states



Outlook

a.) Can we use make use of classical simulability in a training pipeline that learns a ‘quantumly’ interesting state?

b.) Identify and run minimal viable but interesting setting on quantum hardware

IBM **Quantum**

ouf@zurich.ibm.com

Stat-mech mapping

Background: the stat-mech mapping

Assumption: circuit is composed of some entangling gates (either fixed or random) interspersed with random single-qubit gates.

We average over the choice of random single-qubit gates and study statistical properties with respect to these averages, denoted $\mathbb{E}_{\theta}[]$.

First-moment quantities $\mathbb{E}_{\theta}\mathrm{Tr}[\rho(\theta)O]$ can be studied by evaluating $\mathbb{E}_{\theta}[\rho(\theta)] = \frac{I}{2^n}$.

Next interesting quantities are second-moment ones of the form $\mathbb{E}_{\theta}\mathrm{Tr}[\rho(\theta)O]^2$

Handle these via doubled Hilbert space: $\mathbb{E}_{\theta}\mathrm{Tr}[\rho(\theta)O]^2 =$

$\mathrm{Tr}[\mathbb{E}_{\theta}[\rho(\theta)O \otimes \rho(\theta)O]]$

$= \mathrm{Tr}[\mathbb{E}_{\theta}[\rho(\theta) \otimes \rho(\theta)] \cdot O \otimes O]]$

It suffices to keep track of the evolution of

$\mathbb{E}_{\theta}[\rho(\theta) \otimes \rho(\theta)] =: \overline{\rho}$

When averaging the single-qubit gates over a 2-design, it is easy to keep track of this object and derive simple rules describing how it evolves.

Example: Single-qubit state ρ.

Performing the single-qubit average

$\mathbb{E}_U[(U \otimes U)\rho^{\otimes 2}(U^\dagger \otimes U^\dagger)]$ ➔ $\overline{\rho} = aI + bS$

where I = identity operator on doubled Hilbert space (dimension 4) and

$S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, the SWAP operator.

Determine a and b through $\mathrm{Tr}\overline{\rho} = 4a + 2b = \mathrm{Tr}\rho^{\otimes 2}$ and

$$\mathrm{Tr}\overline{\rho}S = 2a + 4b = \mathrm{Tr}[\rho^{\otimes 2}S] = \mathrm{Tr}[\rho^2]$$

For pure states, this results in $\overline{\rho} = \frac{2}{3}\cdot\frac{I}{4} + \frac{1}{3}\cdot\frac{S}{2}$.

Background: the stat-mech mapping

Assumption: circuit is composed of some entangling gates (either fixed or random) interspersed with random single-qubit gates.

We average over the choice of random single-qubit gates and study statistical properties with respect to these averages, denoted $\mathbb{E}_{\vec{\theta}}[]$.

First-moment quantities $\mathbb{E}_{\vec{\theta}}\mathrm{Tr}[\rho(\theta)O]$ can be studied by evaluating $\mathbb{E}_{\vec{\theta}}[\rho(\theta)] = \frac{I}{2^n}$.

Next interesting quantities are second-moment ones of the form $\mathbb{E}_{\vec{\theta}}\mathrm{Tr}[\rho(\theta)O]^2$

Handle these via doubled Hilbert space: $\mathbb{E}_{\vec{\theta}}\mathrm{Tr}[\rho(\theta)O]^2 =$

$\mathrm{Tr}[\mathbb{E}_{\vec{\theta}}[\rho(\theta)O \otimes \rho(\theta)O]]$

$= \mathrm{Tr}[\mathbb{E}_{\vec{\theta}}[\rho(\theta) \otimes \rho(\theta)] \cdot O \otimes O]]$

We want to keep track of the evolution of

$\mathbb{E}_{\vec{\theta}}[\rho(\theta) \otimes \rho(\theta)] =: \overline{\rho}$

When averaging the single-qubit gates over a 2-design, it is easy to keep track of this object and derive simple rules describing how it evolves.

Example: Single-qubit state ρ.
Performing the single-qubit average

$$\mathbb{E}_U[(U \otimes U)\rho^{\otimes 2}(U^\dagger \otimes U^\dagger)] \rightarrow \overline{\rho} = aI + bS$$

where I = identity operator on doubled Hilbert space (dimension 4) and

$S = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, the SWAP operator.

Stat-mech mapping

For any operation one can apply, if the operation is sandwiched by single-qubit random gates, then the two-copy state always starts off and ends in the symmetric subspace spanned by $\{I, S\}^n$.

Two copy state on n qubits may always be written as

$$\bar{\rho} = \sum_{x\in\{0,1\}^n} c_x \left(\frac{I}{4}\right)^{1-x_1} \cdot \left(\frac{S}{2}\right)^{x_1} \otimes \ldots \otimes \left(\frac{I}{4}\right)^{1-x_n} \cdot \left(\frac{S}{2}\right)^{x_n}$$
$$= \sum_{x\in\{0,1\}^n} c_x \mathtt{I}^{1-x_1} \cdot \mathtt{S}^{x_1} \otimes \ldots \otimes \mathtt{I}^{1-x_n} \cdot \mathtt{S}^{x_n}, \text{ in terms of the}$$

trace-1 operators $\mathtt{I} = \frac{I}{4}$ and $\mathtt{S} = \frac{S}{2}$.

Interpret the above as a (quasi)probability distribution over "words" (since $\mathrm{Tr}\bar{\rho} = \sum c_x = 1$).

We can also take partial traces of these quantities and they behave just as expected.

When applying an operation on, say, qubits i and j, it suffices to examine how the reduced density matrix on those two qubits evolves:

$$\mathrm{Tr}_{\{i,j\}^c}\bar{\rho}_{\mathrm{in}} = a\mathtt{I}_i\mathtt{I}_j + b\mathtt{I}_i\mathtt{S}_j + c\mathtt{S}_i\mathtt{I}_j + d\mathtt{S}_i\mathtt{S}_j \rightarrow$$
$$\mathrm{Tr}_{\{i,j\}^c}\bar{\rho}_{\mathrm{out}} = a'\mathtt{I}_i\mathtt{I}_j + b'\mathtt{I}_i\mathtt{S}_j + c'\mathtt{S}_i\mathtt{I}_j + d'\mathtt{S}_i\mathtt{S}_j.$$

$\begin{pmatrix} a' \\ b' \\ c' \\ d' \end{pmatrix} = T \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}$ for a *transfer matrix* T whose rows add to $(1\ldots1)$.

$$T_{\mathrm{Haar}} = \begin{pmatrix} 1 & \frac{4}{5} & \frac{4}{5} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{1}{5} & 1 \end{pmatrix} \quad T_{\mathrm{CZ}} = \begin{pmatrix} 1 & \frac{8}{9} & \frac{8}{9} & 0 \\ 0 & \frac{1}{9} & \frac{-2}{9} & 0 \\ 0 & \frac{-2}{9} & \frac{1}{9} & 0 \\ 0 & \frac{2}{9} & \frac{2}{9} & 1 \end{pmatrix} \quad T_{\mathrm{single-qubit}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

References: [Hunter-Jones arXiv: 1905.12053]
[Dalzell et al., arXiv: 2011.12277] and [arXiv: 2111.14907]
[Napp arXiv:2203.06174]
[Ware et al. arXiv:2305.04954]

IBM

Barren plateaus from the stat-mech model

Let's understand the physics of the stat-mech model

Initial state $\bar{\rho} = \left(\frac{2}{3}\mathrm{I} + \frac{1}{3}\mathrm{S}\right)^{\otimes n}$

Local update: $T_{\mathrm{Haar}} = \begin{pmatrix} 1 & \frac{4}{5} & \frac{4}{5} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{1}{5} & 1 \end{pmatrix}$

Variance of nonidentity Pauli observable supported on subset J is proportional to the probability of ending in a string with all S in J.

Expressivity-induced barren plateaus: For a sufficiently well-connected architecture and deep unitary random circuits, the steady state is

$\bar{\rho}_s = \frac{2^n}{2^n+1}\mathrm{II}\ldots\mathrm{I} + \frac{1}{2^n+1}\mathrm{SS}\ldots\mathrm{S}$ (same as performing a global Haar-average on an arbitrary pure state)

$\Longrightarrow$ for deep random circuits, both local and global cost functions have barren plateaus.

Noise-induced barren plateaus: in the presence of local unital noise, stat-mech model has an additional update rule: $\mathrm{S} \to (1-\gamma)\mathrm{S} + \gamma\mathrm{I}$.
$\Longrightarrow$ noise-induced barren plateaus as long as depth is $\Omega(\mathrm{poly}(n))$.

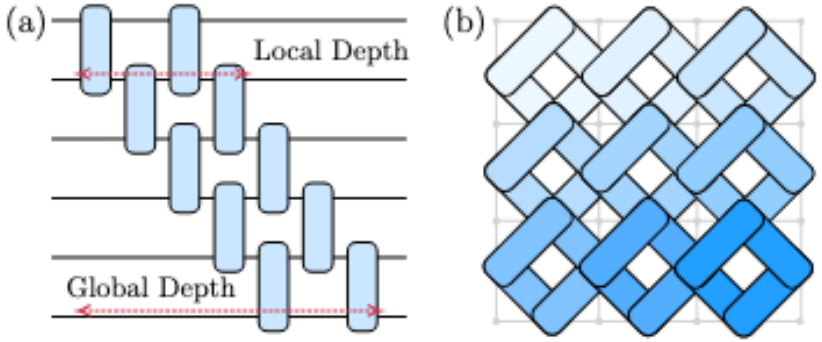
Entanglement-induced barren plateaus: the weight of an S operator is precisely proportional to the average local purity of the resulting state, which in turn is related to the Rényi-2 entropy.

IBM

[Zhang et al., PRL (2024)]

Absence of barren plateaus in finite local-depth circuits with long-range entanglement

https://arxiv.org/pdf/2311.01393



(c)

	FDC	FLDC	GLDC
BP-free	✓	✓	×
Area law	✓	✓	×
LRE	×	✓	✓
Non-classicality	×	✓	✓

FIG. 1. (a) and (b) Typical examples of finite local-depth circuits (FLDC) on 1D and 2D lattices, respectively. Darker colors in (b) indicate later action orders. (c) Compares the class of finite depth circuit (FDC), FLDC, and general linear depth circuit (GLDC) in terms of whether they are in general free from barren plateaus (BP), preserve entanglement area law, generate long-range entanglement (LRE), and can be simulated efficiently to compute local observable expectations by known classical methods (classicality). The inclusion relation is FDC $\subset$ FLDC $\subset$ GLDC.