Estimates of loss function concentration in noisy parametrized quantum circuits

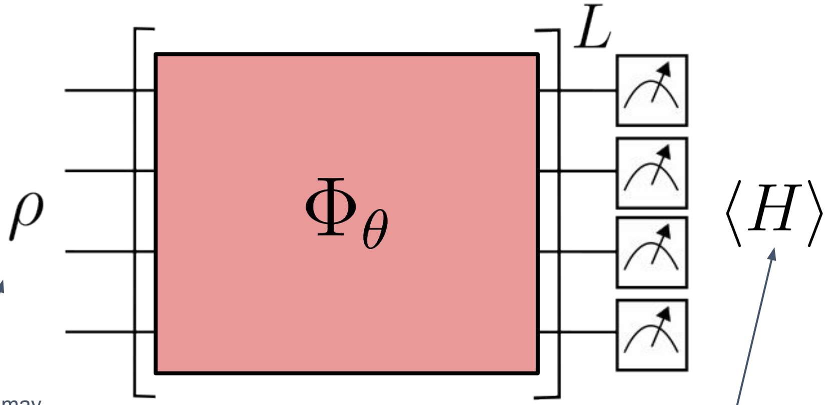
Giulio Crognaletti, Michele Grossi, Angelo Bassi

CERN | QUANTUM TECHNOLOGY INITIATIVE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

Variational Quantum Algorithms in a nutshell

Variational Quantum Algorithms are based on a **hybrid** quantum classical optimization scheme.



Initial state, may be a *reference state* or encode *quantum data*

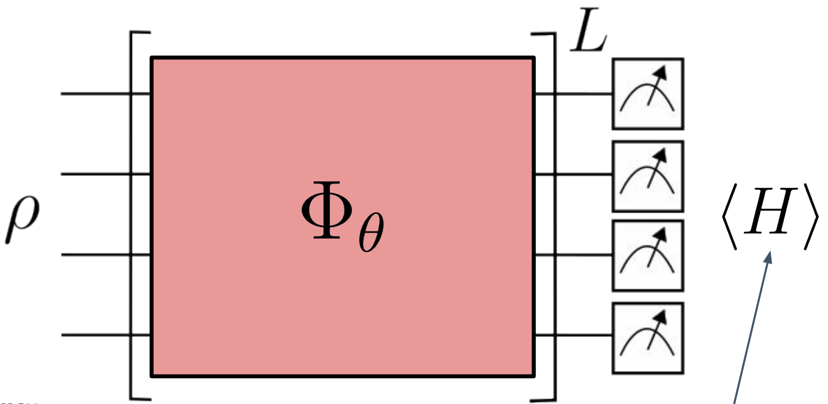
Parameterized quantum channel

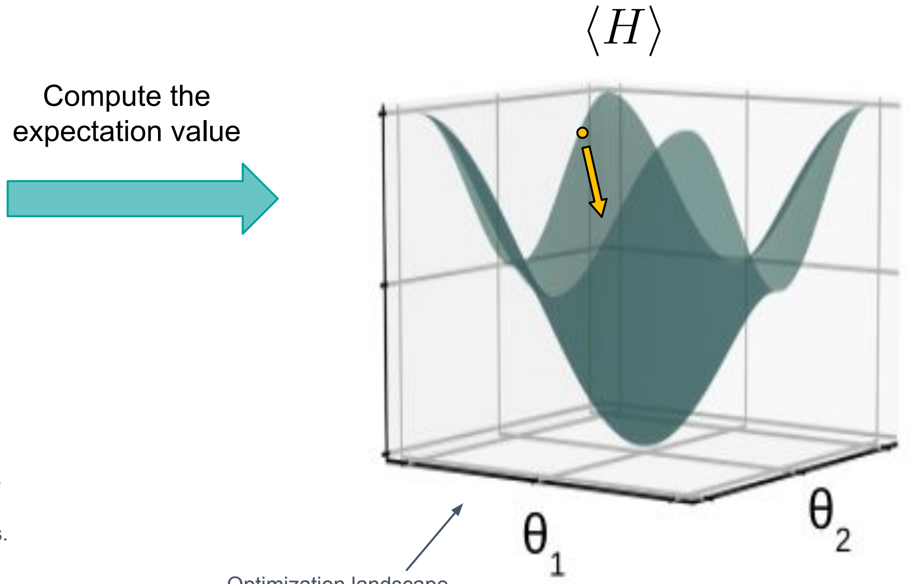
An observable, generally the Hamiltonian for applications to condensed matter physics.

(generally referred to as loss operator in quantum ML)

Variational Quantum Algorithms in a nutshell

Variational Quantum Algorithms are based on a **hybrid** quantum classical optimization scheme.





Initial state, may be a *reference state* or encode *quantum data*

Parameterized quantum channel

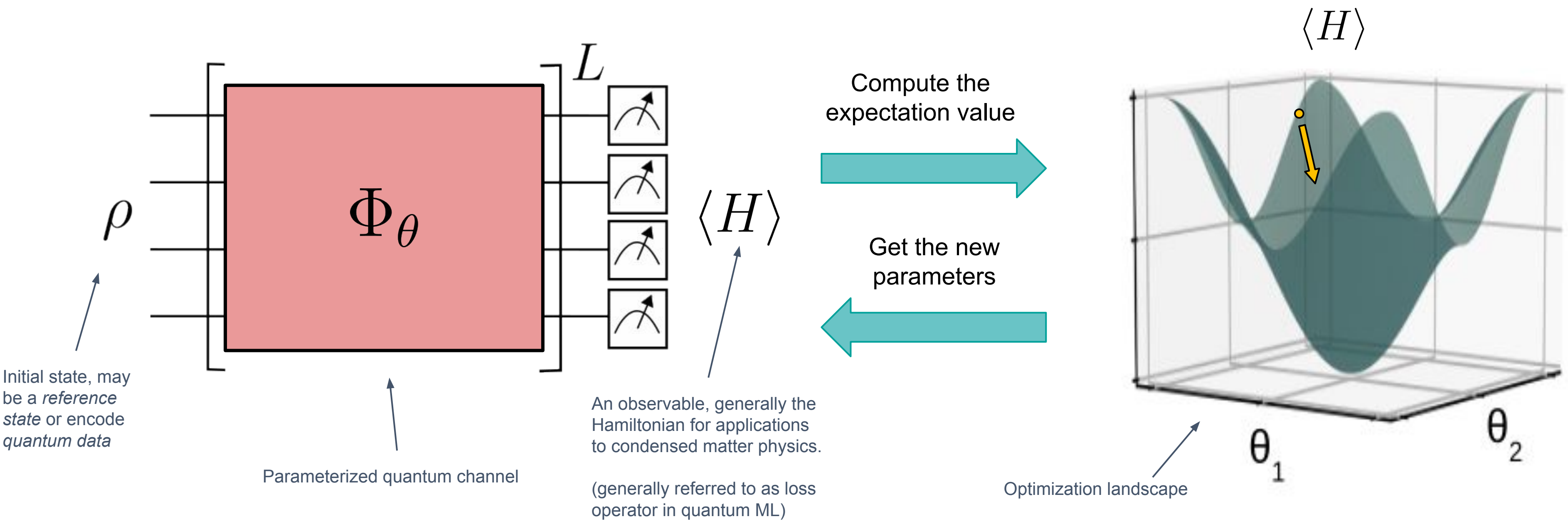
An observable, generally the Hamiltonian for applications to condensed matter physics.

(generally referred to as loss operator in quantum ML)

Optimization landscape

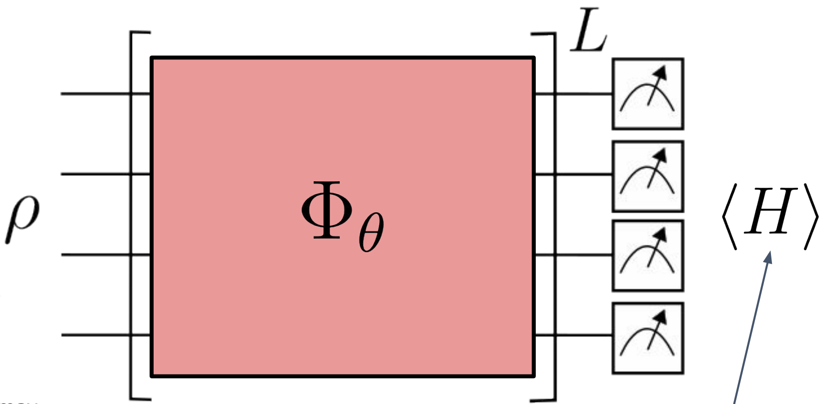
Variational Quantum Algorithms in a nutshell

Variational Quantum Algorithms are based on a **hybrid** quantum classical optimization scheme.



Variational Quantum Algorithms in a nutshell

Variational Quantum Algorithms are based on a **hybrid** quantum classical optimization scheme.



Initial state, may be a *reference state* or encode *quantum data*

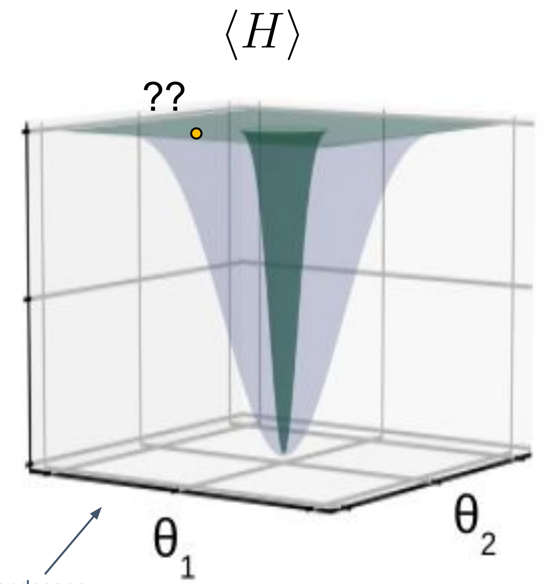
Parameterized quantum channel

An observable, generally the Hamiltonian for applications to condensed matter physics.

(generally referred to as loss operator in quantum ML)

Compute the expectation value

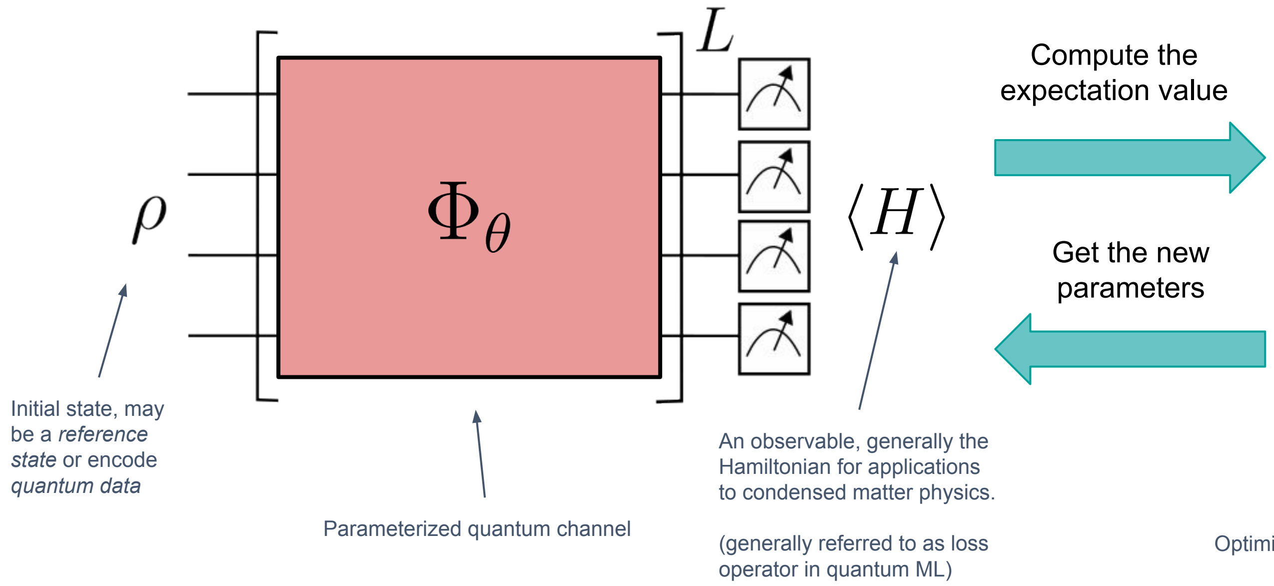
Get the new parameters

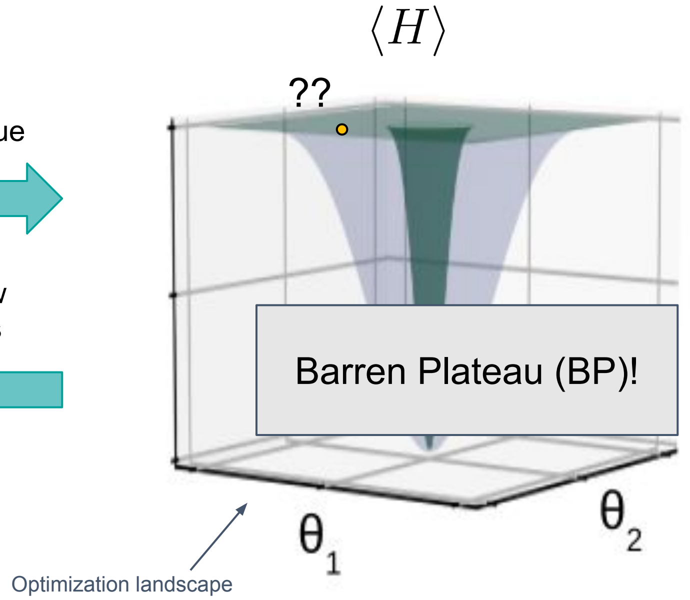


Optimization landscape

Variational Quantum Algorithms in a nutshell

Variational Quantum Algorithms are based on a **hybrid** quantum classical optimization scheme.





Barren Plateaus: a curse of dimensionality

The **exponential** concentration of the loss function, dubbed **barren plateau**, is a major **threat** to the trainability of Parameterized Quantum Circuits (PQCs), and is tightly linked to the **dimensionality** of the **space explorable** by the circuit

Space of **bounded operators** on H

Space of **bounded, traceless operators invariant** under the action of the PQC

$$\mathcal{B} = \bigoplus_{\kappa \in \{0,1\}^M} \mathcal{B}_\kappa$$

E. Fontana et al., Nat. Comm. 2024, M. Ragone et al., Nat. Comm. 2024

Barren Plateaus: a curse of dimensionality

The **exponential** concentration of the loss function, dubbed **barren plateau**, is a major **threat** to the trainability of Parameterized Quantum Circuits (PQCs), and is tightly linked to the **dimensionality** of the **space explorable** by the circuit

Space of **bounded operators** on H

Space of **bounded, traceless operators invariant** under the action of the PQC

$$\mathcal{B} = \bigoplus_{\kappa \in \{0,1\}^M} \mathcal{B}_\kappa$$

Variance wrt θ of the **loss function**

Each subspace contributes **independently** to the variance

$$\mathbb{V}^\infty\{\langle H \rangle_\theta\} = \sum_\kappa \mathbb{V}_\kappa^\infty$$

Each contribution is **inversely proportional** to the dimension of the **corresponding subspace**

$$\mathbb{V}_\kappa^\infty \propto \frac{1}{\dim(\mathcal{B}_\kappa)}$$

E. Fontana et al., Nat. Comm. 2024, M. Ragone et al., Nat. Comm. 2024

Barren Plateaus: a curse of dimensionality

The **exponential** concentration of the loss function, dubbed **barren plateau**, is a major **threat** to the trainability of Parameterized Quantum Circuits (PQCs), and is tightly linked to the **dimensionality** of the **space explorable** by the circuit

Space of **bounded operators** on H

Space of **bounded, traceless operators invariant** under the action of the PQC

$$\mathcal{B} = \bigoplus_{\kappa \in \{0,1\}^M} \mathcal{B}_\kappa$$

Variance wrt θ of the **loss function**

Each subspace contributes **independently** to the variance

$$\mathbb{V}^\infty\{\langle H \rangle_\theta\} = \sum_\kappa \mathbb{V}^\infty_\kappa$$

Each contribution is **inversely proportional** to the dimension of the **corresponding subspace**

$$\mathbb{V}^\infty_\kappa \propto \frac{1}{\dim(\mathcal{B}_\kappa)}$$

Typically, such dimension **scales exponentially** in the number *n* of qubits. However, in the presence of **symmetry**, some of such contributions might be **constant** or **scale polynomially** ⇒ No BP!

E. Fontana et al., Nat. Comm. 2024, M. Ragone et al., Nat. Comm. 2024

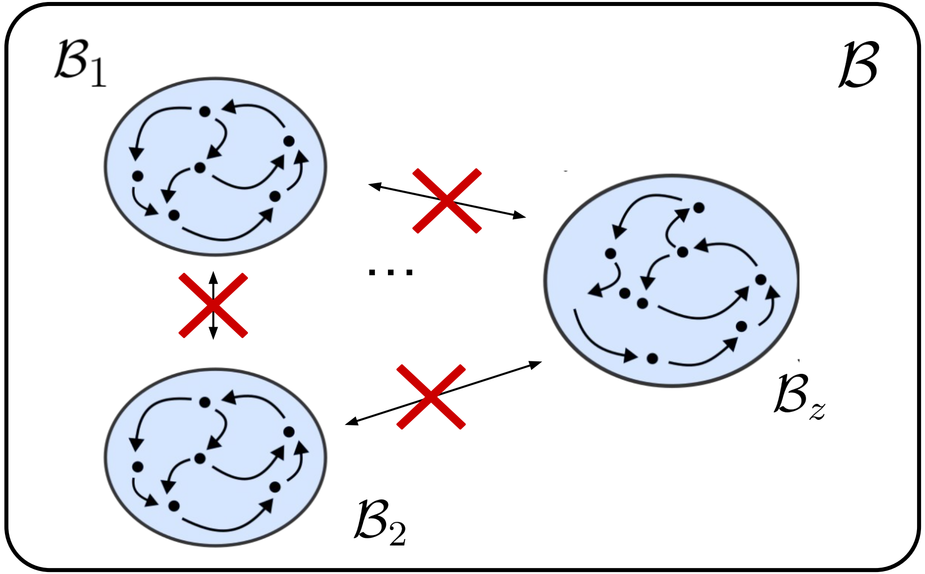
Barren Plateaus in a picture

One can visualize what's going on here using a **graph**. The space of bounded operators is represented here by a set of points, and the action of the PQC by arrows connecting some.

Space of **bounded operators** on H

Space of **bounded, traceless operators invariant** under the action of the PQC

$$\mathcal{B} = \bigoplus_{\kappa \in \{0,1\}^M} \mathcal{B}_\kappa$$



These are generally computed using an **algebraic approach** (Dynamical Lie Algebra)

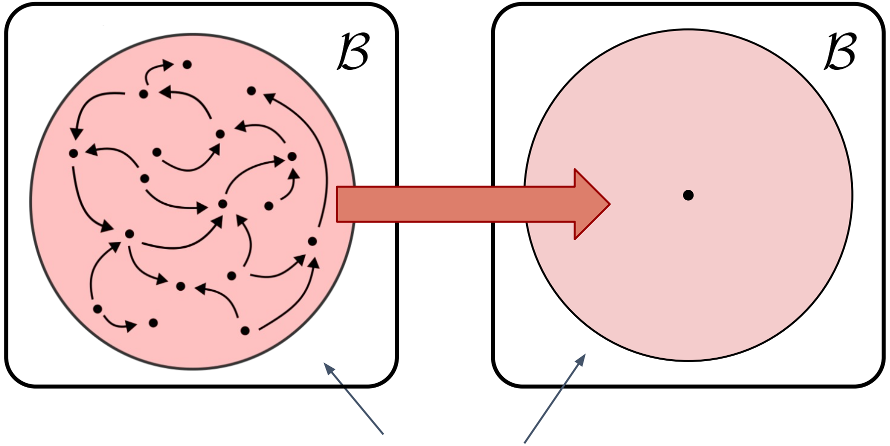
The emergence of Noise-Induced Barren Plateaus (NIBP)

In the presence of **noise**, **quantum information is lost** to the environment. Whenever this happens at a **constant rate**, **throughout the space**, all contributions **vanish** exponentially fast in the deep circuit limit.

S. Wang et al., Nat. Comm. 2021

The emergence of Noise-Induced Barren Plateaus (NIBP)

In the presence of **noise**, **quantum information is lost** to the environment. Whenever this happens at a **constant rate**, **throughout the space**, all contributions **vanish** exponentially fast in the deep circuit limit.

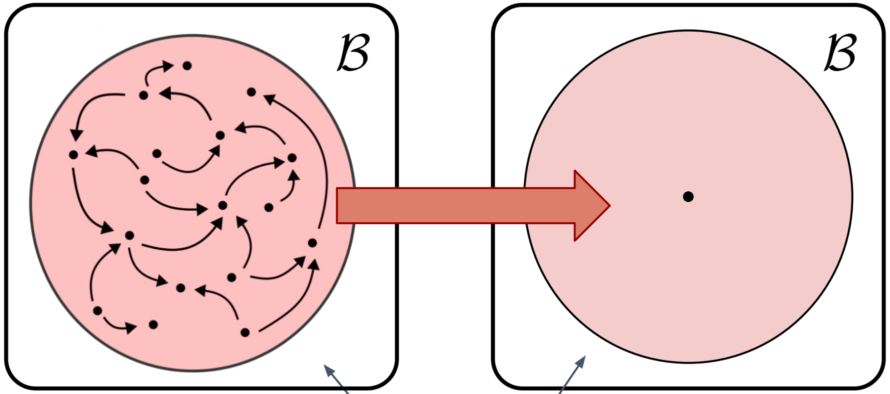


As the quantum information is lost, everything becomes **less and less distinguishable**. In the end, all points collapse into the same one, and all parameters inevitably produce the same output, i.e. the **variance vanishes**.

S. Wang et al., Nat. Comm. 2021

The emergence of Noise-Induced Barren Plateaus (NIBP)

In the presence of **noise**, **quantum information is lost** to the environment. Whenever this happens at a **constant rate**, **throughout the space**, all contributions **vanish** exponentially fast in the deep circuit limit.



As the quantum information is lost, everything becomes **less and less distinguishable**. In the end, all points collapse into the same one, and all parameters inevitably produce the same output, i.e. the **variance vanishes**.

NIBP phenomenon,

$$|\mathbb{V}^L| \in O(e^{\beta L})$$

$$\Downarrow$$

$$\mathbb{V}^{\infty} = 0$$

NIBP phenomenon, expressed in terms of the variance in the **deep circuit limit**

S. Wang et al., Nat. Comm. 2021

A rich spectrum of quantum noises

Generally, there exist a wide variety of **quantum channels**, but we only have rigorous result on the **‘extreme’** cases.

NIBP phenomenon due to a too **strong interaction** with the environment

?

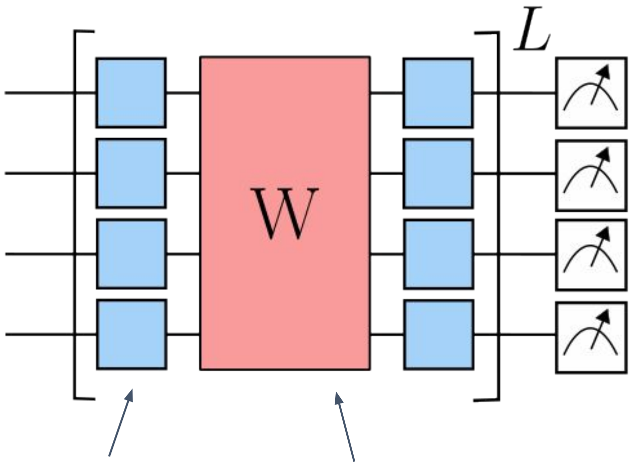
Non-unital noise? More?

BP phenomenon due to an **exponential growth** of the computational space

‘Too much’ noise ← Noise properties → No noise

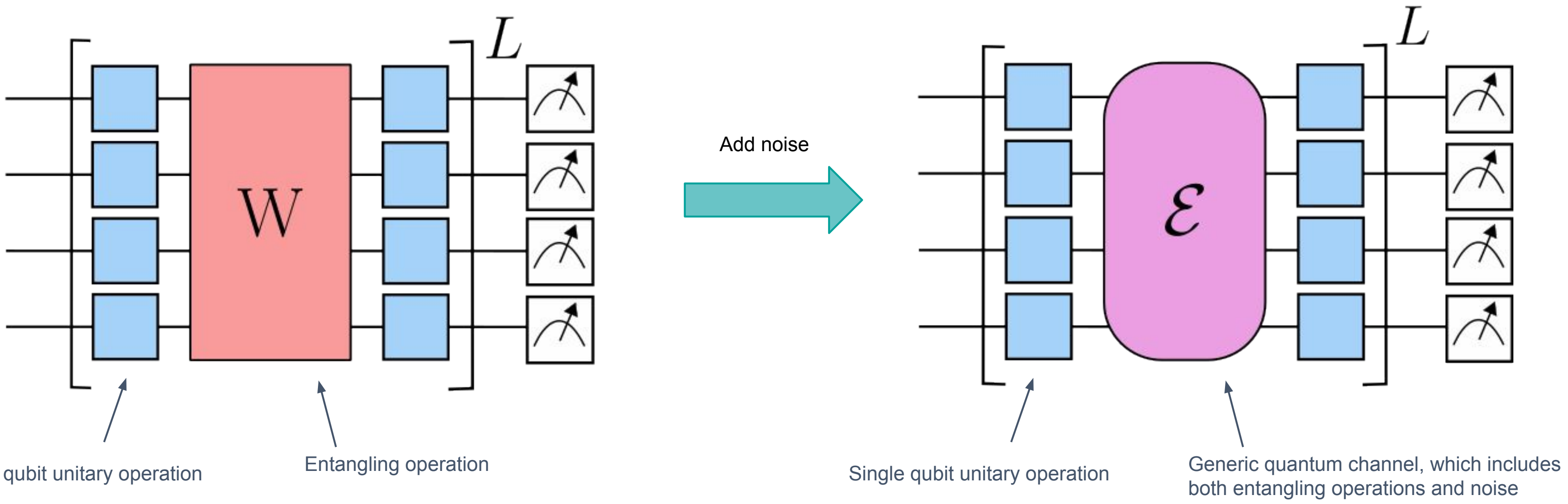
How we plan to study it?

To investigate the main mechanisms at play in this general scenario, we devise a simple yet general model, comprised of **local unitary designs** and **general quantum channels**



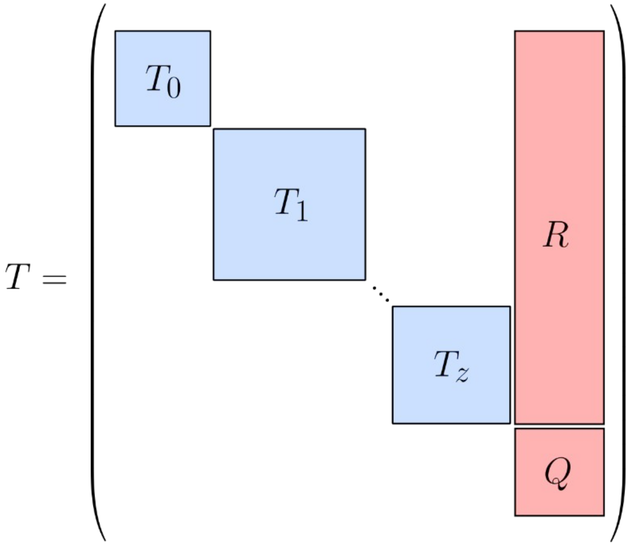
How we plan to study it?

To investigate the main mechanisms at play in this general scenario, we devise a simple yet general model, comprised of **local unitary designs** and **general quantum channels**



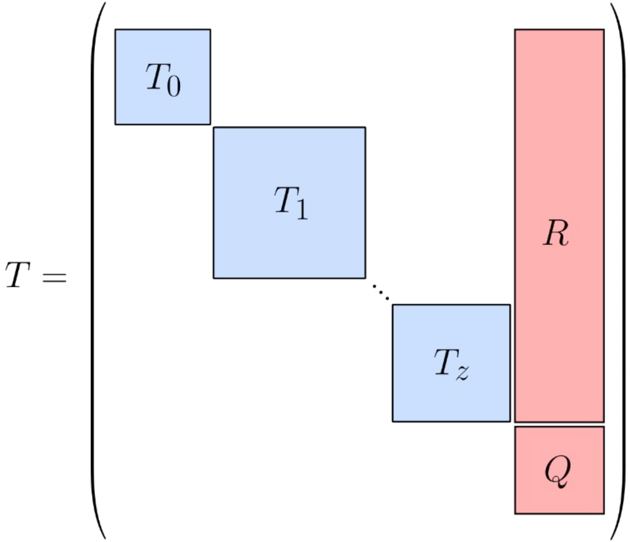
Two main applications

In this way, we can associate a **non-negative** matrix to each **layer** of the channel.
Interestingly, the **spectral properties** of T can give valuable information on the variance of the **full model**.



Two main applications

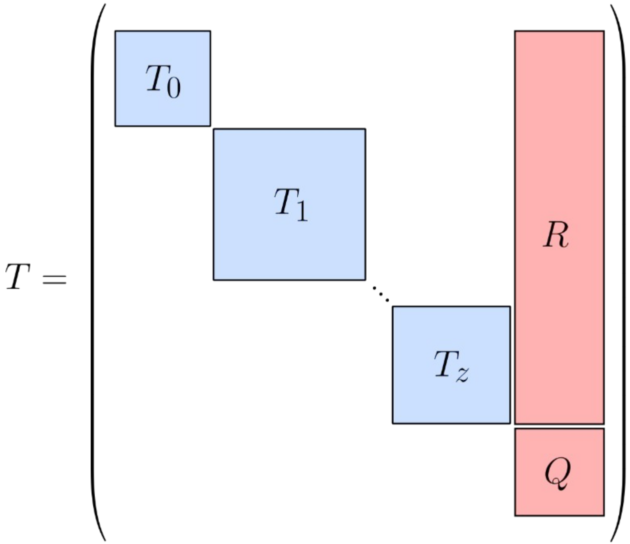
In this way, we can associate a **non-negative** matrix to each **layer** of the channel.
Interestingly, the **spectral properties** of T can give valuable information on the variance of the **full model**.



- **Analytical variance** computation in the **deep** circuit limit (L>>1)

Two main applications

In this way, we can associate a **non-negative** matrix to each **layer** of the channel.
Interestingly, the **spectral properties** of T can give valuable information on the variance of the **full model**.



- **Analytical variance** computation in the **deep** circuit limit (L>>1)

- General formalization of **BP free**, smart initializations as **stochastic unravelling** of **sufficiently weak** noise maps

Analytical variance for general channels

*Holds rigorously for aperiodic circuits. For periodic circuits only holds for the Cesàro sum of all depths L.

For deep circuits, we have **Theorem 1 (Deep circuits):**

The variance converges exponentially fast in the number of layers L to a limiting value, i.e.*

$$\left|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right)$$

Analytical variance for general channels

*Holds rigorously for aperiodic circuits. For periodic circuits only holds for the Cesàro sum of all depths L.

For deep circuits, we have **Theorem 1 (Deep circuits):**

The variance converges exponentially fast in the number of layers L to a limiting value, i.e.*

$$\left|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right)$$

$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z|r_z=1} \frac{(\ell_\rho)_z(\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z(A\ell_H)_z}{d_z}$$

Analytical variance for general channels

*Holds rigorously for aperiodic circuits. For periodic circuits only holds for the Cesàro sum of all depths L.

For deep circuits, we have **Theorem 1 (Deep circuits):**

The variance converges exponentially fast in the number of layers L to a limiting value, i.e.*

$$\left|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right)$$

Variance of loss function in the limit of **large number of layers**

Locality (= purity) of the initial state and observable

Absorption matrix, represents the absorption to the invariant subspaces

$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z|r_z=1} \frac{(\ell_\rho)_z(\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z(A\ell_H)_z}{d_z}$$

Dimension of such subspaces

Sum over suitable **invariant subspaces of both** the single qubit blocks and the intermediate channel, based on the **largest block eigenvalue**

Analytical variance for general channels

*Holds rigorously for aperiodic circuits. For periodic circuits only holds for the Cesàro sum of all depths L.

For deep circuits, we have **Theorem 1 (Deep circuits):**

The variance converges exponentially fast in the number of layers L to a limiting value, i.e.*

$$\left|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right)$$

$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z|r_z=1} \frac{(\ell_\rho)_z(\ell_H)_z}{d_z} + \frac{(\ell_\rho)_z(A\ell_H)_z}{d_z}$$

Variance of loss function in the limit of **large number of layers**

Locality (= purity) of the initial state and observable

Absorption matrix, represents the absorption to the invariant subspaces

Sum over suitable **invariant subspaces of both** the single qubit blocks and the intermediate channel, based on the **largest block eigenvalue**

Unitary contribution

Noise contribution

Dimension of such subspaces

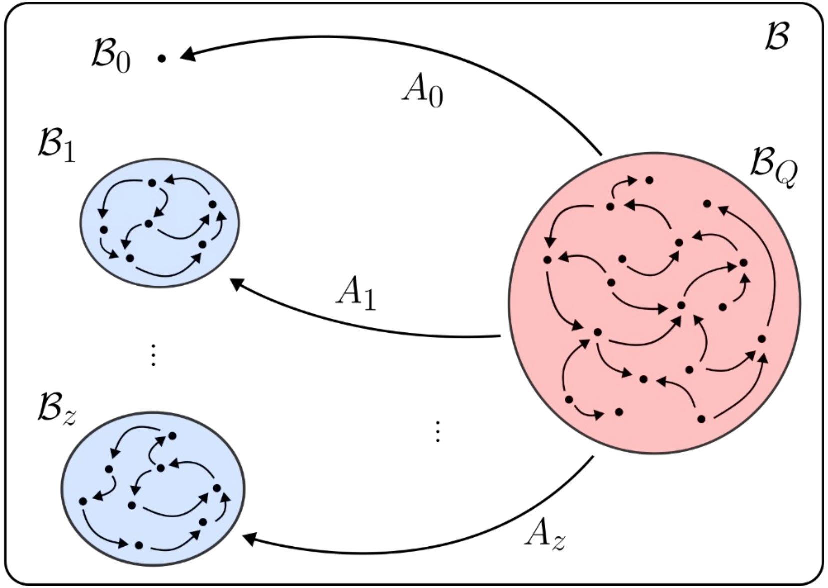
Examples: Unitary circuits

The absorption term A vanishes in this case because unitary dynamics is **reversible**.

Corollary III.1.1 (Deep, unitary circuits).

If the channel is unitary, then the absorption terms vanish.

$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z}$$



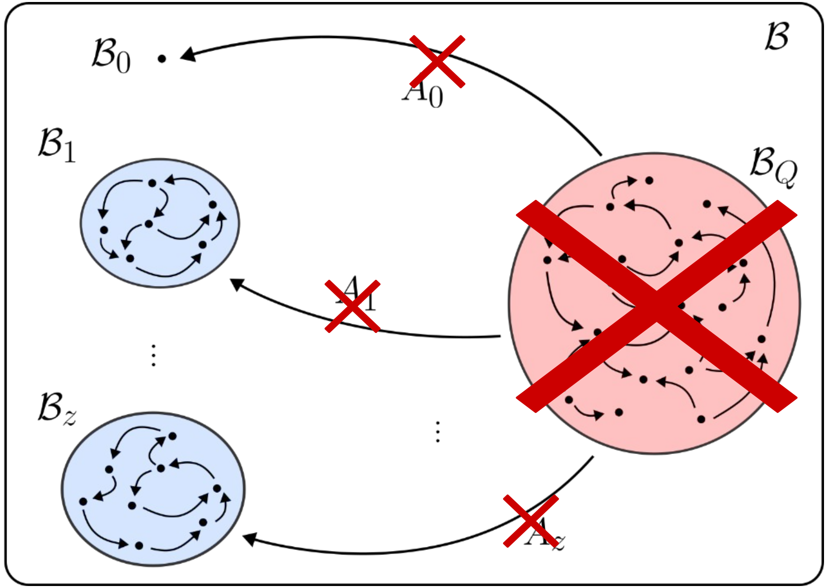
Examples: Unitary circuits

The absorption term A vanishes in this case because unitary dynamics is **reversible**.

Corollary III.1.1 (Deep, unitary circuits).

If the channel is unitary, then the absorption terms vanish.

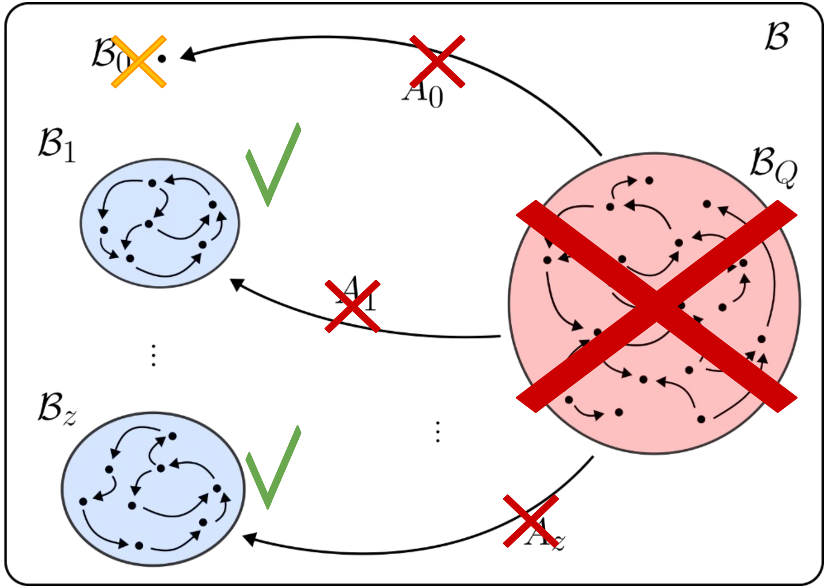
$$\mathbb{V}^{\infty}_{\rho,H} = \sum_{z} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z}$$



Examples: Unitary circuits

The absorption term A vanishes in this case because unitary dynamics is **reversible**.

> **Corollary III.1.1 (Deep, unitary circuits).**
>
> *If the channel is unitary, then the absorption terms vanish.*
>
> $$\mathbb{V}_{\rho,H}^{\infty} = \sum_{z} \frac{(\ell_\rho)_z (\ell_H)_z}{d_z}$$



Examples: Contractive circuits

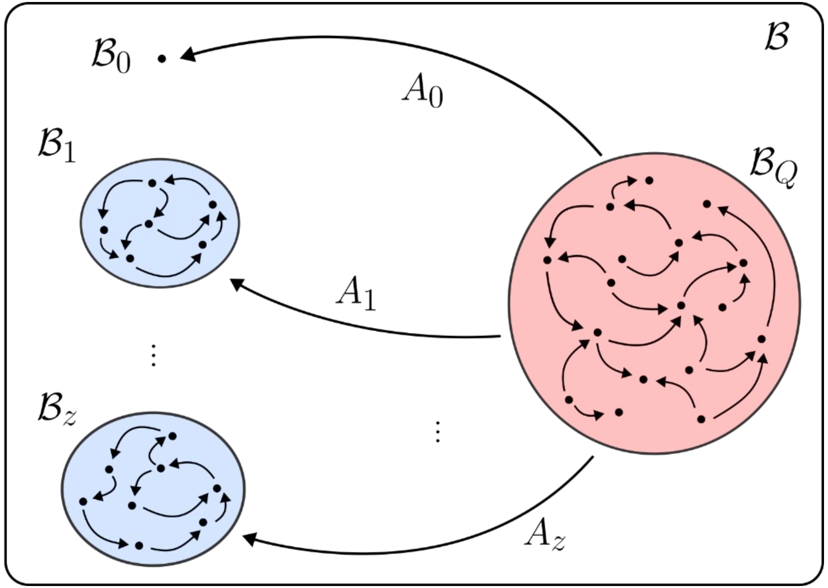
Conversely, since each invariant subspace norm-preserving to '*survive*' the deep circuit limit, in the strictly contractive case **only the absorption terms** remain.

Corollary III.1.2 (Deep, contractive circuits).

If the channel has no unitary subspaces, then only the absorption term survives.

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{(A\ell_H)_0}{d}$$

In particular, if the channel is unital, then $\mathbb{V}^{\infty}_{\rho,H} = 0$.



Examples: Contractive circuits

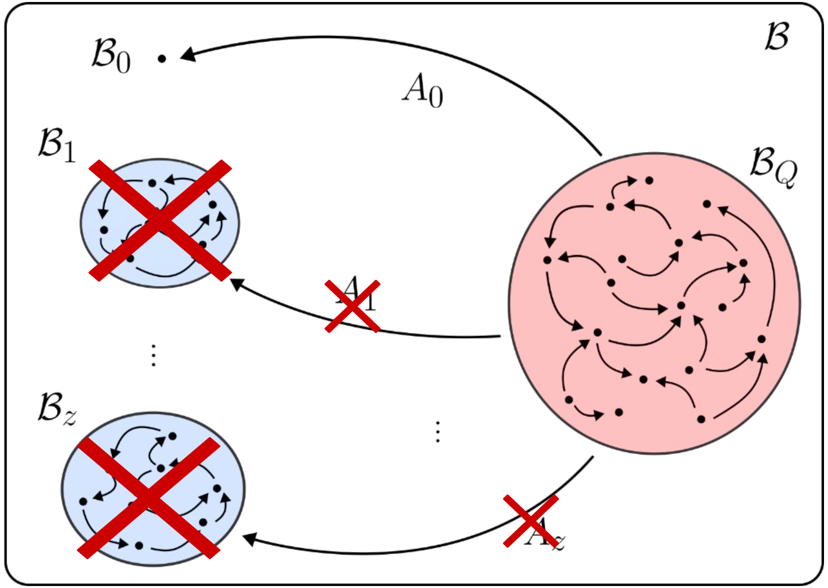
Conversely, since each invariant subspace norm-preserving to '*survive*' the deep circuit limit, in the strictly contractive case **only the absorption terms** remain.

Corollary III.1.2 (Deep, contractive circuits).

If the channel has no unitary subspaces, then only the absorption term survives.

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{(A\ell_H)_0}{d}$$

In particular, if the channel is unital, then $\mathbb{V}^{\infty}_{\rho,H} = 0$.



Examples: Contractive circuits

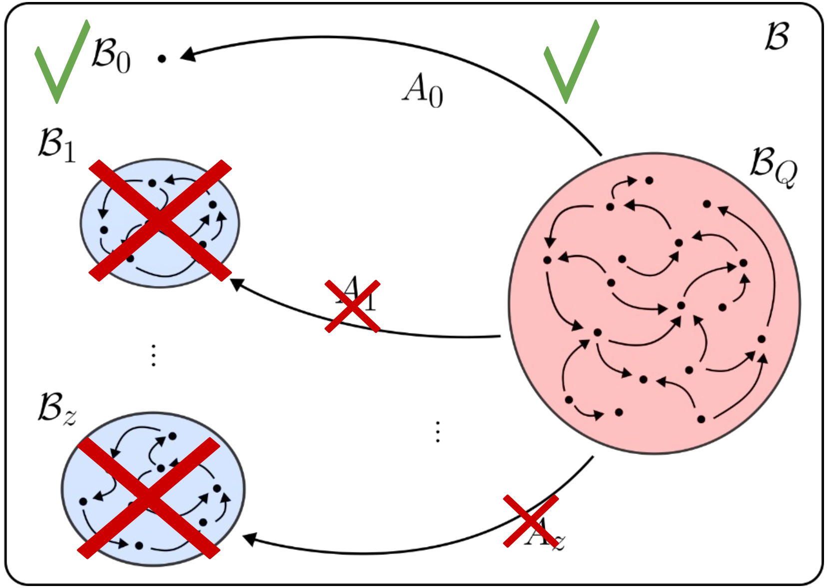
Conversely, since each invariant subspace norm-preserving to '*survive*' the deep circuit limit, in the strictly contractive case **only the absorption terms** remain.

Corollary III.1.2 (Deep, contractive circuits).

If the channel has no unitary subspaces, then only the absorption term survives.

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{(A\ell_H)_0}{d}$$

In particular, if the channel is unital, then $\mathbb{V}^{\infty}_{\rho,H} = 0$.



Examples: Contractive circuits

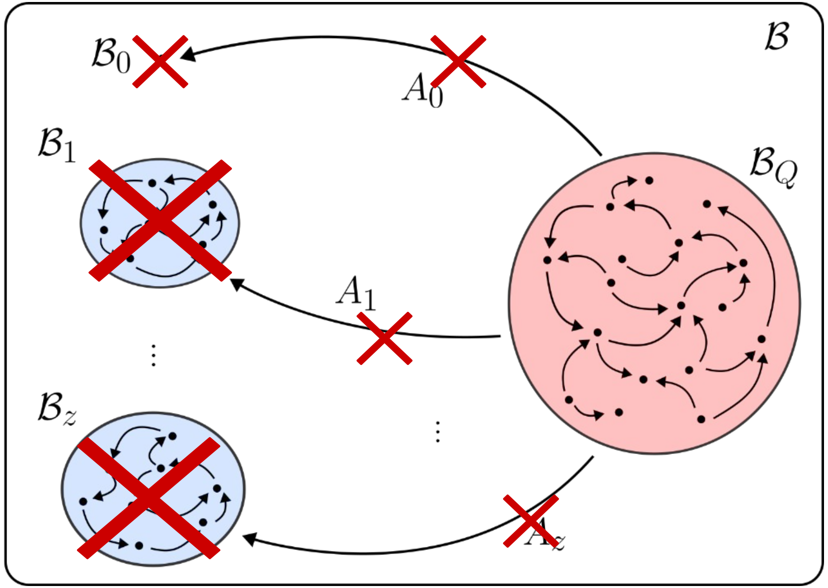
Conversely, since each invariant subspace norm-preserving to ‘*survive*’ the deep circuit limit, in the strictly contractive case **only the absorption terms** remain.

Corollary III.1.2 (Deep, contractive circuits).

If the channel has no unitary subspaces, then only the absorption term survives.

$$\mathbb{V}^{\infty}_{\rho,H} = \frac{(A\ell_H)_0}{d}$$

In particular, if the channel is ***unital****, then* $\mathbb{V}^{\infty}_{\rho,H} = 0$.



Lower-bounding the variance

Recalling that:

The variance converges exponentially fast in the number of layers L to a limiting value, i.e.*

$$\left|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right)$$

It becomes clear that for deep circuits, the only way to **escape** such bound is by enforcing a **slower speed of convergence** to the process, acting on the constant **β**

Lower-bounding the variance

Recalling that:

The variance converges exponentially fast in the number of layers L to a limiting value, i.e.*

$$\left|\mathbb{V}_{\rho,H}^{L} - \mathbb{V}_{\rho,H}^{\infty}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right)$$

It becomes clear that for deep circuits, the only way to **escape** such bound is by enforcing a **slower speed of convergence** to the process, acting on the constant **β**

$$\beta \sim \frac{1}{L} \quad \Rightarrow \quad \left|\mathbb{V}_{\rho,H}^{L} - \mathbb{V}_{\rho,H}^{\infty}\right| \in O\left(\|H\|_2^2\right)$$

Here, lower-bounds can actually be found.

Lower-bounding the variance

Recalling that:

The variance converges exponentially fast in the number of layers L to a limiting value, i.e.*

$$\left|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}\right| \in O\left(e^{-\beta L}\|H\|_2^2\right)$$

It becomes clear that for deep circuits, the only way to **escape** such bound is by enforcing a **slower speed of convergence** to the process, acting on the constant **β**

$$\beta \sim \frac{1}{L} \quad \Rightarrow \quad \left|\mathbb{V}^{L}_{\rho,H} - \mathbb{V}^{\infty}_{\rho,H}\right| \in O\left(\|H\|_2^2\right)$$

This regime is achieved easily by constraining the **second largest eigenvalue** of relevant blocks

Here, lower-bounds can actually be found.

Meaning in terms of channel property

Theorem 2 (Shallow circuit).
If the diagonal entries of T scale as T_{ij}~1-log(n)/L, then β~log(n)/L and the variance is lower bounded by

$$\mathbb{V}^{L}_{\rho,H} \geq F(n)(\ell_{\rho}, \ell_{H})$$

where F(n) decays polynomially with n.

Orthogonality measure between the initial state and observable.

Meaning in terms of channel property

Theorem 2 (Shallow circuit).
If the diagonal entries of T scale as T_{ij}~1-log(n)/L, then β~log(n)/L and the variance is lower bounded by

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_H)$$

where F(n) decays polynomially with n.

Orthogonality measure between the initial state and observable.

Example: **Lindbladian noise**

$$\mathcal{E} = e^{\delta t \mathcal{L}}$$

$$\left(\mathcal{L}(\rho) = \sum_i L_i \rho L_i - \frac{1}{2}\{L_i^2, \rho\} \, , \quad L_i = L_i^\dagger \right)$$

Meaning in terms of channel property

Theorem 2 (Shallow circuit).
If the diagonal entries of T scale as T_{ij}~1-log(n)/L, then β~log(n)/L and the variance is lower bounded by

$$\mathbb{V}^L_{\rho,H} \geq F(n)(\ell_\rho, \ell_H)$$

where F(n) decays polynomially with n.

Orthogonality measure between the initial state and observable.

Example: **Lindbladian noise**

$$\mathcal{E} = e^{\delta t \mathcal{L}}$$

$$\left(\mathcal{L}(\rho) = \sum_i L_i \rho L_i - \frac{1}{2}\{L_i^2, \rho\} \,,\quad L_i = L_i^\dagger \right)$$

$$\delta t \sim \beta \sim \frac{1}{L}$$

Meaning in terms of channel property

Theorem 2 (Shallow circuit).
If the diagonal entries of T scale as T_{ij}~1-log(n)/L, then β~log(n)/L and the variance is lower bounded by

$$\mathbb{V}^{L}_{\rho,H} \geq F(n)(\ell_\rho, \ell_H)$$

where F(n) decays polynomially with n.

Orthogonality measure between the initial state and observable.

Example: **Lindbladian noise**

$$\mathcal{E} = e^{\delta t \mathcal{L}}$$

$$\left(\mathcal{L}(\rho) = \sum_i L_i \rho L_i - \frac{1}{2}\{L_i^2, \rho\}\,, \quad L_i = L_i^\dagger \right)$$

$$\delta t \sim \beta \sim \frac{1}{L}$$

Basically, **weak enough** noise won't produce NIBP

Stochastic unravelings of channels and QResNets

For any given noise map, we are guaranteed to be able to express it a **stochastic unraveling**, i.e. as the **expectation value** of some simpler operators. For maps as before, such operators can even be **unitary**!

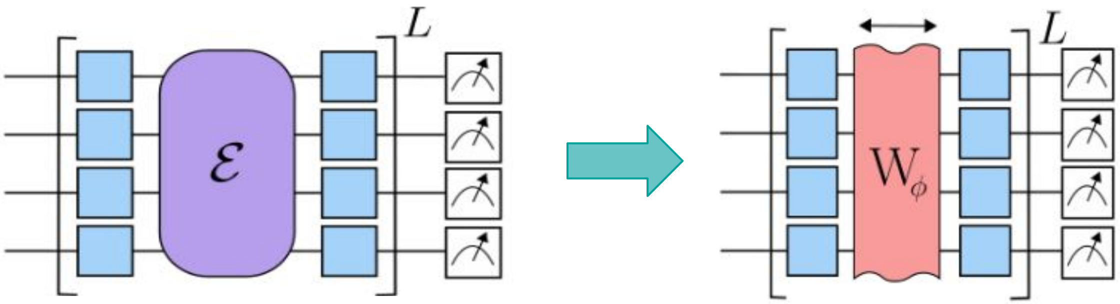
$$\mathcal{E}(\rho) = \mathbb{E}_\phi \left\{ W_\phi \rho W_\phi^\dagger \right\} \qquad \{W_\phi\} \subset U(d)$$

Stochastic unravelings of channels and QResNets

For any given noise map, we are guaranteed to be able to express it a **stochastic unraveling**, i.e. as the **expectation value** of some simpler operators. For maps as before, such operators can even be **unitary**!

$$\mathcal{E}(\rho) = \mathbb{E}_\phi \left\{ W_\phi \rho W_\phi^\dagger \right\} \qquad \{W_\phi\} \subset U(d)$$

$$\Downarrow$$

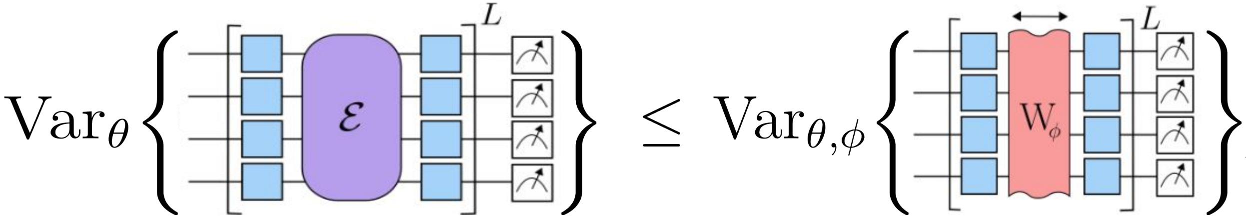


We call this a **QResNet!**

Stochastic unravelings of channels and QResNets

Proposition 4 (informal)
The variance of a channel is always smaller than the variance of the corresponding stochastic unravelling, with equality holding if and only if the channel is unitary.

Quantum channel is **still interpreted as noise**



Quantum channel is used to derive the corresponding **QResNet**, and the used as **trainable anstaz**

Where the probability distribution over ϕ is **defined by the unraveling**.

Conclusions

We introduced a new tool for the analysis of concentration in noisy variational circuits.

In particular, we observed:

- **Concentration** for noisy circuits is more **intricate** than one can naively expect.
- Noise can **not only induce** barren plateaus, but **also prevent it** if engineered correctly.
- **Weak noise** maps are closely related to **small angle-like initialization** strategies.

QUESTIONS?

giulio.crognaletti@phd.units.it