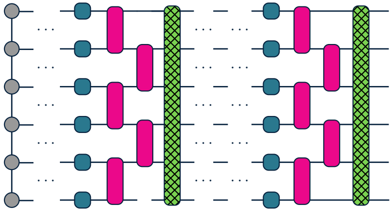
Overparameterization and Expressivity of Realistic Quantum Systems

Matthew Duschenes*, Juan Carrasquilla, Raymond Laflamme,
Diego García-Martín, Martín Larocca, Zoë Holmes, Marco Cerezo

University of Waterloo, Institute for Quantum Computing, ETH Zurich, & Los Alamos National Laboratory

Seminar



arXiv:2401.05500 , arXiv:2405.XXXXX

Fall, 2024

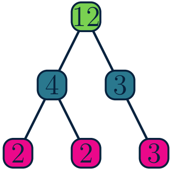
Quantum Tasks Of Interest

Ultimately, we want to do something *useful* with our quantum devices

Quantum Tasks Of Interest

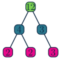
Ultimately, we want to do something *useful* with our quantum devices

- *Quantum algorithms* i.e) Factoring numbers

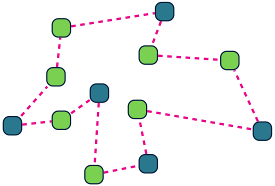


Quantum Tasks Of Interest

Ultimately, we want to do something *useful* with our quantum devices

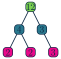
- *Quantum algorithms* i.e) Factoring numbers 

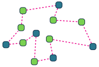
- *Optimization* problems i.e) Travelling Salesman Problem

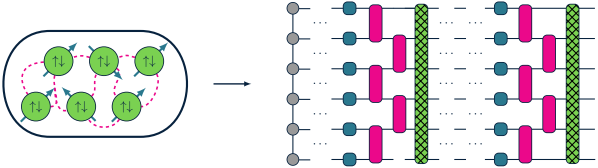


Quantum Tasks Of Interest

Ultimately, we want to do something *useful* with our quantum devices

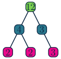
- *Quantum algorithms* i.e) Factoring numbers 

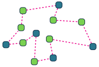
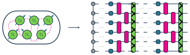
- *Optimization* problems i.e) Travelling Salesman Problem 
- *Compilation* tasks i.e) Form operators U given native gates $\{V\}$

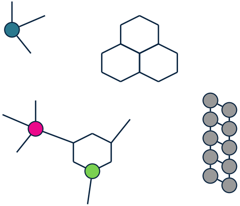


Quantum Tasks Of Interest

Ultimately, we want to do something *useful* with our quantum devices

- *Quantum algorithms* i.e) Factoring numbers 

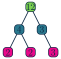
- *Optimization* problems i.e) Travelling Salesman Problem 
- *Compilation* tasks i.e) Form operators U given native gates $\{V\}$ 
- *Simulate* quantum systems i.e) Complicated molecules and chemical reactions



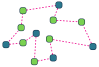
Quantum Tasks Of Interest

Ultimately, we want to do something *useful* with our quantum devices

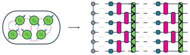
- *Quantum algorithms* i.e) Factoring numbers



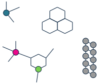
- *Optimization* problems i.e) Travelling Salesman Problem



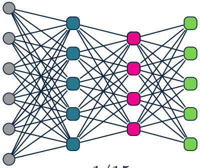
- *Compilation* tasks i.e) Form operators U given native gates $\{V\}$



- *Simulate* quantum systems i.e) Complicated molecules and chemical reactions



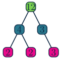
- *Machine learning* functions i.e) Classification, Regression, Generative



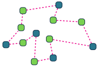
Quantum Tasks Of Interest

Ultimately, we want to do something *useful* with our quantum devices

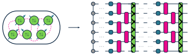
- *Quantum algorithms* i.e) Factoring numbers



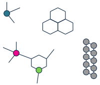
- *Optimization* problems i.e) Travelling Salesman Problem



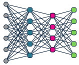
- *Compilation* tasks i.e) Form operators U given native gates $\{V\}$

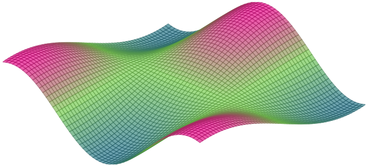


- *Simulate* quantum systems i.e) Complicated molecules and chemical reactions

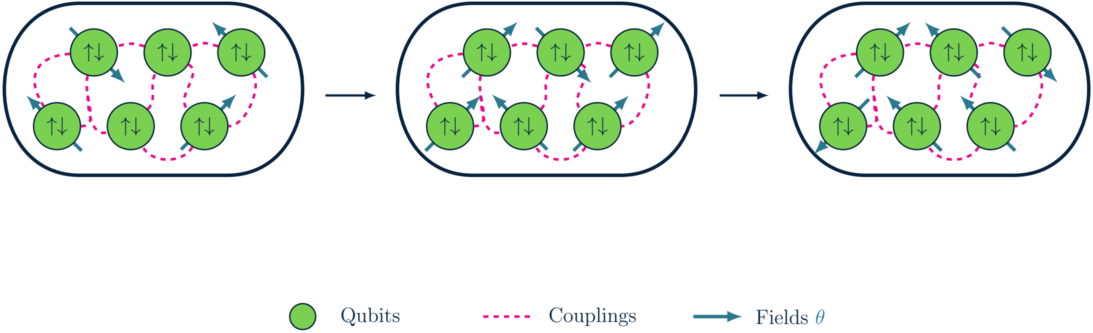


- *Machine learning* functions i.e) Classification, Regression, Generative

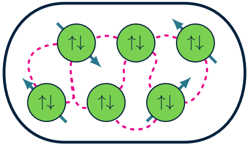




What Are Parameterized Quantum Systems?

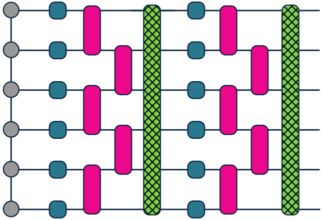


What Are Parameterized Quantum Systems?



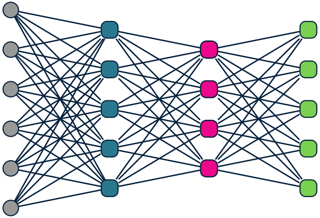
Quantum System

$H_\theta = \sum_\mu \theta_\mu G_\mu$



Quantum Circuit

$U_\theta = \prod_\mu U_\theta^\mu$

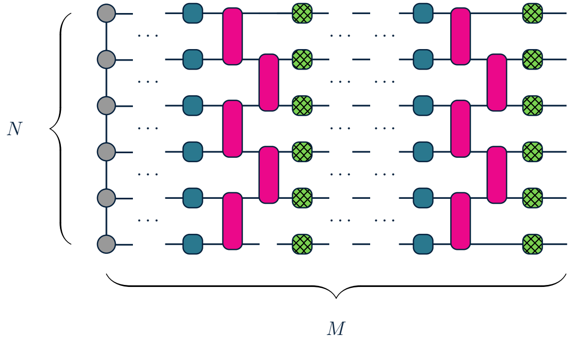


Classical Algorithm

$f_\theta = \circ_\mu f_\theta^\mu$

Tasks of Interest: Unitary Compilation, State Preparation

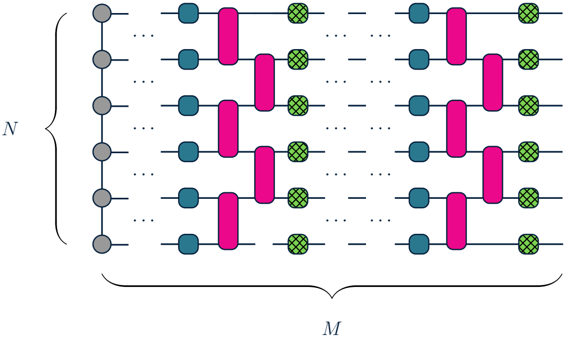
Learning Phenomena of Quantum Systems



Let our parameterized ansatz consist of unitary $\mathcal{U}_\theta$ and noise $\mathcal{N}_\gamma$ components

$$\Lambda_{\theta\gamma} = \bigcirc_{\mu}^{M} \mathcal{N}_{\gamma}^{(\mu)} \circ \mathcal{U}_{\theta}^{(\mu)} \tag{1}$$

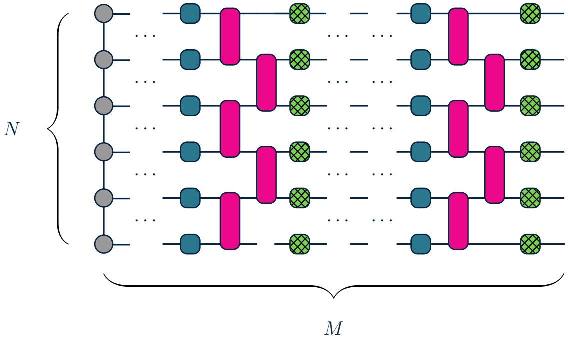
Learning Phenomena of Quantum Systems



Let our parameterized ansatz consist of unitary $\mathcal{U}_\theta$ and noise $\mathcal{N}_\gamma$ components

$$U_\theta^{(\mu)} = e^{-i\delta H_\theta^{(\mu)}} \;:\; H_\theta^{(\mu)} = \sum_\nu \theta_\nu^{(\mu)} G_\nu \tag{2}$$

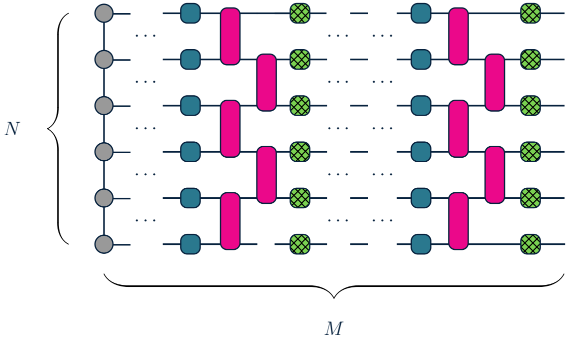
$$\mathcal{N}_\gamma^{(\mu)} = \otimes_i^N \mathcal{N}_{\gamma_i} \;:\; \mathcal{N}_{\gamma_i} = (1-\gamma)\mathcal{I}_i + \gamma \mathcal{K}_i \tag{3}$$



Let our parameterized ansatz consist of unitary $\mathcal{U}_\theta$ and noise $\mathcal{N}_\gamma$ components

$$\Lambda_{\theta\gamma} = \bigcirc_{\mu}^{M} \mathcal{N}_{\gamma}^{(\mu)} \circ \mathcal{U}_{\theta}^{(\mu)} \tag{1}$$

$$\text{NMR: } H_\theta^{(\mu)} = \sum_i^N \theta_i^{x(\mu)} X_i + \sum_i^N \theta_i^{y(\mu)} Y_i + \sum_i^N h_i Z_i + \sum_{i<j}^N J_{ij} Z_i Z_j \tag{4}$$



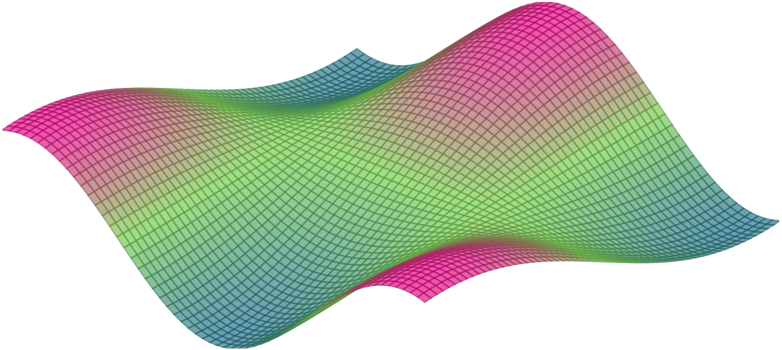
How does the amount of *noise* γ and the *evolution depth* M

of a *constrained* system

affect its classical simulation and optimization, and resulting infidelities

$$\mathcal{L}_{\theta^*_{\gamma}} \; : \; U_{\theta_\gamma} \approx U \;,\; \rho_{\theta_\gamma} \approx \rho \;?$$

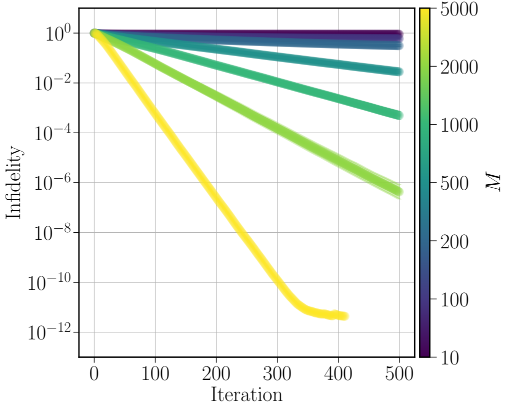
Learning Phenomena of Quantum Systems



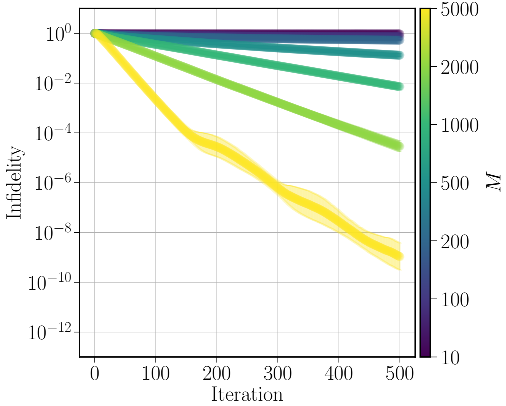
How can we leverage approaches from
quantum optimal control and *learning theory* to describe these relationships?

$$\textit{Infidelity}\text{: } 1 - \operatorname{tr}\left(\rho\rho_{\theta\gamma}\right)\ ,\ \textit{Impurity}\text{: } 1 - \operatorname{tr}\left(\rho_{\theta\gamma}^2\right)\ ,\ \textit{Entropy}\text{: } -\operatorname{tr}\left(\rho_{\theta\gamma}\log\rho_{\theta\gamma}\right)$$

Unconstrained vs. Constrained Optimization

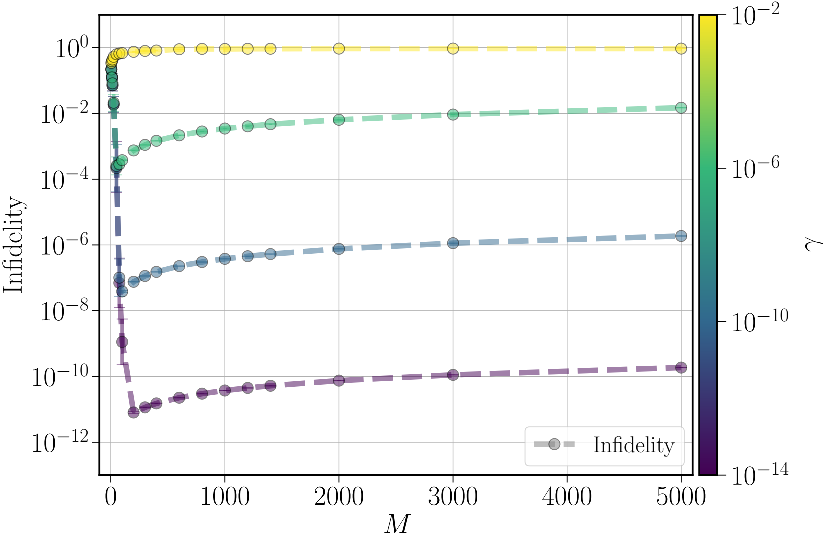


(a) Unconstrained Unitary Compilation



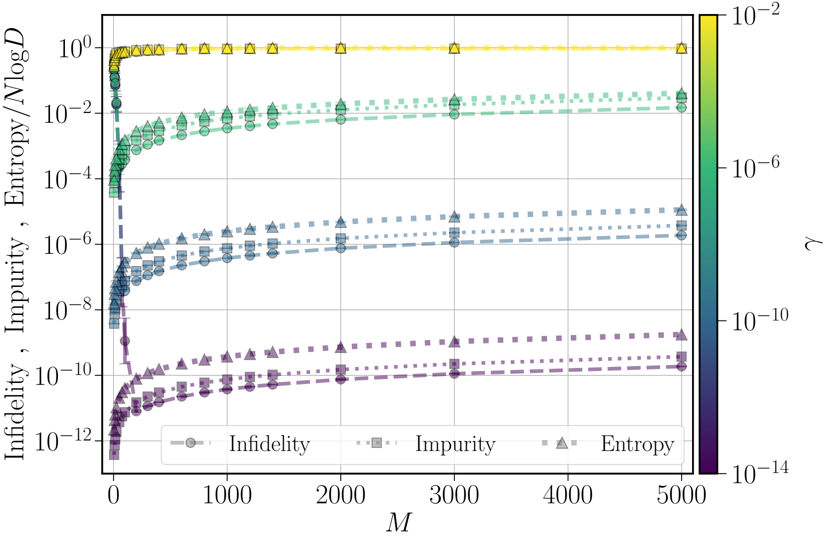
(b) Constrained Unitary Compilation

Regimes of Noisy Optimization



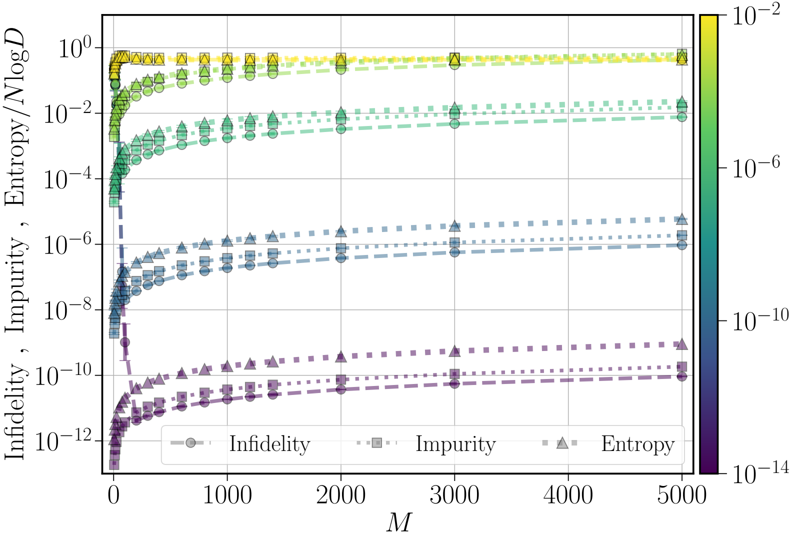
(c) Unital Dephasing for State Preparation

Regimes of Noisy Optimization



(d) Unital Dephasing for State Preparation

Regimes of Noisy Optimization



(e) Non-Unital Amplitude Damping for State Preparation

Noise Induced Critical Depth

Noise induces a critical depth (Fontana *et al.* PRA **104** (2021))

$$M_\gamma \sim \log 1/\gamma \ , \tag{5}$$

meaning the minimum infidelity is *linear-quadratic* $(1 \leq \alpha \leq 2)$ in noise

$$\mathcal{L}_{\theta^*\gamma|M_\gamma} \sim \gamma^\alpha \ , \tag{6}$$

and parameterized noise channels are therefore *robust* tof approximately

$$\bar{M}_\gamma \sim \gamma \log 1/\gamma \quad \text{errors} \ . \tag{7}$$

Noise Induced Critical Depth

Noise induces a critical depth (Fontana *et al.* PRA **104** (2021))

$$M_{\gamma} \sim \log 1/\gamma \ , \tag{5}$$

meaning the minimum infidelity is *linear-quadratic* $(1 \leq \alpha \leq 2)$ in noise

$$\mathcal{L}_{\theta^{*}\gamma|M_{\gamma}} \sim \gamma^{\alpha} \ , \tag{6}$$

and parameterized noise channels are therefore *robust* tof approximately

$$\bar{M}_{\gamma} \sim \gamma \log 1/\gamma \quad \text{errors} \ . \tag{7}$$

Is it possible to derive the M, γ scaling of the optimal $\mathcal{L}_{\theta^{*}\gamma}$ *analytically*?

$$|\mathcal{L}_{\theta_{\gamma}} - \mathcal{L}_{\theta}| \leq 2|(1-\gamma)^{NM} - 1| \tag{8}$$

Representations of Channels and States

- Channels can be represented as *ensembles* of $l \leq K$ non-trivial-error channels

Representations of Channels and States

- Channels can be represented as *ensembles* of $l \leq K$ non-trivial-error channels

$$\Lambda_{\theta\gamma} = \langle \Lambda_{\theta\gamma_l} \rangle_{l \sim p_{K\gamma}} \tag{9}$$

Representations of Channels and States

- Channels can be represented as *ensembles* of $l \leq K$ non-trivial-error channels

$$\Lambda_{\theta\gamma} = \langle \Lambda_{\theta\gamma_l} \rangle_{l \sim p_{K\gamma}} \quad , \quad \Lambda_{\theta\gamma_l} = \frac{1}{\binom{K}{l}} \sum_{\substack{\chi \subseteq [K] \\ |\chi| = l}} \Lambda_{\theta\gamma}^{\chi} \tag{9}$$

Representations of Channels and States

- Channels can be represented as *ensembles* of $l \leq K$ non-trivial-error channels

$$\Lambda_{\theta\gamma} = \langle \Lambda_{\theta\gamma_l} \rangle_{l \sim p_{K\gamma}} \quad , \quad \Lambda_{\theta\gamma_l} = \frac{1}{\binom{K}{l}} \sum_{\substack{\chi \subseteq [K] \\ |\chi| = l}} \Lambda^{\chi}_{\theta\gamma} \quad , \quad p_{K\gamma}(l) = \binom{K}{l} \gamma^l (1-\gamma)^{K-l} \tag{9}$$

Representations of Channels and States

- Channels can be represented as *ensembles* of $l \leq K$ non-trivial-error channels

$$\Lambda_{\theta\gamma} = \langle \Lambda_{\theta\gamma_l} \rangle_{l \sim p_{K\gamma}} \quad , \quad \Lambda_{\theta\gamma_l} = \frac{1}{\binom{K}{l}} \sum_{\substack{\chi \subseteq [K] \\ |\chi| = l}} \Lambda_{\theta\gamma}^{\chi} \quad , \quad p_{K\gamma}(l) = \binom{K}{l} \gamma^l (1-\gamma)^{K-l} \tag{9}$$

- States can be represented as *Bloch* coefficients $\rho_{\theta\gamma} \approx \rho \;\longleftrightarrow\; \lambda_{\theta\gamma} \approx \lambda$

$$\rho_{\theta\gamma} \;=\; \frac{I + \lambda_{\theta\gamma} \cdot \omega}{d} \;=\; (1-\gamma)^K \rho + (1 - (1-\gamma)^K)\epsilon_{\theta\gamma} + \Delta_{\theta\gamma} \tag{10}$$

Representations of Channels and States

- Quantities of interest at *noiseless* optimality scale remarkably similarly

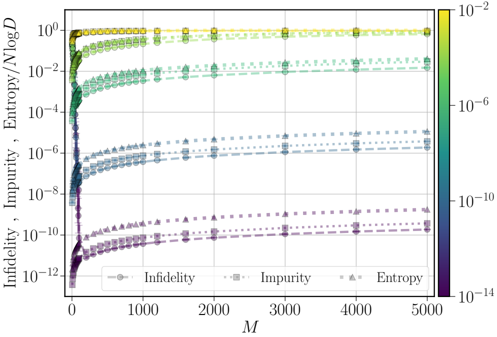
$$\mathcal{L}^{\rho}_{\theta\gamma} \sim \frac{1}{2}\mathcal{I}_{\theta\gamma} \sim \boxed{K\gamma\frac{d-1}{d}\left(1-\frac{\lambda\cdot\varepsilon_{\theta\gamma}}{\lambda^2}\right)} + O(\tbinom{K}{2}\gamma^2) \tag{11}$$

$$\mathcal{S}_{\theta\gamma} \sim \mathcal{D}^{\rho}_{\theta\gamma} \sim O(K\gamma) \tag{12}$$

Representations of Channels and States

Noise phenomena dominates at $M \geq M_{\gamma}$:
The scale of optimization and entropic driven infidelities *intersect*

$$\mathcal{L}^{\rho}_{\theta^*_{\gamma}\gamma} \sim e^{-\alpha M}|_{M_{\gamma}} \approx \mathcal{L}^{\rho}_{\theta^*{\gamma}} \sim NM\gamma|_{M_{\gamma}} \ , \tag{13}$$



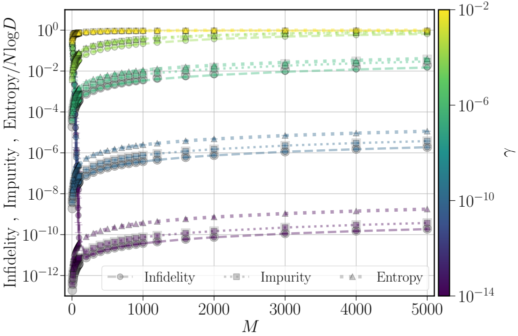
Representations of Channels and States

Noise phenomena dominates at $M \geq M_\gamma$:
The scale of optimization and entropic driven infidelities *intersect*

$$\mathcal{L}^{\rho}_{\theta^*_\gamma \gamma} \sim e^{-\alpha M}|_{M_\gamma} \approx \mathcal{L}^{\rho}_{\theta^* \gamma} \sim NM\gamma|_{M_\gamma} \ , \tag{13}$$

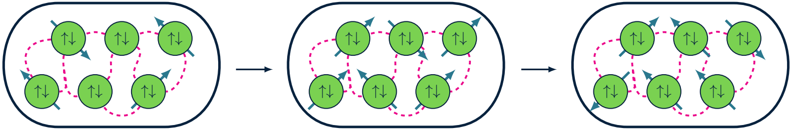
and we recover our numerically predicted noise-induced critical depth!

$$M_\gamma \sim \log 1/\gamma \tag{14}$$

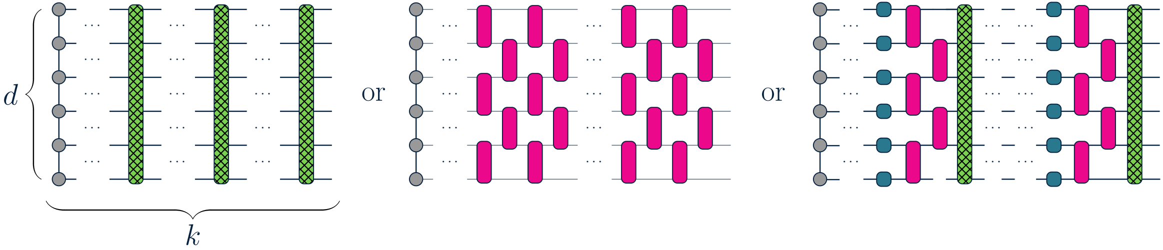


What Have We Learned About Noisy Overparameterization?

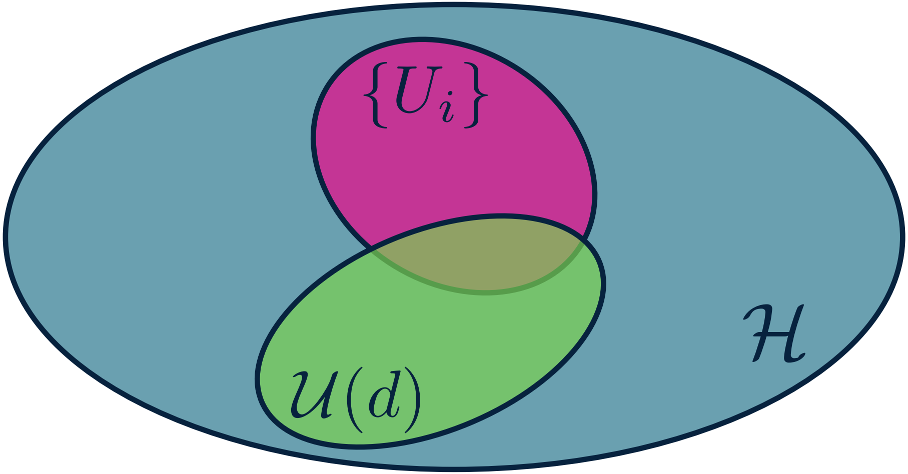
- Overparameterization is *robust* to constraints
- Accumulation of noise induces a *critical* depth M_γ that prevents convergence
- Fidelities, purities, entropies highly *correlated* in $\gamma \ll 1, M \gg 1$ regime
- What are other *noise-induced* effects on *trainability* versus *expressivity*?



How Expressive are our Ansatze?

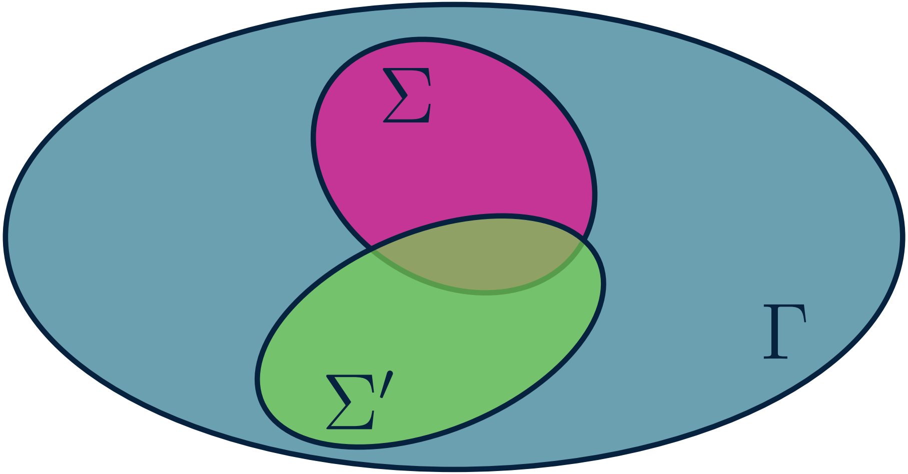


How Expressive are our Ansatze?

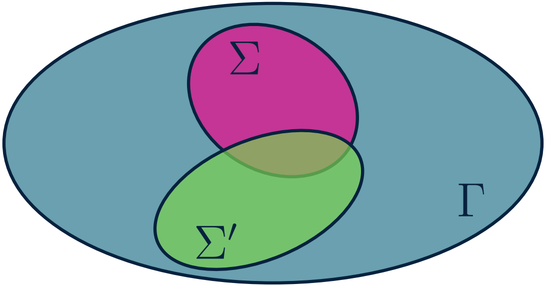


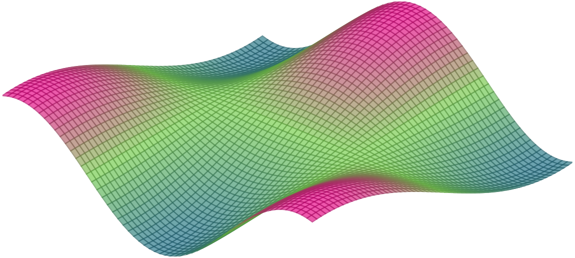
[1] Holmes, Z. *et al.*, *PRX Quantum* **3**, 010313 (2022)

How Expressive are our Ansatze?



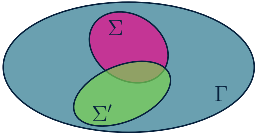
How Expressive are our Ansatze?





How Expressive are our Ansatze?

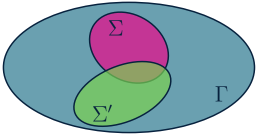
- Expressivity and trainability of *unitary ansatze* are well understood [1]



[1] Holmes, Z. *et al.*, *PRX Quantum* **3**, 010313 (2022)

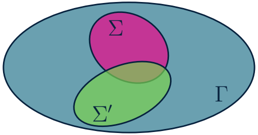
How Expressive are our Ansatze?

- Expressivity and trainability of *unitary ansatze* are well understood
- How does an ansatz compare to a *maximally expressive* reference ansatz?



How Expressive are our Ansatze?

- Expressivity and trainability of *unitary ansatze* are well understood
- How does an ansatz compare to a *maximally expressive* reference ansatz?
- How do generalized expressivity measures for channels depend on:
  - Noise induced phenomena
  - Underlying (parameterized) unitary evolution
  - Coupling with the environment



Expressivity Measures

- Let an *ensemble* of channels $\Lambda \sim \Sigma$ over a space $\mathcal{H}$ of dimension d have an average behaviour defined by the t-order *twirl*

$$\mathcal{T}_{\Sigma}^{(t)} = \int_{\Sigma} d\Lambda \; \Lambda^{\otimes t} \tag{15}$$

Expressivity Measures

- Let an *ensemble* of channels $\Lambda \sim \Sigma$ over a space $\mathcal{H}$ of dimension d have an average behaviour defined by the t-order *twirl*

$$\mathcal{T}_{\Sigma}^{(t)} = \int_{\Sigma} d\Lambda \; \Lambda^{\otimes t} \tag{15}$$

- This allows us to define an *expressivity* measure between ensembles

$$\mathcal{E}_{\Sigma\Sigma'}^{(t)2} = \|\mathcal{T}_{\Sigma}^{(t)} - \mathcal{T}_{\Sigma'}^{(t)}\|^2 \;\sim\; \|\mathcal{T}_{\Sigma}^{(t)}\|^2 \;-\; \|\mathcal{T}_{\Sigma'}^{(t)}\|^2 \;+\; \cdots \tag{16}$$

Expressivity Measures

- Let an *ensemble* of channels $\Lambda \sim \Sigma$ over a space $\mathcal{H}$ of dimension d have an average behaviour defined by the t-order *twirl*

$$\mathcal{T}_{\Sigma}^{(t)} = \int_{\Sigma} d\Lambda \; \Lambda^{\otimes t} \tag{15}$$

- This allows us to define an *expressivity* measure between ensembles

$$\mathcal{E}_{\Sigma\Sigma'}^{(t)2} = \|\mathcal{T}_{\Sigma}^{(t)} - \mathcal{T}_{\Sigma'}^{(t)}\|^2 \;\sim\; \|\mathcal{T}_{\Sigma}^{(t)}\|^2 \;-\; \|\mathcal{T}_{\Sigma'}^{(t)}\|^2 \;+\; \cdots \tag{16}$$

- Twirls are crucially *trace-preserving*, with *ensemble-dependent* expansions

$$\mathcal{T}_{\Sigma}^{(t)}(\cdot) \;=\; \boxed{\underbrace{\frac{\mathrm{tr}\,(\cdot)}{d^t} I}_{\text{Depolarizing}}} \;+\; \boxed{\underbrace{\Delta_{\Sigma}^{(t)}(\cdot)}_{\text{Deviations}}} \tag{17}$$

Expressivity Measures

- Let an *ensemble* of channels $\Lambda \sim \Sigma$ over a space $\mathcal{H}$ of dimension d have an average behaviour defined by the t-order *twirl*

$$\mathcal{T}_{\Sigma}^{(t)} = \int_{\Sigma} d\Lambda \; \Lambda^{\otimes t} \tag{15}$$

- This allows us to define an *expressivity* measure between ensembles

$$\mathcal{E}_{\Sigma\Sigma'}^{(t)2} = \|\mathcal{T}_{\Sigma}^{(t)} - \mathcal{T}_{\Sigma'}^{(t)}\|^2 \;\sim\; \|\mathcal{T}_{\Sigma}^{(t)}\|^2 \;-\; \|\mathcal{T}_{\Sigma'}^{(t)}\|^2 \;+\; \cdots \tag{16}$$

- Twirls are crucially *trace-preserving*, with *ensemble-dependent* expansions

$$\boxed{\mathcal{E}_{\Sigma}^{(t)} = \underbrace{\|\Delta_{\Sigma}^{(t)}\|}_{\text{Deviations}}} \tag{18}$$

Reference Ensembles

Channels present several choices for reference ensembles

Reference Ensembles

Channels present several choices for reference ensembles

- *Haar* ∼ Unitary Haar measure (uniformly random unitaries) [1]

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Haar}}}(\rho)} = \int_{\mathcal{U}(d_{\mathcal{H}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \; U^{\otimes t\,\dagger} \tag{19}$$

[1] Holmes, Z. *et al.*, *PRX Quantum* **3**, 010313 (2022)

Reference Ensembles

Channels present several choices for reference ensembles

- *Haar* ∼ Unitary Haar measure (uniformly random unitaries)

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Haar}}}(\rho)} = \int_{\mathcal{U}(d_{\mathcal{H}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \; U^{\otimes t\,\dagger} \tag{19}$$

- *cHaar* ∼ Stinespring Unitary Haar measure (random channels) [2]

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\mathrm{cHaar}}}(\rho)} = \mathrm{tr}_{\mathcal{E}}\left(\int_{\mathcal{U}(d_{\mathcal{H}} d_{\mathcal{E}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \otimes \; \nu_{\mathcal{E}}^{\otimes t} \; U^{\otimes t\,\dagger}\right) \tag{20}$$

[2] Kukulski, R. *et al.*, *J. Math. Phys.* **62**, 062201 (2021)

Reference Ensembles

Channels present several choices for reference ensembles

- *Haar* $\sim$ Unitary Haar measure (uniformly random unitaries)

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Haar}}}(\rho)} = \int_{\mathcal{U}(d_{\mathcal{H}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \; U^{\otimes t \,\dagger} \tag{19}$$

- *cHaar* $\sim$ Stinespring Unitary Haar measure (random channels)

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}(\rho)} = \operatorname{tr}_{\mathcal{E}}\left(\int_{\mathcal{U}(d_{\mathcal{H}} d_{\mathcal{E}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \otimes \; \nu_{\mathcal{E}}^{\otimes t} \; U^{\otimes t \,\dagger}\right) \tag{20}$$

- *Depolarizing* $\sim$ Maximally Depolarizing (single channel)

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Depolarize}}}(\rho)} = \frac{\operatorname{tr}\left(\rho^{\otimes t}\right)}{d_{\mathcal{H}}^{t}} \; I^{\otimes t} \tag{21}$$

Behaviour of Random Quantum Channels

The t-order Haar, cHaar, Depolarize ensembles are related by

$$\lim_{d_{\mathcal{E}} \to 1} \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}} \to \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Haar}}}} \qquad \lim_{\frac{d_{\mathcal{H}}}{d_{\mathcal{E}}} \to \infty} \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}} \to \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Depolarize}}}} \tag{22}$$

Behaviour of Random Quantum Channels

The t-order Haar, cHaar, Depolarize ensembles are related by

$$\lim_{d_{\mathcal{E}}\to 1} \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}} \to \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Haar}}}} \qquad \lim_{\frac{d_{\mathcal{H}}}{d_{\mathcal{E}}}\to\infty} \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}} \to \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Depolarize}}}} \tag{22}$$

The k-concatenated, t-order cHaar ensemble is *depolarizing* and *non-unital* [3]

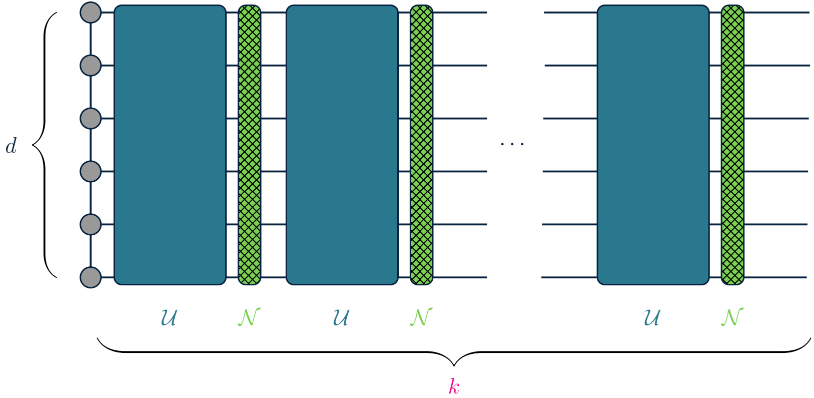
$$\lim_{\frac{d_{\mathcal{H}}}{d_{\mathcal{E}}}\to\infty} \mathcal{T}^{(t)k}_{\Sigma_{\text{cHaar}}}(\rho) = \boxed{\underbrace{\frac{\operatorname{tr}\left(\rho^{\otimes t}\right)}{d^t_{\mathcal{H}}} I^{\otimes t}}_{\text{Depolarize}}} + \boxed{\underbrace{O\left(\frac{1}{d^2_{\mathcal{H}} d_{\mathcal{E}}}\right) \sum_{P\neq I^{\otimes t}} P}_{\text{Non-Unital}}} \tag{23}$$

[3] Bai, J. *et al.*, *Quantum Information Processing* **23**, 1–18 (2024)

Relationships between Noise and Expressivity

Analytical *expressivities* for k layers of channels

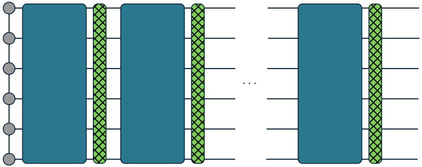
$$\Lambda_{\mathcal{U}_\gamma}^{(k)} = (\mathcal{N}_\gamma \circ \mathcal{U})^k = \frac{\operatorname{tr}(\cdot)}{d} I + \Delta_\gamma^{(k)}(\cdot) \tag{24}$$



Relationships between Noise and Expressivity

Haar Random Unitaries + Fixed *Unital* Noise: *Increases* Expressivity

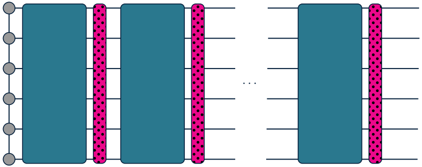
$$\mathcal{E}_{\mathcal{U}_{\gamma}}^{(t,k)2} = \binom{t}{2} (1-\gamma)^{4k} + O\left((1-\gamma)^{6k}\right) \tag{25}$$



Relationships between Noise and Expressivity

Haar Random Unitaries + Fixed *Non-Unital* Noise: *Decreases* Expressivity

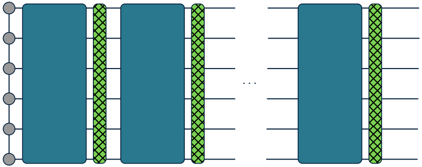
$$\mathcal{E}_{\mathcal{U}\gamma\eta}^{(t,k)2} = t\left(d^2 - 1\right)\ \eta^2\ +\ O\left(\eta^4\right) \tag{26}$$



Relationships between Noise and Expressivity

Parameterized Random Unitaries + Fixed *Unital* Noise: *Increases* Expressivity

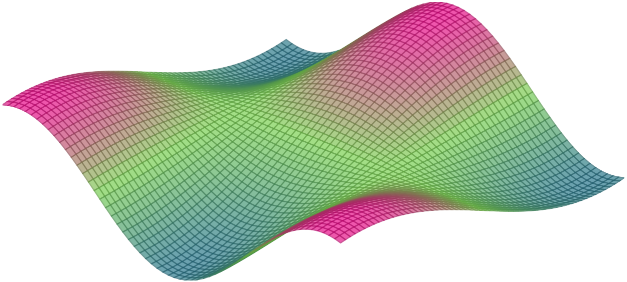
$$\mathcal{E}_{\mathcal{G}\gamma}^{(t,k)2} = t|\mathcal{S}_G \setminus \{I\}| \left(1-\gamma\right)^{2k} + O\left(\left(1-\gamma\right)^{2k+2}\right) \tag{27}$$



Relationships between Noise and Expressivity

Objective $\mathcal{L}$ and Gradient $\partial\mathcal{L}$ *Concentration*: $p\left(|\mathcal{L} - \mu_{\mathcal{L}}| \geq \epsilon\right) \leq \sigma_{\mathcal{L}}^2/\epsilon^2$

$$\mathcal{L}(\rho, O) = \mathrm{tr}\left(O\Lambda(\rho)\right)$$

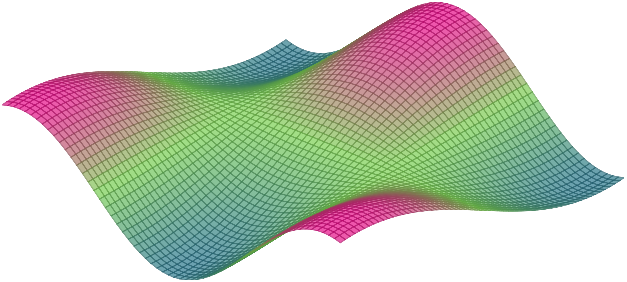


Relationships between Noise and Expressivity

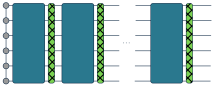
Objective $\mathcal{L}$ and Gradient $\partial\mathcal{L}$ *Concentration*: $p\left(|\mathcal{L}-\mu_{\mathcal{L}}|\geq\epsilon\right)\leq\sigma_{\mathcal{L}}^2/\epsilon^2$

$$\mathcal{L}(\rho,O)=\mathrm{tr}\left(O\Lambda(\rho)\right)$$

$$\sigma_{\mathcal{L}}^2\sim O\left(\frac{1}{d_{\mathcal{H}}d_{\mathcal{E}}}\right)\;+\;\|O\|_2^2\,\mathcal{E}_{\Sigma\Sigma'}^{(2)}(\rho)\quad(\text{with } \textit{caveats} \text{ on } \Sigma',\rho,O \text{ locality})\tag{28}$$



Relationships between Noise and Expressivity



- *Haar* Random Unitaries + Fixed *Unital* Noise:

$$\mathcal{E}_{\mathcal{U}\gamma}^{(t,k)2} = \binom{t}{2}(1-\gamma)^{4k} + O\left((1-\gamma)^{6k}\right) \tag{25}$$

- *Haar* Random Unitaries + Fixed *Non-Unital* Noise:

$$\mathcal{E}_{\mathcal{U}\gamma\eta}^{(t,k)2} = t\left(d^2-1\right)\eta^2 + O\left(\eta^4\right) \tag{26}$$

- *Parameterized* Random Unitaries + Fixed *Unital* Noise:

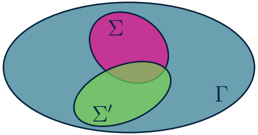
$$\mathcal{E}_{\mathcal{G}\gamma}^{(t,k)2} = t|\mathcal{S}_G\backslash\{I\}|(1-\gamma)^{2k} + O\left((1-\gamma)^{2k+2}\right) \tag{27}$$

- Objective $\mathcal{L}$ and Gradient $\partial\mathcal{L}$ *Concentration*: $p\left(|\mathcal{L}-\mu_{\mathcal{L}}| \geq \epsilon\right) \leq \sigma_{\mathcal{L}}^2/\epsilon^2$

$$\sigma_{\mathcal{L}}^2 \sim O\left(\frac{1}{d_{\mathcal{H}}d_{\mathcal{E}}}\right) + \|O\|_2^2\, \mathcal{E}_{\Sigma\Sigma'}^{(2)}[\rho] \tag{28}$$

Channels versus Unitary Ensembles

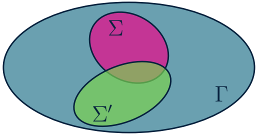
Many *subtle* differences between ensembles of channels and unitaries



Channels versus Unitary Ensembles

Many *subtle* differences between ensembles of channels and unitaries

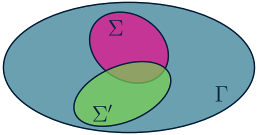
- Twirls are quasi-projections (quasi-commutant may defined via *dilation*)



Channels versus Unitary Ensembles

Many *subtle* differences between ensembles of channels and unitaries

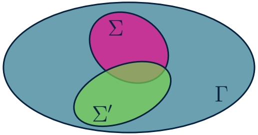
- Twirls are quasi-projections (quasi-commutant may defined via *dilation*)
- Open systems have degrees of freedom (environment *size*, *coupling* strength)



Channels versus Unitary Ensembles

Many *subtle* differences between ensembles of channels and unitaries

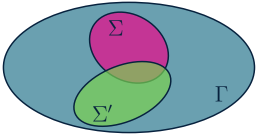
- Twirls are quasi-projections (quasi-commutant may defined via *dilation*)
- Open systems have degrees of freedom (environment *size*, *coupling* strength)
- Adjoint channels are not strictly *physical* channels (concentration *caveats*)



Channels versus Unitary Ensembles

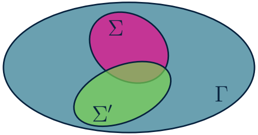
Many *subtle* differences between ensembles of channels and unitaries

- Twirls are quasi-projections (quasi-commutant may defined via *dilation*)
- Open systems have degrees of freedom (environment *size*, *coupling* strength)
- Adjoint channels are not strictly *physical* channels (concentration *caveats*)
- Subtleties in realizing channel t-designs in *practice*



Channels versus Unitary Ensembles

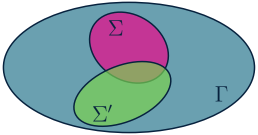
Ensembles of channels have inherently different *interpretations*



Channels versus Unitary Ensembles

Ensembles of channels have inherently different *interpretations*

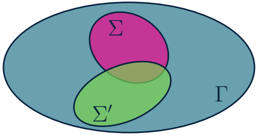
- *Uniformly Random*: cHaar channels are a *uniform* random ensemble



Channels versus Unitary Ensembles

Ensembles of channels have inherently different *interpretations*

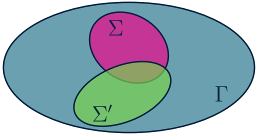
- *Uniformly Random*: cHaar channels are a *uniform* random ensemble
- *Capacities*: Depolarizing channels maximize environment *exchange entropy*



Channels versus Unitary Ensembles

Ensembles of channels have inherently different *interpretations*

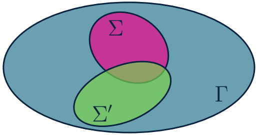
- *Uniformly Random*: cHaar channels are a *uniform* random ensemble
- *Capacities*: Depolarizing channels maximize environment *exchange entropy*
- *Tomography*: Depolarizing channels maximize *uncertainty* in measurements



Channels versus Unitary Ensembles

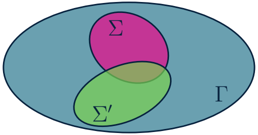
Ensembles of channels have inherently different *interpretations*

- *Uniformly Random*: cHaar channels are a *uniform* random ensemble
- *Capacities*: Depolarizing channels maximize environment *exchange entropy*
- *Tomography*: Depolarizing channels maximize *uncertainty* in measurements
- *Scrambling*: Depolarizing channels maximally *scramble* information



Operational Meaning of Channel Expressivity Measures

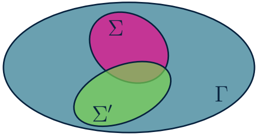
- *Noise induced* phenomena are actually channel *expressivity* phenomena!



Operational Meaning of Channel Expressivity Measures

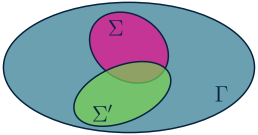
- *Noise induced* phenomena are actually channel *expressivity* phenomena!

- Channel expressivity is more subtly related to *usefulness* or *capability*



Operational Meaning of Channel Expressivity Measures

- *Noise induced* phenomena are actually channel *expressivity* phenomena!
- Channel expressivity is more subtly related to *usefulness* or *capability*
- Are there relationships between channel expressivity and their *simulability*?



Appendix

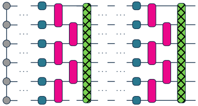
How May We Control Quantum Systems?

- Represented as *channels* $\Lambda_{\theta_\gamma} = \mathcal{N}_\gamma \circ \mathcal{U}_\theta$ with unitary evolution $\mathcal{U}_\theta$, and noise $\mathcal{N}_\gamma$
- Evolution generated by Hamiltonians with localized generators $\{G_\mu\}$

$$H_\theta^{(\mu)} = \sum_\nu \theta_\nu^{(\mu)} G_\nu \;\rightarrow\; U_\theta \approx \prod_{\mu,\nu}^{M} U_\theta^{(\mu,\nu)} \;:\; U_\theta^{(\mu,\nu)} = e^{-i\delta H_\theta^{(\mu,\nu)}} \approx e^{-i\delta\theta_\nu^{(\mu)} G_\nu} \tag{29}$$

i.e) *NMR* with variable transverse fields and constant longitudinal fields (Peterson *et al.* , PRA **13** (2020)) (Coloured in circuit ↘)

$$H_\theta^{(\mu)} = \sum_i \theta_i^{x(\mu)} X_i + \sum_i \theta_i^{y(\mu)} Y_i + \sum_i h_i Z_i + \sum_{i<j} J_{ij} Z_i Z_j \tag{30}$$

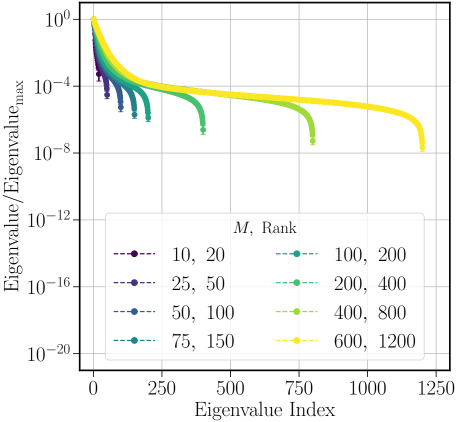


- Noise generated by constant *Kraus* operators $\{\mathcal{K}_{\gamma_\alpha}\}$
  i.e) Dephasing $\{\sqrt{1-\gamma}\; I,\; \sqrt{\gamma}\; Z\}$

$$\rho \;\rightarrow\; \rho_{\Lambda_{\theta_\gamma}} = \circ_\mu^M \mathcal{N}_\gamma \circ \mathcal{U}_\theta^{(\mu)}(\rho) = \circ_\mu^M \left[\sum_\alpha \mathcal{K}_{\gamma_\alpha}\, U_\theta^{(\mu)}\, \rho\, {U_\theta^{(\mu)}}^\dagger\, {\mathcal{K}_{\gamma_\alpha}}^\dagger\right] \tag{31}$$

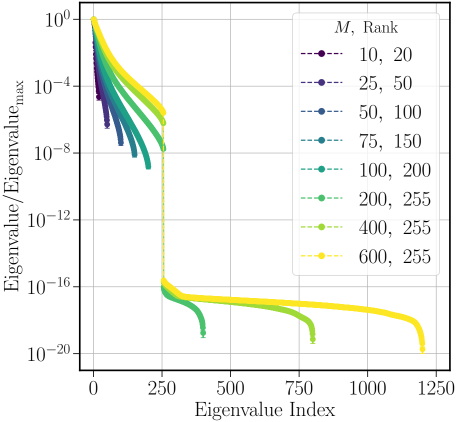
Overparameterization Phenomena

- *Overparameterized* regime is reached with constraints for sufficient depth $M > O(G)$ (Dynamical Lie Algebra $\mathcal{G}_{\mathrm{NMR}}$, with dimension $G = 2^{2N} - 1$)



(h) Hessian Rank Saturation

$$\mathcal{H}_{\mu\nu} = \partial_{\mu\nu}\mathcal{L}_\theta$$

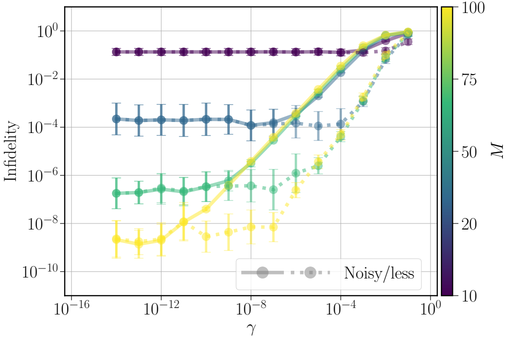


(i) Fisher Information Rank Saturation

$$\mathcal{F}_{\mu\nu} = \frac{1}{d}\mathrm{tr}\left(\partial_\mu U_\theta^\dagger \partial_\nu U_\theta\right) - \frac{1}{d^2}\mathrm{tr}\left(\partial_\mu U_\theta^\dagger U_\theta\right)\mathrm{tr}\left(U_\theta^\dagger \partial_\nu U_\theta\right)$$

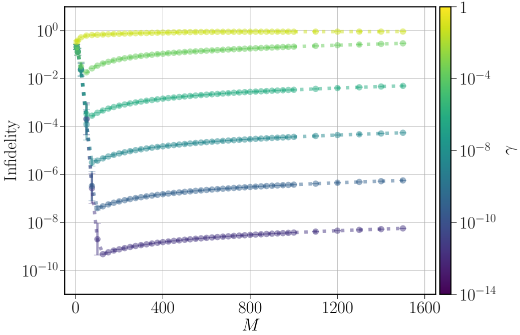
Noisy Optimization

- Haar random state preparation for $N = 4$ qubits, with independent dephasing



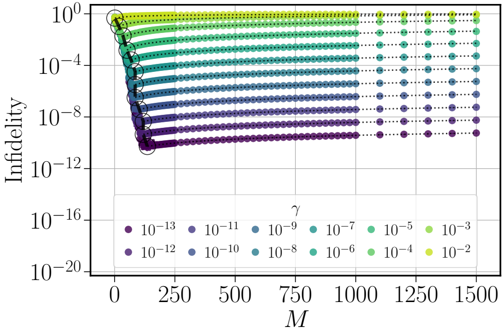
(j) Trained Noisy Infidelity, and Tested Infidelity of Noisy Parameters in Noiseless Ansatz

$\partial\mathcal{L}_{\theta\gamma} \sim \sum_\eta \alpha_\eta \, \mathcal{L}_{\theta+\eta\ \gamma}$

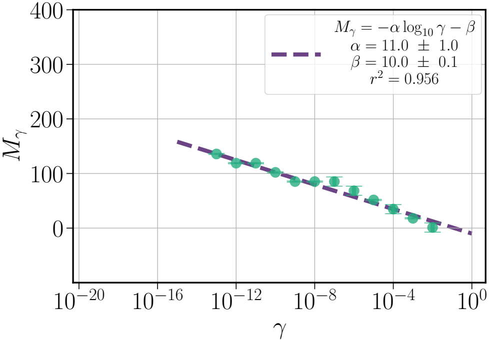


(k) Critical Depth for Noisy Infidelity

Noise Induced Critical Depth



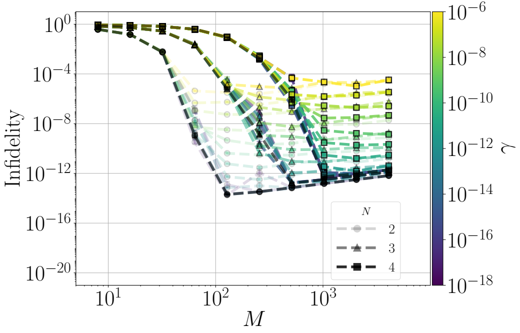
(l) Piecewise Fit of Noisy Infidelity



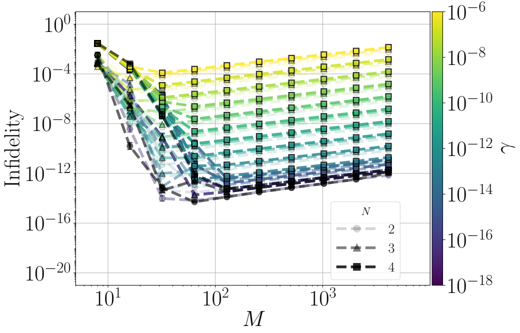
(m) Linear-Log Fit of Critical Depth

Universal Effects of Noise

- Effects of infidelities on noise for Haar random targets in $d = D^N$ dimensions



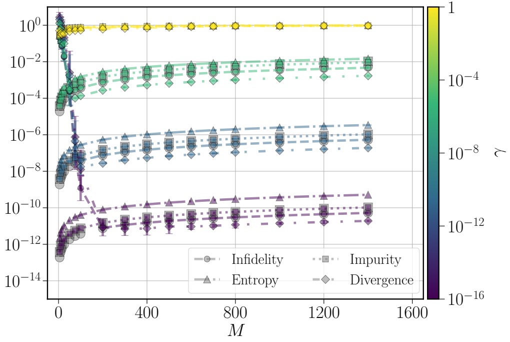
(n) Classical floating point noise for unitary compilation
$|\mathcal{L}_{\theta_\gamma} - \mathcal{L}_\theta| \leq |(1+\gamma)^{O(2NM)} - 1|$



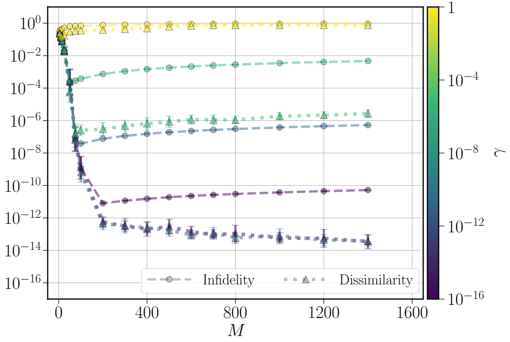
(o) Quantum dephasing noise for state preparation
$|\mathcal{L}_{\theta_\gamma} - \mathcal{L}_\theta| \leq 2|(1-\gamma)^{NM} - 1|$

Correlated Quantities

- Haar random state preparation for $N = 4$ qubits, with independent dephasing

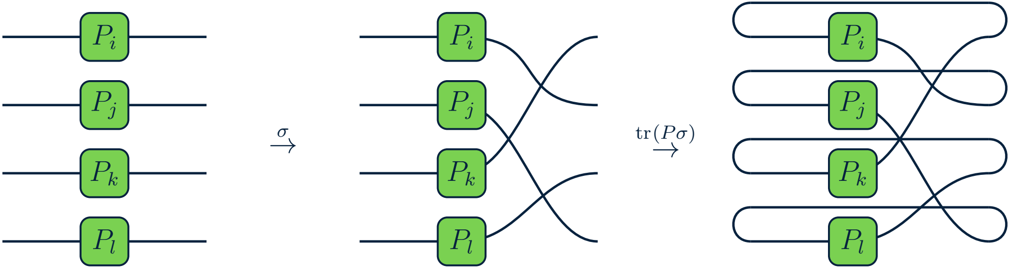


(p) Impurity, Entropy, Divergence



(q) Cosine Dissimilarity

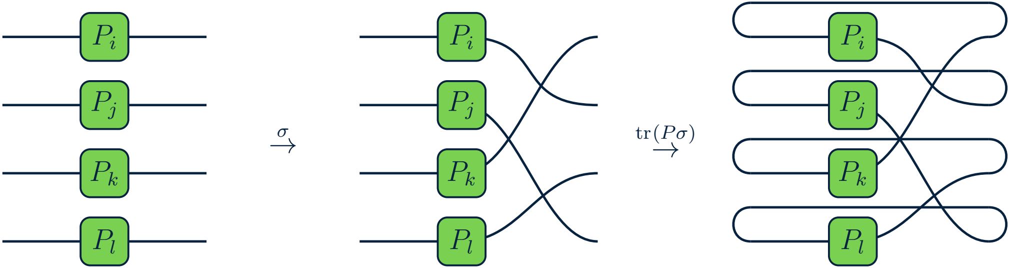
Diagrammatic Expansions of Permutations



$$P = P_i \otimes P_j \otimes P_k \otimes P_l \qquad \sigma = (ijlk) \qquad \mathrm{tr}\left(P\sigma\right) = \mathrm{tr}\left(\prod_{\mu\in\sigma} P_\mu\right)$$

$$\boxed{\tau = \frac{1}{d}\sum_{P\in\mathcal{P}_d} P\otimes P^{-1}} \quad \rightarrow \quad \boxed{\sigma = \frac{1}{d^{|\sigma|}}\sum_{P\in\mathcal{P}_d^{(\sigma)}} P} \tag{32}$$

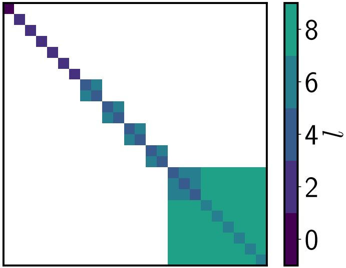
Diagrammatic Expansions of Permutations



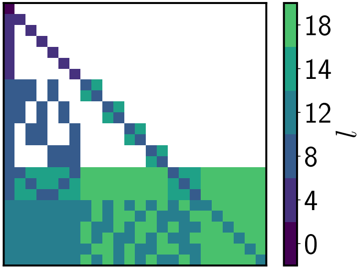
$$P = P_i \otimes P_j \otimes P_k \otimes P_l \qquad \sigma = (ijlk) \qquad \mathrm{tr}\,(P\sigma) = \mathrm{tr}\left(\prod\nolimits_{\mu\in\sigma} P_\mu\right)$$

$$\mathcal{T}_{\Sigma}^{(t)} = \boxed{\frac{1}{d^t}\sum_{\sigma,\pi\in\mathcal{S}_{\Sigma}^{(t)}} \tau_d^{(t)}(\sigma,\pi)|\sigma\rangle\langle\pi|} = \boxed{\frac{1}{d^t}|I\rangle\langle I| + \frac{1}{d^t}\sum_{P\in\mathcal{P}_d^{(\mathcal{S}_{\Sigma}^{(t)})}\setminus\{I\}} \tau_d^{(t)}(P,S)\,|P\rangle\langle S|} \tag{33}$$

Twirl Expansion Coefficients



(r) Haar Twirl Cycle Operator Coefficients $\tau^{(t)}_{\mathbb{U}(d)}(P,S) \sim O(1/d^l)$ for $t=4$



(r) cHaar Twirl Cycle Operator Coefficients $\tau_{\mathbb{E}(d_{\mathcal{H}},d_{\mathcal{E}})}(P,S) \sim O(1/d^l)$ for $t=4$

$$\mathcal{T}^{(t)}_{\Sigma} = \frac{1}{d^t} \sum_{P,S \in \mathcal{S}^{(t)}_{\Sigma}} \tau^{(t)}_{\Sigma}(P,S)|P\rangle\langle S| \tag{34}$$

Haar, cHaar, and Depolarizing Ensembles

$\Sigma \backslash t$	1	2
Haar	$\frac{1}{d_{\mathcal{H}}}\|I\rangle\!\langle I\|$	$\frac{1}{d_{\mathcal{H}}^2}\|I\rangle\!\langle I\| + \frac{1}{d_{\mathcal{H}}^2}\frac{1}{d_{\mathcal{H}}^2-1}\sum_{P,S\in\mathcal{P}_{d_{\mathcal{H}}}^{(\tau)}\setminus\{I\}}\|P\rangle\!\langle S\|$
cHaar	$\frac{1}{d_{\mathcal{H}}}\|I\rangle\!\langle I\|$	$\frac{1}{d_{\mathcal{H}}^2}\|I\rangle\!\langle I\| + \frac{1}{d_{\mathcal{H}}^2}\frac{d_{\mathcal{E}}-1}{d_{\mathcal{H}}^2 d_{\mathcal{E}}^2-1}\sum_{P\in\mathcal{P}_{d_{\mathcal{H}}}^{(\tau)}\setminus\{I\}}\|P\rangle\!\langle I\| + \frac{1}{d_{\mathcal{H}}^2}\frac{d_{\mathcal{E}}}{d_{\mathcal{H}}^2 d_{\mathcal{E}}^2-1}\sum_{P,S\in\mathcal{P}_{d_{\mathcal{H}}}^{(\tau)}\setminus\{I\}}\|P\rangle\!\langle S\|$
Depolarize	$\frac{1}{d_{\mathcal{H}}^t}\|I\rangle\!\langle I\|$	

Table 1: Twirls $\mathcal{T}_{\Sigma}^{(t)}$ for various ensembles and moments

Monotonic Convergence and Hierarchy of cHaar Twirl Norms

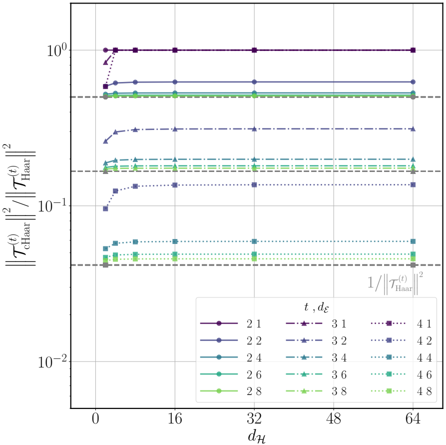


Figure 10: cHaar t-order twirl norms convergence with $d_{\mathcal{H}}, d_{\mathcal{E}}$ towards $1/\|\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Haar}}}\|^2$.

$$\boxed{1 = \|\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Depolarize}}}\|^2 \leq \|\mathcal{T}^{(t)k}_{\Sigma_{\mathrm{cHaar}}}\|^2 \leq \|\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Haar}}}\|^2 = |\mathcal{S}_t|} \tag{35}$$

Subtleties with Concentration of Channel Ansatz

Given $\Lambda \sim \mathcal{U}_\theta$ for $U_\theta = e^{-i\theta G}$, with involutory generators G and pure inputs ρ: Objective $\mathcal{L}_\Lambda$ variance concentrates as

$$\sigma^2_{\mathcal{L}} \sim O\left(\frac{1}{d_{\mathcal{H}} d_{\mathcal{E}}}\right) + \|O\|_p^2 \, \mathcal{E}^{(2|q)}_{\Sigma\Sigma'}[\rho] \tag{36}$$

$$\sigma^2_{\mathcal{L}_\Lambda|\Sigma}[\rho, O] \leq \begin{cases} O\left(\frac{d_{\mathcal{O}}}{d_{\mathcal{E}}} \frac{1}{d^2_{\mathcal{H}}}\right) \; + \; \min_{\frac{1}{p}+\frac{1}{q}=1} \|O\|_q^2 \; \mathcal{E}^{(2|p)}_{\Sigma\Sigma'} & \{O_{\text{Pauli}} \; , \; \Sigma'_{\text{cHaar}}\} \\ O\left(\frac{1}{d^2_{\mathcal{H}}}\right) \; + \; \min_{\frac{1}{p}+\frac{1}{q}=1} \|O\|_q^2 \; \mathcal{E}^{(2|p)}_{\Sigma\Sigma'} & \{O_{\text{Projector}} \; , \; \Sigma'_{\text{cHaar}}\} \\ \min_{\frac{1}{p}+\frac{1}{q}=1} \|O\|_q^2 \; \mathcal{E}^{(2|p)}_{\Sigma\Sigma'} & \left\{O_{\text{Pauli}} \; , \; \Sigma'_{\text{Depolarize}}\right\} \\ O\left(\frac{1}{d^2_{\mathcal{H}}}\right) \; + \; \min_{\frac{1}{p}+\frac{1}{q}=1} \|O\|_q^2 \; \mathcal{E}^{(2|p)}_{\Sigma\Sigma'} & \left\{O_{\text{Projector}} \; , \; \Sigma'_{\text{Depolarize}}\right\} \end{cases} . \tag{37}$$

Subtleties with Concentration of Channel Ansatz

Given $\Lambda \sim \mathcal{U}_\theta$ for $U_\theta = e^{-i\theta G}$, with involutory generators G and pure inputs ρ:
Objective gradient $\partial_\mu \mathcal{L}_\Lambda$ variance concentrates as

$$\sigma^2_{\partial_\mu \mathcal{L}} \sim O\left(\frac{1}{d_{\mathcal{H}} d_{\mathcal{E}}}\right) + O\left(\mathcal{E}^{(2|p^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho]\ \mathcal{E}^{(2|\dagger q^*)}_{\Sigma_{\mu_L}\Sigma'_{\mu_L}}[O]\right) \tag{38}$$

$$\sigma^2_{\partial_\mu \mathcal{L}_\Lambda|\Sigma\Sigma'_{RL}}[\rho, O] \leq \sigma^{2RL}_{\partial_\mu \mathcal{L}_\Lambda|\Sigma'_{\mu_{RL}}}[\rho, O] + \begin{cases} \min_{\frac{1}{p}+\frac{1}{q}=1} O\left(\frac{1}{d^2_{\mathcal{H}} d^2_{\mathcal{E}}}\|S\|_p\right) \mathcal{E}^{(2|q^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho] + & \{O_{\text{Orthogonal}}\ ,\ \Sigma'_{\text{cHaar}}\} \\ O\left(\frac{1}{d^2_{\mathcal{H}} d_{\mathcal{E}}}\|S\|_p\right) \mathcal{E}^{(2|q^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[O] + 4\ \mathcal{E}^{(2|p^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho]\ \mathcal{E}^{(2|\dagger q^*)}_{\Sigma_{\mu_L}\Sigma'_{\mu_L}}[O] & \\ \min_{\frac{1}{p}+\frac{1}{q}=1} O\left(\frac{1}{d^3_{\mathcal{H}} d^2_{\mathcal{E}}}\|S\|_p\right) \mathcal{E}^{(2|q^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho] + & \{O_{\text{Projector}}\ ,\ \Sigma'_{\text{cHaar}}\} \\ O\left(\frac{1}{d^2_{\mathcal{H}} d_{\mathcal{E}}}\|S\|_p\right) \mathcal{E}^{(2|q^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[O] + 4\ \mathcal{E}^{(2|p^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho]\ \mathcal{E}^{(2|\dagger q^*)}_{\Sigma_{\mu_L}\Sigma'_{\mu_L}}[O] & \\ \min_{\frac{1}{p}+\frac{1}{q}=1} 4\ \mathcal{E}^{(2|p^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho]\ \mathcal{E}^{(2|\dagger q^*)}_{\Sigma_{\mu_L}\Sigma'_{\mu_L}}[O] & \left\{\Sigma'_{\text{Depolarize}}\right\} \end{cases} \tag{39}$$

where the *left* (L) and *right* (R) 2-design gradient variance is

$$\sigma^{2RL}_{\partial_\mu \mathcal{L}_\Lambda|\Sigma'_{\mu_{RL}}}[\rho, O] = \begin{cases} O\left(\frac{1}{d_{\mathcal{H}} d^2_{\mathcal{E}}}\right) & \{O_{\text{Orthogonal}}\ ,\ \Sigma'_{\text{cHaar}}\} \\ O\left(\frac{1}{d^2_{\mathcal{H}} d^2_{\mathcal{E}}}\right) & \{O_{\text{Projector}}\ ,\ \Sigma'_{\text{cHaar}}\} \\ 0 & \left\{\Sigma'_{\text{Depolarize}}\right\} \end{cases} \tag{40}$$