Non-Unitary Measures of Expressivity and their Operational Meaning

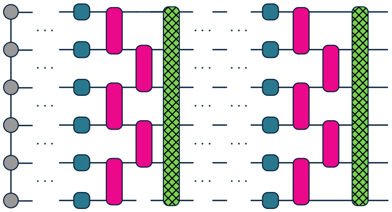
Matt Duschenes*, Diego García-Martín, Martín Larocca, Zoë Holmes, Marco Cerezo

Los Alamos National Laboratory

Fall, 2024

arXiv:2409.XXXXX

Seminar



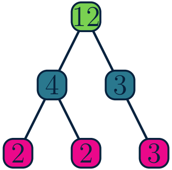
Quantum Tasks Of Interest

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

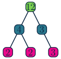
- *Quantum algorithms* i.e) Factoring numbers



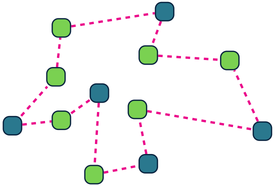
Quantum Tasks Of Interest

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

- *Quantum algorithms* i.e) Factoring numbers



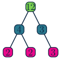
- *Optimization* problems i.e) Travelling Salesman Problem



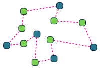
Quantum Tasks Of Interest

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

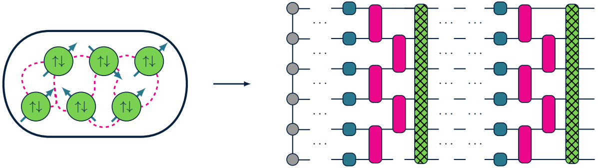
- *Quantum algorithms* i.e) Factoring numbers



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

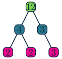


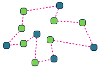
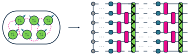
- *Compilation* tasks i.e) Form operators U given native gates $\{V\}$

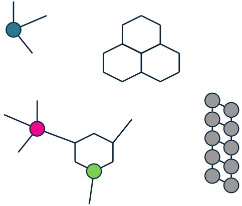


Quantum Tasks Of Interest

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

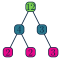
- *Quantum algorithms* i.e) Factoring numbers 

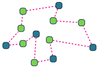
- *Optimization* problems i.e) Travelling Salesman Problem 
- *Compilation* tasks i.e) Form operators U given native gates $\{V\}$ 
- *Simulate* quantum systems i.e) Complicated molecules and chemical reactions

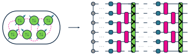


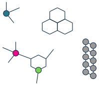
Quantum Tasks Of Interest

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

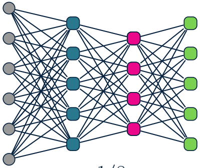
- *Quantum algorithms* i.e) Factoring numbers 

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

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

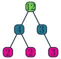
- *Simulate* quantum systems i.e) Complicated molecules and chemical reactions 

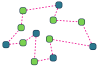
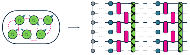
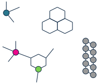
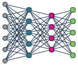
- *Machine learning* functions i.e) Classification, Regression, Generative

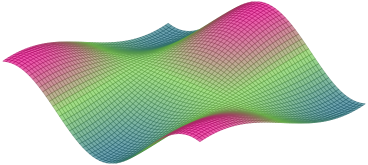


Quantum Tasks Of Interest

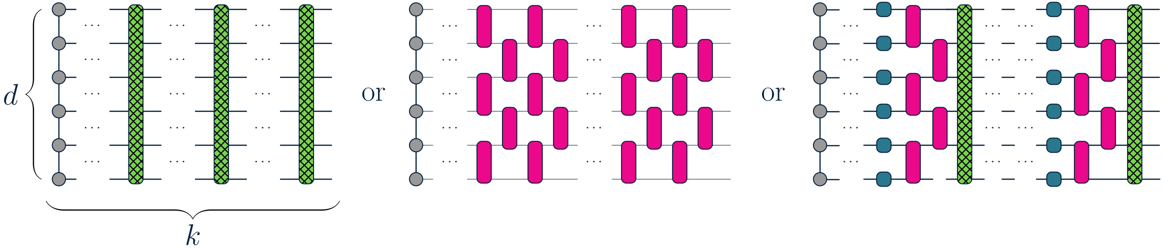
Ultimately, we want to do something *useful* with our quantum devices

- *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 

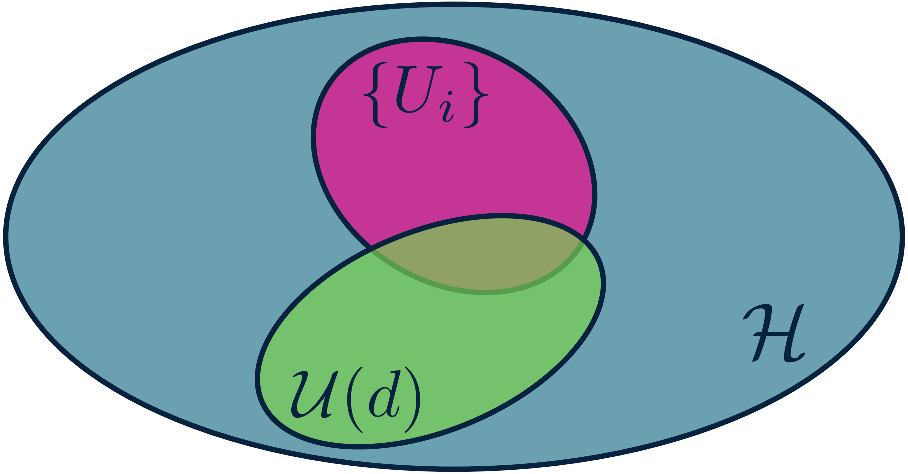


How Expressive are our Ansatze?



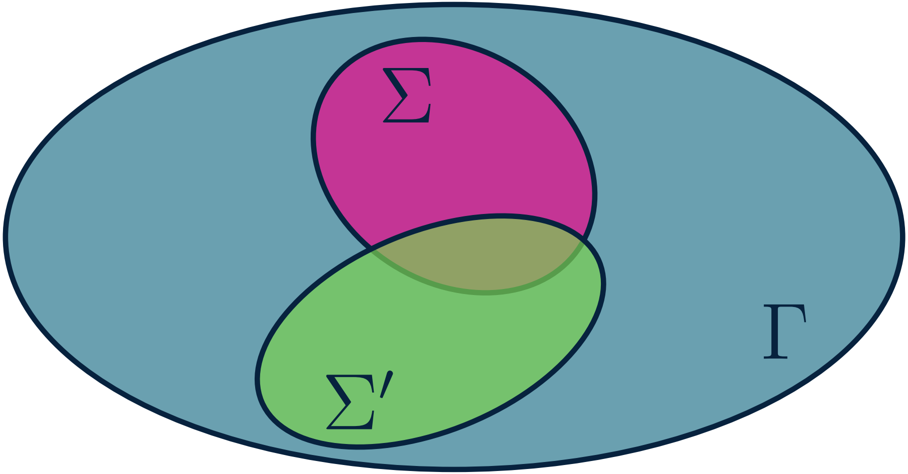
Los Alamos NATIONAL LABORATORY EST. 1943

How Expressive are our Ansatze?

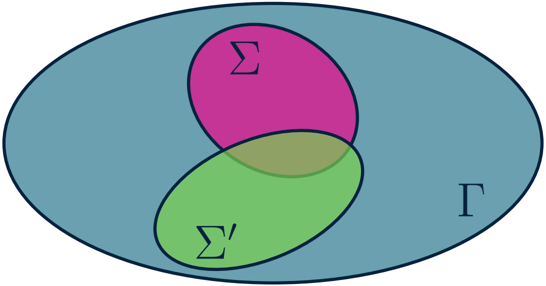


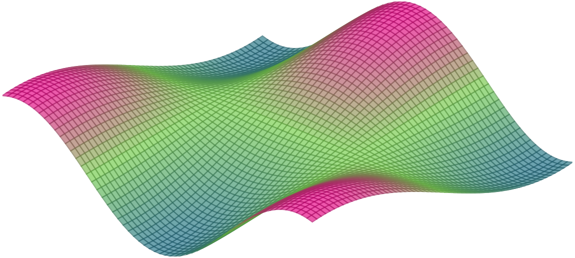
[1] Holmes, Z. *et al.*, *PRX Quantum* **3**, 010313 (2022)

How Expressive are our Ansatze?



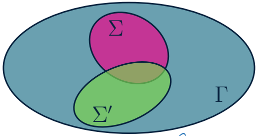
How Expressive are our Ansatze?





How Expressive are our Ansatze?

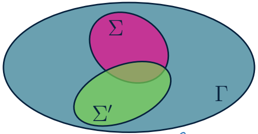
- Expressivity and trainability of *unitary ansatze* are well understood [1]



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

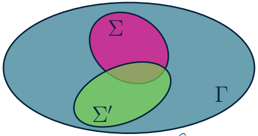
How Expressive are our Ansatze?

- 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*

$$\mathcal{T}_{\Sigma}^{(t)} = \int_{\Sigma} d\Lambda \; \Lambda^{\otimes t} \tag{1}$$

Los Alamos National Laboratory EST. 1943

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*

$$\mathcal{T}_{\Sigma}^{(t)} = \int_{\Sigma} d\Lambda\ \Lambda^{\otimes t} \tag{1}$$

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

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

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*

$$\mathcal{T}_{\Sigma}^{(t)} = \int_{\Sigma} d\Lambda \; \Lambda^{\otimes t} \tag{1}$$

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

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

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

$$\mathcal{T}_{\Sigma}^{(t)}(\cdot) \;=\; \boxed{\underbrace{\frac{\mathrm{tr}\,(\cdot)}{d^t} I}_{\text{Depolarizing}}} \;+\; \boxed{\underbrace{\Delta_{\Sigma}^{(t)}(\cdot)}_{\text{Deviations}}} \tag{3}$$

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*

$$\mathcal{T}_{\Sigma}^{(t)} = \int_{\Sigma} d\Lambda \; \Lambda^{\otimes t} \tag{1}$$

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

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

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

$$\boxed{\mathcal{E}_{\Sigma}^{(t)} = \underbrace{\|\Delta_{\Sigma}^{(t)}\|}_{\text{Deviations}}} \tag{4}$$

Reference Ensembles

Channels present several choices for reference ensembles

Reference Ensembles

Channels present several choices for reference ensembles

- *Haar* $\sim$ Unitary Haar measure (uniformly random unitaries) [1]

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Haar}}}(\rho)} = \int_{\mathcal{U}(d_{\mathcal{H}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \; U^{\otimes t \, \dagger} \tag{5}$$

[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)

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Haar}}}(\rho)} = \int_{\mathcal{U}(d_{\mathcal{H}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \; U^{\otimes t\,\dagger} \tag{5}$$

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

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\mathrm{cHaar}}}(\rho)} = \mathrm{tr}_{\mathcal{E}}\left(\int_{\mathcal{U}(d_{\mathcal{H}} d_{\mathcal{E}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \otimes \; \nu_{\mathcal{E}}^{\otimes t} \; U^{\otimes t\,\dagger}\right) \tag{6}$$

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

Reference Ensembles

Channels present several choices for reference ensembles

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

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Haar}}}(\rho)} = \int_{\mathcal{U}(d_{\mathcal{H}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \; U^{\otimes t \, \dagger} \tag{5}$$

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

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}(\rho)} = \operatorname{tr}_{\mathcal{E}} \left(\int_{\mathcal{U}(d_{\mathcal{H}} d_{\mathcal{E}})} dU \; U^{\otimes t} \; \rho^{\otimes t} \otimes \; \nu_{\mathcal{E}}^{\otimes t} \; U^{\otimes t \, \dagger} \right) \tag{6}$$

- *Depolarizing* ∼ Maximally Depolarizing (single channel)

$$\boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Depolarize}}}(\rho)} = \frac{\operatorname{tr}\left(\rho^{\otimes t}\right)}{d_{\mathcal{H}}^{t}} \; I^{\otimes t} \tag{7}$$

Behaviour of Random Quantum Channels

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

$$\lim_{d_{\mathcal{E}}\to 1} \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}} \to \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Haar}}}} \qquad \lim_{\frac{d_{\mathcal{H}}}{d_{\mathcal{E}}}\to\infty} \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}} \to \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Depolarize}}}} \tag{8}$$

Behaviour of Random Quantum Channels

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

$$\lim_{d_{\mathcal{E}} \to 1} \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}} \to \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Haar}}}} \qquad \lim_{\frac{d_{\mathcal{H}}}{d_{\mathcal{E}}} \to \infty} \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{cHaar}}}} \to \boxed{\mathcal{T}^{(t)}_{\Sigma_{\text{Depolarize}}}} \tag{8}$$

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

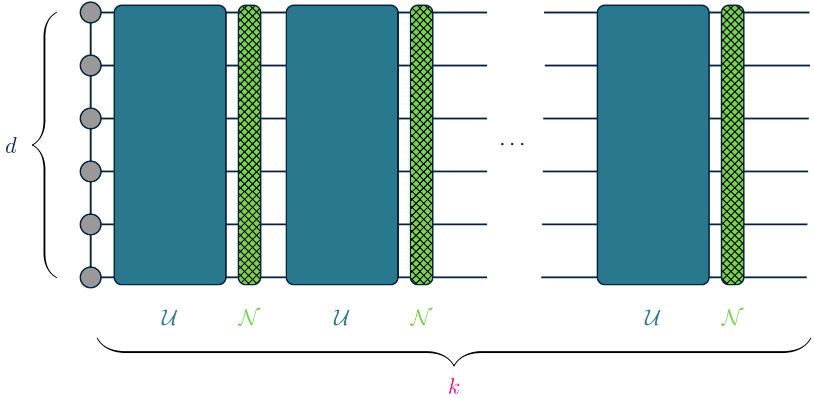
$$\lim_{\frac{d_{\mathcal{H}}}{d_{\mathcal{E}}} \to \infty} \mathcal{T}^{(t)k}_{\Sigma_{\text{cHaar}}}(\rho) = \boxed{\underbrace{\frac{\operatorname{tr}\left(\rho^{\otimes t}\right)}{d_{\mathcal{H}}^{t}} I^{\otimes t}}_{\text{Depolarize}}} + \boxed{\underbrace{O\left(\frac{1}{d_{\mathcal{H}}^{2} d_{\mathcal{E}}}\right) \sum_{P \neq I^{\otimes t}} P}_{\text{Non-Unital}}} \tag{9}$$

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

Relationships between Noise and Expressivity

Analytical *expressivities* for k layers of specific channel *ansatze*

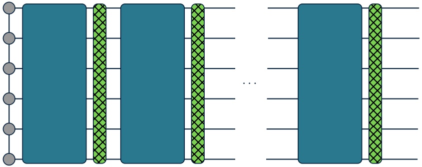
$$\Lambda^{(k)}_{\mathcal{U}_\gamma}(\rho) = (\mathcal{N}_\gamma \circ \mathcal{U})^k(\rho) = \frac{\operatorname{tr}(\rho)}{d} I + \Delta^{(k)}_\gamma(\rho) \tag{10}$$



Relationships between Noise and Expressivity

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

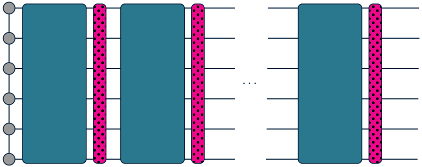
$$\mathcal{E}_{\mathcal{U}_\gamma}^{(t,k)} = O\left((1-\gamma)^{2k}\right) \tag{11}$$



Relationships between Noise and Expressivity

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

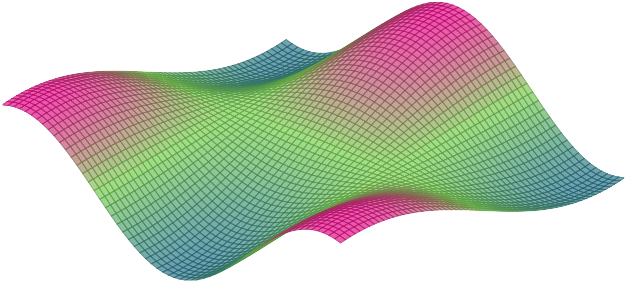
$$\mathcal{E}^{(t,k)}_{\mathcal{U}\gamma\eta} = O\left(\eta\right) \tag{12}$$



Relationships between Noise and Expressivity

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

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

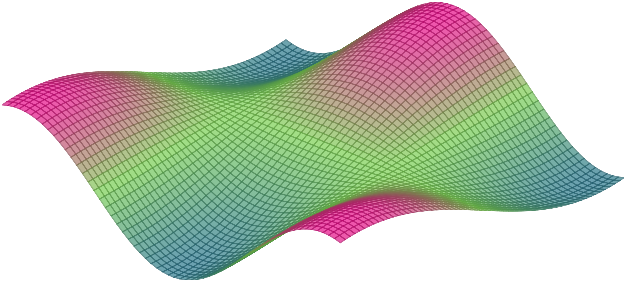


Relationships between Noise and Expressivity

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

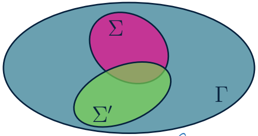
$$\mathcal{L}(\rho, O) = \mathrm{tr}\left(O\Lambda(\rho)\right)$$

$$\sigma_{\mathcal{L}}^2 \sim O\left(\frac{1}{d_{\mathcal{H}} d_{\mathcal{E}}}\right) \; + \; \|O\|_2^2 \, \mathcal{E}_{\Sigma\Sigma'}^{(2)}(\rho) \quad \text{(with } \textit{caveats} \text{ on } \Sigma', \rho, O \text{ locality)} \tag{13}$$



Channels versus Unitary Ensembles

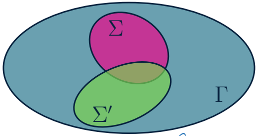
Many *subtle* differences between ensembles of channels and unitaries



Channels versus Unitary Ensembles

Many *subtle* differences between ensembles of channels and unitaries

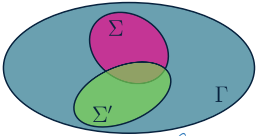
- Twirls are quasi-projections (quasi-commutant may defined via *dilation*)



Channels versus Unitary Ensembles

Many *subtle* differences between ensembles of channels and unitaries

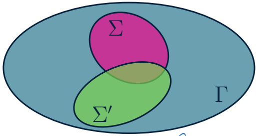
- Twirls are quasi-projections (quasi-commutant may defined via *dilation*)
- Open systems have degrees of freedom (environment *size*, *coupling* strength)



Channels versus Unitary Ensembles

Many *subtle* differences between ensembles of channels and unitaries

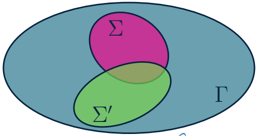
- Twirls are quasi-projections (quasi-commutant may defined via *dilation*)
- Open systems have degrees of freedom (environment *size*, *coupling* strength)
- Adjoint channels are not strictly *physical* channels (concentration *caveats*)



Channels versus Unitary Ensembles

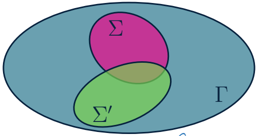
Many *subtle* differences between ensembles of channels and unitaries

- 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

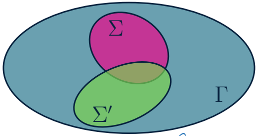
Ensembles of channels have inherently different *interpretations*



Channels versus Unitary Ensembles

Ensembles of channels have inherently different *interpretations*

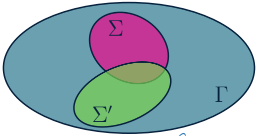
- *Uniformly Random*: cHaar channels are a *uniform* random ensemble



Channels versus Unitary Ensembles

Ensembles of channels have inherently different *interpretations*

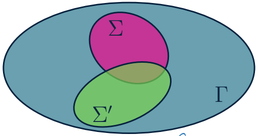
- *Uniformly Random*: cHaar channels are a *uniform* random ensemble
- *Capacities*: Depolarizing channels maximize environment *exchange entropy*



Channels versus Unitary Ensembles

Ensembles of channels have inherently different *interpretations*

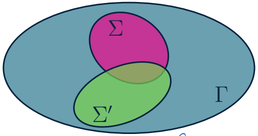
- *Uniformly Random*: cHaar channels are a *uniform* random ensemble
- *Capacities*: Depolarizing channels maximize environment *exchange entropy*
- *Tomography*: Depolarizing channels maximize *uncertainty* in measurements



Channels versus Unitary Ensembles

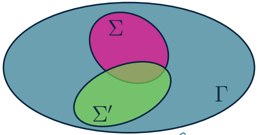
Ensembles of channels have inherently different *interpretations*

- *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

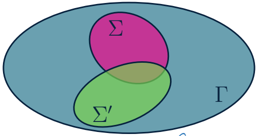
- *Noise induced* phenomena are actually channel *expressivity* phenomena!



Operational Meaning of Channel Expressivity Measures

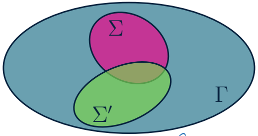
- *Noise induced* phenomena are actually channel *expressivity* phenomena!

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



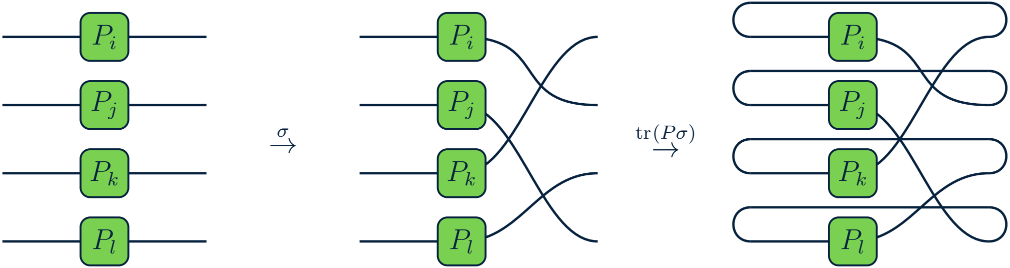
Operational Meaning of Channel Expressivity Measures

- *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

Diagrammatic Expansions of Permutations



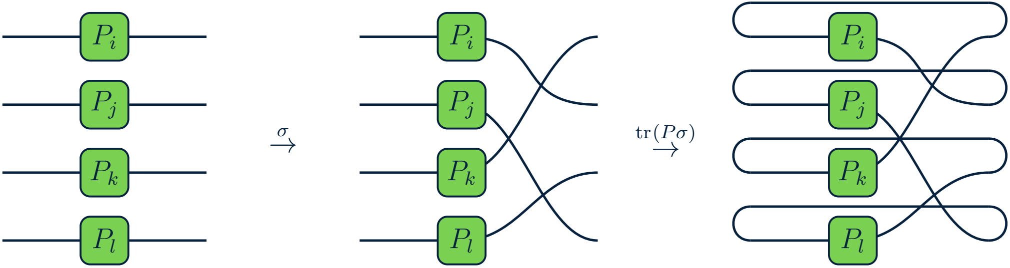
$P = P_i \otimes P_j \otimes P_k \otimes P_l$

$\sigma = (ijlk)$

$\mathrm{tr}\,(P\sigma) = \mathrm{tr}\left(\prod_{\mu\in\sigma} P_\mu\right)$

$$\boxed{\tau = \frac{1}{d}\sum_{P\in\mathcal{P}_d} P\otimes P^{-1}} \quad \rightarrow \quad \boxed{\sigma = \frac{1}{d^{|\sigma|}}\sum_{P\in\mathcal{P}_d^{(\sigma)}} P} \tag{14}$$

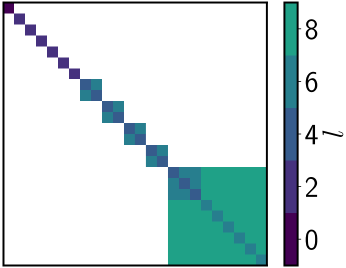
Diagrammatic Expansions of Permutations



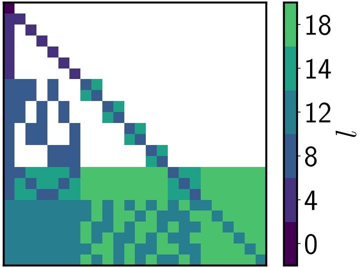
$$P = P_i \otimes P_j \otimes P_k \otimes P_l \qquad \sigma = (ijlk) \qquad \mathrm{tr}\,(P\sigma) = \mathrm{tr}\left(\prod_{\mu\in\sigma} P_\mu\right)$$

$$\mathcal{T}_\Sigma^{(t)} = \boxed{\frac{1}{d^t}\sum_{\sigma,\pi\in\mathcal{S}_\Sigma^{(t)}} \tau_d^{(t)}(\sigma,\pi)|\sigma\rangle\langle\pi|} = \boxed{\frac{1}{d^t}|I\rangle\langle I| + \frac{1}{d^t}\sum_{P\in\mathcal{P}_d^{(\mathcal{S}_\Sigma^{(t)})}\setminus\{I\}} \tau_d^{(t)}(P,S)\,|P\rangle\langle S|} \tag{15}$$

Twirl Expansion Coefficients



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



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

$$\mathcal{T}_{\Sigma}^{(t)} = \frac{1}{d^t} \sum_{P,S \in \mathcal{S}_{\Sigma}^{(t)}} \tau_{\Sigma}^{(t)}(P,S) |P\rangle\langle S| \tag{16}$$

Haar, cHaar, and Depolarizing Ensembles

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

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

Monotonic Convergence and Hierarchy of cHaar Twirl Norms

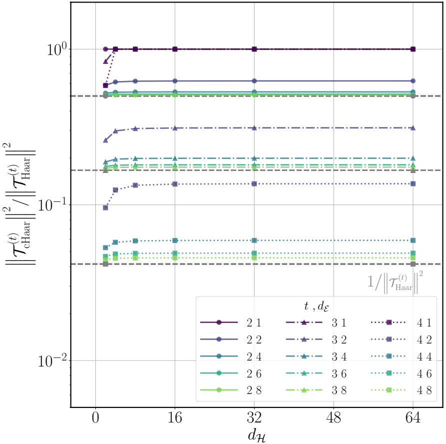


Figure 2: cHaar t-order twirl norms convergence with $d_{\mathcal{H}}, d_{\mathcal{E}}$ towards $1/\|\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Haar}}}\|^2$.

$$\boxed{1 = \|\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Depolarize}}}\|^2 \leq \|\mathcal{T}^{(t)k}_{\Sigma_{\mathrm{cHaar}}}\|^2 \leq \|\mathcal{T}^{(t)}_{\Sigma_{\mathrm{Haar}}}\|^2 = |\mathcal{S}_t|} \tag{17}$$

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

$$\sigma_{\mathcal{L}}^2 \sim O\left(\frac{1}{d_{\mathcal{H}} d_{\mathcal{E}}}\right) + \|O\|_p^2 \, \mathcal{E}_{\Sigma\Sigma'}^{(2|q)}[\rho] \tag{18}$$

$$\sigma^2_{\mathcal{L}_\Lambda|\Sigma}[\rho, O] \leq \begin{cases} O\left(\frac{d_{\mathcal{O}}}{d_{\mathcal{E}}} \frac{1}{d_{\mathcal{H}}^2}\right) + \min_{\frac{1}{p}+\frac{1}{q}=1} \|O\|_q^2 \, \mathcal{E}_{\Sigma\Sigma'}^{(2|p)} & \{O_{\text{Pauli}} \; , \; \Sigma'_{\text{cHaar}}\} \\ O\left(\frac{1}{d_{\mathcal{H}}^2}\right) + \min_{\frac{1}{p}+\frac{1}{q}=1} \|O\|_q^2 \, \mathcal{E}_{\Sigma\Sigma'}^{(2|p)} & \{O_{\text{Projector}} \; , \; \Sigma'_{\text{cHaar}}\} \\ \min_{\frac{1}{p}+\frac{1}{q}=1} \|O\|_q^2 \, \mathcal{E}_{\Sigma\Sigma'}^{(2|p)} & \left\{O_{\text{Pauli}} \; , \; \Sigma'_{\text{Depolarize}}\right\} \\ O\left(\frac{1}{d_{\mathcal{H}}^2}\right) + \min_{\frac{1}{p}+\frac{1}{q}=1} \|O\|_q^2 \, \mathcal{E}_{\Sigma\Sigma'}^{(2|p)} & \left\{O_{\text{Projector}} \; , \; \Sigma'_{\text{Depolarize}}\right\} \end{cases} . \tag{19}$$

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

$$\sigma^2_{\partial_\mu \mathcal{L}} \sim O\left(\frac{1}{d_{\mathcal{H}} d_{\mathcal{E}}}\right) + O\left(\mathcal{E}^{(2|p^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho]\ \mathcal{E}^{(2|\dagger q^*)}_{\Sigma_{\mu_L}\Sigma'_{\mu_L}}[O]\right) \tag{20}$$

$$\sigma^2_{\partial_\mu \mathcal{L}_\Lambda | \Sigma\Sigma'_{RL}}[\rho, O] \leq \sigma^{2RL}_{\partial_\mu \mathcal{L}_\Lambda | \Sigma'_{\mu_{RL}}}[\rho, O] + \begin{cases} \min_{\frac{1}{p}+\frac{1}{q}=1} O\left(\frac{1}{d_{\mathcal{H}}^2 d_{\mathcal{E}}^2}\|S\|_p\right) \mathcal{E}^{(2|q^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho] \ + & \{O_{\text{Orthogonal}}\ ,\ \Sigma'_{\text{cHaar}}\} \\ \qquad O\left(\frac{1}{d_{\mathcal{H}}^2 d_{\mathcal{E}}}\|S\|_p\right) \mathcal{E}^{(2|q^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[O] \ + \ 4\ \mathcal{E}^{(2|p^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho]\ \mathcal{E}^{(2|\dagger q^*)}_{\Sigma_{\mu_L}\Sigma'_{\mu_L}}[O] & \\ \min_{\frac{1}{p}+\frac{1}{q}=1} O\left(\frac{1}{d_{\mathcal{H}}^3 d_{\mathcal{E}}^2}\|S\|_p\right) \mathcal{E}^{(2|q^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho] \ + & \{O_{\text{Projector}}\ ,\ \Sigma'_{\text{cHaar}}\} \\ \qquad O\left(\frac{1}{d_{\mathcal{H}}^2 d_{\mathcal{E}}}\|S\|_p\right) \mathcal{E}^{(2|q^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[O] \ + \ 4\ \mathcal{E}^{(2|p^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho]\ \mathcal{E}^{(2|\dagger q^*)}_{\Sigma_{\mu_L}\Sigma'_{\mu_L}}[O] & \\ \min_{\frac{1}{p}+\frac{1}{q}=1} 4\ \mathcal{E}^{(2|p^*)}_{\Sigma_{\mu_R}\Sigma'_{\mu_R}}[\rho]\ \mathcal{E}^{(2|\dagger q^*)}_{\Sigma_{\mu_L}\Sigma'_{\mu_L}}[O] & \left\{\Sigma'_{\text{Depolarize}}\right\} \end{cases} \tag{21}$$

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

$$\sigma^{2RL}_{\partial_\mu \mathcal{L}_\Lambda | \Sigma'_{\mu_{RL}}}[\rho, O] = \begin{cases} O\left(\frac{1}{d_{\mathcal{H}} d_{\mathcal{E}}^2}\right) & \{O_{\text{Orthogonal}}\ ,\ \Sigma'_{\text{cHaar}}\} \\ O\left(\frac{1}{d_{\mathcal{H}}^2 d_{\mathcal{E}}^2}\right) & \{O_{\text{Projector}}\ ,\ \Sigma'_{\text{cHaar}}\} \\ 0 & \left\{\Sigma'_{\text{Depolarize}}\right\} \end{cases} \tag{22}$$