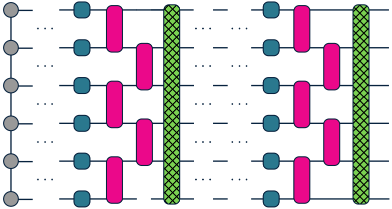
Overparameterization of Realistic Quantum Systems

Matthew Duschenes*, Juan Carrasquilla, Raymond Laflamme
University of Waterloo, Institute for Quantum Computing, & Vector Institute

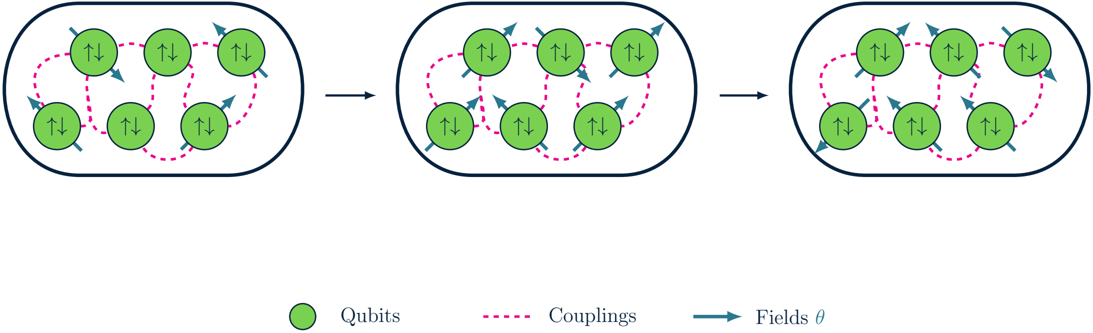
Winter, 2024

arXiv:2401.05500

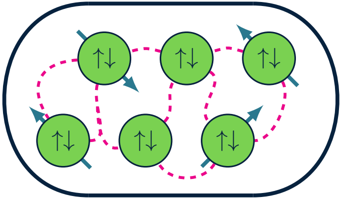
Seminar



What Are Parameterized Quantum Systems?

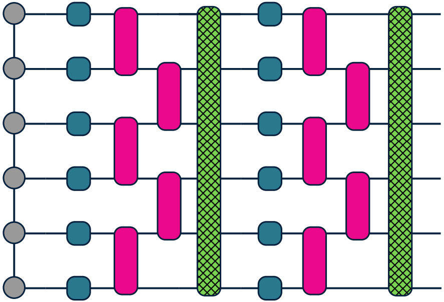


What Are Parameterized Quantum Systems?



Quantum System

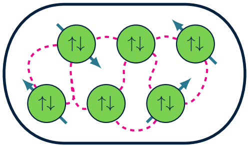
$H_\theta = \sum_\mu \theta_\mu G_\mu$



Quantum Circuit

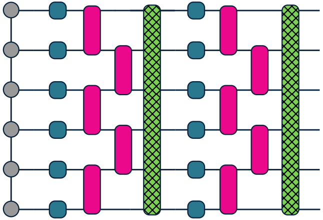
$U_\theta = \prod_\mu U_\theta^\mu$

What Are Parameterized Quantum Systems?



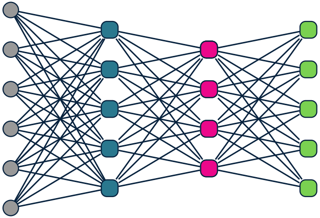
Quantum System

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



Quantum Circuit

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

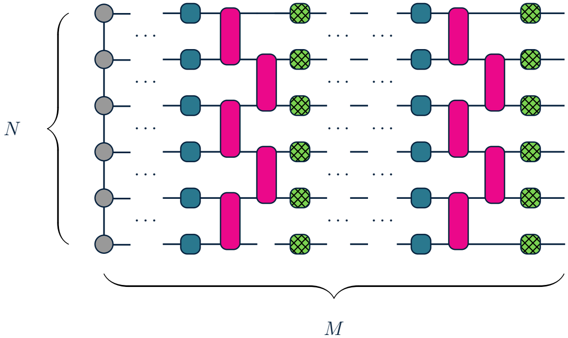


Classical Algorithm

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

Tasks of Interest: Unitary Compilation, State Preparation

Learning Phenomena of Quantum Systems



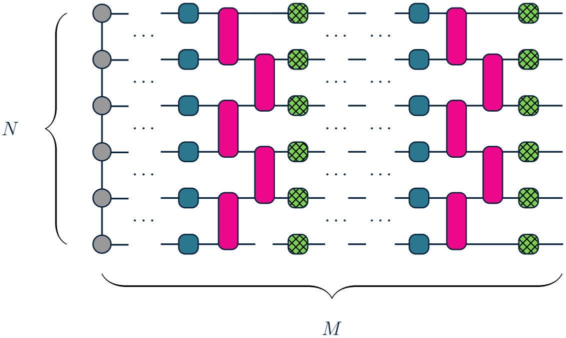
How does the amount of *noise* γ and the *evolution depth* M

of a *constrained* system

affect its classical simulation and optimization, and resulting infidelities

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

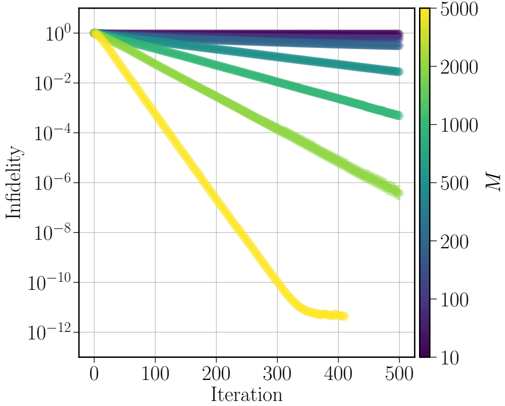
Learning Phenomena of Quantum Systems



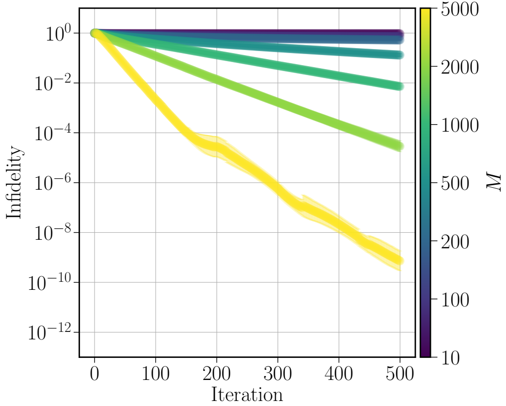
How can we leverage approaches from
quantum optimal control and *learning theory* to describe these relationships?

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

Unconstrained vs. Constrained Optimization

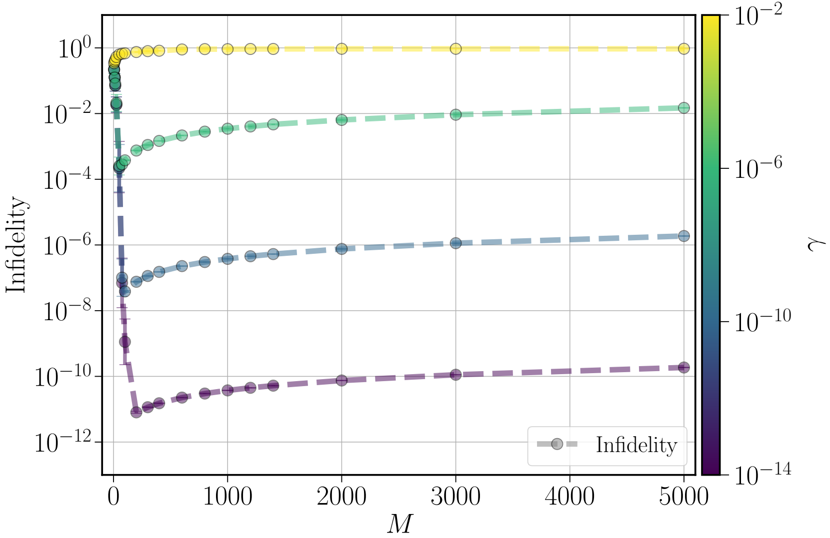


(a) Unconstrained Unitary Compilation



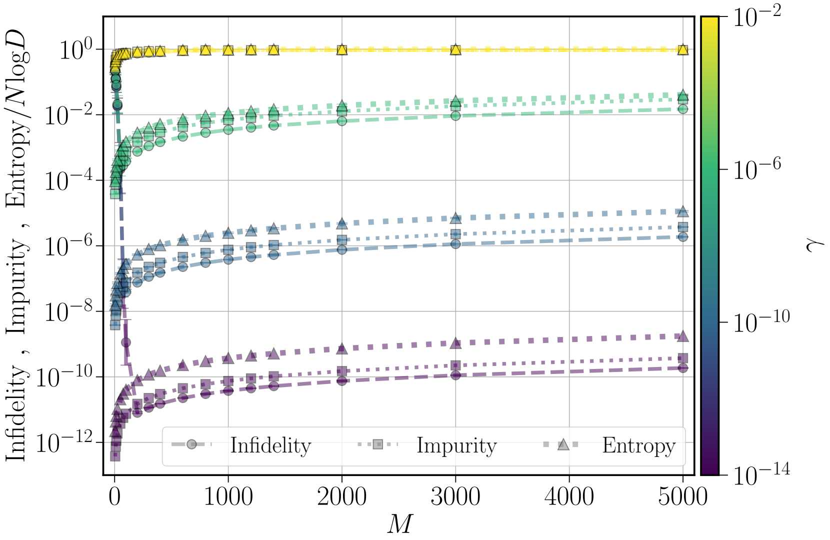
(b) Constrained Unitary Compilation

Regimes of Noisy Optimization



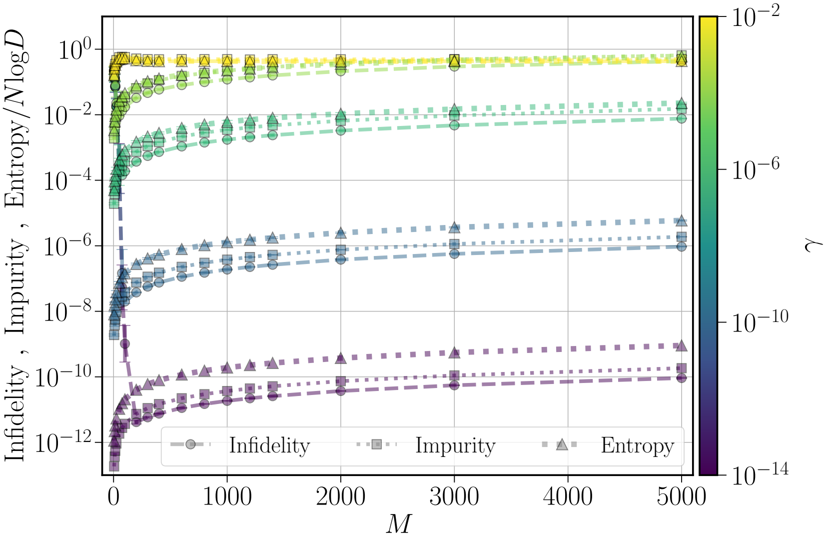
(c) Unital Dephasing for State Preparation

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))

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

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

$$\mathcal{L}_{\theta^*\gamma|M_{\gamma}} \sim \gamma^{\alpha} \ , \tag{2}$$

and parameterized noise channels can therefore *mitigate* approximately

$$\bar{M}_{\gamma} \sim \gamma \log 1/\gamma \quad \text{errors} \ . \tag{3}$$

Noise Induced Critical Depth

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

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

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

$$\mathcal{L}_{\theta^*\gamma|M_\gamma} \sim \gamma^\alpha \ , \tag{2}$$

and parameterized noise channels can therefore *mitigate* approximately

$$\bar{M}_\gamma \sim \gamma \log 1/\gamma \quad \text{errors} \ . \tag{3}$$

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

$$|\mathcal{L}_{\theta_\gamma} - \mathcal{L}_\theta| \leq 2\Big|(1-\gamma)^{NM} - 1\Big| \tag{4}$$

Representations of Channels and States

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

Representations of Channels and States

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

$$\Lambda_{\theta\gamma} = \left\langle \Lambda_{\theta\gamma_k} \right\rangle_{k \sim p_{K\gamma}} \tag{5}$$

Representations of Channels and States

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

$$\Lambda_{\theta\gamma} = \langle \Lambda_{\theta\gamma_k} \rangle_{k \sim p_{K\gamma}} \quad , \quad \Lambda_{\theta\gamma_k} = \frac{1}{\binom{K}{k}} \sum_{\substack{\chi \subseteq [K] \\ |\chi| = k}} \Lambda^{\chi}_{\theta\gamma} \tag{5}$$

Representations of Channels and States

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

$$\Lambda_{\theta\gamma} = \langle \Lambda_{\theta\gamma_k} \rangle_{k \sim p_{K\gamma}} \quad , \quad \Lambda_{\theta\gamma_k} = \frac{1}{\binom{K}{k}} \sum_{\substack{\chi \subseteq [K] \\ |\chi| = k}} \Lambda_{\theta\gamma}^{\chi} \quad , \quad p_{K\gamma}(k) = \binom{K}{k} \gamma^k (1-\gamma)^{K-k} \qquad (5)$$

Representations of Channels and States

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

$$\Lambda_{\theta\gamma} = \langle \Lambda_{\theta\gamma_k} \rangle_{k \sim p_{K\gamma}} \quad , \quad \Lambda_{\theta\gamma_k} = \frac{1}{\binom{K}{k}} \sum_{\substack{\chi \subseteq [K] \\ |\chi| = k}} \Lambda_{\theta\gamma}^{\chi} \quad , \quad p_{K\gamma}(k) = \binom{K}{k} \gamma^k (1-\gamma)^{K-k} \tag{5}$$

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

$$\rho_{\theta\gamma} = \frac{I + \lambda_{\theta\gamma} \cdot \omega}{d} = (1-\gamma)^K \rho + (1 - (1-\gamma)^K) \epsilon_{\theta\gamma} + \Delta_{\theta\gamma} \tag{6}$$

Representations of Channels and States

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

$$\mathcal{L}^{\rho}_{\theta_\gamma} \sim \frac{1}{2}\mathcal{I}_{\theta_\gamma} \sim \boxed{K\gamma\frac{d-1}{d}\left(1-\frac{\lambda\cdot\varepsilon_{\theta_\gamma}}{\lambda^2}\right)} + O(\tbinom{K}{2}\gamma^2) \tag{5}$$

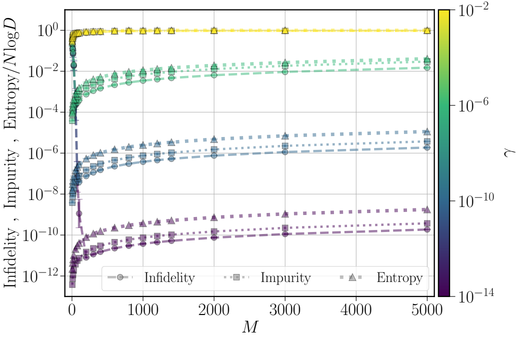
$$\mathcal{S}_{\theta_\gamma} \sim \mathcal{D}^{\rho}_{\theta_\gamma} \sim O(K\gamma) \tag{6}$$

Representations of Channels and States

Noise phenomena dominates at $M \geq M_\gamma$:

The scale of optimization and entropic driven scales *intersect*

$$\mathcal{L}^{\rho}_{\theta^*_\gamma \gamma} \sim e^{-\alpha M}|_{M_\gamma} \approx \mathcal{L}^{\rho}_{\theta^* \gamma} \sim N M \gamma|_{M_\gamma} \ , \tag{9}$$



Representations of Channels and States

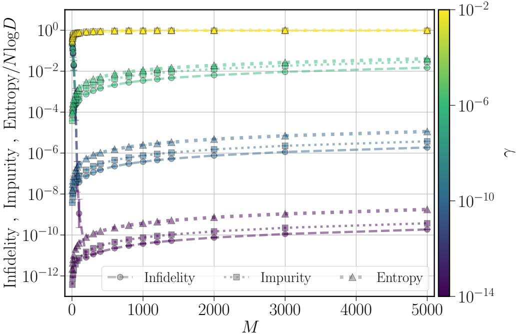
Noise phenomena dominates at $M \geq M_\gamma$:

The scale of optimization and entropic driven scales *intersect*

$$\mathcal{L}^{\rho}_{\theta^*_\gamma \gamma} \sim e^{-\alpha M}|_{M_\gamma} \approx \mathcal{L}^{\rho}_{\theta^* \gamma} \sim NM\gamma|_{M_\gamma} \ , \tag{9}$$

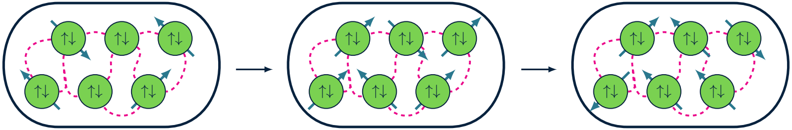
and we recover our numerically predicted noise-induced critical depth!

$$M_\gamma \sim \log 1/\gamma \tag{10}$$



What Have We Learned About Noisy Overparameterization?

- Overparameterization is *robust* to constraints; requires $\sim O(N)$ greater depth
- 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
- How can parameterized systems be applied to entropy mitigation?



Appendix

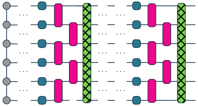
How May We Control Quantum Systems?

- 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 $\{G_\mu\}$

$$H_\theta^{(\mu)} = \sum_\nu \theta_\nu^{(\mu)} G_\nu \;\rightarrow\; U_\theta \approx \prod_{\mu,\nu}^{M} U_\theta^{(\mu,\nu)} \;:\; U_\theta^{(\mu,\nu)} = e^{-i\delta H_\theta^{(\mu,\nu)}} \approx e^{-i\delta\theta_\nu^{(\mu)} G_\nu} \tag{11}$$

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

$$H_\theta^{(\mu)} = \sum_i \theta_i^{x(\mu)} X_i + \sum_i \theta_i^{y(\mu)} Y_i + \sum_i h_i Z_i + \sum_{i<j} J_{ij} Z_i Z_j \tag{12}$$

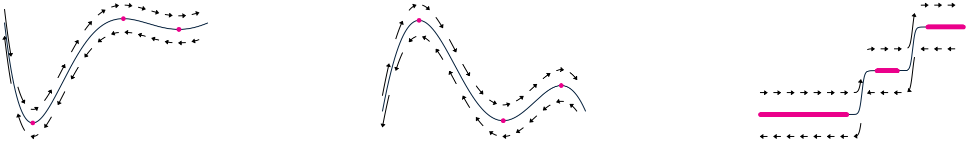


- Noise generated by constant *Kraus* operators $\{\mathcal{K}_{\gamma_\alpha}\}$
  i.e) Dephasing $\{\sqrt{1-\gamma}\; I,\; \sqrt{\gamma}\; Z\}$

$$\rho \;\rightarrow\; \rho_{\Lambda_{\theta_\gamma}} = \circ_\mu^M \mathcal{N}_\gamma \circ \mathcal{U}_\theta^{(\mu)}(\rho) = \circ_\mu^M \left[\sum_\alpha \mathcal{K}_{\gamma_\alpha}\, U_\theta^{(\mu)}\, \rho\, {U_\theta^{(\mu)}}^\dagger\, {\mathcal{K}_{\gamma_\alpha}}^\dagger \right] \tag{13}$$

Learning Phenomena

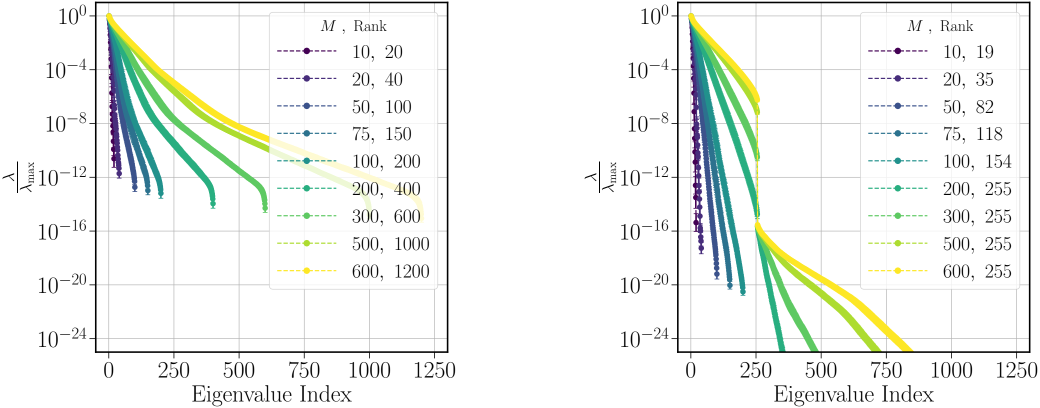
- How do optimization algorithms *learn*, and traverse the *objective landscape*?



- Learning can converge exponentially quickly in the *overparameterized* regime
- Dimensionality of *dynamical Lie algebra* spanned by Hamiltonian, determines *expressivity* (Larocca *et al.* arXiv:2109.11676 (2021))
- Optimal control pulses must evolve according to a *quantum speed limit* (Deffner *et al.* J. Phys. A, **50** (2017))

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

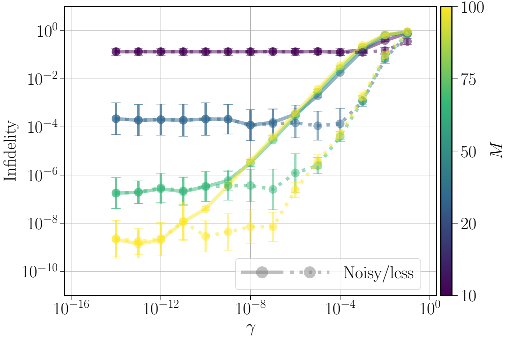
$$\mathcal{H}_{\mu\nu} = \partial_{\mu\nu}\mathcal{L}_\theta$$

(i) Fisher Information Rank Saturation

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

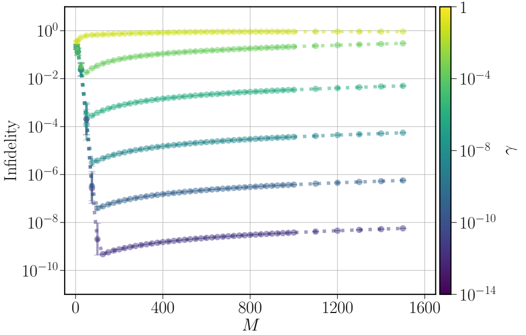
Noisy Optimization

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



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

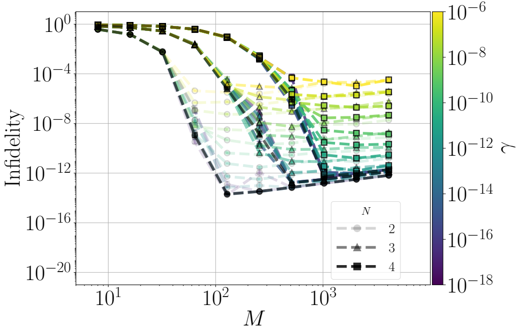
$$\partial \mathcal{L}_{\theta_\gamma} \sim \sum_\eta \alpha_\eta \, \mathcal{L}_{\theta+\eta \ \gamma}$$



(k) Critical Depth for Noisy Infidelity

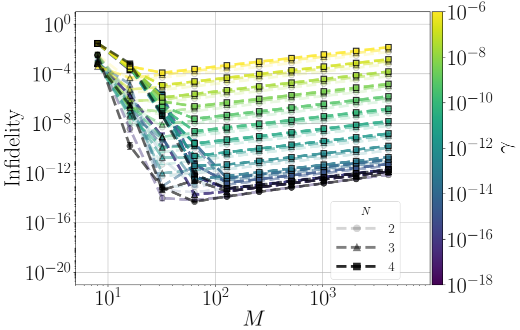
Universal Effects of Noise

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



(l) Classical floating point noise for unitary compilation

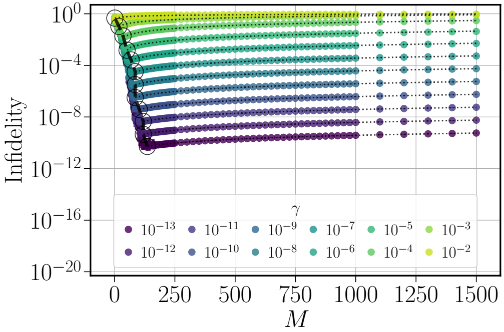
$$|\mathcal{L}_{\theta_\gamma} - \mathcal{L}_\theta| \leq n^{O(NM)}\left|(1+\gamma/n)^{O(NM)} - 1\right|$$



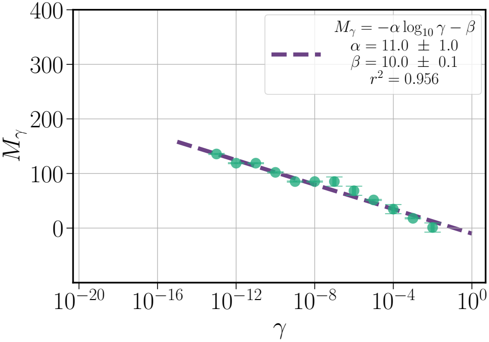
(m) Quantum dephasing noise for state preparation

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

Noise Induced Critical Depth



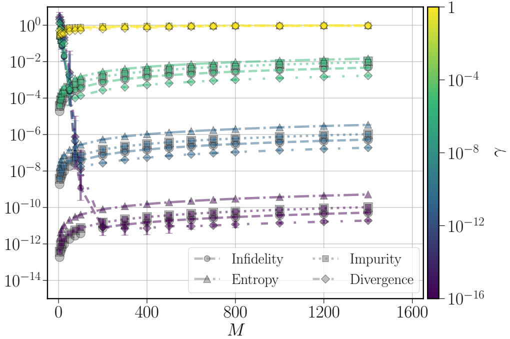
(n) Piecewise Fit of Noisy Infidelity



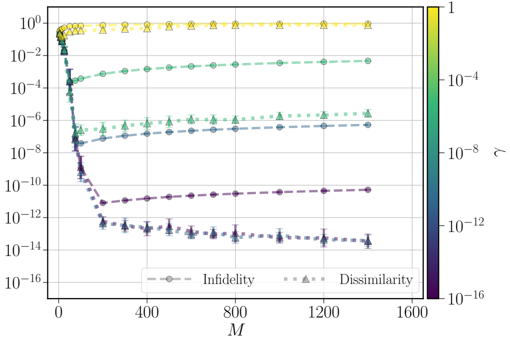
(o) Linear-Log Fit of Critical Depth

Correlated Quantities

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



(p) Impurity, Entropy, Divergence



(q) Cosine Dissimilarity