Efficient Self-consistent Learning of Gate Set Pauli Noise

Senrui Chen (U of Chicago)

APQC 2024, Oct 7

with Zhihan Zhang, Liang Jiang, and Steve Flammia

Manuscript available at csenrui.github.io and on arXiv tonight

Introduction

- Learning noise in quantum circuits:
  - Important for benchmarking, error suppression/mitigation/correction, etc.

- **Pauli channels** are widely used to model quantum noise
  - Randomized Compiling[1]: generic noise -> Pauli channels
  - Well-developed learning protocols: ACES, Cycle Error Reconstruction, …
  - Applications in error mitigation, optimized decoder, …

- This work: A framework of learning Pauli noise channels over a gate set
  - Criteria: **self-consistent, complete**, and **scalable**.

[1] Wallman and Emerson, PRA, 2016

Pauli channels

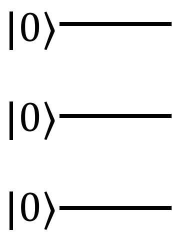
- n-qubit Pauli channel:

$$\Lambda(\rho) = \sum_{a \in \mathrm{P}^n} p_a P_a \rho P_a = \sum_{b \in \mathrm{P}^n} \lambda_b P_b \mathrm{Tr}(P_b \rho)/2^n$$

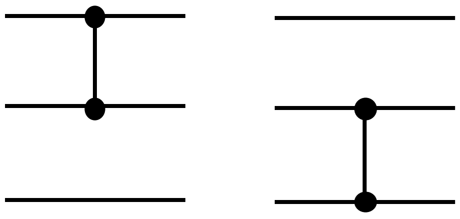
  - $\mathrm{P}^n = \{I, X, Y, Z\}^{\otimes n}$ - n-qubit Pauli group
  - $\{p_a\}_a$ – Pauli error rates
  - $\{\lambda_b\}_b$ – Pauli fidelities or Pauli eigenvalues, $\Lambda(P_b) = \lambda_b P_b$

- Throughout this work, we assume $\lambda_b > 0$ for all Pauli channels
  - Define $x_b \coloneqq -\log(\lambda_b)$ as the log Pauli fidelities.

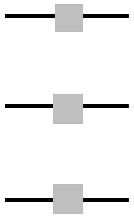
- Given an n-qubit system with the following set of operations (**gate set**)



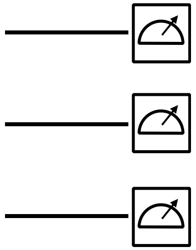
Initialization to $|0\rangle^{\otimes n}$



Layer of multi-qubit Clifford gates $\mathcal{G}$ from a set $\mathfrak{G}$

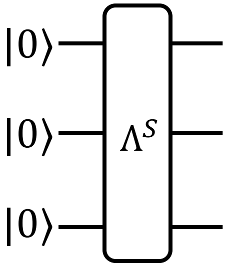


Layer of arbitrary 1-qubit gates $\bigotimes_j \mathcal{U}_j$

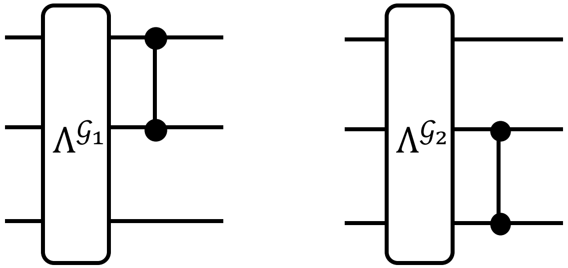


Computational-basis measurement $\{E_j\}_j$

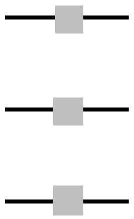
- Assume the following **Pauli noise model** (All Λ's are Pauli channels)



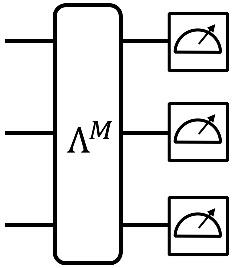
$\tilde{\rho} = \Lambda^S(|0\rangle\langle 0|^{\otimes n})$



$\tilde{\mathcal{G}} = \Lambda^{\mathcal{G}} \circ \mathcal{G}$

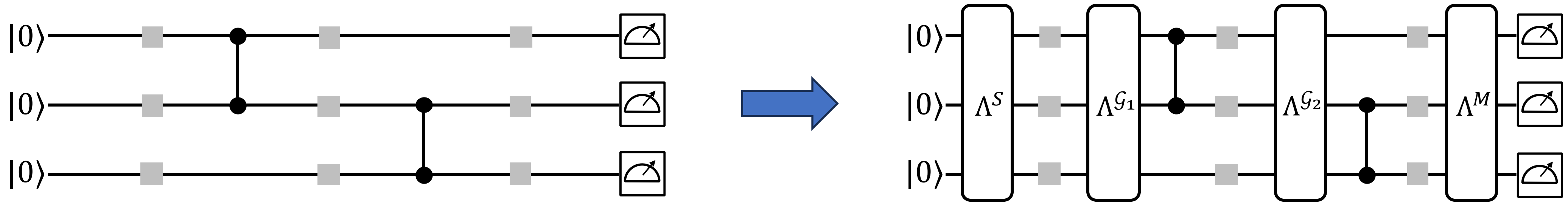


$\bigotimes_j \mathcal{U}_j$ noise negligible



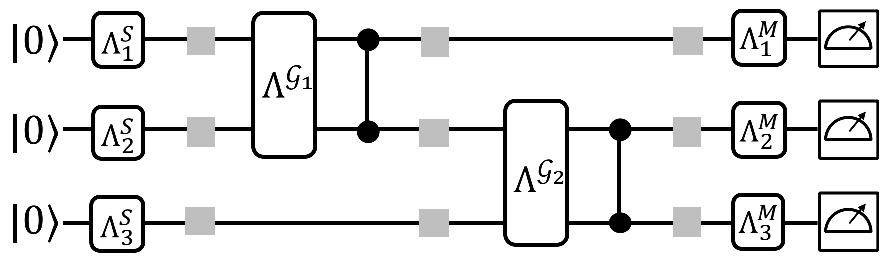
$\tilde{E}_j = \Lambda^M(E_j)$

  - Can be enforced via Randomized Compiling [Wallman and Emerson, PRA, 2016]
  - Ignore time-correlation, non-Markovianity, leakage, …

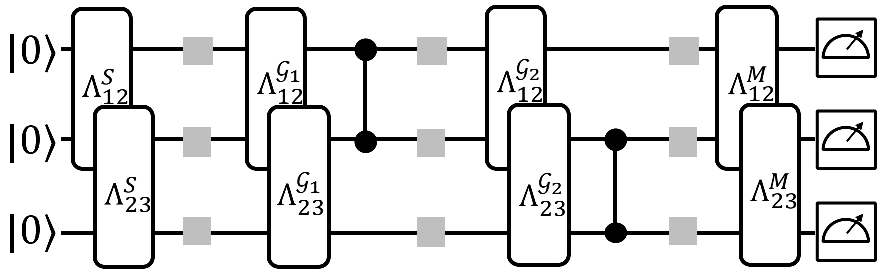


Ideal noisy

- We refer to this as a **complete** Pauli noise model
  - Parametrized by n-qubit Pauli channels $\{\Lambda^S, \Lambda^M, \Lambda^{\mathcal{G}} : \mathcal{G} \in \mathfrak{G}\}$
- Generic Λ has exp many parameters. Need an **efficient noise ansatz** in practice
  - Which is parametrized by a set of (polynomially many) reduced parameters



Fully-local



Nearest-neighbor

- We refer to these as **reduced** Pauli noise models

Main Question

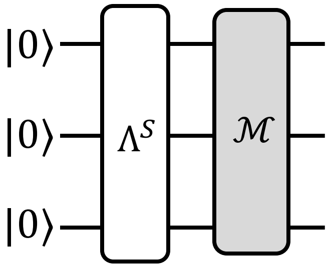
- How to **self-consistently** learn a Pauli noise model?
  - Self-consistent ⇔ learning with operations only from the noisy gate set themselves

- Generically, Λ's can be fully-determined self-consistently[1,2].
  - One exception: Λ^{id} for idle gate $\mathcal{G} = \mathrm{id}$

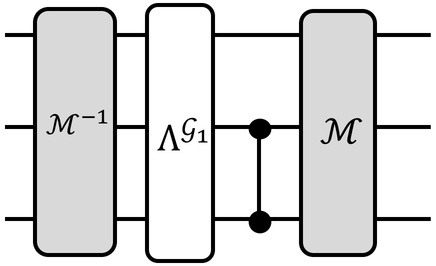
- This can be seen using **gauge transformations** as follows

[1] Huang, Flammia, Preskill. (2022).
[2] Chen, Liu, et al. Nat. Commun. (2023).

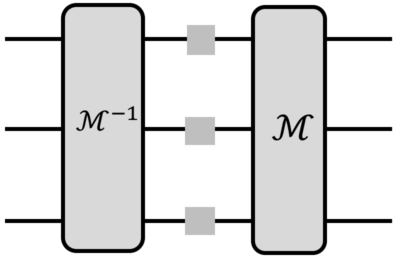
- Consider the following transformation of model params with an invertible map $\mathcal{M}$



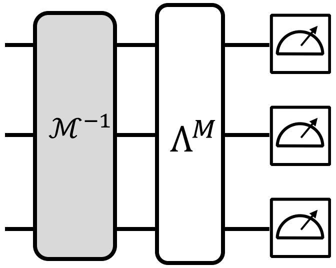
$\tilde{\rho} \mapsto \mathcal{M}(\tilde{\rho})$



$\tilde{\mathcal{G}} \mapsto \mathcal{M} \circ \tilde{\mathcal{G}} \circ \mathcal{M}^{-1}$

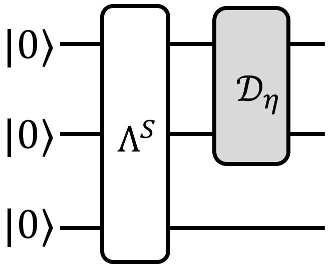


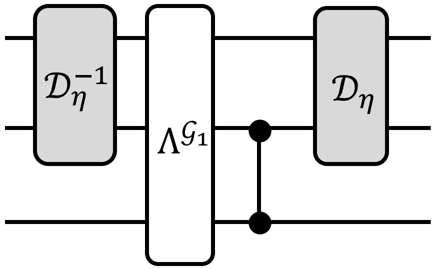
$\bigotimes_i \mathcal{U}_i \mapsto \mathcal{M} \circ \bigotimes_i \mathcal{U}_i \circ \mathcal{M}^{-1}$

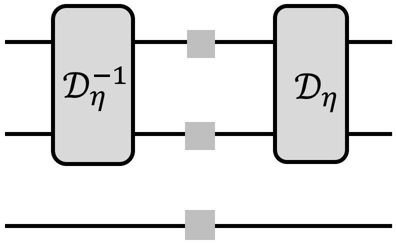


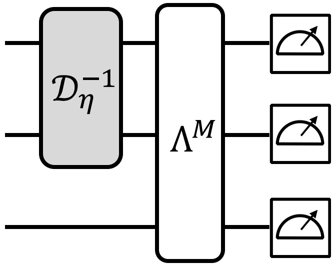
$\tilde{E}_j \mapsto (\mathcal{M}^{-1})^*(E_j)$

- By choosing $\mathcal{M}$ appropriately, one can preserve the Pauli noise model
  - E.g., Depolarizing channels on any subset of qubits $\mathcal{D}_\eta(\rho) = \eta\, I/d + (1-\eta)\, \rho$

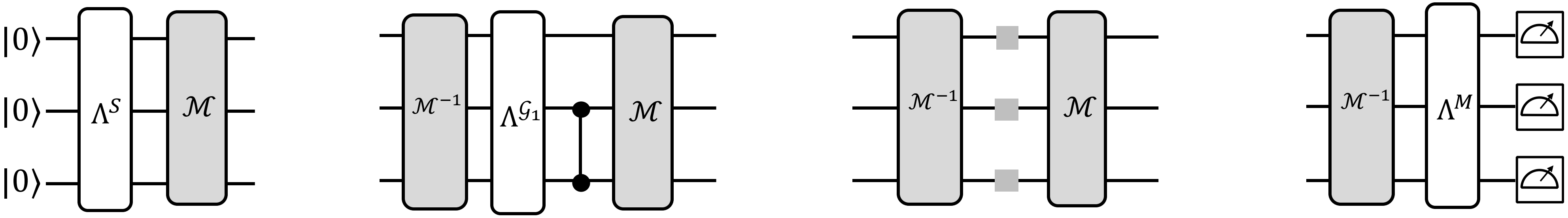






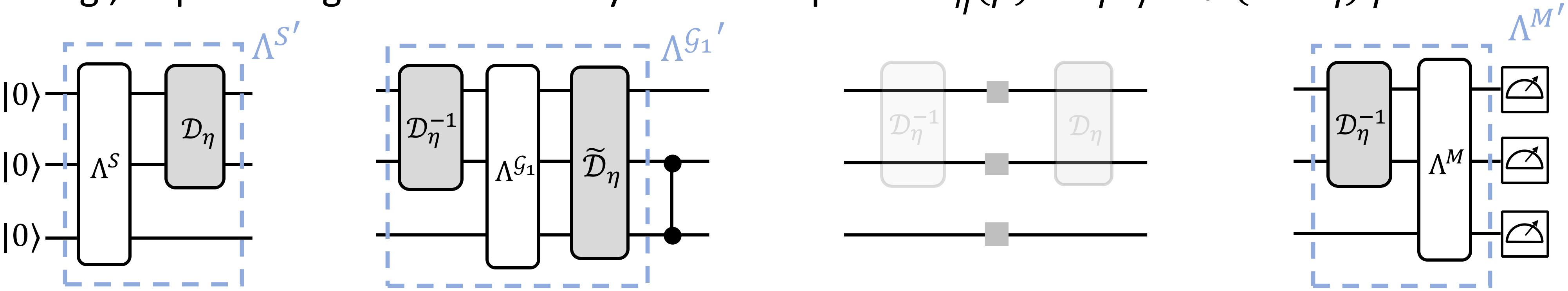


- Consider the following transformation of model params with an invertible map $\mathcal{M}$

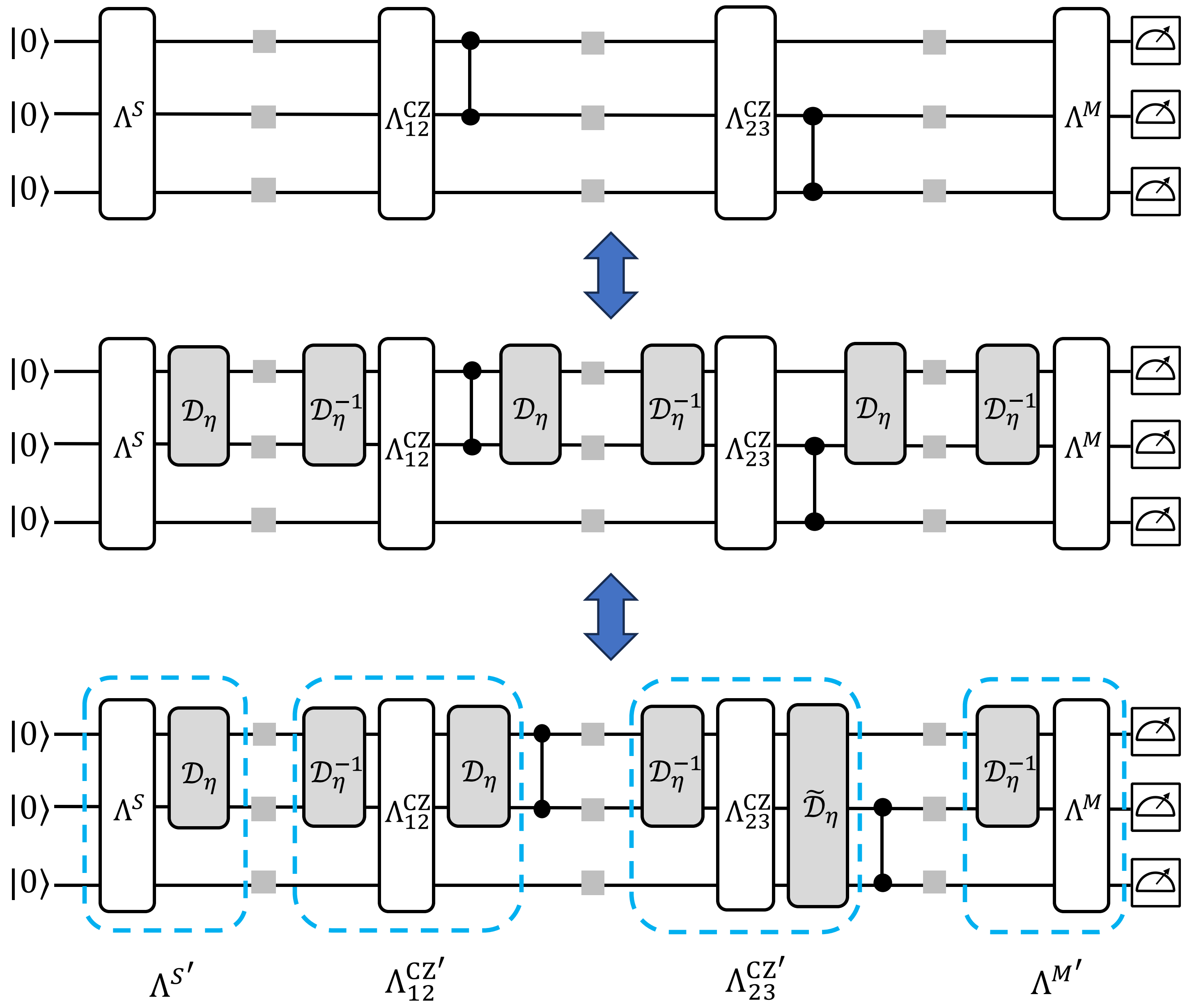


$\tilde{\rho} \mapsto \mathcal{M}(\tilde{\rho})$ $\quad$ $\tilde{\mathcal{G}} \mapsto \mathcal{M} \circ \tilde{\mathcal{G}} \circ \mathcal{M}^{-1}$ $\quad$ $\otimes_i \mathcal{U}_i \mapsto \mathcal{M} \circ \otimes_i \mathcal{U}_i \circ \mathcal{M}^{-1}$ $\quad$ $\tilde{E}_j \mapsto (\mathcal{M}^{-1})^*(E_j)$

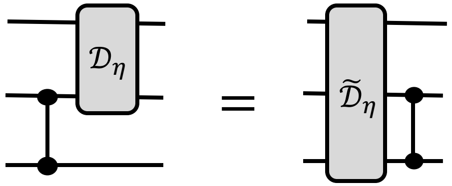
- By choosing $\mathcal{M}$ appropriately, one can preserve the Pauli noise model
  - E.g., Depolarizing channels on any subset of qubits $\mathcal{D}_\eta(\rho) = \eta\, I/d + (1-\eta)\,\rho$

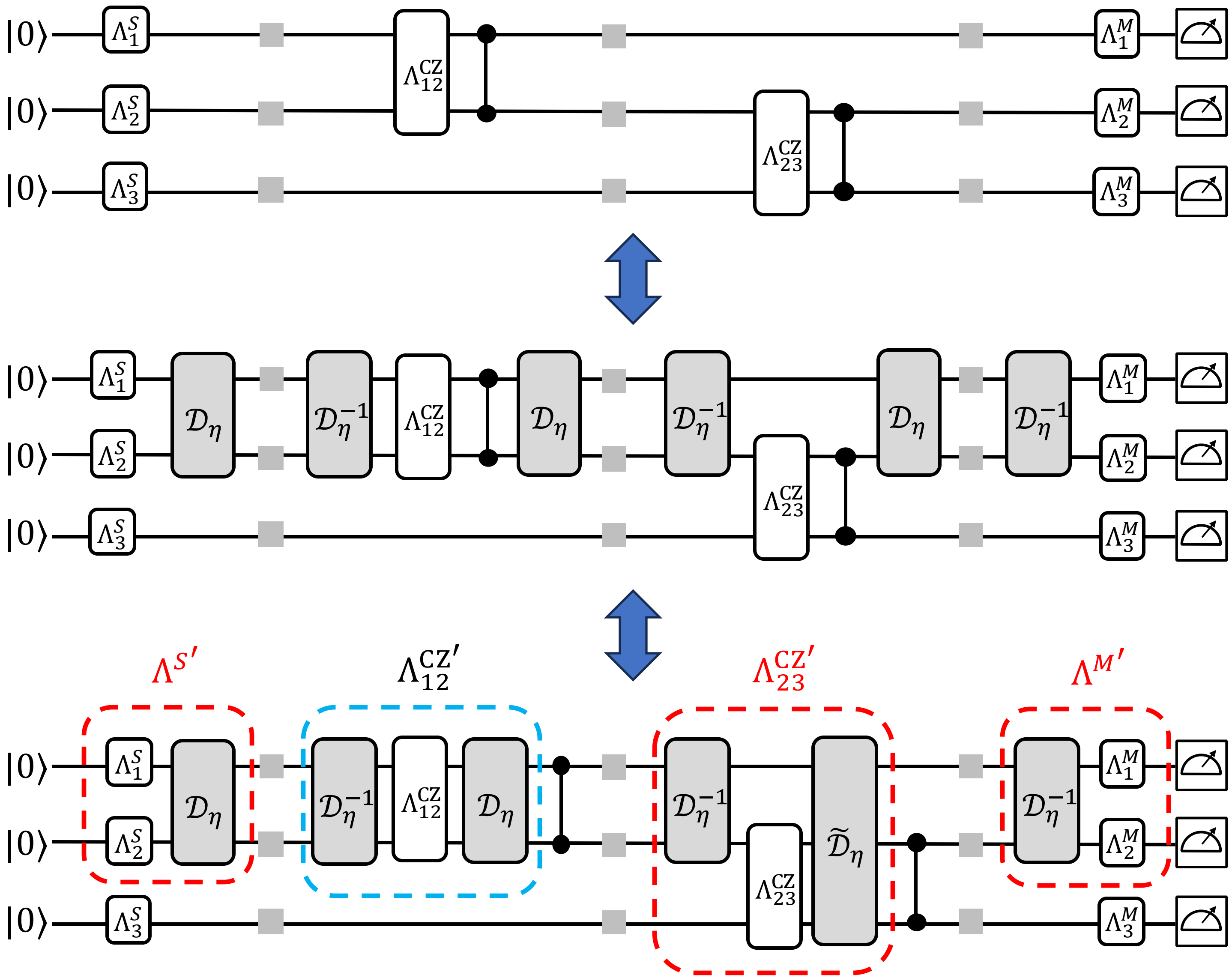


  - This yields a Pauli noise model with different parameters



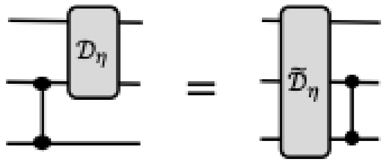
- Two choices of parameters related by a gauge transformation is **indistinguishable** self-consistently.
- If a **function** of noise parameters changes under a gauge transformation, it is **not** self-consistently learnable.
- Can we characterize all **learnable functions** and **gauges** of a Pauli noise model?





Reduced model

- Certain gauge transformations become **invalid** due to additional noise ansatz.
- In this example, the transformed $\Lambda^{S'}, \Lambda_{23}^{CZ'}, \Lambda^{M'}$ are not within the fully-local ansatz. Thus, the gauge is invalid.
- Can we characterize all **learnable functions** and **gauges** of a generic reduced Pauli noise model?



Pauli noise model in linear space

- Start with the **complete** model
- Let's make a few linear-algebraic definitions
  1. $\boldsymbol{x}$ is a vectorization of all log Pauli fidelities of $\Lambda's$
     - **Parameter space** X: The real linear space where $\boldsymbol{x}$ lives in
  2. Experiment $\boldsymbol{F}_{\mathcal{C}}$: $X \mapsto \mathbb{R}^{2^n}$ such that $\boldsymbol{F}_{\mathcal{C}}[j] = \mathrm{Tr}(\tilde{E}_j \tilde{\mathcal{C}}(\tilde{\rho}_0))$ for gate sequence $\mathcal{C}^{**}$
  3. Learnable linear function $\boldsymbol{f} \in X^*$: $\boldsymbol{f}(\boldsymbol{x})$ can be determined by a set of experiments $\boldsymbol{F}_k(\boldsymbol{x})$
     - **Learnable space** L: subspace of all learnable functions
  4. Gauge vectors $\boldsymbol{\mathfrak{d}} \in X$: For any experiments $\boldsymbol{F}$ and $\boldsymbol{x} \in X$, $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{x} + \boldsymbol{\mathfrak{d}})$.
     - **Gauge space** T: subspace of all gauge vectors

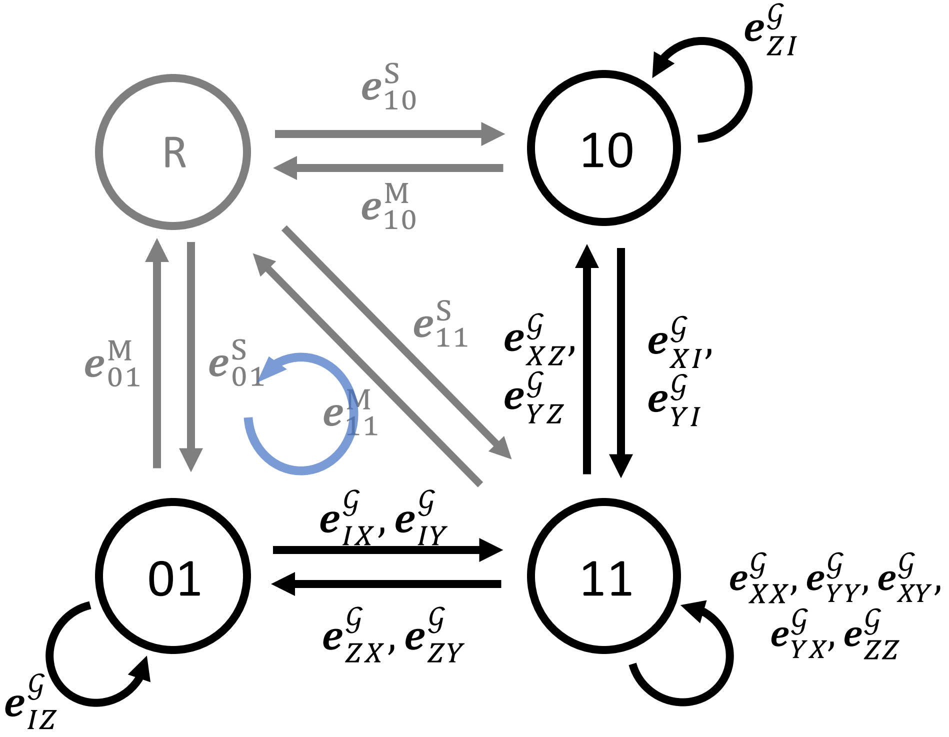
- Fact: $L \perp T$, i.e., $\boldsymbol{f}(\boldsymbol{\mathfrak{d}}) \equiv \boldsymbol{f} \cdot \boldsymbol{\mathfrak{d}} = 0$ for all learnable $\boldsymbol{f}$ and gauge $\boldsymbol{\mathfrak{d}}$.
- What else can we say about L and T?

** Extend the definitions of experiments to non-physical states for mathematical convenience

Pattern transfer graph

- Define Pauli **pattern transfer graph** (PTG)[1]
  - $2^n - 1$ Pauli pattern nodes + 1 Root node
  - Each fidelities parameter is assigned a unique edge:
    - SP fidelities $\boldsymbol{e}_t^S$: From Root to pattern t
    - M fidelities $\boldsymbol{e}_t^M$: From pattern t to Root
    - Gate fidelities $\boldsymbol{e}_a^{\mathcal{G}}$: From pattern$(a)$ to pattern$(\mathcal{G}(a))$
- Linear spaces on graph:
  - **Edge space** E: spanned by all edges $\{\boldsymbol{e_i}\}$
  - **Cycle space** Z: spanned by all cycle vectors
    - e.g. $\boldsymbol{e}_{01}^S + \boldsymbol{e}_{IX}^{\mathcal{G}} + \boldsymbol{e}_{11}^M$
  - **Cut space** U: spanned by all cut vectors
    - e.g. $\boldsymbol{e}_{01}^S + \boldsymbol{e}_{11}^S + \boldsymbol{e}_{XI}^{\mathcal{G}} + \boldsymbol{e}_{XI}^{\mathcal{G}} - \boldsymbol{e}_{01}^M - \boldsymbol{e}_{11}^M - \boldsymbol{e}_{XZ}^{\mathcal{G}} - \boldsymbol{e}_{YZ}^{\mathcal{G}}$
- $E = Z \oplus U$ and $Z \perp U$

Pattern: $\mathrm{pt}(XIZYI) \mapsto 10110$

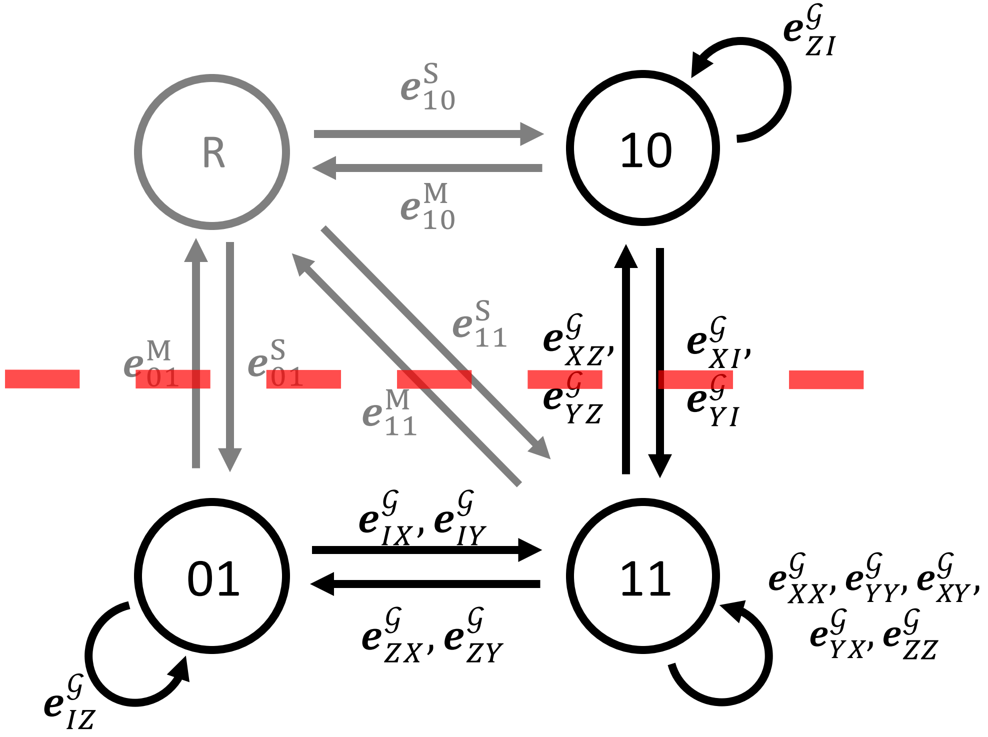


$\mathfrak{G} := \{\mathcal{G} = \mathrm{CZ}\}$

[1] Chen, Liu, et al. "The learnability of Pauli noise". Nat. Commun. (2023).

Pattern transfer graph

- Define Pauli **pattern transfer graph** (PTG)[1]
  - $2^n - 1$ Pauli pattern nodes + 1 Root node
  - Each fidelities parameter is assigned a unique edge:
    - SP fidelities $\boldsymbol{e}_t^S$: From Root to pattern t
    - M fidelities $\boldsymbol{e}_t^M$: From pattern t to Root
    - Gate fidelities $\boldsymbol{e}_a^{\mathcal{G}}$: From pattern$(a)$ to pattern$(\mathcal{G}(a))$
- Linear spaces on graph:
  - **Edge space** E: spanned by all edges $\{\boldsymbol{e_i}\}$
  - **Cycle space** Z: spanned by all cycle vectors
    - e.g. $\boldsymbol{e}_{01}^S + \boldsymbol{e}_{IX}^{\mathcal{G}} + \boldsymbol{e}_{11}^M$
  - **Cut space** U: spanned by all cut vectors
    - e.g. $\boldsymbol{e}_{01}^S + \boldsymbol{e}_{11}^S + \boldsymbol{e}_{XI}^{\mathcal{G}} + \boldsymbol{e}_{XI}^{\mathcal{G}} - \boldsymbol{e}_{01}^M - \boldsymbol{e}_{11}^M - \boldsymbol{e}_{XZ}^{\mathcal{G}} - \boldsymbol{e}_{YZ}^{\mathcal{G}}$
- $E = Z \oplus U$ and $Z \perp U$



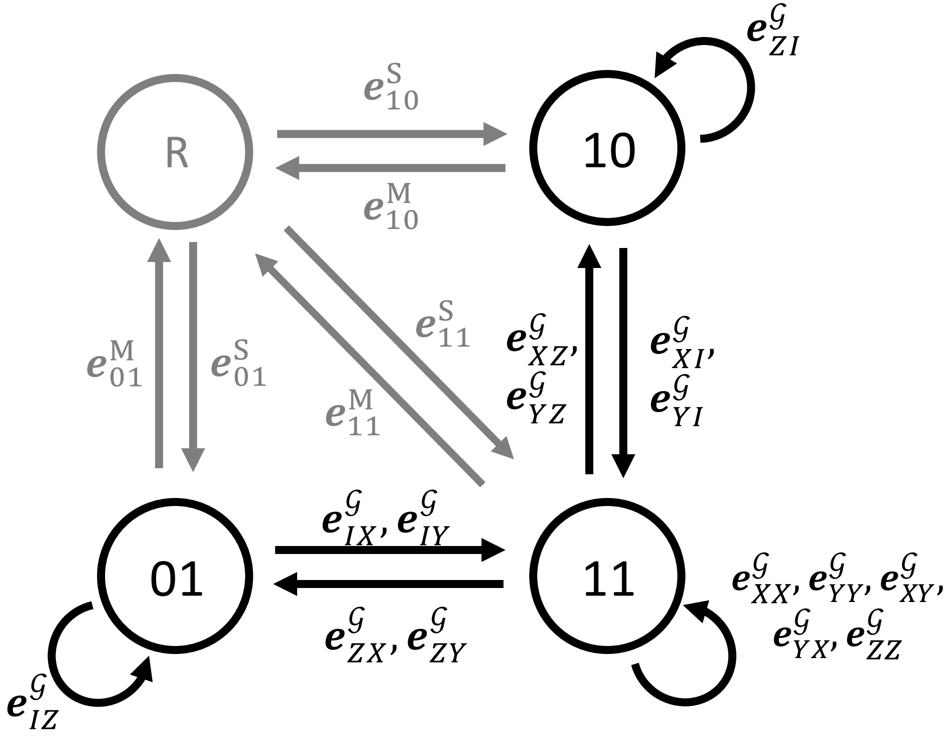
[1] Chen, Liu, et al. "The learnability of Pauli noise". Nat. Commun. (2023).

Learnability of complete Pauli noise

- Theorem:

E	$=$	Z	$\oplus^\perp$	U
(edge)		(cycle)		(cut)
‖		‖		‖
X	$=$	L	$\oplus^\perp$	T
(parameter)		(learnable)		(gauge)

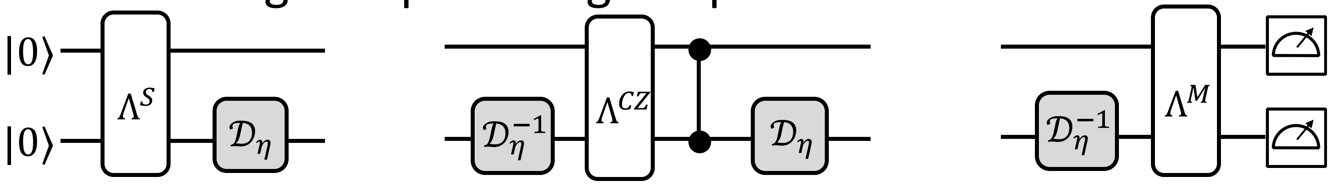
Note: $\oplus^\perp$ stands for orthogonal complement

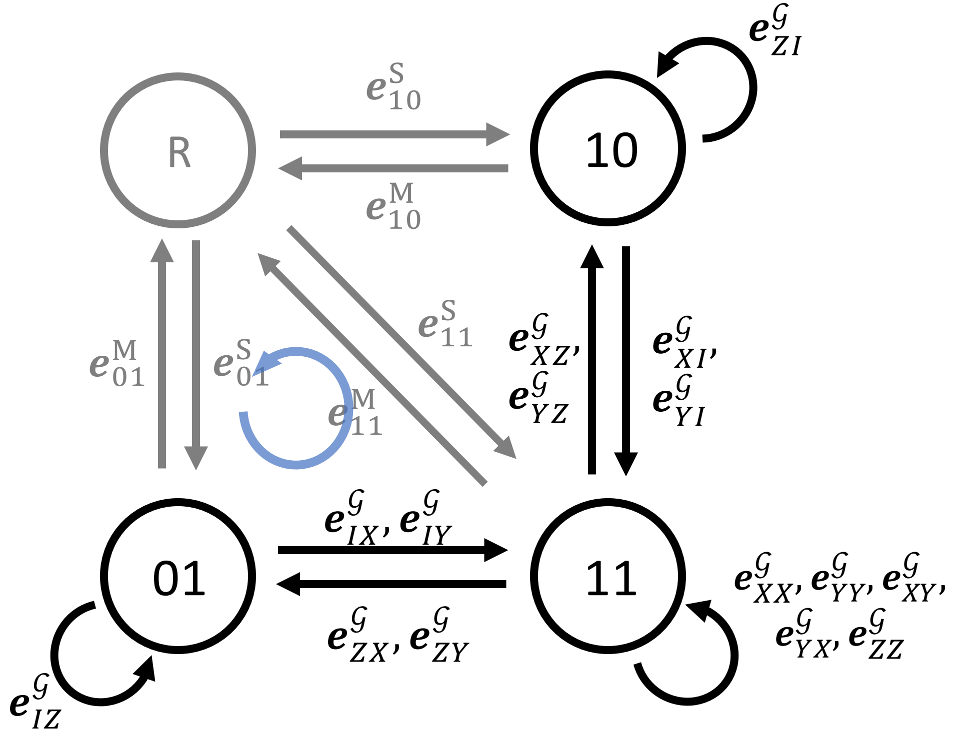


[1] Chen, Liu, et al. "The learnability of Pauli noise". Nat. Commun. (2023).

Key proof ideas

- Any rooted cycle implies an **experiment**
  - Clifford gates sequence + Pauli measurements
  - $\mathrm{Tr}(\tilde{P}\tilde{\mathcal{C}}(\tilde{\rho}_0)) = \lambda^{S}_{\mathrm{pt}(a_1)}\lambda^{\mathcal{G}_1}_{a_1}\lambda^{\mathcal{G}_2}_{a_2}\cdots\lambda^{\mathcal{G}_M}_{a_M}\lambda^{M}_{\mathrm{pt}(a_M)}$
    $= \exp\left(-\left(x^{S}_{\mathrm{pt}(a_1)} + \cdots + x^{M}_{\mathrm{pt}(a_M)}\right)\right)$
  - Rooted cycles span cycle space
- Any cut vector is a gauge vector
  - Cut: {R, 10} / {01, 11} ⇔
  - Gauge: Depolarizing on qubit 2



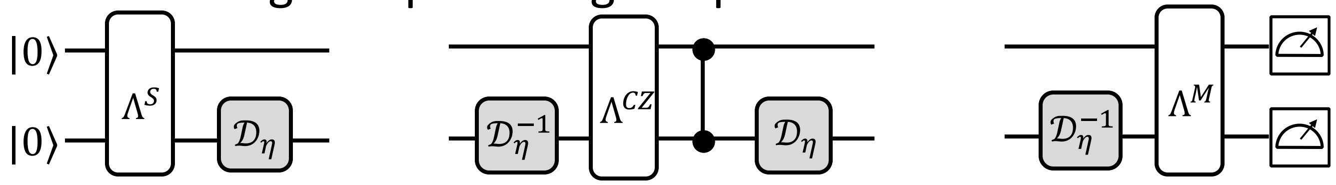


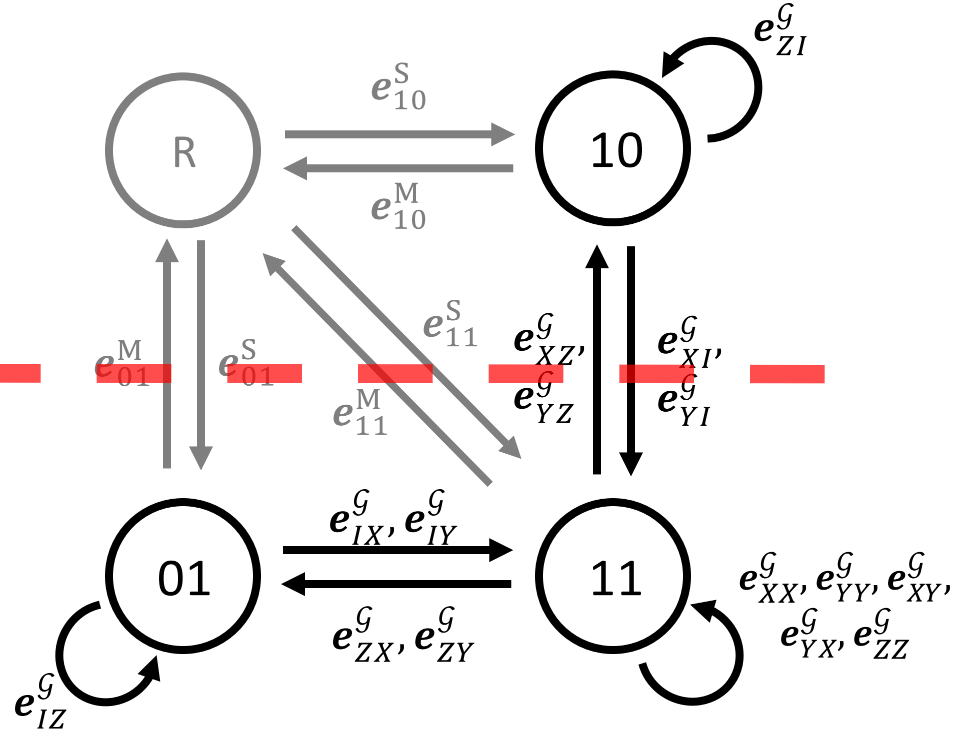
  - Corollary: Subset depolarizing gauge generates all gauges

[1] Chen, Liu, et al. "The learnability of Pauli noise". Nat. Commun. (2023).

Key proof ideas

- Any rooted cycle implies an **experiment**
  - Clifford gates sequence + Pauli measurements
  - $\mathrm{Tr}(\tilde{P}\tilde{\mathcal{C}}(\tilde{\rho}_0)) = \lambda^{S}_{\mathrm{pt}(a_1)}\lambda^{\mathcal{G}_1}_{a_1}\lambda^{\mathcal{G}_2}_{a_2}\cdots\lambda^{\mathcal{G}_M}_{a_M}\lambda^{M}_{\mathrm{pt}(a_M)}$
    $= \exp\left(-\left(x^{S}_{\mathrm{pt}(a_1)} + \cdots + x^{M}_{\mathrm{pt}(a_M)}\right)\right)$
  - Rooted cycles span cycle space
- Any cut vector is a gauge vector
  - Cut: {R, 10} / {01, 11} ⇔
  - Gauge: Depolarizing on qubit 2



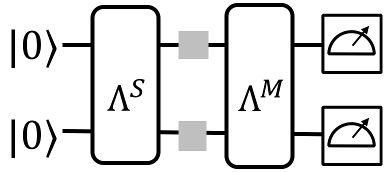


- Corollary: Subset depolarizing gauge generates all gauges

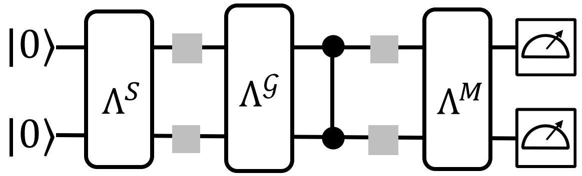
[1] Chen, Liu, et al. "The learnability of Pauli noise". Nat. Commun. (2023).

Learning the complete model

- Learnable space = Cycle space; Rooted cycles yield concrete experiments
- Find a **rooted cycle basis** and learn all of them.
- $\{\boldsymbol{e}_t^S + \boldsymbol{e}_t^M\} \quad \cup \quad \{\boldsymbol{e}_{\mathrm{pt}(a)}^S + \boldsymbol{e}_a^{\mathcal{G}} + \boldsymbol{e}_{\mathrm{pt}(\mathcal{G}(a))}^M\}$

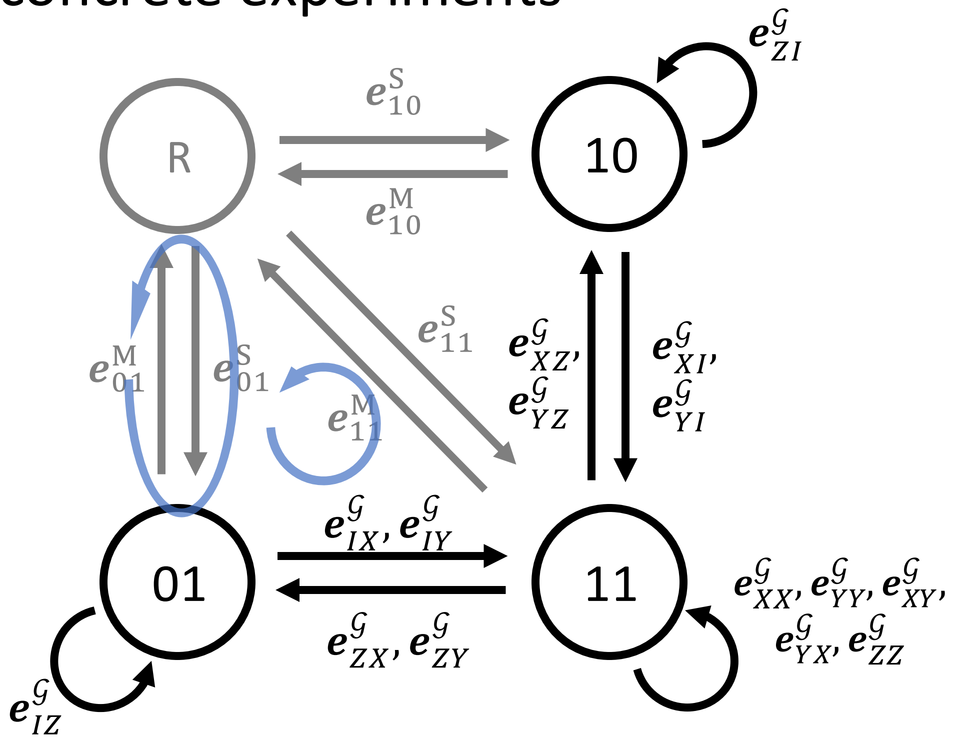


Depth-0



Depth-1

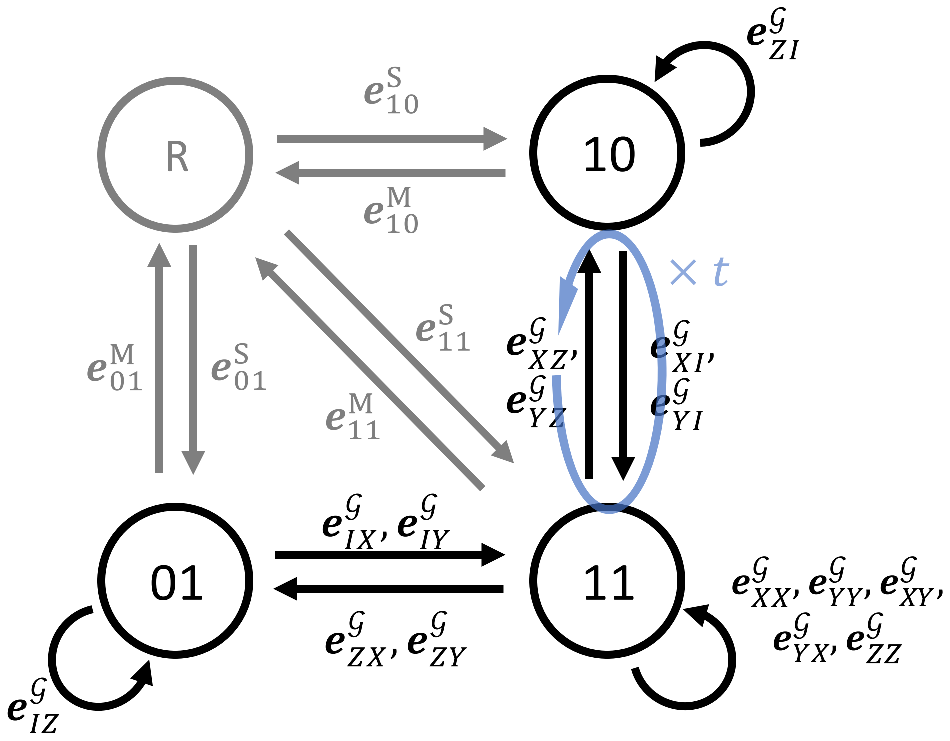
  - Learn $\mathcal{G} \in \mathfrak{G}$ one-by-one. No concatenation needed.
  - Only gives additive precision estimation



Learning to relative precision

- One often hopes to learn noise parameters to relative precision
  - With a small number of measurements
- Key: **Amplify** certain noise parameters using **concatenation** $\lambda^{S}_{\mathrm{pt}(a_1)}(\lambda^{\mathcal{G}_1}_{a_1}\lambda^{\mathcal{G}_2}_{a_2}\cdots\lambda^{\mathcal{G}_M}_{a_M})^t\,\lambda^{M}_{\mathrm{pt}(a_M)}$
  $$= \exp(-(\boldsymbol{x}_{SPAM} + \boldsymbol{x}^t_{\mathrm{cycles}})$$
- Theorem (informally):

  Any cycle consisting only of gate noise params can by amplified and learned via concatenation



- Cycles (of gate noise params) ≈ Germs or Tuples in GST

Definitions for reduced model

- Reduced noise model $(X_R, \mathcal{Q})$:
    1. **Reduced parameter space** X_R: where the reduced set of parameters $\boldsymbol{r}$ lives in
    2. Embedding map $\mathcal{Q}$: describes how $\boldsymbol{r}$ defines the Pauli channels. $\boldsymbol{x} = \mathcal{Q}(\boldsymbol{r})$

       We require $\mathcal{Q}$ to be **linear** and **injective**

- Similar Linear algebraic definitions:
    1. Experiments on reduced models $X_R \mapsto \mathbb{R}^{2^n}$ $\boldsymbol{F}_R(\boldsymbol{r}) \coloneqq \boldsymbol{F}\,(\mathcal{Q}(\boldsymbol{r}))$
    2. Reduced learnable function $\boldsymbol{f} \in X_R^*$: can be determined by a set of experiments
        - **Reduced learnable space** L_R: subspace of all reduced learnable functions
    3. Reduced gauge vectors $\boldsymbol{\mathfrak{d}} \in X_R$: For any experiments $\boldsymbol{F}_R$ and $\boldsymbol{r} \in X_R$, $\boldsymbol{F}_R(\boldsymbol{r}) = \boldsymbol{F}_R(\boldsymbol{r} + \boldsymbol{\mathfrak{d}})$.
        - **Reduced gauge space** T_R: subspace of all reduced gauge vectors

Learnability of reduced model

- Theorem:
- $L_R = \mathcal{Q}^{\mathrm{T}}(L) \equiv \{f(\mathcal{Q}(\cdot)) \mid \forall f \in L\}$
  - Reduced learnable space are complete learnable space projected by $\mathcal{Q}^{\mathrm{T}}$
- $T_R = \mathcal{Q}^{-1}(T) \equiv \{\mathfrak{d} \in X_R \mid \mathcal{Q}(\mathfrak{d}) \in T\}$
  - Reduced gauge space are the preimage of complete gauge space via $\mathcal{Q}$
  - More intuitively, $\mathcal{Q}(T_R) = T \cap \mathrm{Im}\mathcal{Q}$, a gauge is allowed iff it is in the image of embedding

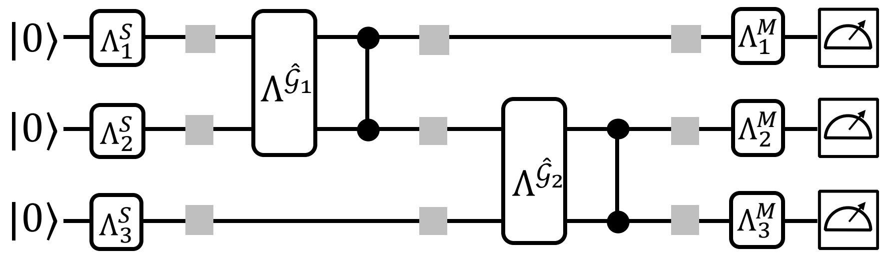
- Putting everything together:

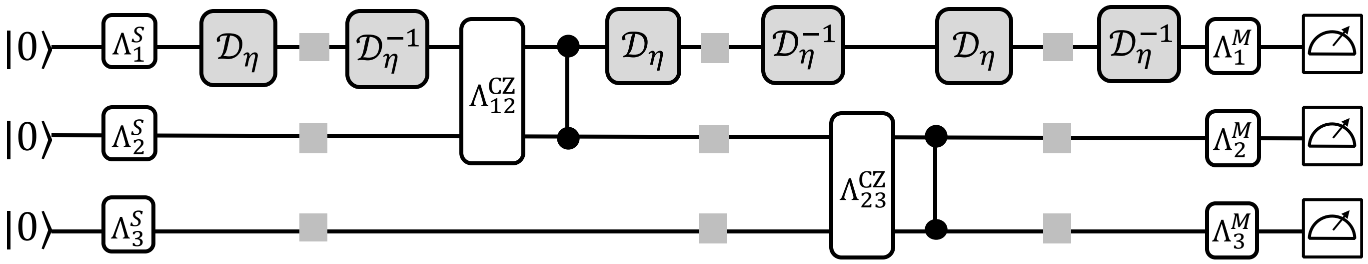
$$\begin{array}{ccccccc} E & = & Z & \oplus^{\perp} & U \\ \text{(edge)} & & \text{(cycle)} & & \text{(cut)} \\ \parallel & & \parallel & & \parallel \\ X & = & L & \oplus^{\perp} & T \\ \text{(parameter)} & & \text{(learnable)} & & \text{(gauge)} \\ \uparrow \mathcal{Q} & & \downarrow \mathcal{Q}^{\mathrm{T}} & & \downarrow \mathcal{Q}^{-1} \\ X_R & = & L_R & \oplus^{\perp} & T_R \\ \text{(reduced param)} & & \text{(reduced learnable)} & & \text{(Reduced gauge)} \end{array}$$

- Now we have a linear algebra procedure to determine L_R and T_R.

* $\mathcal{Q}^{\mathrm{T}}$ denotes the conjugate map of $\mathcal{Q}$. Here it can basically be understood as transposition.

Example: Fully-local model



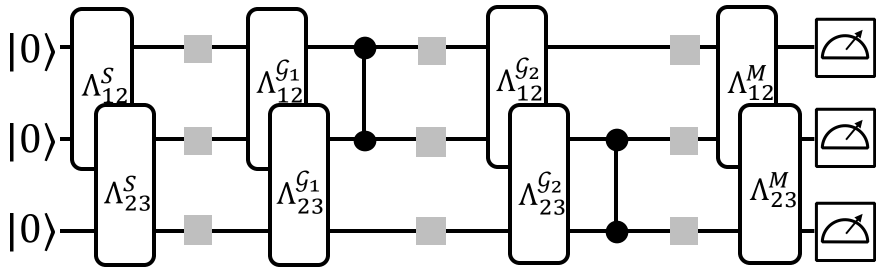


- Noise ansatz:
- Local SPAM noise: Product of 1q noise
- Local Gate noise: within gate support

- Thm: The allowed gauge space $\mathcal{Q}(T_R)$ is spanned by the 1-qubit depolarizing gauges
  - n gauge params

- Efficient learning discussed in paper
  - Local experiments suffices, usually

Example: Quasi-local model



$\Omega^* = \{\{1,2\},\{2,3\}\}$

- Quasi-local Pauli channel:
- Composition of Pauli channels on subsets of qubits Ω^*
- Equivalent to Pauli-Lindblad models[1] and others[2]

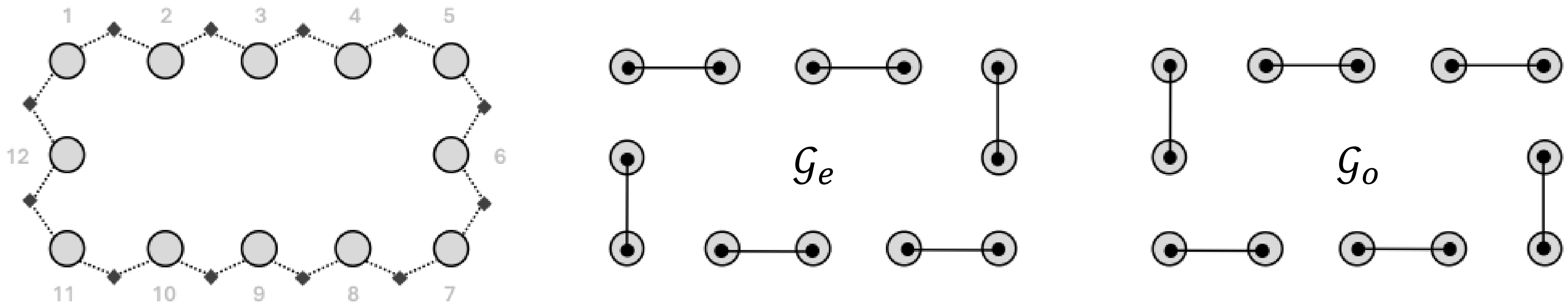
- Thm: Under certain "covariant" conditions, $\mathcal{Q}(T_R)$ are spanned by subset depolarizing gauge consistent with Ω^*.

- Details and efficient learning discussed in paper

[1] Ewout et al. Nat. Phys. 2023. [2] Wagner et al. PRL 2024.

Example: 2-local noise of parallel CZ gates

- We analyzed a nearest-neighbor 2-local model from [Ewout et al. Nat. Phys. 2023]
  - Two layers of parallel CZ gates on a 1D ring
  - Gates and SPAM Pauli noise assumed to be 2-local
- We can explicitly compute L_R, T_R in this case
  - Specifically, 1-qubit depolarizing gauges spans $\mathcal{Q}(T_R)$, n gauge parameters
- One can efficiently and self-consistently learn this noise model
  - Enables error mitigation without “symmetry assumptions” (on-going)



(a) (b) (c)

Summary

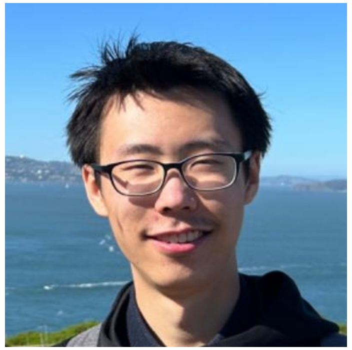
- We develop a framework of efficient self-consistent gate set Pauli noise learning

1. Characterization of learnable/gauge space for flexible gate set/noise ansatz
2. Efficient learning algorithms, additive vs. relative precision
3. Applications to concrete examples

- Outlook:

1. Graph-theoretical techniques beyond Pauli noise model
2. Including MCMs *[Zhang et al. 2406.02669, Hines et al. 2406.09299]*, extending to FTQC learning.
3. Applications to efficient self-consistent quantum error mitigation
4. Optimal experiment design, fine-grained complexity analysis

Thank you!



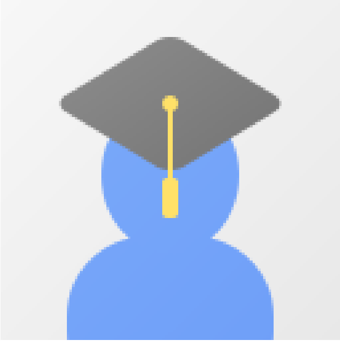
Zhihan Zhang
(Tsinghua University)

Liang Jiang
(University of Chicago)

Steve Flammia
(Virginia Tech & Phasecraft)

Special thanks --->

Alireza Seif
(IBM Quantum)



Yunchao Liu
(Berkeley->Harvard)

Scan for slides & manuscripts
Also available on csenrui.github.io
and on arXiv tonight

NTT

the David & Lucile Packard FOUNDATION