Automated quantum error mitigation based on probabilistic error reduction

Ben McDonough, *Yale University*
Andrea Mari, *Unitary Fund*
Nathan Shammah, *Unitary Fund*
Nathaniel Stemen, *Unitary Fund*
Misty Wahl, *Unitary Fund*
William Zeng, *Unitary Fund, Goldman Sachs*
Peter Orth, *Iowa State University*

SC22 Dallas, TX hpc accelerates.

Motivation & Summary

Quantum error mitigation techniques focused on removing noise-induced bias from expectation value are

- Applicable to relevant algorithms
- Viable on NISQ devices

An automated error mitigation framework based on probabilistic error reduction could extend the use cases of NISQ devices by making error mitigation easier to implement and less resource intensive.

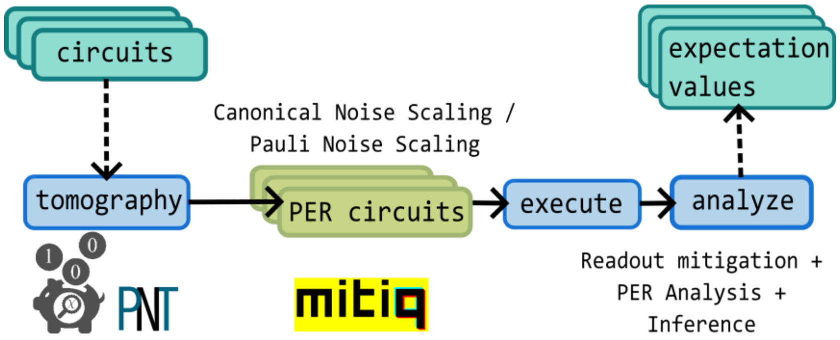
To that end, we have developed

- Notebooks demonstrating the use of existing software for this purpose
- Open source, publicly available Python toolkit for such a framework using Pauli noise tomography and noise scaling

https://github.com/benmcdonough20/AutomatedPERTools

Motivation & Summary

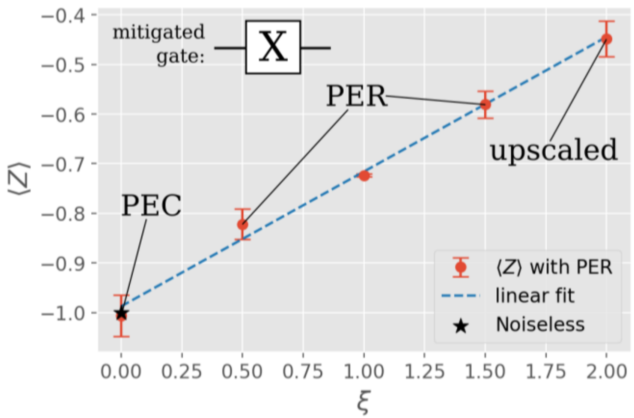
Automation



Input: *list of circuits and observables*

Output: *error-mitigated expectation values*

Improving efficiency



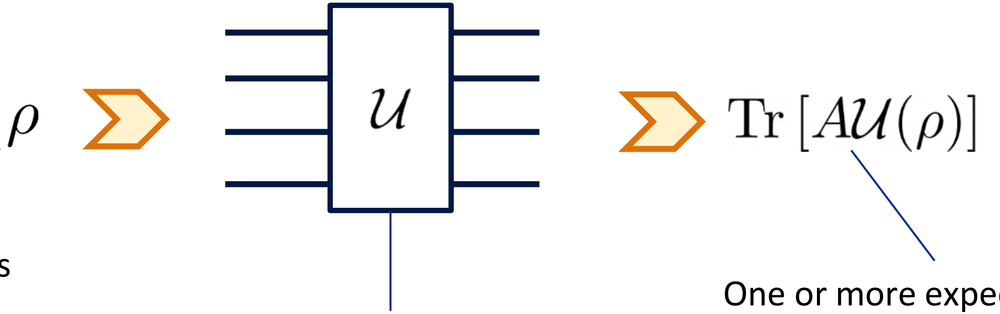
Probabilistic error reduction and **virtual zero-noise extrapolation**

- Noise can be scaled predictably *above or below* hardware level
- Zero-noise limit does not need to be evaluated directly

Probabilistic Error Cancellation

Applicable to algorithms in which the figure of merit is an expectation value averaged over many shots of a unitary circuit

ideal circuit:



Input state is prepared

A unitary circuit is run

One or more expectation values are obtained

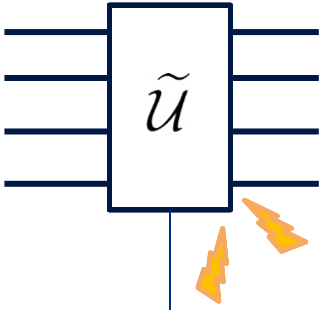
[1] Temme, Kristan, Sergey Bravyi, and Jay M. Gambetta. "Error mitigation for short-depth quantum circuits." Physical review letters 119, no. 18 (2017): 180509.
[2] Endo, Suguru, Simon C. Benjamin, and Ying Li. "Practical quantum error mitigation for near-future applications." *Physical Review X* 8, no. 3 (2018): 031027.
[3] Cai, Zhenyu, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. "Quantum error mitigation." arXiv preprint arXiv:2210.00921 (2022).

Probabilistic Error Cancellation

Noise introduces bias into the estimator of this expectation value.

noisy circuit:

$\widetilde{\rho}$



$\mathrm{Tr}\left[\widetilde{A}\widetilde{\mathcal{U}}(\widetilde{\rho})\right]$

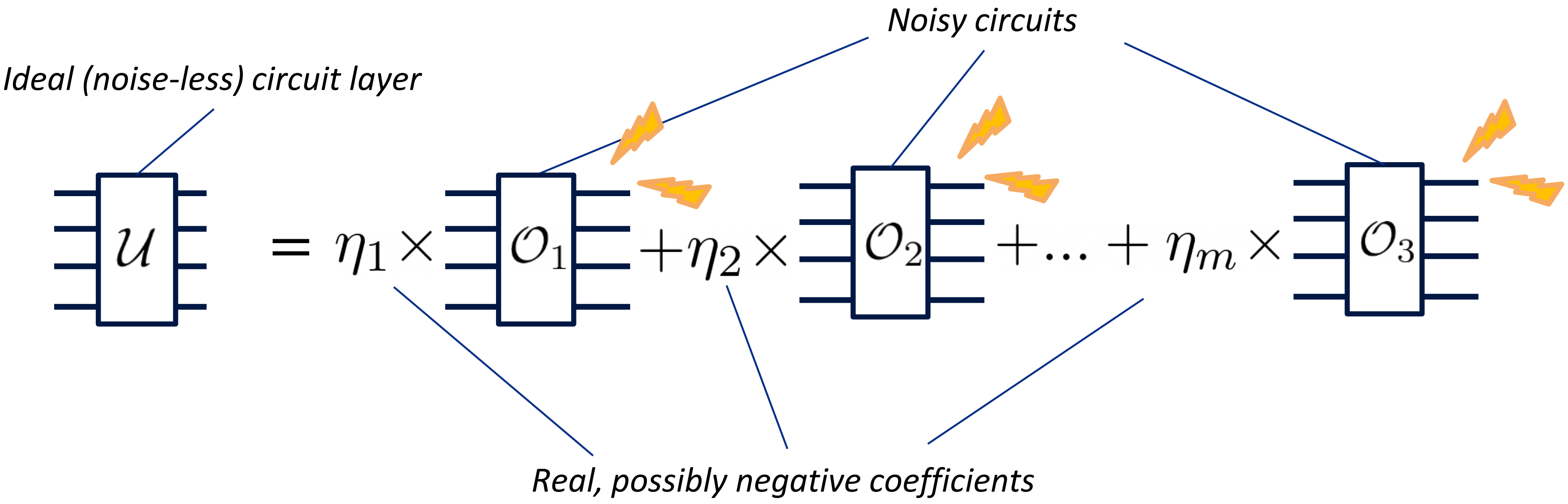
State preparation error affects input state

Circuit is non-unitary

Measurement error also adds to noise in final expectation value

Probabilistic Error Cancellation

After characterizing the noisy operations $\{\mathcal{O}_\alpha\}$, the ideal circuit is decomposed into $\mathcal{U}(\rho) = \sum_\alpha \eta_\alpha \mathcal{O}_\alpha(\rho)$ where the η_α are real.

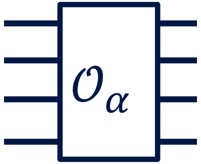


The linearity of the expectation value allows the ideal value to be written as $\langle A \rangle_{ideal} = \sum_\alpha \eta_\alpha \langle A \rangle_\alpha$

Probabilistic Error Cancellation

Number of terms grows exponentially with circuit depth…

The linear combination can be converted into a quasi-probability distribution (QPD) and sampled.

Sample 

with probability $\frac{|\eta_\alpha|}{\gamma}$

Scale expectation value by $\gamma\, \mathrm{sgn}(\eta_\alpha)$, where $\gamma = \sum_\alpha |\eta_\alpha|$

The sign problem:

- Since η can be negative, $\gamma > 1$. The magnitude of γ is generally determined by the noise strength
- γ determines the variance of the estimator, related to the strength of noise on the hardware
- If γ_i is the overhead of a single layer, then the total overhead is $\gamma = \prod_i \gamma_i \sim \gamma_i^n$

Gate Set Tomography and Canonical Noise Scaling

Gate Set Tomography

Gate set tomography (GST): characterize noisy set of operations $\{|\rho\rangle\rangle, \langle\langle E|, G_0, \ldots, G_K\}$

- SPAM errors are self-consistently included in the model
- Superoperator elements of noisy gates are learned up to a gauge choice [1]

PyGSTi: open-source software for *modeling and characterizing noisy quantum information processors* (QIPs) [2]

Generates tomography circuits → *User runs circuits on QPU and records results* → *Pauli transfer matrices of noisy gates*

[1] Greenbaum, Daniel. "Introduction to quantum gate set tomography." arXiv preprint arXiv:1509.02921 (2015).
[2] Nielsen, Erik, Kenneth Rudinger, Timothy Proctor, Antonio Russo, Kevin Young, and Robin Blume-Kohout. "Probing quantum processor performance with pyGSTi." Quantum science and technology 5, no. 4 (2020): 044002.

GST on Rigetti Aspen-11

Superoperator results of long-sequence GST on Rigetti Aspen-11 for the gate set $\{RX\left(\frac{\pi}{2}\right), RZ\left(\frac{\pi}{2}\right)\}$:

$$\widetilde{RX}\left(\frac{\pi}{2}\right) =$$

```
[[ 0.504+0.j   , -0.015-0.488j, -0.015+0.488j,  0.493+0.j   ],
 [-0.014-0.489j,  0.502-0.03j ,  0.489+0.003j,  0.021+0.487j],
 [-0.014+0.489j,  0.489-0.003j,  0.502+0.03j ,  0.021-0.487j],
 [ 0.496+0.j   ,  0.015+0.488j,  0.015-0.488j,  0.507+0.j   ]]
```

$$\widetilde{RZ}\left(\frac{\pi}{2}\right) =$$

```
[[ 0.998+0.j   , -0.015+0.015j, -0.015-0.015j,  0.002+0.j   ],
 [-0.015+0.015j, -0.001+0.998j,  0.   +0.002j,  0.015-0.015j],
 [-0.015-0.015j,  0.   -0.002j, -0.001-0.998j,  0.015+0.015j],
 [ 0.002+0.j   ,  0.015-0.015j,  0.015+0.015j,  0.998+0.j   ]]
```

- Number of circuits run: 679 @ 1000 shots each
- Time on Rigetti QPU: 15 minutes
- Moving to a two-qubit gate set: 14834

PEC with Mitiq

Mitiq: *an open-source toolkit for implementing error mitigation techniques [1].*

Constrained optimization algorithm for finding QPD representation with least overhead:

```
mitiq.pec.representations.optimal.find_optimal_representation
```

- Set of superoperators with maximum span is constructed using sequences of noisy gates.

$$\tilde{G}_1, \tilde{G}_2 \rightarrow \tilde{G}_1\tilde{G}_1\tilde{G}_1, \tilde{G}_1\tilde{G}_1\tilde{G}_2, \ldots, \tilde{G}_2\tilde{G}_2\tilde{G}_1, \tilde{G}_2\tilde{G}_2\tilde{G}_2$$

Set of noisy operations → **mitiq** → QPD representation of the ideal gate

$$RX_{PEC}\left(\frac{\pi}{2}\right) = \begin{bmatrix} 0.4999900 + 0.0000000j & 0.0000099 + -0.4999900j & 0.0000099 + 0.4999900j & 0.4999900 + 0.0000000j \\ 0.0000100 + -0.5000059j & 0.4999900 + -0.0000092j & 0.4999900 + 0.0000090j & 0.0000100 + 0.4999941j \\ 0.0000100 + 0.5000059j & 0.4999900 + -0.0000090j & 0.4999900 + 0.0000092j & 0.0000100 + -0.4999941j \\ 0.5000100 + -0.0000000j & -0.0000099 + 0.4999900j & -0.0000099 + -0.4999900j & 0.5000100 + -0.0000000j \end{bmatrix}$$

[1] LaRose, R., A. Mari, N. Shammah, P. Karalekas, and W. Zeng. "Mitiq: A software package for error mitigation on near-term quantum computers." (2020).

Canonical Noise Scaling

PER Representation:

$$\mathcal{U}(\rho) = \sum_\alpha \eta_\alpha \mathcal{O}_\alpha(\rho) = \gamma^+\Phi^+(\rho) - \gamma^-\Phi^-(\rho)\,,$$

- Positive volume, overhead: $\Phi^+ = \sum_{\eta_\alpha>0} \frac{|\eta_\alpha|}{\gamma^+}\mathcal{O}_\alpha(\rho)$
- Negative volume, overhead: $\Phi^- = \sum_{\eta_\alpha<0} \frac{|\eta_\alpha|}{\gamma^-}\mathcal{O}_\alpha(\rho)$

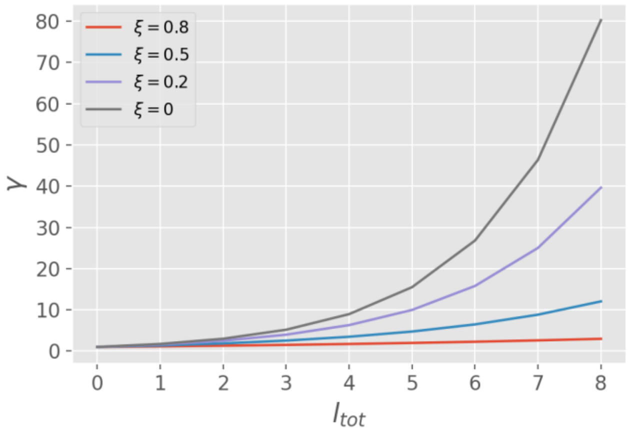
Canonical noise scaling [1]:

$$\mathcal{U}^{(\xi)}(\rho) = (\gamma^+ - \xi\gamma^-)\Phi^+(\rho) - (1-\xi)\gamma^-\Phi^-(\rho)$$

Controlling ξ interpolates between

- $\mathcal{U}^{(0)}(\rho) = \mathcal{U}(\rho)$
- $\mathcal{U}^{(1)}(\rho) = \tilde{\mathcal{U}}(\rho)$

Overhead is reduced to $\gamma^{(\xi)} = \gamma - \xi(\gamma - 1)$ for $\xi \in [0,1]$



The overhead interpolates between the hardware noise value $\gamma^{(1)} = 1$ and the noiseless value $\gamma^{(0)} = \gamma$.

Number of circuits for fixed precision $\propto \gamma^2$

- At $\xi = 0, \gamma = 1.73$. At depth $8, \gamma \approx 80$
- At $\xi = 0.5, \gamma = 1.37$. At depth $8, \gamma \approx 12$
- By combining estimates at $\xi = .5, \xi = 1$, and $\xi = 2$, the total number of circuits $\approx 14/\delta^2$

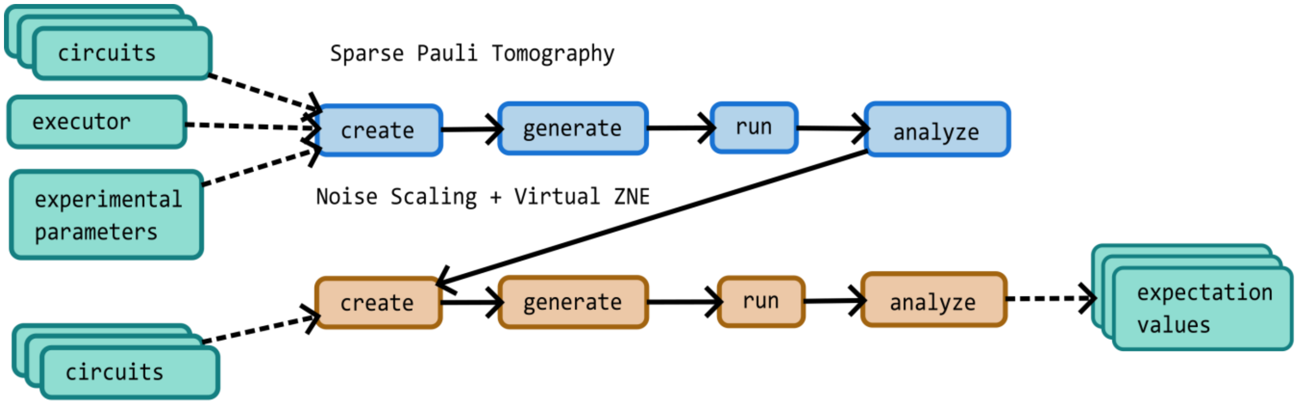
[1] Mari, Andrea, Nathan Shammah, and William J. Zeng. "Extending quantum probabilistic error cancellation by noise scaling." *Physical Review A* 104, no. 5 (2021): 052607.

Pauli Noise Tomography and Pauli Noise Scaling

Pauli Noise Tomography

- We implement a previously described procedure to characterize a Pauli-noise channel under Pauli twirling [1],[2].
- We extend this procedure to carry out PER, and we develop software to automate this process.

custom toolchain

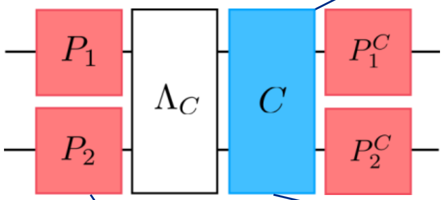


[1] Berg, Ewout van den, Zlatko K. Minev, Abhinav Kandala, and Kristan Temme. "Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors." arXiv preprint arXiv:2201.09866 (2022).
[2] Flammia, Steven T., and Joel J. Wallman. "Efficient estimation of Pauli channels." ACM Transactions on Quantum Computing 1, no. 1 (2020): 1-32.

Pauli Twirling

- Pauli operator P_a is an eigenvector of twirled noise channel with eigenvector f_a [2]

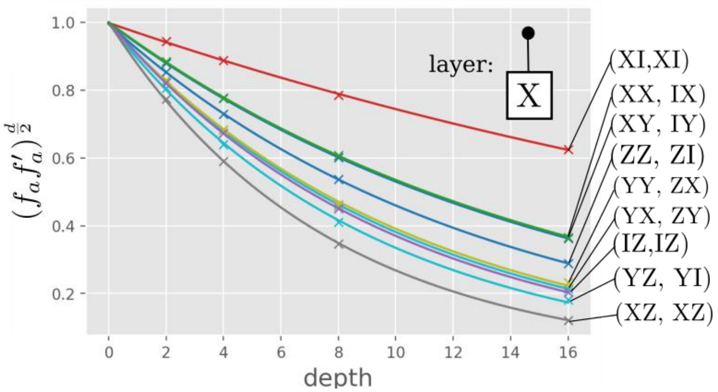
Randomly sampled Pauli gates *Conjugate Pauli operators*



Noise associated with layer *Clifford layer*

- Even number of repetitions of noisy Clifford layer
- Noise associated with dressed Clifford layers is learned

- Self-adjoint Clifford layers result in fidelity pairs $\sqrt{f_a f_a{}'}$ [1]



Simulated with random Pauli noise + amplitude damping noise

[1] Berg, Ewout van den, Zlatko K. Minev, Abhinav Kandala, and Kristan Temme. "Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors." arXiv preprint arXiv:2201.09866 (2022).
[2] Flammia, Steven T., and Joel J. Wallman. "Efficient estimation of Pauli channels." ACM Transactions on Quantum Computing 1, no. 1 (2020): 1-32.

Sparse Pauli-Lindblad Model

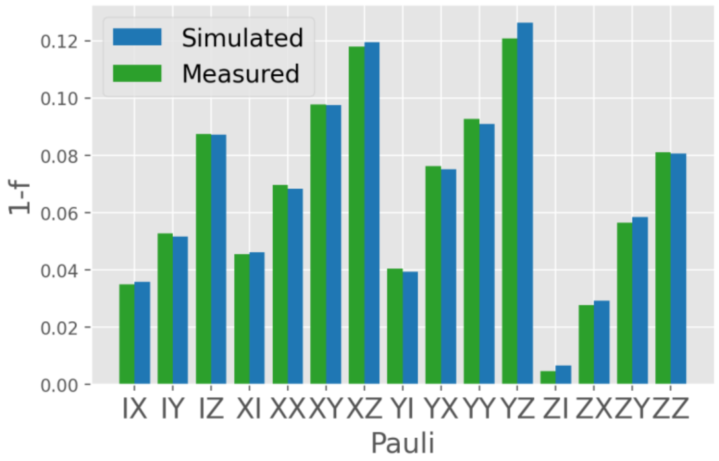
- Even repetitions of noisy Clifford layer result in exponential decays giving fidelity products:

$$\frac{1}{d}Tr\left(P_a\Lambda(P_a)\right) = (f_a f_a')^{\frac{d}{2}}$$

- Several direct fidelity measurements can lift the degeneracy for pairs where $P_a \neq P_a'$ [1]

Simulation of learning procedure

- random Pauli noise and amplitude damping channel.
- software automatically generates benchmark circuits based on input circuit
- data is automatically processed to obtain model estimate.



[1] Berg, Ewout van den, Zlatko K. Minev, Abhinav Kandala, and Kristan Temme. "Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors." arXiv preprint arXiv:2201.09866 (2022).

Noise Scaling with Sparse Pauli-Lindblad Model

Sparse model simplification: choose P_k that have support only on physically connected qubits [1]

Partial inversion of the noise channel: $\Lambda^{(\xi)} = \gamma^{(\xi)} \prod_k \omega_k^{(\xi)} \mathcal{I} - \left(1 - \omega_k^{(\xi)}\right) \mathcal{P}_k$

- $\omega_k^{(\xi)}$ *can be found using measured fidelities*
- *Efficient QPD sampling by sampling from each term in the product and recording the sign*
- *Partial inverse is* ***exact*** *with no numerical estimation*

Similar properties to canonical noise scaling!

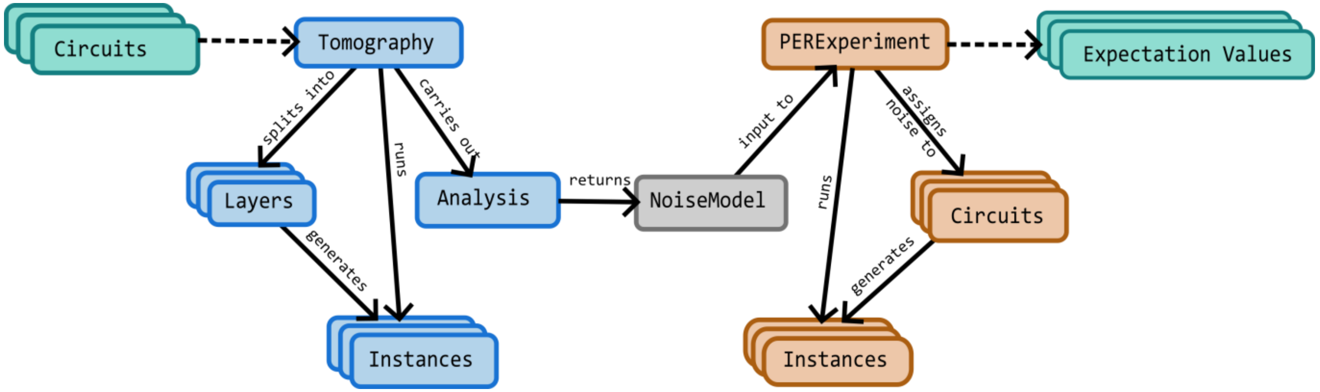
For $\xi \in [0,1]$, overhead reduces exponentially: $\gamma^{(\xi)} = \gamma^{1-\xi}$

Complexity has constant scaling $\mathcal{O}(1)$ in circuit depth and qubit number [1]

[1] Berg, Ewout van den, Zlatko K. Minev, Abhinav Kandala, and Kristan Temme. "Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors." arXiv preprint arXiv:2201.09866 (2022).

Software Package

Design



Functionality

Automates tomography and scaling process

- Breaks user-defined circuit into layers
- Generates & runs tomography circuits (skip some of these)
- Runs vZNE and extracts expectation values

Built on a multi-platform wrapper

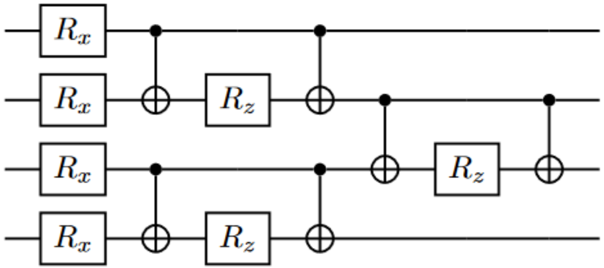
- Processor
- Circuit
- Instruction
- Pauli

Example Circuit

1D Transverse-Field Ising Model

$$H = -J \sum_i Z_i Z_{i+1} + h \sum_i X_i$$

Trotter step



$$R_x = RX(2h\delta t)$$

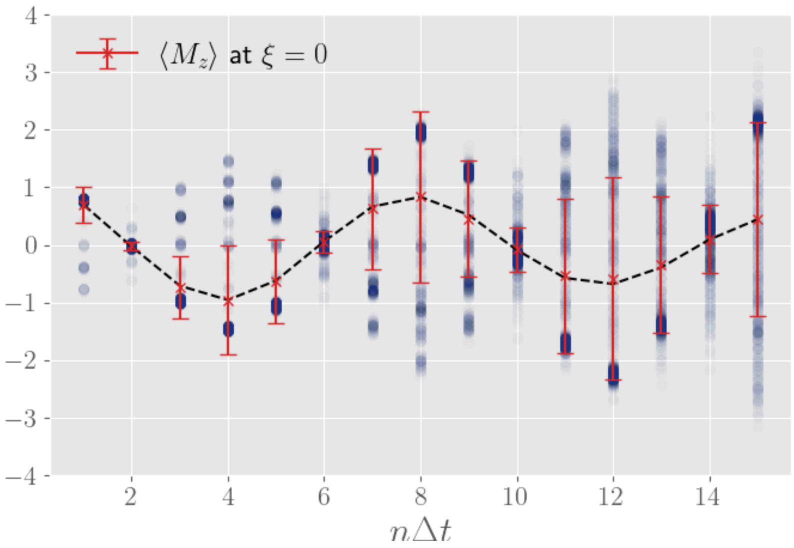
$$R_z = RZ(-2J\delta t)$$

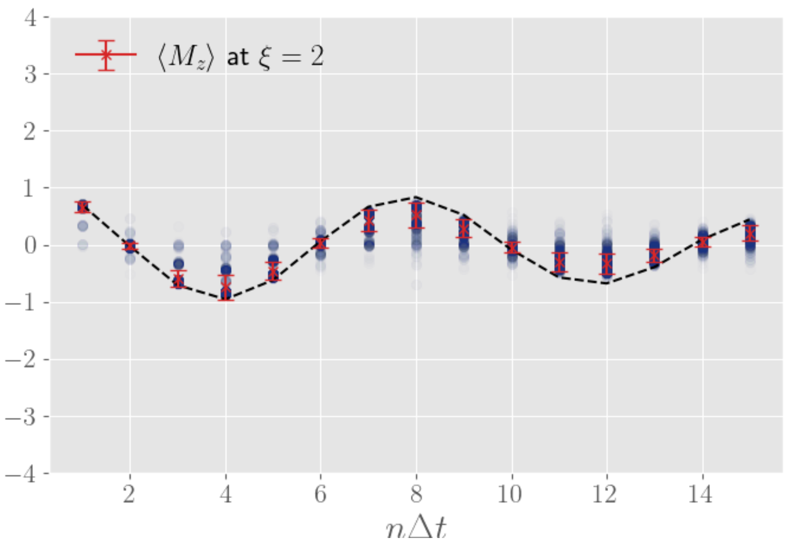
Quantity to measure

Global Z-Magnetization:

$$M_z = \frac{1}{N} \sum_i \langle Z_i \rangle$$

Visualization of PER





The tradeoff between bias and variance controlled by ξ

- Each blue dot represents 1024 shots of a single PER circuit
- Scaling down the noise incurs a higher variance
- This overhead is a symptom of the negativity of the distribution

Parameters

- 1000 PER samples
- 1024 shots per circuit
- Simulator: *FakeVigo*

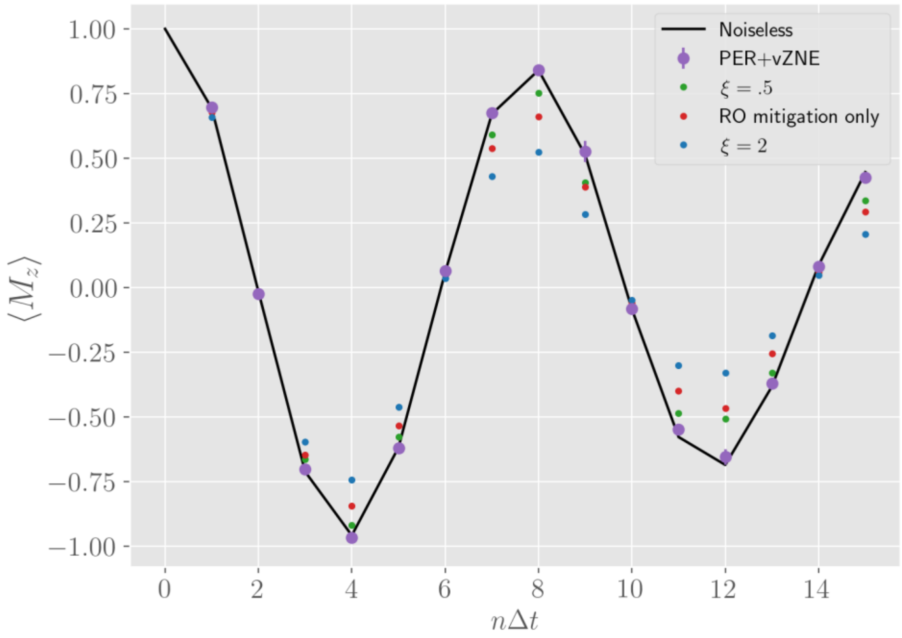
SUPERCONDUCTING QUANTUM MATERIALS & SYSTEMS CENTER

Results

Simulation of PER carried out on Z-magnetization of TFIM on four qubits

- Red, blue, and green dots show how the estimator converges with ξ
- The purple dots show extrapolation using $\xi = 0.5, 1,$ and 2
- Extrapolation shows good agreement with noiseless simulation

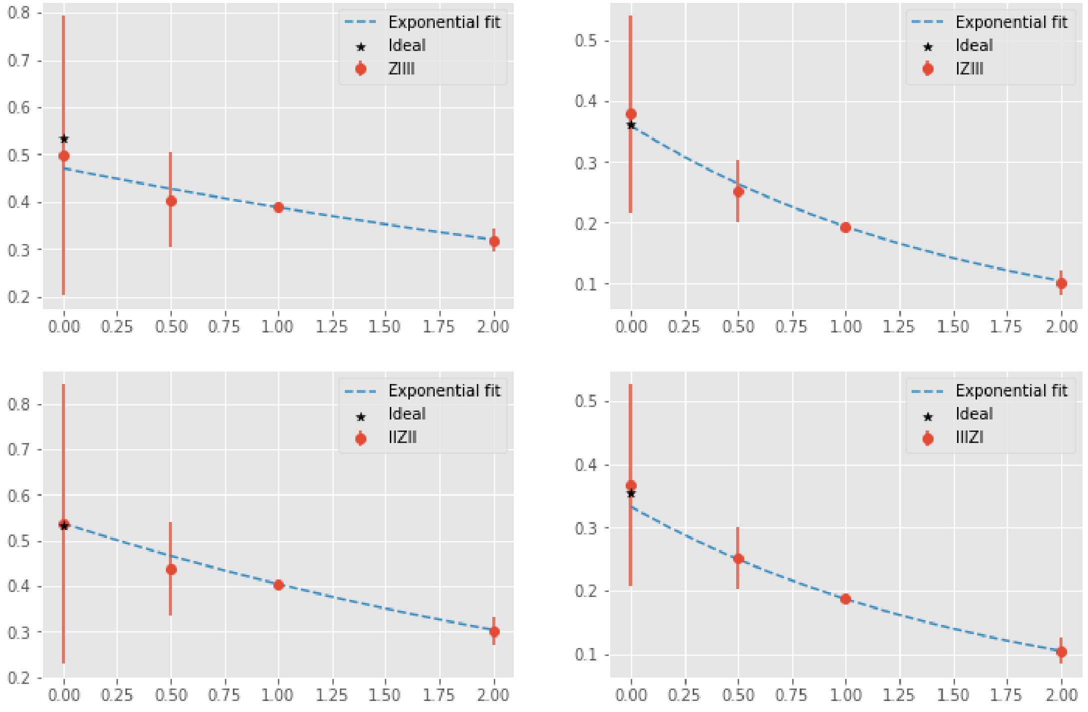
Parameters:
- 1000 PER samples
- 1024 shots
- $\Delta t = 0.2$
- $J = 0.15$
- $h = 1$



Results

Convergence of the individual estimators

- The convergence is exhibited by each expectation value $\langle Z_n \rangle$
- The extrapolation shown uses only the values $\xi \in \{.5,1,2\}$.
- Extrapolation using an exponential fit can yield results with accuracy similar to PEC without evaluating the expectation at $\xi = 0$.



Summary

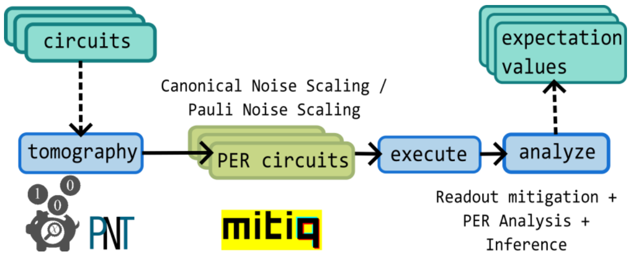
Summary

- ➢ Demonstrated usage of PyGSTi and Mitiq to combine GST and canonical noise scaling
- ➢ Created method for using sparse Pauli-Lindblad model for noise scaling
- ➢ Wrote software to automatically implement Pauli Noise Tomography and apply PER and vZNE with noise scaling

All code is publicly available. Please contact us if you are interested in contributing!

- ➢ Implementing wrappers for more commonly used API's
- ➢ Using memoization to decrease the circuits needed for PER
- ➢ Decreasing compile times through parametric compilation

Thank You!



- https://github.com/benmcdonough20/AutomatedPERTools
- https://github.com/unitaryfund/mitiq
- G. Mari, Andrea, Nathan Shammah, and William J. Zeng. "Extending quantum probabilistic error cancellation by noise scaling." *Physical Review A* 104, no. 5 (2021): 052607.

SUPERCONDUCTING QUANTUM MATERIALS & SYSTEMS CENTER

Two-Qubit Clifford Gates

Layer $\mathcal{U}$ gets broken into layer of self-adjoint Clifford gates and single-qubit gates:

$$\hat{\mathcal{U}} = \Lambda\mathcal{C}\mathcal{S}$$

Twirl is performed through layer using conjugate Pauli gates:

$$\begin{gathered}
\Lambda_s(\mathcal{P}_n) \circ \Lambda(\mathcal{C}\mathcal{P}_n^{\mathcal{C}}\mathcal{P}_{n-1}) \circ \cdots \circ \Lambda\left(\mathcal{C}\mathcal{P}_2^{\mathcal{C}}\mathcal{P}_1\right) \circ \Lambda\left(\mathcal{C}\mathcal{P}_1^{\mathcal{C}}\right) \\
= \Lambda_s(\mathcal{P}_n\Lambda\mathcal{P}_n)\mathcal{C} \circ \cdots \circ (\mathcal{P}_2\Lambda\mathcal{P}_2)\mathcal{C} \circ (\mathcal{P}_1\Lambda\mathcal{P}_1)\mathcal{C} \\
\to \Lambda_s\left(\Lambda^{\mathbb{P}^n}\mathcal{C}\right)^d
\end{gathered}$$

Basis change gates and readout gates are composed into first and last layer:

$$\Lambda_s(\mathcal{R}\mathcal{B}_a^{\dagger}\mathcal{P}_n) \circ \Lambda(\mathcal{C}\mathcal{P}_n^{\mathcal{C}}\mathcal{P}_{n-1}) \circ \cdots \circ \Lambda\left(\mathcal{C}\mathcal{P}_2^{\mathcal{C}}\mathcal{P}_1\right) \circ \Lambda\left(\mathcal{C}\mathcal{P}_1^{\mathcal{C}}\mathcal{B}_a\right)$$

Two-Qubit Clifford Gates

Layer of self-adjoint Clifford gates results in fidelity pairs:

$$\widetilde{\Lambda}\mathcal{C}(P_a) = \widetilde{\Lambda}(P_a^C) = f_a' P_a^C \qquad \widetilde{\Lambda}\mathcal{C}(P_a^C) = \widetilde{\Lambda}(P_a) = f_a P_a$$

$$\left(\widetilde{\Lambda}\mathcal{C}\right)^{2d}(P_a) = (f_a f_a')^d$$

Since $f_a f_a' = (\frac{1}{b} f_a)(b f_a')$, this results in degeneracy. The degeneracy can be lifted with measurements of the form

$$\Lambda_s(\mathcal{R}\mathcal{B}_a\mathcal{P})\Lambda(\mathcal{C}\mathcal{P}^C\mathcal{B}_a)$$

Can be used to obtain f_a. When P_a and P_a^C have different weights, this measurement is not robust to ground state preparation errors, but these are relatively low with active reset.

There are $6N$ such terms if N is the number of self-adjoint two-qubit Clifford gates in the layer. Fidelities of terms with support on disjoint gates can be measured simultaneously, resulting in 6 measurement bases. Measurements still scale as $\mathcal{O}(1)$ in n.

Gate Set Tomography

Advantages

- Complete characterization
  - Full superoperators are reconstructed
  - No noise transformation necessary
- Robust to SPAM
  - SPAM errors are incorporated into GST
  - SPAM gauge invariance does not affect PEC

Disadvantages

- Highly unscalable
  - $\sim 16^n$ superoperator elements to reconstruct
  - Accuracy is better for smaller gate sets
- Harder to produce QPD gate representations
  - Few degrees of freedom
  - Can result in high overhead

Pauli fidelity measurement

For any two Pauli eigenstates $|x\rangle, |y\rangle$ of $a \in \mathbb{P}^n$ satisfy $|y\rangle = b|x\rangle$ for some b $\in \mathbb{P}^n$. Since Pauli channels commute,

$$tr(\Lambda(a\Lambda(|y\rangle\langle y|))) = tr(\Lambda(a\Lambda(b|x\rangle\langle x|)) = tr(\Lambda(ab\Lambda(|x\rangle\langle x|)b)) = (-1)^{\langle a,b\rangle} tr(a\Lambda(|x\rangle\langle x|))$$

Pauli matrices are diagonal in their eigenbasis:

$$P = \sum_{b\in\{0,1\}^{\otimes n}} (-1)^{\langle a,b\rangle} |b\rangle\langle b|$$

If $a|x\rangle = -ab|x\,\rangle$, then $\langle a, b\rangle = 1$. Labeling eigenstates with binary numbers,

$$\frac{1}{\sqrt{2}} tr\left(p\Lambda\left(\sum_{x\in\{0,1\}^{\otimes n}} (-1)^x |x\rangle\langle x|\right)\right) = (-1)^y tr(p\Lambda(|y\rangle\langle y|))$$

Noise Inverse

Choosing the jump operators $L_k = \sqrt{\lambda_k} P_k$, the master equation becomes

$$\dot{\rho} = \sum_k \lambda_k (P_k \rho P_k - \rho)$$

This equation has the solution

$$\Lambda(\rho) = \prod_k \omega_k \mathcal{I} + (1 - \omega_k) \mathcal{P}_k$$

The inverse can be written down:

$$\Lambda^{-1}(\rho) = \gamma \prod_k \omega_k \mathcal{I} - (1 - \omega_k) \mathcal{P}_k$$

Where $\omega_k = 1/2\left(1 + e^{-2\lambda_k}\right)$ and the associated overhead is

$$\gamma = \exp\left(2 \sum_k \lambda_k\right)$$

Partial Inverse for Noise Tuning

$$\mathcal{L}_{\vec{\lambda}} = \sum_k \lambda_k (\mathcal{P}_k - \mathcal{I})$$

Individual Lindbladian terms commute,

$$[\mathcal{P}_j - \mathcal{I}, \mathcal{P}_k - \mathcal{I}] = 0$$

Therefore a generalized inverse can be written down in the form

$$\exp(\mathcal{L}_{\vec{\phi}-\vec{\lambda}}) \exp(\mathcal{L}_{\vec{\lambda}}) = \exp\left(\mathcal{L}_{\vec{\phi}}\right)$$

This can be implemented probabilistically to produce a noise model characterized by $\vec{\phi}$

Noise Tuning Sampling Procedure

for all $k \in [1, N]$:

- If $\phi_k < \lambda_k$:
  - Sample I with probability $\frac{1}{2}\left(1 + e^{-2(\lambda_k - \phi_k)}\right)$ and P_k otherwise, and compose into circuit
  - Scale result by $-e^{2(\lambda_k - \phi_k)}$
- otherwise:
  - Sample I with probability $\frac{1}{2}\left(1 + e^{-2(\phi_k - \lambda_k)}\right)$ and P_k otherwise, and compose into circuit

Average results