Efficient information recovery from Pauli noise via classical shadow

Yifei Chen[1], Zhan Yu[1,2], Chenghong Zhu[1,3], and Xin Wang[1,3]

1:Baidu Research

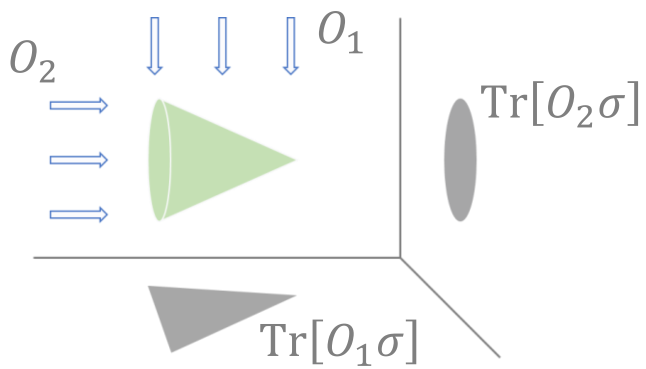
2:National University of Singapore

3:HKUST (Guangzhou)

arXiv: 2305.04148

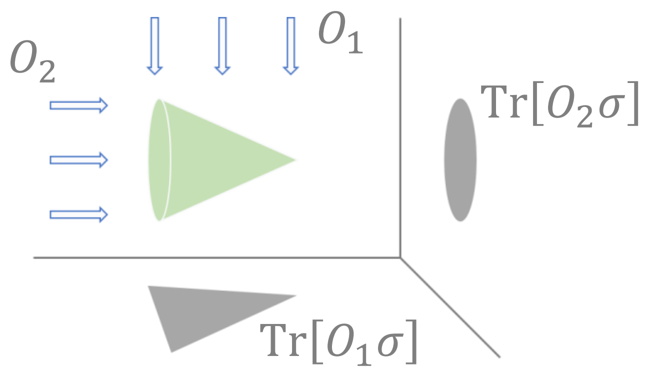
Shadow Information

- Results we usually want from quantum computing are still **classical information**
- **Shadow information** is the expectation value of some chosen observable: $\mathrm{Tr}[O\sigma]$



Shadow Information

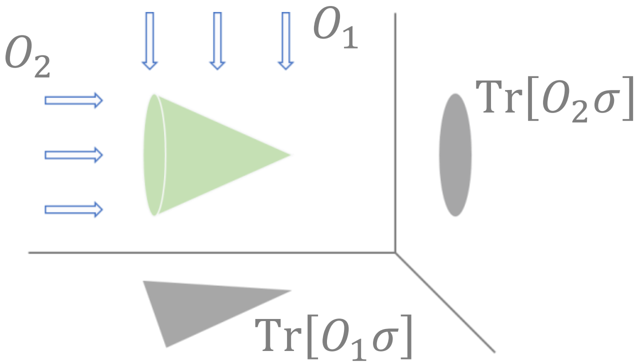
- Results we usually want from quantum computing are still **classical information**
- **Shadow information** is the expectation value of some chosen observable: $\mathrm{Tr}[O\sigma]$



- Shadow information is the target of many quantum algorithms

Shadow Information

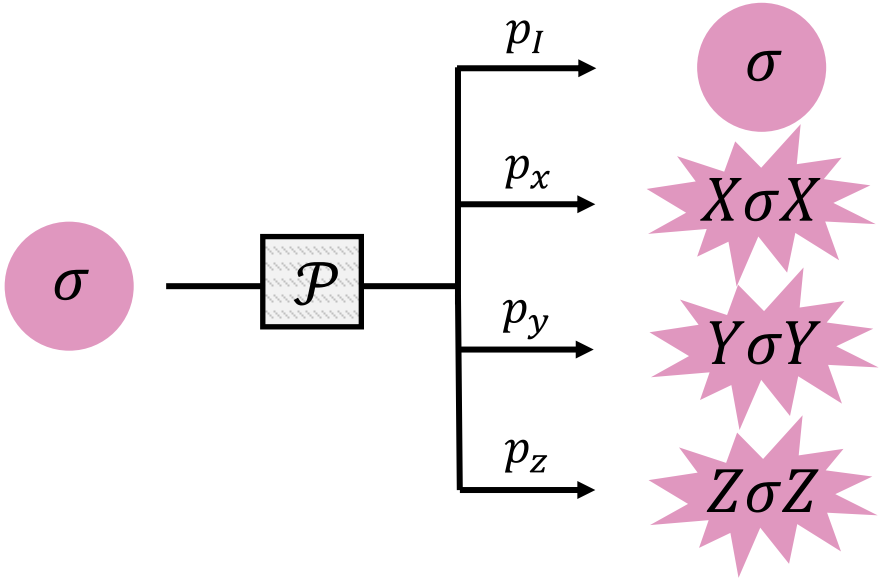
- Results we usually want from quantum computing are still **classical information**
- **Shadow information** is the expectation value of some chosen observable: $\mathrm{Tr}[O\sigma]$



- Shadow information is the target of many quantum algorithms
- Estimation of the information $\mathrm{Tr}[O\sigma]$ is obtained via making **measurements** with the corresponding observable O and post-processing.

Pauli noise

- One of the most standard theoretical models for quantum noise in the study of quantum error correction and mitigation is Pauli noise.



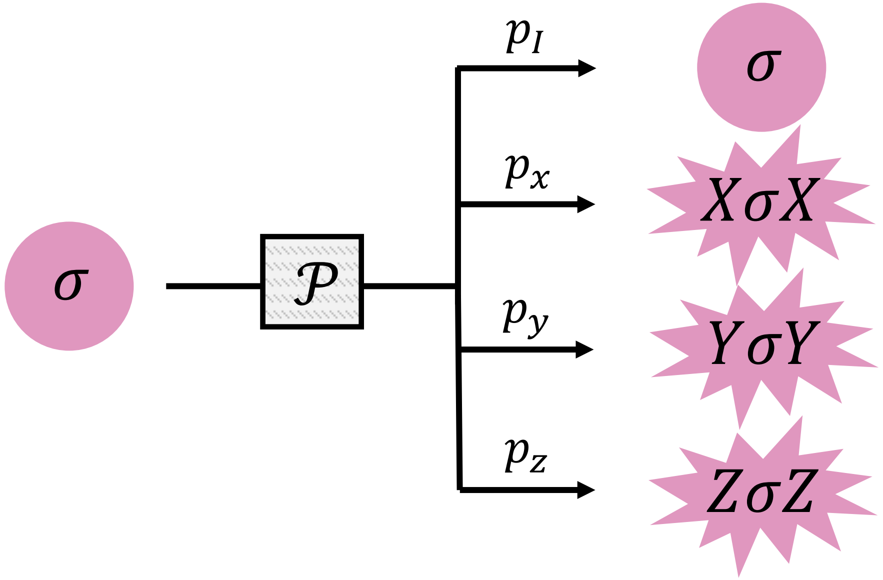
Pauli noise Kraus representation:

$$\mathcal{P}(\sigma) = p_I\sigma + p_x X\sigma X + p_y Y\sigma Y + p_z Z\sigma Z,$$

where X, Y, Z are Pauli matrices and $p_I + p_x + p_y + p_z = 1$.

Pauli noise

- One of the most standard theoretical models for quantum noise in the study of quantum error correction and mitigation is Pauli noise.



Pauli noise Kraus representation:

$$\mathcal{P}(\sigma) = p_I\sigma + p_x X\sigma X + p_y Y\sigma Y + p_z Z\sigma Z,$$

where X, Y, Z are Pauli matrices and $p_I + p_x + p_y + p_z = 1$.

- Pauli twirling can be applied to converts any noise channel to a Pauli channel

Problem of recovering from Pauli noise

- Given access to an unknown Pauli channel $\mathcal{P}$ and a noisy state $\mathcal{P}(\sigma)$.

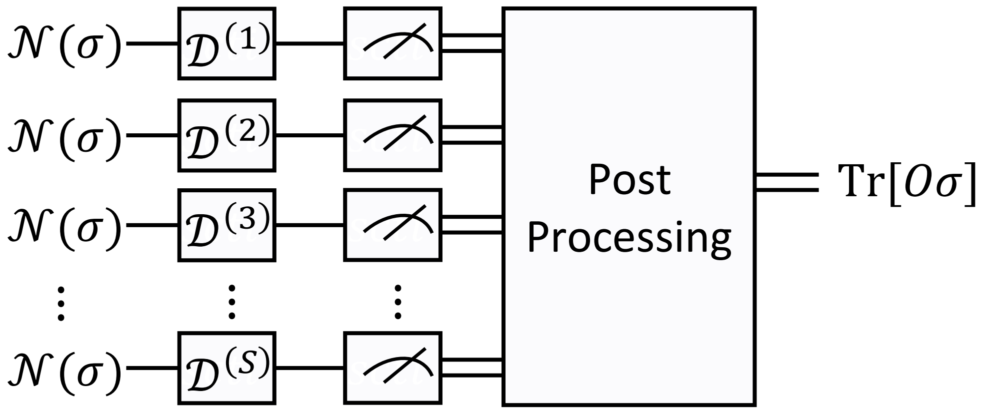
Problem of recovering from Pauli noise

- Given access to an unknown Pauli channel $\mathcal{P}$ and a noisy state $\mathcal{P}(\sigma)$.
- For a known k-local observable O, $\|O\| = 1$, we want a function $f(\mathcal{P}(\sigma), O)$ that can predict the ideal value of $tr(O\sigma)$ up to accuracy ϵ, i.e.

$$|f(\mathcal{P}(\sigma), O) - tr(O\sigma)| \leq \epsilon$$

Known method

- Probabilistic Error Cancellation
  - Given description of the channel $\mathcal{P}$, find the map $\mathcal{D}$ such that $tr\big(O\ \mathcal{D} \circ \mathcal{P}(\sigma)\big) = tr(O\sigma)$.
  - $\mathcal{D}$ is not CPTP but can be written as a linear combination of CPTP maps.
  - Simulate action of $\mathcal{D}$ by probabilistic sampling.

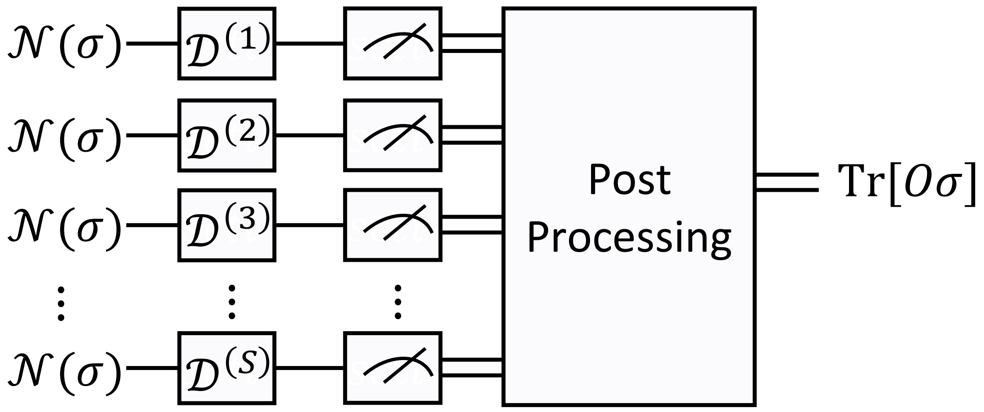


Zhao, Xuanqiang, et al. "Information recoverability of noisy quantum states." Quantum 7 (2023): 978.

Known method

- Probabilistic Error Cancellation
  - Given description of the channel $\mathcal{P}$, find the map $\mathcal{D}$ such that $tr\big(O\ \mathcal{D} \circ \mathcal{P}(\sigma)\big) = tr(O\sigma)$.
  - $\mathcal{D}$ is not CPTP but can be written as a linear combination of CPTP maps.
  - Simulate action of $\mathcal{D}$ by probabilistic sampling.

- Description of the channel required
  - Tomography – computational inefficient and still an approximation
  - Typically requires number of copies exponentially in number of qubits

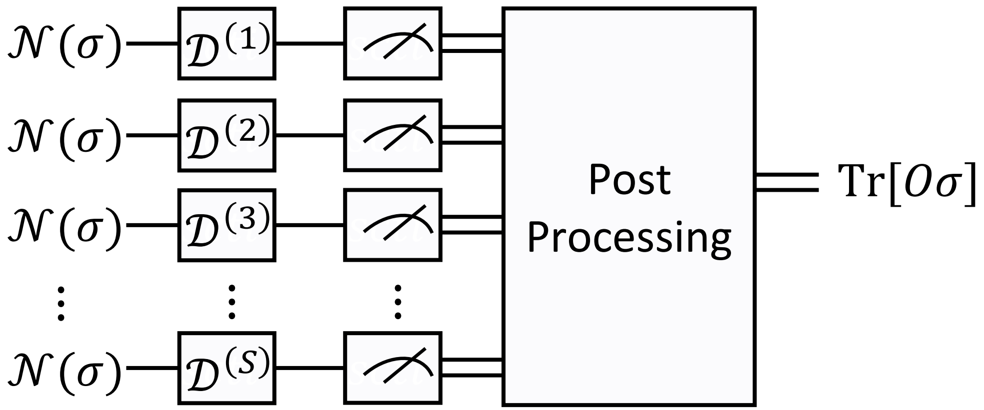


Zhao, Xuanqiang, et al. "Information recoverability of noisy quantum states." Quantum 7 (2023): 978.

Known method

- Probabilistic Error Cancellation
  - Given description of the channel $\mathcal{P}$, find the map $\mathcal{D}$ such that $tr\big(O\ \mathcal{D} \circ \mathcal{P}(\sigma)\big) = tr(O\sigma)$.
  - $\mathcal{D}$ is not CPTP but can be written as a linear combination of CPTP maps.
  - Simulate action of $\mathcal{D}$ by probabilistic sampling.

- Description of the channel required
  - Tomography – computational inefficient and still an approximation
  - Typically requires number of copies exponentially in number of qubits
- Implementation of arbitrary CPTP map required



Zhao, Xuanqiang, et al. "Information recoverability of noisy quantum states." Quantum 7 (2023): 978.

Main result

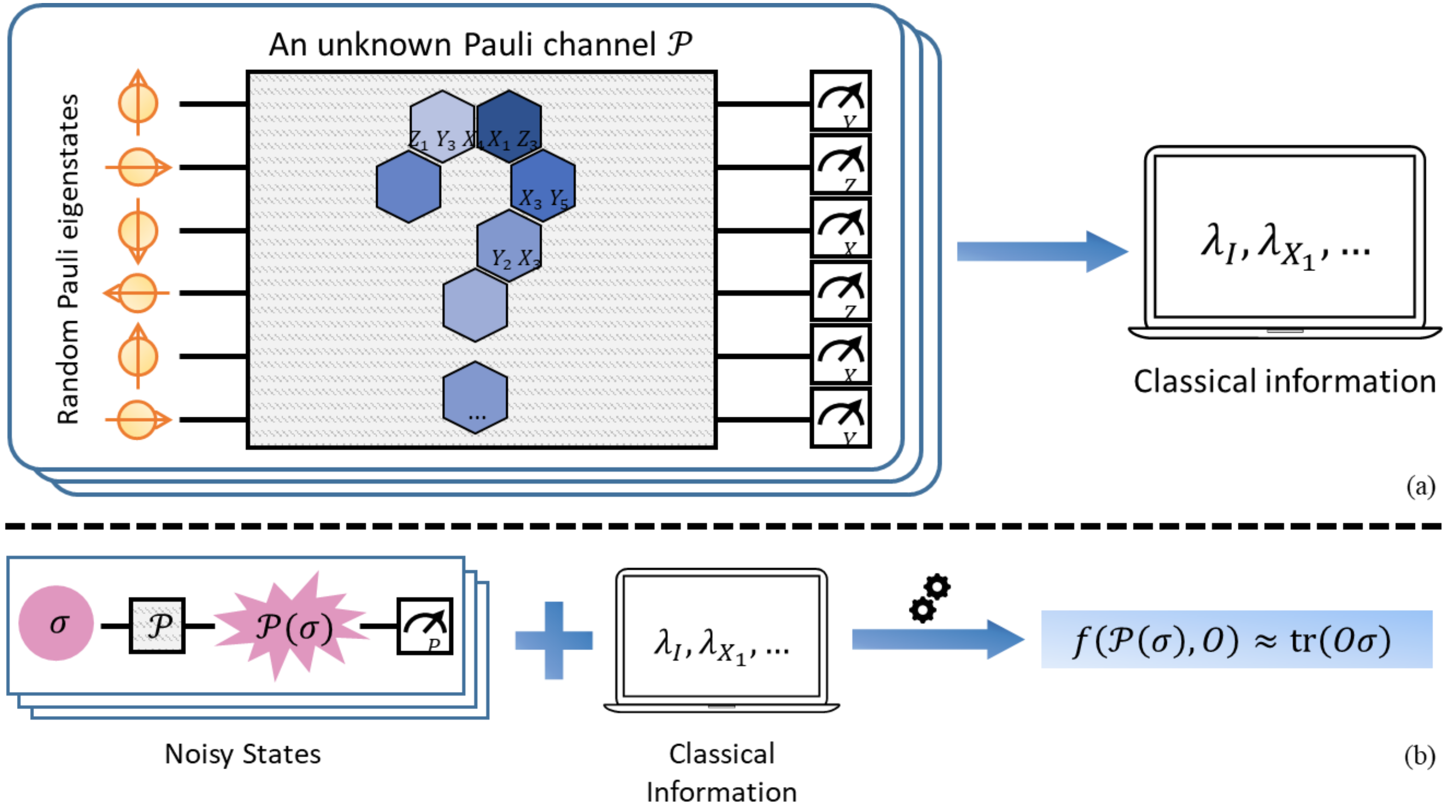
- Efficient algorithm for information recovery from Pauli noise for local observable

Main result

- Efficient algorithm for information recovery from Pauli noise for local observable
- Channel access $\mathcal{O}(\log(n)/\epsilon^2)$
- State copies $\mathcal{O}(\log(n)/\epsilon^2)$
- Computation time $\mathcal{O}\left(n^k \log(n)\ /\epsilon^2\right)$

Main result

- Efficient algorithm for information recovery from Pauli noise for local observable
  - Channel access $\mathcal{O}(\log(n)/\epsilon^2)$
  - State copies $\mathcal{O}(\log(n)/\epsilon^2)$
  - Computation time $\mathcal{O}\left(n^k \log(n)/\epsilon^2\right)$

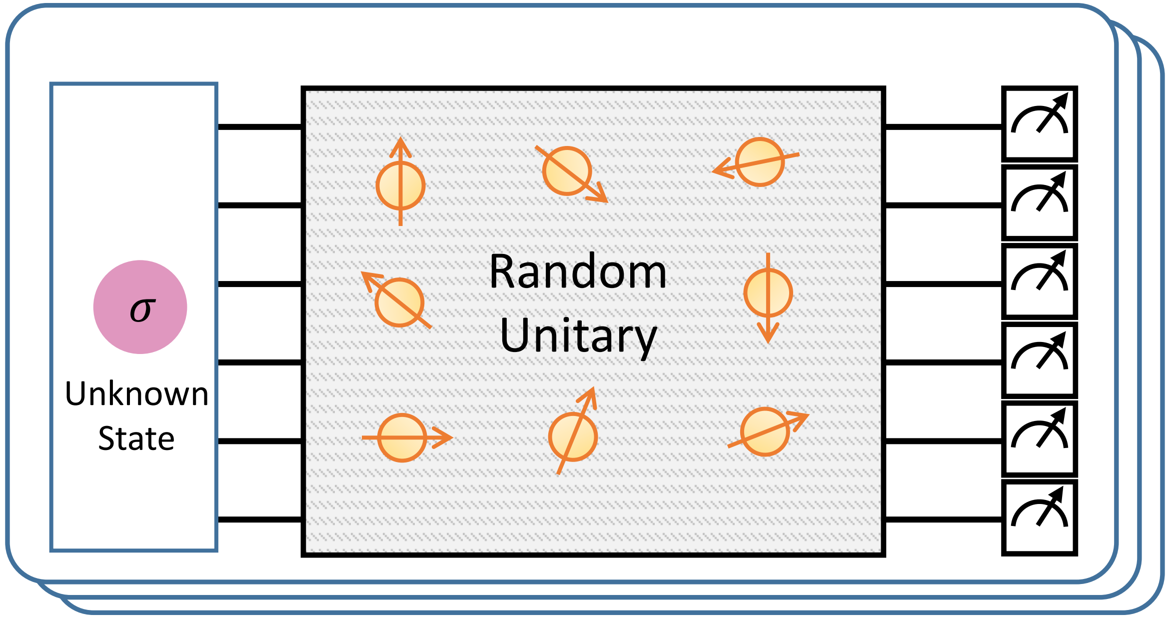


Learning classical information

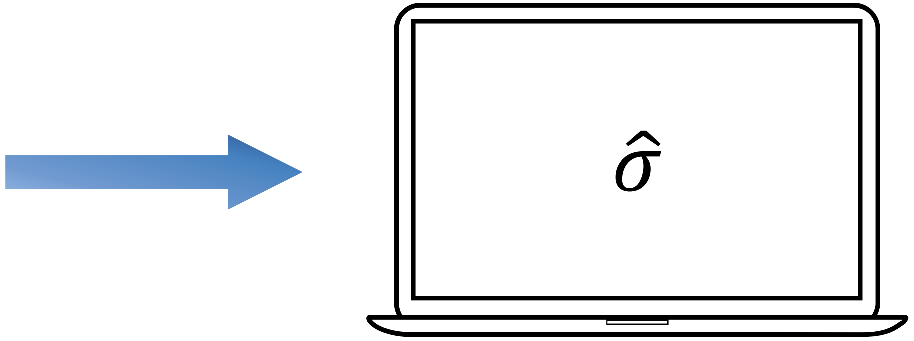
Post-processing measurement of noisy states to recover true shadow information

Classical shadow

- Used to predict $tr(O\sigma)$ for a set of O **simultaneously**



Repeated for a few times

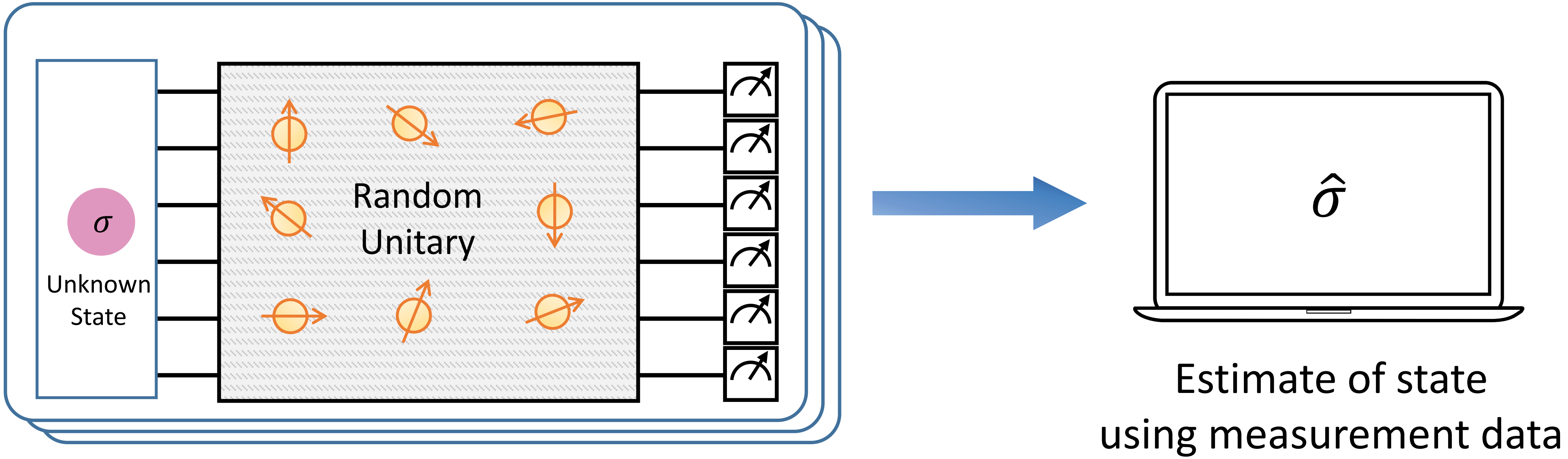


Estimate of state using measurement data

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." Nature Physics 16.10 (2020): 1050-1057.

Classical shadow

- Used to predict $tr(O\sigma)$ for a set of O **simultaneously**



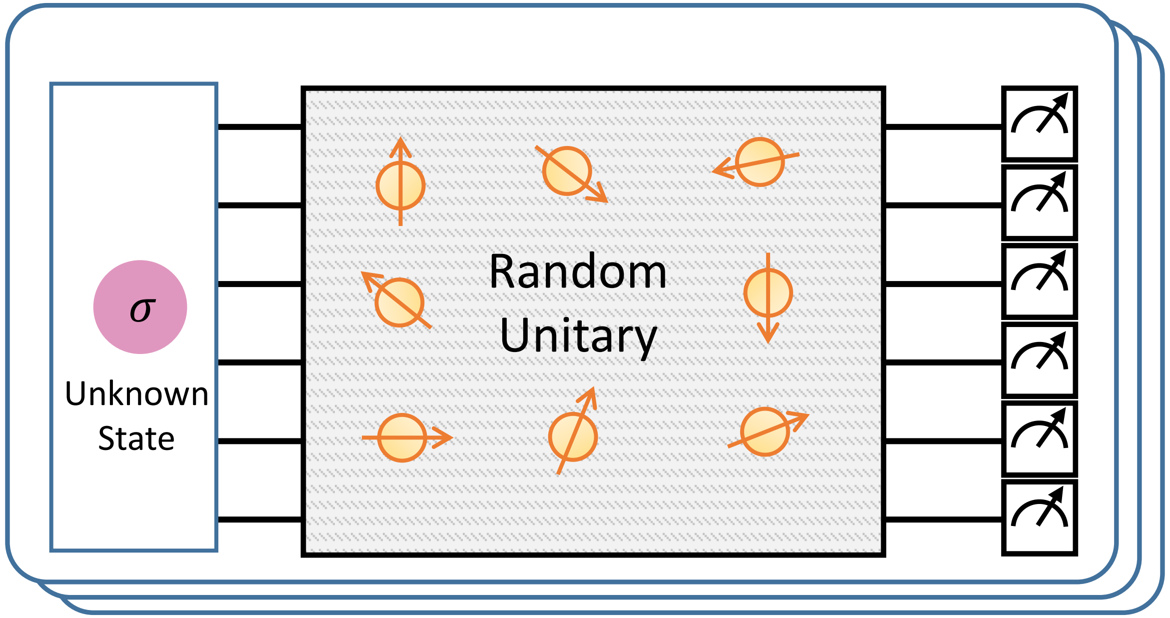
Repeated for a few times

$$\hat{\sigma} = \otimes_i \left(3U_i^{\dagger}|b_i\rangle\langle b_i|U_i - \mathbb{I}\right) = \otimes_i \left(3|t_i\rangle\langle t_i| - \mathbb{I}\right)$$

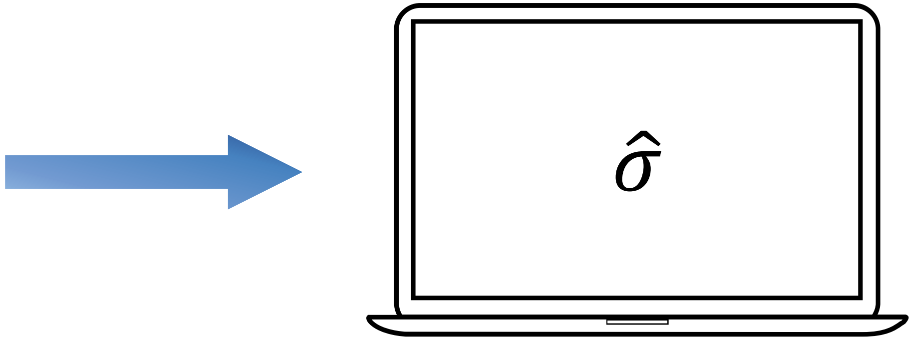
Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." Nature Physics 16.10 (2020): 1050-1057.

Classical shadow

- Used to predict $tr(O\sigma)$ for a set of O **simultaneously**



Repeated for a few times



Estimate of state using measurement data

$$\hat{\sigma} = \otimes_i \left(3U_i^\dagger |b_i\rangle\langle b_i| U_i - \mathbb{I}\right) = \otimes_i \left(3|t_i\rangle\langle t_i| - \mathbb{I}\right)$$

Can estimate shadow information for M k-local observables to accuracy ϵ with $\boldsymbol{\mathcal{O}(\log(M)/\epsilon^2)}$

Huang, Hsin-Yuan, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements." Nature Physics 16.10 (2020): 1050-1057.

Learning the Pauli channel

- Observation 1: $\mathrm{tr}\big(O\mathcal{P}(\sigma)\big) = tr\big(\mathcal{P}^\dagger(O)\sigma\big)$, where $\mathcal{P}^\dagger$ is the adjoint map of $\mathcal{P}$.
- Observation 2: Pauli operators are eigenoperators of $\mathcal{P}^\dagger$, i.e.,

$$\mathcal{P}^\dagger(Q) = \lambda_Q \mathrm{Q},$$

Learning the Pauli channel

- Observation 1: $\mathrm{tr}\big(O\mathcal{P}(\sigma)\big) = tr\big(\mathcal{P}^\dagger(O)\sigma\big)$, where $\mathcal{P}^\dagger$ is the adjoint map of $\mathcal{P}$.
- Observation 2: Pauli operators are eigenoperators of $\mathcal{P}^\dagger$, i.e.,

$$\mathcal{P}^\dagger(Q) = \lambda_Q \mathrm{Q},$$

- How can we estimate λ_Q?

  For any Pauli operator Q, we have that

$$\mathbb{E}_{\rho\sim\mathcal{D}^0}\, tr\big(Q\mathcal{P}(\rho)\big)tr(Q\rho) = \left(\frac{1}{3}\right)^{|Q|} \lambda_Q$$

where $\mathcal{D}^0$ is the uniform distribution over product Pauli eigenstates.

Learning the Pauli channel

- Observation 1: $\mathrm{tr}\big(O\mathcal{P}(\sigma)\big) = tr\big(\mathcal{P}^\dagger(O)\sigma\big)$, where $\mathcal{P}^\dagger$ is the adjoint map of $\mathcal{P}$.
- Observation 2: Pauli operators are eigenoperators of $\mathcal{P}^\dagger$, i.e.,

$$\mathcal{P}^\dagger(Q) = \lambda_Q \mathrm{Q},$$

- How can we estimate λ_Q?

  For any Pauli operator Q, we have that

$$\mathbb{E}_{\rho\sim\mathcal{D}^0}\, tr\big(Q\mathcal{P}(\rho)\big)tr(Q\rho) = \left(\frac{1}{3}\right)^{|Q|} \lambda_Q$$

where $\mathcal{D}^0$ is the uniform distribution over product Pauli eigenstates.

$$\hat{\lambda}_Q = 3^{|Q|} \cdot \frac{1}{N}\sum_{i=1}^{N}\prod_{j=1}^{n} tr\left(Q_j(3|t_{ij}\rangle\langle t_{ij}| - \mathbb{I})\right) tr(Q_j|s_{ij}\rangle\langle s_{ij}|)$$

Recover shadow information

Recover shadow information

- For any $\overleftarrow{O} = \sum_Q \overleftarrow{\alpha}_Q Q$,

$$\mathrm{tr}\left(\overleftarrow{O}\mathcal{P}(\sigma)\right) = \sum_Q \overleftarrow{\alpha}_Q \lambda_Q \mathrm{tr}(Q\sigma)$$

Recover shadow information

- For any $\overleftarrow{O} = \sum_Q \overleftarrow{\alpha}_Q Q$,

$$\mathrm{tr}\left(\overleftarrow{O}\mathcal{P}(\sigma)\right) = \sum_Q \overleftarrow{\alpha}_Q \lambda_Q \mathrm{tr}(Q\sigma)$$

- Compare with $\mathrm{tr}(O\sigma) = \sum_Q \alpha_Q \mathrm{tr}(Q\sigma)$, simply set $\overleftarrow{\alpha}_Q = \frac{\alpha_Q}{\lambda_Q}$.

Recover shadow information

- For any $\overleftarrow{O} = \sum_Q \overleftarrow{\alpha}_Q Q$,

$$\mathrm{tr}\left(\overleftarrow{O}\mathcal{P}(\sigma)\right) = \sum_Q \overleftarrow{\alpha}_Q \lambda_Q \mathrm{tr}(Q\sigma)$$

- Compare with $\mathrm{tr}(O\sigma) = \sum_Q \alpha_Q \mathrm{tr}(Q\sigma)$, simply set $\overleftarrow{\alpha}_Q = \frac{\alpha_Q}{\lambda_Q}$.

$$\sum_{Q:|Q|\leq k} \frac{\alpha_Q}{\lambda_Q} tr\big(Q\mathcal{P}(\sigma)\big) = tr(O\sigma)$$

Recover shadow information

- For any $\overleftarrow{O} = \sum_Q \overleftarrow{\alpha}_Q Q$,

$$\mathrm{tr}\left(\overleftarrow{O}\mathcal{P}(\sigma)\right) = \sum_Q \overleftarrow{\alpha}_Q \lambda_Q \mathrm{tr}(Q\sigma)$$

- Compare with $\mathrm{tr}(O\sigma) = \sum_Q \alpha_Q \mathrm{tr}(Q\sigma)$, simply set $\overleftarrow{\alpha}_Q = \frac{\alpha_Q}{\lambda_Q}$.

$$\sum_{Q:|Q|\leq k} \frac{\alpha_Q}{\lambda_Q} tr\left(Q\mathcal{P}(\sigma)\right) = tr(O\sigma)$$

$$\sum_{Q:|Q|\leq k} \frac{\alpha_Q}{\hat{\lambda}_Q} \widehat{tr}\left(Q\mathcal{P}(\sigma)\right)$$

Algorithm for information recovery

1. Prepare N random n-fold product Pauli eigenstates $\{\rho_i = \otimes_j |s_{ij}\rangle\langle s_{ij}|\}$.
2. Apply the unknown channel $\mathcal{P}$ on the N random states and perform random Pauli measurements on each qubit, obtaining data $\{\otimes_j |t_{ij}\rangle\langle t_{ij}|\}$.
3. For each n-qubit Pauli operator Q with $|Q| \leq k$, compute $\hat{\lambda}_Q$.

Learning

---

4. For $O = \sum_{Q:|Q|\leq k} \alpha_Q Q$, let $\overleftarrow{\alpha}_Q = \alpha_Q / \hat{\lambda}_Q$ for $|Q| \leq k$.
5. Use classical shadows to estimate $tr\big(Q\mathcal{P}(\sigma)\big)$ for $|Q| \leq k$.
6. Construct estimator for $tr(O\sigma)$ as

$$f(\mathcal{P}(\sigma), O) = \sum_{Q:|Q|\leq k} \overleftarrow{\alpha}_Q \, \widehat{tr}\big(Q\mathcal{P}(\sigma)\big)$$

Post-processing

Complexity of the algorithm

- Three quantities to consider: number of channel used (N_1), number of noisy states used (N_2) and total computation time

Complexity of the algorithm

- Three quantities to consider: number of channel used (N_1), number of noisy states used (N_2) and total computation time

- ϵ_1: accuracy of estimating each λ_Q

$$N_1 = \mathcal{O}(\log\left(n^k\right) / \epsilon_1^2)$$

Complexity of the algorithm

- Three quantities to consider: number of channel used (N_1), number of noisy states used (N_2) and total computation time

- ϵ_1: accuracy of estimating each λ_Q

$$N_1 = \mathcal{O}(\log(n^k) / \epsilon_1^2)$$

- ϵ_2: accuracy of estimating each $tr(Q\mathcal{P}(\sigma))$

$$N_2 = \mathcal{O}(\log(n^k) / \epsilon_2^2)$$

Complexity of the algorithm

- Three quantities to consider: number of channel used (N_1), number of noisy states used (N_2) and total computation time

- ϵ_1: accuracy of estimating each λ_Q

$$N_1 = \mathcal{O}(\log(n^k)/\epsilon_1^2) \qquad \epsilon_1 = C\lambda_{min}^{(k)}\epsilon$$

- ϵ_2: accuracy of estimating each $tr(Q\mathcal{P}(\sigma))$

$$N_2 = \mathcal{O}(\log(n^k)/\epsilon_2^2) \qquad \epsilon_2 = \mathcal{O}(\epsilon(1-\epsilon))$$

Complexity of the algorithm

- Three quantities to consider: number of channel used (N_1), number of noisy states used (N_2) and total computation time

- ϵ_1: accuracy of estimating each λ_Q

  $N_1 = \mathcal{O}(\log(n^k)/\epsilon_1^2)$ $\quad$ $\epsilon_1 = C\lambda_{min}^{(k)}\epsilon$

- ϵ_2: accuracy of estimating each $tr(Q\mathcal{P}(\sigma))$

  $N_2 = \mathcal{O}(\log(n^k)/\epsilon_2^2)$ $\quad$ $\epsilon_2 = \mathcal{O}(\epsilon(1-\epsilon))$

$$\mathcal{O}(\log(n)/\epsilon^2)$$

Complexity of the algorithm

- Three quantities to consider: number of channel used (N_1), number of noisy states used (N_2) and total computation time

- ϵ_1: accuracy of estimating each λ_Q

$$N_1 = \mathcal{O}(\log(n^k)/\epsilon_1^2) \qquad \epsilon_1 = C\lambda_{min}^{(k)}\epsilon$$

- ϵ_2: accuracy of estimating each $tr(Q\mathcal{P}(\sigma))$

$$N_2 = \mathcal{O}(\log(n^k)/\epsilon_2^2) \qquad \epsilon_2 = \mathcal{O}(\epsilon(1-\epsilon))$$

$$\mathcal{O}(\log(n)/\epsilon^2)$$

Computational complexity: $\mathcal{O}(n^k \log(n)/\epsilon^2)$

Numerical experiments

- Two-qubit product Pauli channel with p_I around 0.75
- Observable is a Heisenberg-typed Hamiltonian: $O = \sum_{j=1}^{n-1}\left(J_x\sigma_j^x\sigma_{j+1}^x + J_y\sigma_j^y\sigma_{j+1}^y + J_z\sigma_j^z\sigma_{j+1}^z + h\sigma_j^z\right)$
- To see how much our algorithm improves estimation, for 500 random states, $\{\sigma_1, \dots, \sigma_{500}\}$ subject them to noise. We compute

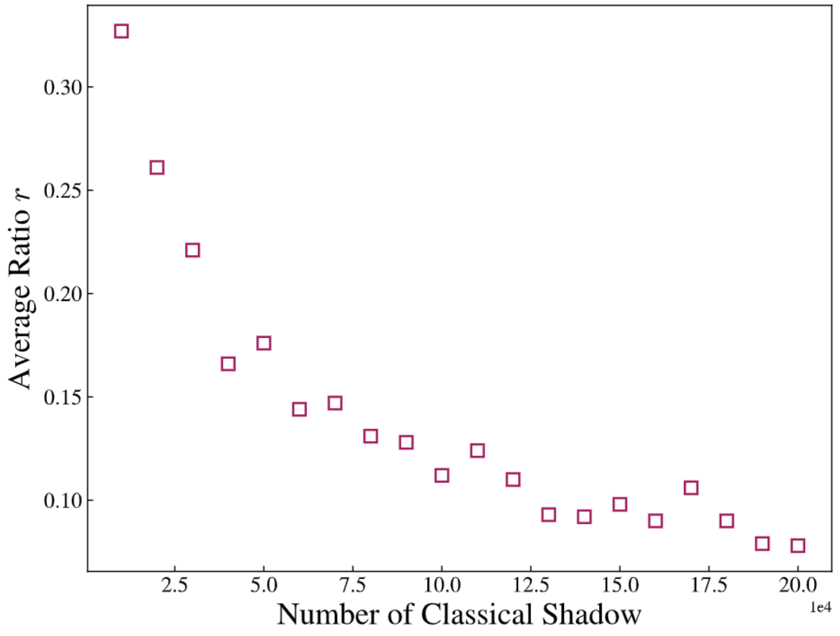
$$\text{MAE with no post-processing} = \frac{1}{500}\sum_{i=1}^{500}\left|tr\left(O\mathcal{P}(\sigma_i)\right) - tr(O\sigma_i)\right| \qquad \text{MAE with post-processing} = \frac{1}{500}\sum_{i=1}^{500}\left|f(\mathcal{P}(\sigma_i), O) - tr(O\sigma_i)\right|$$

- This allows us to compute the ratio between the two MAEs

$$r = \frac{\text{MAE with post–processing}}{\text{MAE with no post–processing}}$$

- We repeat this 10 times for N ranging from 10,000 to 200,000

As number of samples increases, more precise classical Information we are extracting, accuracy improves with samples.



Application in mitigating Pauli errors in Clifford circuits

Application in mitigating Pauli errors in Clifford circuits

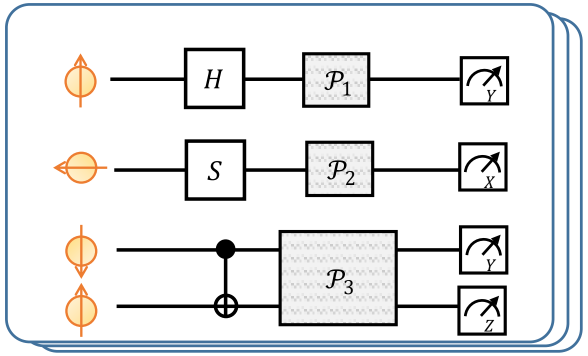
- Mitigate Pauli errors in Clifford circuit, consisting of $H, S, CNOT$ gates.

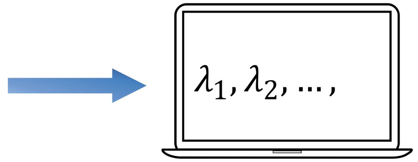
Application in mitigating Pauli errors in Clifford circuits

- Mitigate Pauli errors in Clifford circuit, consisting of $H, S, CNOT$ gates.
  - The ideal circuit is $\mathcal{C}$ and the noisy circuit is $\widetilde{\mathcal{C}}$, assuming that Pauli noise after each type of gate is the same.

Application in mitigating Pauli errors in Clifford circuits

- Mitigate Pauli errors in Clifford circuit, consisting of $H, S, CNOT$ gates.
  - The ideal circuit is $\mathcal{C}$ and the noisy circuit is $\widetilde{\mathcal{C}}$, assuming that Pauli noise after each type of gate is the same.
  - The goal is to obtain the information $\mathrm{tr}\big(O\mathcal{C}(\sigma)\big)$ for some observable O.
- Method:
  - Learn eigenvalues of the three noise channels

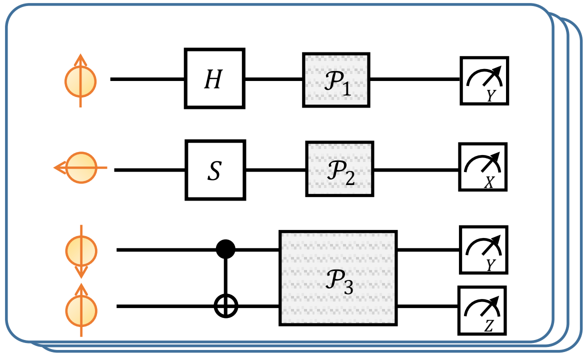


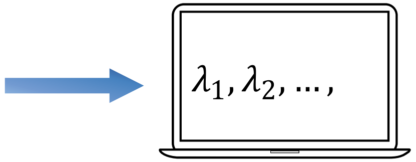


Classical information

Application in mitigating Pauli errors in Clifford circuits

- Mitigate Pauli errors in Clifford circuit, consisting of $H, S, CNOT$ gates.
  - The ideal circuit is $\mathcal{C}$ and the noisy circuit is $\widetilde{\mathcal{C}}$, assuming that Pauli noise after each type of gate is the same.
  - The goal is to obtain the information $\mathrm{tr}\big(O\mathcal{C}(\sigma)\big)$ for some observable O.
- Method:
  - Learn eigenvalues of the three noise channels





Classical information

  - Each α_Q is scaled by product of eigenvalues

Extensions of the algorithm

Extensions of the algorithm

- Can the algorithm work for other types of channel while remain $\mathcal{O}(poly(n))$?

Extensions of the algorithm

- Can the algorithm work for other types of channel while remain $\mathcal{O}(poly(n))$?

- Sufficient condition:

  $\mathcal{N}^\dagger$ is **weight contracting**, i.e. $\left|\mathcal{N}^\dagger(Q)\right| \leq |Q|$

Extensions of the algorithm

- Can the algorithm work for other types of channel while remain $\mathcal{O}(poly(n))$?

- Sufficient condition:

  $\mathcal{N}^\dagger$ is **weight contracting**, i.e. $\left|\mathcal{N}^\dagger(Q)\right| \leq |Q|$

- Updated algorithm for weight contracting channel
- Channel access $\mathcal{O}(\boldsymbol{n^{2k}} \log(n)/\epsilon^2)$
- State copies $\mathcal{O}(\log(n)/\epsilon^2)$
- Computation time $\mathcal{O}\left(\boldsymbol{n^{4k}} \log(n)\ /\epsilon^2\right)$

Conclusion

- Main results:
  - Scalable algorithm for information recovery from Pauli noise for local observable
  - $\mathcal{O}(\log(n))$ sample complexity, $\mathcal{O}(n^k \log(n))$ time complexity
  - Application in Clifford circuit error mitigation
  - Extension to weight contracting channel
- Questions
  - Complete characterisation of weight contracting channel
  - What other groups of observables can the algorithm be applied to

Conclusion

- Main results:
  - Scalable algorithm for information recovery from Pauli noise for local observable
  - $\mathcal{O}(\log(n))$ sample complexity, $\mathcal{O}(n^k \log(n))$ time complexity
  - Application in Clifford circuit error mitigation
  - Extension to weight contracting channel
- Questions
  - Complete characterisation of weight contracting channel
  - What other groups of observables can the algorithm be applied to

Thank you for listening

arXiv: 2305.04148