- Quantum error mitigation
  - General principles
  - ZNE
  - PEC
  - Measurement error mitigation

Follow Cai et al., RMP (2022).

Definition of QEM:

Algorithmic scheme that reduces noise induced bias on estimator of expectation value O by (classical) post-processing outputs of ensemble of circuit runs, executed at the same noise level as the unmitigated ones or above.

In short: post-processing algorithm that reduces noise induced bias of estimator $\hat{O}$ using output of noisy circuit runs.

- no threshold needed (unlike QEC)
- in practice QEM only works if total fault rate is below O(1). limits max. circuit size as function of error rate p.

- goal is to translate improvements in hardware (smaller p, more qubits, etc) into improvements for quantum information processing.

Desirable features of QEM algorithms

- small or no qubit overhead
- simple & straightforwardly implementable on hardware
- accuracy guarantees, ideally formal error bounds as function of noise level and classical resources

Error-mitigated estimators of observable O

Primary circuit $\mathcal{U}_p$: $\mathcal{U}_p \rho_{init} = U_p \rho_{init} U_p^\dagger = \rho_0$

Noiseless ideal output state : ρ_0

Noisy output state : ρ

Goal: estimate $\mathrm{Tr}[O \rho_0]$ (target parameter)

Quality of an estimator $\hat{O}$ is given by the mean squared error

$$\mathrm{MSE}[\hat{O}] = E\left[\left(\hat{O} - \mathrm{Tr}[O\rho_0]\right)^2\right]$$

Goal of QME: reduce MSE using fixed number of classical & quantum resources.

Decompose MSE into bias and variance:

$$\mathrm{MSE}[\hat{O}] = \mathrm{Bias}[\hat{O}]^2 + \mathrm{Var}[\hat{O}]$$

with

$$\mathrm{Bias}[\hat{O}] = E[\hat{O}] - \mathrm{Tr}[O\rho_0]$$

$$\mathrm{Var}[\hat{O}] = E(\hat{O}^2) - E(\hat{O})^2$$

Simplest estimator: measure O on noisy output state ρ using N_{cir} number of circuit executions (also called shots).

Denote measurement output of a single circuit execution by random variable $\hat{O}_\rho$.

Take N_{cir} measurements to obtain sample of random noisy outputs $\{ O_{\rho,1}, O_{\rho,2}, \dots, O_{\rho,N_{cir}} \}$.

Sample mean estimator

$$\bar{O}_\rho = \frac{1}{N_{cir}} \sum_{j=1}^{N_{cir}} O_{\rho,j}$$

Sample variance

$$\mathrm{Var}[\hat{O}_\rho] = \frac{1}{N_{cir}} \sum_{j=1}^{N_{cir}} \left(O_{\rho,j} - \bar{O}_\rho \right)^2$$

Variance of the mean (IID measurements)

$$\mathrm{Var}[\bar{O}_\rho] = \mathrm{Var}\left[\frac{1}{N_{cir}} \sum_j O_{\rho,j} \right] = \frac{1}{N_{cir}} \mathrm{Var}[\hat{O}_\rho]$$

Has MSE given by (error of the mean)

$$\mathrm{MSE}[\bar{O}_\rho] = \underbrace{\left(\mathrm{Tr}[O\rho] - \mathrm{Tr}[O\rho_0]\right)^2}_{= \mathrm{Bias}[\bar{O}_\rho] \overset{!}{=} \mathrm{Bias}[\hat{O}_\rho]}$$

$$+ \underbrace{\frac{1}{N_{\mathrm{cir}}}\left[\mathrm{Tr}[O^2\rho] - \mathrm{Tr}[O\rho]^2\right]}_{= \mathrm{Var}[\bar{O}_\rho] = \frac{\mathrm{Var}[\hat{O}_\rho]}{N_{\mathrm{cir}}}}$$

variance of the mean goes to zero as $N_{\mathrm{cir}} \to \infty$ (shot noise).

$\Rightarrow$ for large N_{cir}, the $\mathrm{MSE}[\bar{O}_\rho]$ is limited by the bias.

<u>Goal</u>: QEM constructs more sophisticated estimator $\hat{O}_{\mathrm{QEM}}$ such that with the same N_{cir}, the bias is reduced:

$$|\mathrm{Bias}[\bar{O}_{\mathrm{QEM}}]| \leq |\mathrm{Bias}[\bar{O}_\rho]|$$

But this comes with a tradeoff of increased variance

$$\mathrm{Var}[\bar{O}_{QEM}] \geq \mathrm{Var}[\bar{O}_{g}].$$

Bias-variance tradeoff is illustrated in the figure from Cai et al., arxiv: 2210:00921.

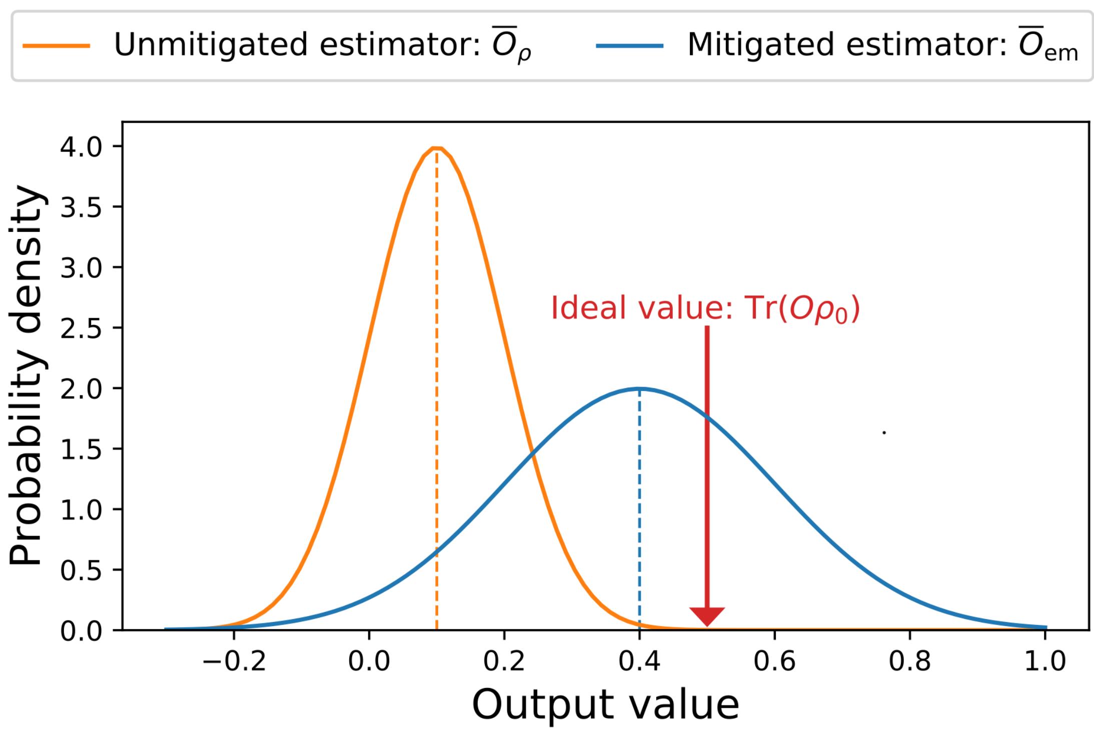

FIG. 1 The probability density distributions of the unmitigated estimator and the error-mitigated estimator. We see a decrease of bias and an increase of variance after performing error mitigation.

Let's quantify the sampling overhead that comes with the increased variance.

Two ways:

① using increased variance $\mathrm{Var}[\bar{O}_{QEM}] \geq \mathrm{Var}[\bar{O}_{\rho}]$

② using increased range of estimator $R[\hat{O}_{QEM}] \geq R[\hat{O}_{\rho}]$.

①

Define „one-shot" error-mitigated estimator $\hat{O}_{QEM}$ that satisfies

$$E[\hat{O}_{QEM}] = E[\bar{O}_{QEM}]$$

$$\mathrm{Var}[\hat{O}_{QEM}] = N_{cir}\,\mathrm{Var}[\bar{O}_{QEM}]$$

Number of shots needed for shot noise level (standard dev.) to be below ε given by $N^{\varepsilon}_{shot}(\hat{X}) = \dfrac{\mathrm{Var}[\hat{X}]}{\varepsilon^2}$.

simply $\varepsilon = \sqrt{\frac{\mathrm{Var}[\hat{X}]}{N_{shot}^{\varepsilon}}}$

Thus, to reach same shot noise level ε with QEM estimator $\hat{O}_{QEM}$ and simple noisy $\hat{O}_g$, we need to average over more circuit runs using QEM:

$$C_{QEM} = \frac{N_{shot}^{\varepsilon}[\hat{O}_{QEM}]}{N_{shot}^{\varepsilon}[\hat{O}_g]} = \frac{\mathrm{Var}[\hat{O}_{QEM}]}{\mathrm{Var}[\hat{O}_g]}$$

This is called the sampling overhead.

(2) C_{QEM} can also be estimated using the range $R[\hat{X}]$ of the estimator using Hoeffding's inequality. The smallest number of samples

required to estimate $E[\hat{X}]$ to ε-precision with $1-\delta$ probability using Hoeffding's inequality is

$$N^{\varepsilon,\delta}_{\mathrm{Hoeff}}(\hat{X}) = \frac{\ln(2/\delta)}{2\varepsilon^2} R[\hat{X}]^2$$

difference b/w max. & min value taken by $\hat{X}$.

An increase in $R[\hat{O}_{\mathrm{QEM}}]$ compared to $R[\hat{O}_g]$ thus requires more samples to achieve the same ε-precision for $E[\hat{O}_{\mathrm{QEM}}]$ as for $E[\hat{O}_g]$.

Specifically,

$$C_{\mathrm{QEM}} = \frac{N^{\varepsilon}_{\mathrm{se}}[\hat{O}_{\mathrm{QEM}}]}{N^{\varepsilon}_{\mathrm{se}}[\hat{O}_g]} \sim \frac{N^{\varepsilon,\delta}_{\mathrm{Hoeff}}[\hat{O}_{\mathrm{QEM}}]}{N^{\varepsilon,\delta}_{\mathrm{Hoeff}}[\hat{O}_g]} = \frac{R[\hat{O}_{\mathrm{QEM}}]^2}{R[\hat{O}_g]^2}.$$

Now, we discuss two concrete QEM methods.
One noise-agnostic method (ZNE) and one noise-aware one (PEC/PER). They can also be combined: PER + vZNE.

Zero-noise extrapolation (ZNE):

- Temme et al., (2017)
- Li, Benjamin (2017)

Execute same unitary U using different circuits that experience different noise levels λ. Can be achieved by (i) stretching gate pulse in real-time to increase gate errors, or (ii) by unitary folding $U \to U U^\dagger U$ or inserting identities $G_i^\dagger G_i = I$ into the circuit.

This yields noisy output states ρ_λ and noisy expectation values $\mathrm{Tr}[O \rho_\lambda]$.

Set of data points $\{\lambda_m, \mathrm{Tr}[O\rho_{\lambda_m}]\}$ with $\lambda_1 < \lambda_2 < \ldots < \lambda_M$, where λ_1 is minimal machine noise level that we can achieve (w/o folding).

Model data using (fit) function

$f(\lambda; \vec{\theta})$ and obtain optimal parameters $\vec{\theta}^*$ by fitting.

The error mitigated estimator of $\mathrm{Tr}[O\rho_0]$ is then the zero-noise extrapolated value

$$E[\hat{O}_{QEM}] = f(\lambda=0, \vec{\theta}^*)$$

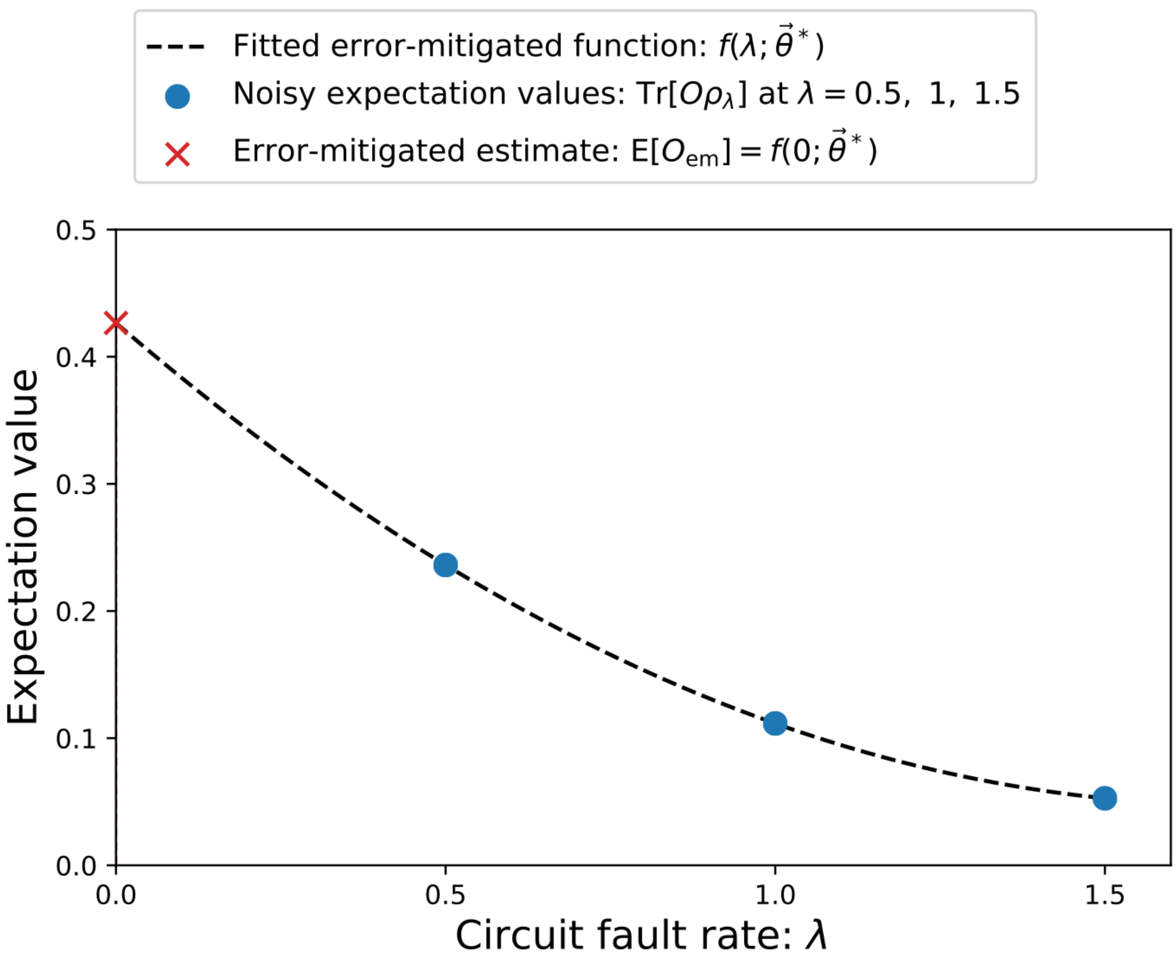

FIG. 2 Obtaining the error-mitigated estimate using zero-noise extrapolation. Here we are performing extrapolation using three noisy expectation values at the circuit fault rate of $\lambda = 0.5,\ 1,\ 1.5$, where 0.5 is the lowest circuit fault rate we can achieve and the other two are obtained through boosting the noise in the device.

From Cai et al., arxiv: 2210:00921.

Different fit functions can be used:

① Taylor expansion of $f(\lambda, \vec{\theta})$ for small λ (assuming weak noise):

$$f(\lambda, \vec{\theta}) = \sum_{l=0}^{M-1} \theta_l \frac{\lambda^l}{l!}$$

- Linear extrapolation is $M=2$: $f(\lambda, \vec{\theta}) = \theta_0 + \theta_1 \lambda$.
- Richardson extrapolation uses M data points (minimally) → perfect interpolation of the data points such that for $N_{cir} \to \infty$ the error becomes of $O(\lambda^M)$:

$$E[\hat{O}_{QEM}] = \theta_0^* = \sum_{m=1}^{M} \mathrm{Tr}[O \rho_{\lambda_m}] \prod_{k \neq m} \frac{\lambda_k}{\lambda_k - \lambda_m}$$

(Lagrange polynomial used for perfect interpolation).

The variance, however, increases exponentially with M.

Using that the range of the estimator is increased and $C_{QEM} \sim R[\hat{O}_{QEM}]^2 / R[\hat{O}_\rho]$ by Hoeffding's inequality we find:

$$C_{QEM} \sim \left(\sum_{m=1}^{M} \left| \prod_{n \neq m} \frac{\lambda_n}{\lambda_n - \lambda_m} \right| \right)^2$$

Assuming equal gap $\lambda_m = m \lambda_1$ for simplicity, we find

$$C_{QEM} \sim \left(\sum_{m=1}^{M} \underbrace{\left| \prod_{n \neq m} \frac{n}{n-m} \right|} \right)^2$$

$$= \frac{M!}{m!\,(M-m)!} = \frac{M \cdot (M-1) \cdots 1}{1 \cdot 2 \cdots m \cdot M \cdot (M-1) \cdots (M-m)}$$

$$= \frac{M \cdot (M-1) \cdots (M-m+1)}{1 \cdot 2 \cdots m}$$

Example:

$M = 5$

$m = 3$

$$\frac{1 \cdot 2 \cdot 4 \cdot 5}{(-2)(-1)\,1 \cdot 2} = \frac{M!}{m!\,(m-n)!}$$

$M = 10$

$m = 3$

$$\frac{1 \cdot 2 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9 \cdot 10}{(-2)(-1)\,1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7} = \frac{10 \cdot 9 \cdot 8}{(-2)(-1)} = \frac{M \cdots (M-m+1)}{1 \cdot 2 \cdots m} =$$

$$= \frac{M!}{m!\,(M-m)!}$$

Thus,

$$C_{QEM} \sim \left[\sum_{m=1}^{M} \binom{M}{m} \right]^2 =$$

$$\binom{M}{0} = \frac{M!}{0!\, M!} = 1$$

$$= \Bigg[\underbrace{\sum_{m=0}^{M} \binom{M}{m}}_{= 2^M \text{ by } (a+b)^M = \sum_{m=0}^{M} \binom{M}{m} a^m b^{M-m}} - 1 \Bigg]^2 = \left[2^M - 1 \right]^2 .$$

(2) Exponential fit function

$$f(\lambda, \vec{\theta}) = \theta_0 + \theta_1 e^{-\theta_2 \lambda}$$

or more general ansätze containing exponentials

- $f(\lambda, \vec{\theta}) = \theta_{-1} \pm e^{-g(\lambda, \vec{\theta})}$, $g(\lambda, \vec{\theta}) = \sum_{\ell=0}^{d} \theta_\ell \lambda^\ell$

  (poly-exponential), Giurgica-Tiron et al., (2020).

- $f(\lambda, \vec{\theta}) = \sum_{\ell=0}^{2d} \theta_{2\ell}\, e^{-\theta_{2\ell+1} \lambda}$ (multi-exponential)

  Cai, (2021).

These are consistent with fact that $\langle O \rangle \xrightarrow[\lambda \to \infty]{} 0$ under Pauli noise $\Rightarrow$ more appropriate in large noise regime (when method is not really well justified though).

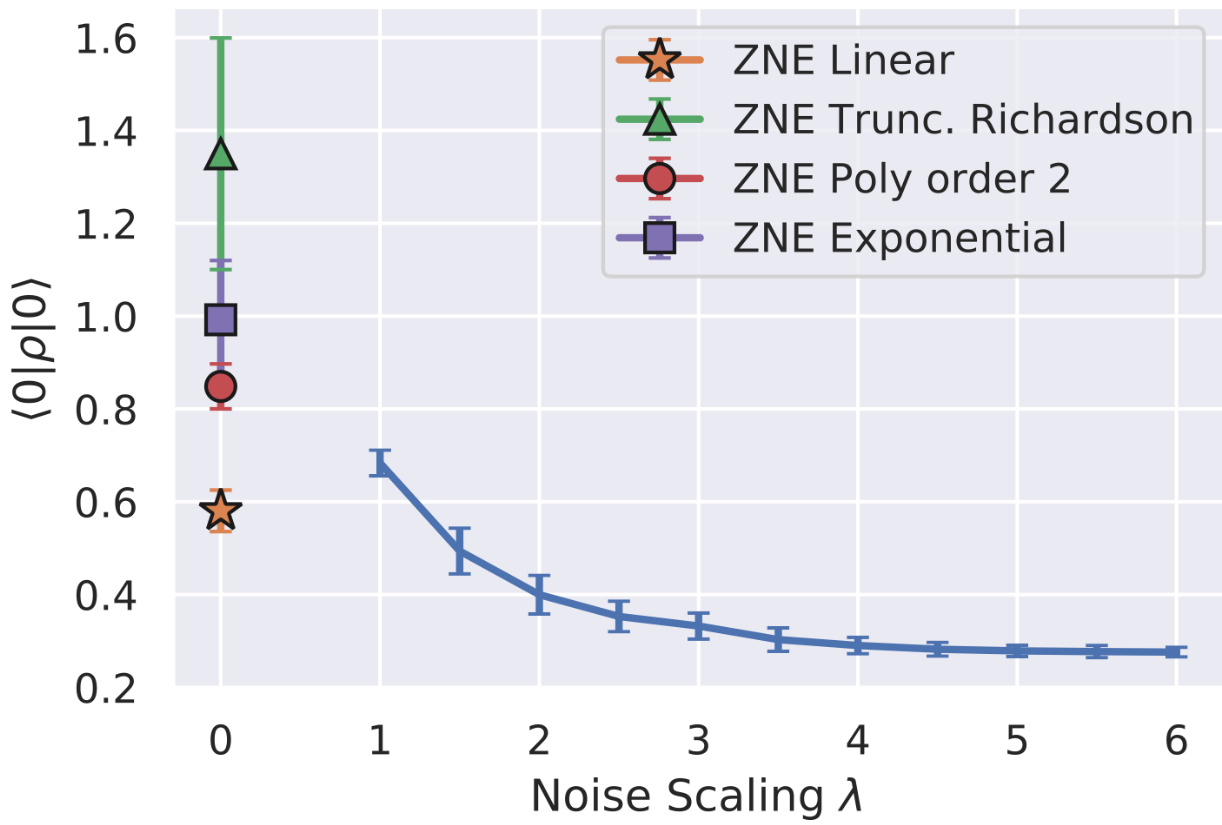

Fig. 6: Comparison of extrapolation methods averaged over 50 two-qubit randomized benchmarking circuits executed on IBMQ's "London" five-qubit chip. The circuits had, on average, 97 single qubit gates and 17 two-qubit gates. The true zero-noise value is $\langle 0|\rho|0\rangle = 1$ and different markers show extrapolated values from different fitting techniques.

From Giurgica-Tiron et al., (2020).

Exponential & 2-degree polynomial outperform Richardson extrapolation.

(3) adaptive schemes, where values of λ are chosen adaptively

One can also use learning based fitting models (such as neural networks).

Probabilistic error cancellation (PEC)

Temme et al., (2017)

- Systematically removes bias.
- Incurs exponential sampling overhead $C_{QEM} \sim e^{\lambda}$
- Requires full knowledge of noise model
  - either perform GST (gate set tomography) & neglect longer-range noise correlations
  - or transform noise into Pauli noise by Pauli twirling

Write noisy gate as composition of ideal gate followed by error channel: $\mathcal{U}_{noisy} = \mathcal{E} \circ \mathcal{U}_{ideal}$

$$\Rightarrow \mathcal{E}(\rho) = \sum_a M_a \rho M_a^{\dagger} \xrightarrow[\text{Pauli twirl}]{} \sum_m a_m P_m \rho P_m^{\dagger}$$

$$\frac{1}{|P|} \sum_{m,a} P_m^{\dagger} M_a \rho M_a^{\dagger} P_m$$

becomes Pauli channel

Proof for Clifford gates U_{ideal}:

$$\mathcal{E}(\rho) = \sum_a M_a \rho M_a^\dagger, \quad \sum_a M_a M_a^\dagger = I$$

Write superoperator also as $\widehat{\mathcal{E}}$.

Clifford gate C with superoperator $\widehat{C}$. Assume C is 2-qubit gate for concreteness (such as CNOT gate).

Pauli twirl dressing of gate:

$$C \longrightarrow \frac{1}{16} \sum_{\alpha,\beta=0}^{4} \sigma_1^\alpha \sigma_2^\beta C \sigma_1^\gamma \sigma_2^\delta$$

Choose $\sigma_1^\gamma \sigma_2^\delta$ such that

$$\sigma_1^\gamma \sigma_2^\delta = C^\dagger \sigma_1^\alpha \sigma_2^\beta C$$

and in practice it is sufficient to select α, β at random and average over sufficiently many random instances.

As a result

$$\widehat{C}_{noisy}\, \rho = \widehat{\mathcal{E}} \circ \widehat{C}_{ideal}\, \rho = \sum_a M_a C \rho C^\dagger M_a^\dagger$$

$$\xrightarrow[\text{of noisy Clifford gate}]{\text{Pauli twirl}} \frac{1}{16}\sum_a \sum_{\alpha,\beta} \sigma_1^\alpha \sigma_2^\beta M_a C \sigma_1^\gamma \sigma_2^\delta \rho \, \sigma_1^\gamma \sigma_2^\delta C^\dagger M_a^\dagger \sigma_1^\alpha \sigma_2^\beta =$$

$$= \frac{1}{16}\sum_a \sum_{\alpha,\beta} \sigma_1^\alpha \sigma_2^\beta M_a \sigma_1^\alpha \sigma_2^\beta C \rho C^\dagger \sigma_1^\alpha \sigma_2^\beta M_a^\dagger \sigma_1^\alpha \sigma_2^\beta$$

$$\sigma_1^\gamma \sigma_2^\delta = C^\dagger \sigma_1^\alpha \sigma_2^\beta C$$

$$= \sum_a \overline{M}_a C \rho C^\dagger \overline{M}_a^\dagger = \overline{\mathcal{E}} \circ \overline{C} \rho ,$$

where $\overline{\mathcal{E}} = p I + \sum_{(\alpha,\beta)\neq(0,0)} p_{\alpha\beta} \overline{\sigma_1^\alpha \sigma_2^\beta}$

is the Pauli twirled channel, which is a Pauli channel by construction.

<u>Key idea</u>: write noiseless gate superoperators as linear combination of noisy implementable ones

$$\overline{\mathcal{U}}_{ideal} = \sum_n \alpha_n \overline{B}_n$$

with real, but potentially negative coefficients α_n.

Assume that you can implement a set of noisy operations $\overline{B}_n$ that forms a (normalized) basis on the Hilbert-Schmidt space.

Now, the ideal expectation value of observable O can be recovered as

$$\mathrm{Tr}[O\rho_0] = \mathrm{Tr}[O\,\overline{\mathcal{U}}_{ideal}\,\rho_{init}] =$$

$$= \langle\langle O|\,\overline{\mathcal{U}}\,|\rho_{init}\rangle\rangle = \sum_n \alpha_n \underbrace{\langle\langle O|\,\overline{B}_n\,|\rho_{init}\rangle\rangle}_{\text{can be computed on noisy hardware}}.$$

need full information of noisy gates to obtain α_n beforehand

It holds that

$$\sum_m \alpha_m = 1$$ due to fact that $\overline{U}_{ideal}$ is

unitary and $\{\overline{B}_m\}$ is normalized basis (may be over-complete), so we can estimate $\sum_m \alpha_m \langle\langle 0 | B_m | \rho_{init} \rangle\rangle$ using

Monte-Carlo sampling.

In the MC sampling a basis element $\overline{B}_m$ is chosen in the noisy circuit with probability $\frac{|\alpha_m|}{Q}$, where $Q = \sum_m |\alpha_m| \geq 1$.

The circuit output $\langle\langle 0 | B_m | \rho_{init} \rangle\rangle$ is multiplied with $Q\,\mathrm{sign}(\alpha_m)$ before averaging to obtain the expected value of the estimator

$$\boxed{E[\hat{O}_{QEM}] = Q \sum_m \mathrm{sign}(\alpha_m) \frac{|\alpha_m|}{Q} \langle\langle 0 | B_m | \rho_{init} \rangle\rangle}$$

However, $\sum_m |\alpha_m| \geq 1$ as some $\alpha_m < 0$, which means that

- $\sum_m \alpha_m \overline{B}_m$ may not be a convex combination of channels.

TABLE I. Sixteen basis operations. Gates $[R_x]$ and $[R_y]$ can be derived from $[H]$ and $[S]$, and other operations can be derived from $[\pi]$, $[R_x]$, and $[R_y]$.

1	$[\mathbb{1}]$ (no operation)
2	$[\sigma^x] = [R_x]^2$
3	$[\sigma^y] = [R_x]^2[R_z]^2$
4	$[\sigma^z] = [R_z]^2$
5	$[R_x] = [(1/\sqrt{2})(\mathbb{1} + i\sigma^x)] = [H][S]^3[H]$
6	$[R_y] = [(1/\sqrt{2})(\mathbb{1} + i\sigma^y)] = [R_z]^3[R_x][R_z]$
7	$[R_z] = [(1/\sqrt{2})(\mathbb{1} + i\sigma^z)] = [S]^3$
8	$[R_{yz}] = [(1/\sqrt{2})(\sigma^y + \sigma^z)] = [R_x][R_z]^2$
9	$[R_{zx}] = [(1/\sqrt{2})(\sigma^z + \sigma^x)] = [R_z][R_x][R_z]$
10	$[R_{xy}] = [(1/\sqrt{2})(\sigma^x + \sigma^y)] = [R_x]^2[R_z]$
11	$[\pi_x] = [\frac{1}{2}(\mathbb{1} + \sigma^x)] = [R_z]^3[R_x]^3[\pi][R_x][R_z]$
12	$[\pi_y] = [\frac{1}{2}(\mathbb{1} + \sigma^y)] = [R_x][\pi][R_x]^3$
13	$[\pi_z] = [\frac{1}{2}(\mathbb{1} + \sigma^z)] = [\pi]$
14	$[\pi_{yz}] = [\frac{1}{2}(\sigma^y + i\sigma^z)] = [R_z]^3[R_x]^3[\pi][R_x]^3[R_z]$
15	$[\pi_{zx}] = [\frac{1}{2}(\sigma^z + i\sigma^x)] = [R_x][\pi][R_x]^3[R_z]^2$
16	$[\pi_{xy}] = [\frac{1}{2}(\sigma^x + i\sigma^y)] = [\pi][R_x]^2$

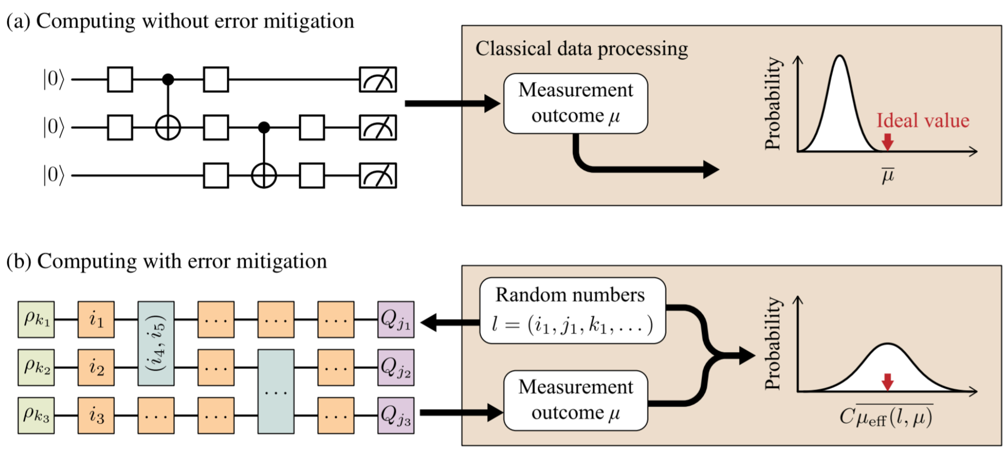

FIG. 1. Quantum computing of the expected value of an observable (a) without quantum error mitigation (QEM) and (b) with QEM. In QEM circuit (b), each operation (including the memory operation) in the original circuit (a) is replaced by an operation depending on the corresponding random numbers [see Fig. 2(a)].

From Endo et al., PRX (2018).

- α_m must be interpreted as a quasi-probability distribution (QPD) with negative probability weight. The negativity $\sum_{\substack{m \\ \alpha_m<0}} |\alpha_m| = Q^-$ quantifies a sampling overhead.

Specifically, since the circuit output is being multiplied with Q, the range increases by a factor of $Q \geq 1$. The variance thus increases by Q^2 and the sampling overhead $C_{QEM} \sim Q^2 = \left(\sum_m |\alpha_m|\right)^2$.

Let's show that more explicitly.

Monte Carlo sampling

In PEC the error mitigated expectation value $E[\hat{O}_{QEM}]$ is a sum of the expectation value of K random variables $\hat{O}_m$, the noisy circuit outputs: $E[\hat{O}_m] = \langle\langle O | B_m | \rho_{init} \rangle\rangle$:

$$E[\hat{O}_{QEM}] = \sum_{m=1}^{K} \alpha_m E[\hat{O}_m] \quad , \alpha_m \in \mathbb{R}$$

error mitigated estimator

$$\hat{O}_{QEM} = \sum_m \alpha_m \hat{O}_m \quad \Rightarrow \text{Range of } \hat{O}_m : R[\hat{O}_m] = 1$$

(Pauli observables typically).

$$\Rightarrow R[\hat{O}_{QEM}] = \sum_m |\alpha_m| = Q$$

$$\Rightarrow \boxed{C_{QEM} = \frac{R[\hat{O}_{QEM}]^2}{R[\hat{O}_\rho]^2} = Q^2} = (1 + 2Q^-)^2 .$$

$$Q^- = \sum_{\substack{m \\ \alpha_m < 0}} |\alpha_m|$$

sampling overhead.

Since $\sum_m \alpha_m = Q^+ - Q^- = 1 \Rightarrow Q^+ = 1 + Q^-$

$$\Rightarrow \sum_m |\alpha_m| = Q^+ + Q^- = Q \Rightarrow \boxed{Q = 1 + 2Q^-} ,$$

C_{QEM} is determined by negativity of QPD.

Explicit example of PEC for depolarizing noise:

$$\mathcal{E}(\rho) = (1-p)\rho + p\mathcal{N} \qquad \left(\mathcal{N} = \frac{I}{2}\right)$$

$\Rightarrow$ noisy unitary gate becomes

$$\mathcal{U}_p = \mathcal{E} \circ \mathcal{U} = (1-p)\,\mathcal{U} + p\mathcal{N}$$

We can thus write the ideal noiseless gate as

$$\mathcal{U} = \frac{1}{1-p}\,\mathcal{U}_p - \frac{p}{1-p}\,\mathcal{N}$$

Let's assume we can implement $\mathcal{N}$ on the device for simplicity. Otherwise, need to expand $\mathcal{N} = \sum_{w_n} B_n$ in noisy basis.

$$\Rightarrow \alpha_1 = \frac{1}{1-p} \;;\; \alpha_2 = -\frac{p}{1-p}$$

$$\Rightarrow Q = \sum_n |\alpha_n| = \frac{1+p}{1-p}$$

Sampling overhead (for a single gate $\mathcal{U}$ in the circuit):

$$C_{QEM} \sim Q^2 = \left(\frac{1+p}{1-p}\right)^2 \geq 1.$$

Note that the same holds for Pauli noise.

For a full circuit composed of M gates, this becomes

$$\langle\langle O | \rho_0 \rangle\rangle = \langle\langle O | \prod_{m=1}^{M} \mathcal{U}_m | \rho_{init} \rangle\rangle$$

ideal noiseless gates

The QEM estimator of this expectation value reads

$$\langle\langle O | \rho_0 \rangle\rangle = E[\hat{O}_{QEM}] = \langle\langle O | \prod_m \left(\sum_{n_m} \alpha^{(m)}_{n_m} \mathcal{B}_{n_m} \right) | \rho_{init} \rangle\rangle =$$

$$= \sum_{\vec{n}} \alpha_{\vec{n}} \langle\langle O | \mathcal{B}_{\vec{n}} | \rho_{init} \rangle\rangle$$

with $\vec{n} = (n_1, n_2, \ldots, n_M)$ and $\mathcal{B}_{\vec{n}} = \prod_m \mathcal{B}_{n_m}$,

$\alpha_{\vec{n}} = \prod_m \alpha_{n_m}$ and $Q = \sum_{\vec{n}} |\alpha_{\vec{n}}|$.

Note that $\{\mathcal{B}_n\}$ contains the basis for all gates in the circuit and is thus overcomplete.

Overall sampling overhead is product of sampling overhead of individual gates $\Rightarrow$ scales exponentially with number of gates.

Assuming uniform error model for simplicity:

$$C_{QEM} \sim \prod_{m=1}^{M} \left(\frac{1+p}{1-p} \right)^2 \approx \left(\frac{1+p}{1-p} \right)^{2M}$$

Expand in small p:

$$C_{QEM} \sim \left(\frac{1+p}{1-p} \right)^{2M} \approx (1+p)^{4M} \approx e^{4Mp} = e^{4\lambda}$$

with circuit fault rate $\lambda = Mp$.

We see why QEM is restricted to $\lambda \simeq O(1)$ as the sampling overhead increases exponentially.

Note that if we reduce the noise (from $p \to p' < p$) rather than cancel it completely, the sampling overhead is reduced to

$$C_{QEM}^{(PER)} \sim e^{4M(p-p')}$$

This is called probabilistic error reduction.

Illustration how PEC works in practice:

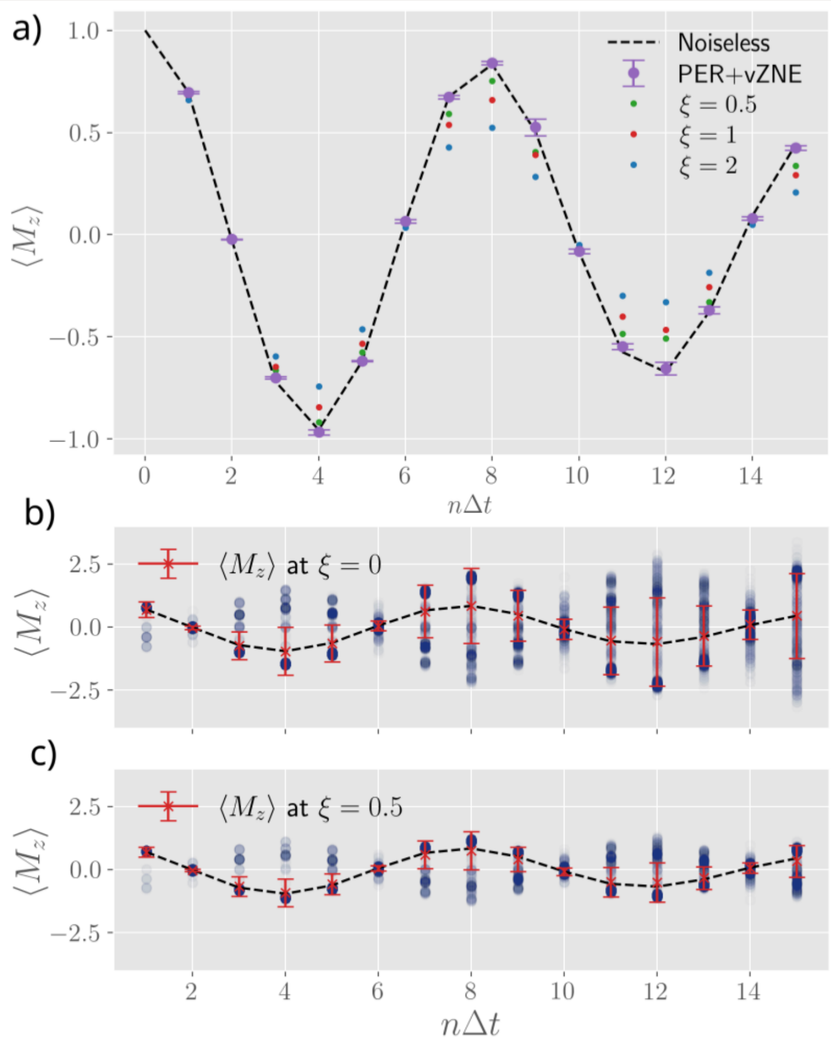

Fig. 8: PER results of total magnetization $\langle M_z(t) \rangle$ of the TFIM, prepared in the $|0\rangle$ initial state and evolved with under Hamiltonian with parameters $J = 0.15$, $h = 1$. The Trotter dynamics are simulated on the IBM noisy simulator `FakeVigoV2` using a Trotter stepsize $\Delta t = 0.2$. We simulate 1000 PER circuits, each is evaluated with 1024 shots. Panel (a) shows that vZNE with the noise levels $\xi \in [0.5, 1, 2]$ yields excellent agreement with the noiseless Trotter result. Readout error mitigation is used at all noise levels. Panel (b) shows the individual estimators (blue) and their average with standard deviation (red crosses) at $\xi = 0$ (upper plot) and $\xi = 0.5$ (lower plot). The variance increases as the noise strength approaches zero, and the estimator values approach the noiseless value. The overhead at $\xi = 0$ after 15 Trotter steps is $\gamma^{(0)} = 7.25$, while it is only $\gamma^{(0.5)} = 2.69$ at $\xi = 0.5$.

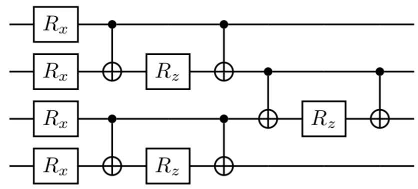

Fig. 7: The realization of a single Trotter step as a quantum circuit. Here we have defined $R_x \equiv RX(-2h\delta t)$ and $R_z \equiv RZ(2J\delta t)$

From:

McDonough et al., (2022).

Measurement error mitigation

Important since measurement errors can be large [e.g. on SC qubits].

Assume you want to perform a projective measurement in the Z-basis at the end of the circuit execution.

Described by projectors $\{ \langle\langle x | \}$ with $x \in \{0,1\}^{\otimes n}$ (bit strings).

Due to noise, this is transformed into a POVM $\{ \langle\langle E_x | = \langle\langle x | A \}$ with different output statistic.

One can mitigate these errors by first measuring the confusion or transition matrix

$$\boxed{A_{xy} = \langle\langle x | A | y \rangle\rangle = \langle\langle E_x | y \rangle\rangle}$$

assuming no state preparation errors

Then, we can insert A_{xy} to obtain

$$\langle\langle E_x| = \sum_y \langle\langle E_x|y\rangle\rangle\langle\langle y| = \sum_y A_{xy}\langle\langle y|$$

$$\Rightarrow \boxed{\langle\langle y| = \sum_x \left(A^{-1}\right)_{yx}\langle\langle E_x|}$$

We thus obtain the error mitigated expectation value from the noisy output $\vec{P}_{noisy} = \{\langle\langle E_x|\rho\rangle\rangle\}$ as

$$\langle\langle O|\rho\rangle\rangle = \sum_x O_x \langle\langle x|\rho\rangle\rangle$$

$$= \sum_x O_x \sum_y \left(A^{-1}\right)_{yz}\langle\langle E_z|\rho\rangle\rangle =$$

$$= \vec{O}^T A^{-1} \vec{P}_{noisy} .$$

Note that for larger systems, one cannot perform full readout error mitigation as A grows exponentially with system size.

One solution: assume uncoupled readout errors and write

$A = \bigotimes_m A_m$ (tensor product of smaller A matrices).

There are many more accurate approaches (see Cai et al., RMP; Sec III.C.)