SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Mitigating Quantum Errors via Truncated Neumann Series

Kun WANG[1*], Yu-Ao CHEN[1] & Xin WANG[1]

[1]*Institute for Quantum Computing, Baidu Research, Beijing 100193, China*

Appendix A Proof of Eq. (3)

Proof. By the definition of $\eta^{(1)}$, we have

$$\eta^{(1)} = \frac{1}{M}\sum_{m=1}^{M} O(\boldsymbol{s}^m) \tag{A1}$$

$$= \frac{1}{M}\sum_{m=1}^{M}\sum_{\boldsymbol{x}\in\{0,1\}^n} O(\boldsymbol{x})\langle\boldsymbol{x}|\boldsymbol{s}^m\rangle \tag{A2}$$

$$= \sum_{\boldsymbol{x}\in\{0,1\}^n} O(\boldsymbol{x})\langle\boldsymbol{x}|\left(\frac{1}{M}\sum_{m=1}^{M}|\boldsymbol{s}^m\rangle\right). \tag{A3}$$

The expectation value can be evaluated as

$$E^{(1)} := \mathbb{E}[\eta^{(1)}] = \mathbb{E}\left[\sum_{\boldsymbol{x}\in\{0,1\}^n} O(\boldsymbol{x})\langle\boldsymbol{x}|\left(\frac{1}{M}\sum_{m=1}^{M}|\boldsymbol{s}^m\rangle\right)\right] \tag{A4}$$

$$= \sum_{\boldsymbol{x}\in\{0,1\}^n} O(\boldsymbol{x})\langle\boldsymbol{x}|\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M}|\boldsymbol{s}^m\rangle\right] \tag{A5}$$

$$= \sum_{\boldsymbol{x}\in\{0,1\}^n} O(\boldsymbol{x})\langle\boldsymbol{x}|A\,\mathrm{vec}(\mathcal{N}(\rho)). \tag{A6}$$

Appendix B Proof of Proposition 1

Proof. We first recall the definition of Neumann series [1, Theorem 4.20]. Let $A \in \mathbf{M}_d$ be a real matrix. If $\lim_{k\to\infty}(I-A)^k = 0$, then

$$A^{-1} = \sum_{k=0}^{\infty}(I-A)^k. \tag{B1}$$

Furthermore, a sufficient condition for $\lim_{k\to\infty}(I-A)^k = 0$ is that $\|I-A\| < 1$ in some consistent norm.

Multiplying both sides of Eq. (B1) with A and rearranging the elements, we have

$$I = AA^{-1} = A\left(\sum_{k=0}^{K}(I-A)^k\right) + A\left(\sum_{k=K+1}^{\infty}(I-A)^k\right). \tag{B2}$$

The above yields

$$A\left(\sum_{k=K+1}^{\infty}(I-A)^k\right) = I - A\left(\sum_{k=0}^{K}(I-A)^k\right) \tag{B3}$$

$$= I - \frac{(I-A)^0 \times [I-(I-A)^{K+1}]}{I-(I-A)}A \tag{B4}$$

$$= (I-A)^{K+1}, \tag{B5}$$

where the second equality follows from the closed-form formula of geometric series. On the other hand, using the recurrence relation for binomial coefficients we can show that

$$\sum_{k=0}^{K}(I-A)^k = \sum_{k=0}^{K}\binom{K+1}{k+1}(-A)^k = \sum_{k=0}^{K}c_K(k)A^k, \tag{B6}$$

* Corresponding author (email: wangkun28@baidu.com)

where the coefficient function c_K is defined in Eq. (5). Putting all pieces together, we get

$$I = \sum_{k=1}^{K+1} c_K(k-1)A^k + (I-A)^{K+1}, \tag{B7}$$

where we change the variable so that the summation begins with $k=1$. Thus, using the linearity of function f we conclude that

$$f(I) = \sum_{k=1}^{K+1} c_K(k-1) f\left(A^k\right) + f\left((I-A)^{K+1}\right). \tag{B8}$$

Appendix C Pauli transfer matrix representation

The four Pauli operators in the qubit space are defined as

$$I \equiv \sigma_0 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X \equiv \sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{C1a}$$

$$Y \equiv \sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z \equiv \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{C1b}$$

Their normalizations provide an orthonormal basis for the qubit linear operators, i.e., arbitrary qubit linear operator can be decomposed with respect to this basis. For the n-qubit case, one can construct a set of (normalized) Pauli operators, which we call the (normalized) Pauli set, as

$$\boldsymbol{P}_n := \left\{ \bigotimes_{k=1}^{n} \frac{\sigma_{j_k}}{\sqrt{2}} : j_k = 0,1,2,3 \right\} \equiv \left\{ \frac{I}{\sqrt{2}}, \frac{X}{\sqrt{2}}, \frac{Y}{\sqrt{2}}, \frac{Z}{\sqrt{2}} \right\}^{\otimes n}. \tag{C2}$$

It is easy to verify that $|\boldsymbol{P}_n| = 4^n$, where $|\cdot|$ denotes the size of the set.

The Pauli transfer matrix (PTM) representation expresses states and evolutions in terms of the Pauli basis $\boldsymbol{P}_n$ (C2), since $\boldsymbol{P}_n$ forms a basis for the n-qubit operators. Specially, an n-qubit quantum state ρ can be written in the vector form by decomposing it into the Pauli basis:

$$|\rho\rangle\rangle := \begin{bmatrix} \vdots \\ \rho_{\boldsymbol{i}} \\ \vdots \end{bmatrix}, \tag{C3}$$

where the vector element is $\rho_{\boldsymbol{i}} := \mathrm{Tr}[\mathrm{P_i}\rho] \in \mathbb{R}$ and $P_{\boldsymbol{i}} \in \boldsymbol{P}_n$. Since $|\boldsymbol{P}_n| = 4^n$, $|\rho\rangle\rangle$ is a 4^n-dimensional real column vector. In this representation, a basis for the vector space is

$$\left\{ |P_{\boldsymbol{i}}\rangle\rangle : \boldsymbol{i} \in \{0,1,2,3\}^n \right\}. \tag{C4}$$

Similarly, a Hermitian observable O can also be expressed as a 4^n-dimensional real row vector via

$$\langle\langle O| := [\cdots \quad O_{\boldsymbol{i}} \quad \cdots], \tag{C5}$$

where the vector element is $O_{\boldsymbol{i}} := \mathrm{Tr}[\mathrm{OP_i}]$ and $P_{\boldsymbol{i}} \in \boldsymbol{P}_n$. That is to say, O can be written as a linear combination of Pauli operators $O = \sum_{\boldsymbol{i}} O_{\boldsymbol{i}} P_{\boldsymbol{i}}$. By definition, one has $\| \langle\langle O| \|_\infty = \sum_{\boldsymbol{i}} |O_{\boldsymbol{i}}|$, where $\|\cdot\|_\infty$ is the matrix ∞-norm. In quantum computation, a common choice of the observable is $O = Z^{\otimes n}$. In this case, one has $\left\| \langle\langle Z^{\otimes n}| \right\|_\infty = \sqrt{2^n}$.

A quantum channel $\mathcal{N}$ (with n-qubit input and output systems) can be expressed as a $4^n \times 4^n$ real square matrix $[\mathcal{N}]$ whose $\boldsymbol{i}$-th row and $\boldsymbol{j}$-th column element is defined via

$$[\mathcal{N}]_{\boldsymbol{i},\boldsymbol{j}} := \mathrm{Tr}\left[\mathrm{P_i}\mathcal{N}(\mathrm{P_j})\right]. \tag{C6}$$

Here, $\boldsymbol{i},\boldsymbol{j} \in \{0,1,2,3\}^n$. Note that the identity channel id has identity Pauli transfer matrix, i.e., $[\mathrm{id}] = \mathbb{1}$, as evident from the relation $\mathrm{Tr}[\mathrm{P_i P_j}] = \delta_{\mathrm{i,j}}$, where $\delta_{x,y}$ is the Kronecker delta function. By definition, if $\sigma = \mathcal{N}(\rho)$, then $|\sigma\rangle\rangle = [\mathcal{N}]|\rho\rangle\rangle$.

A desirable property of the Pauli transfer matrix representation (PTM) is that quantum channel composition is multiplicative in this representation. Specifically, let $\mathcal{M}$ and $\mathcal{N}$ be two n-qubit quantum channels, then

$$[\mathcal{M} \circ \mathcal{N}] = [\mathcal{M}]\,[\mathcal{N}]. \tag{C7}$$

The proof is given as follow.

Proof. Let $\boldsymbol{i},\boldsymbol{j} \in \{0,1,2,3\}^n$. By definition,

$$[\mathcal{M} \circ \mathcal{N}]_{\boldsymbol{i},\boldsymbol{j}} = \mathrm{Tr}\left[\mathrm{P_i}\mathcal{M} \circ \mathcal{N}(\mathrm{P_j})\right] \tag{C8}$$

$$= \mathrm{Tr}\left[\mathrm{P_i}\mathcal{M}\left(\sum_{\mathrm{m}} [\mathcal{N}]_{\mathrm{m,j}} \mathrm{P_m}\right)\right] \tag{C9}$$

$$= \mathrm{Tr}\left[\mathrm{P_i} \cdot \sum_{\mathrm{m,n}} [\mathcal{N}]_{\mathrm{m,j}} [\mathcal{M}]_{\mathrm{n,m}} \mathrm{P_n}\right] \tag{C10}$$

$$= \mathrm{Tr}\left[\mathrm{P_i}\sum_{\mathrm{n}}\left(\sum_{\mathrm{m}}[\mathcal{M}]_{\mathrm{n,m}}[\mathcal{N}]_{\mathrm{m,j}}\right)\mathrm{P_n}\right] \tag{C11}$$

$$= \sum_{\boldsymbol{n}}\left([\mathcal{M}][\mathcal{N}]\right)_{\boldsymbol{n},\boldsymbol{j}}\mathrm{Tr}\left[\mathrm{P_iP_n}\right] \tag{C12}$$

$$= \left([\mathcal{M}][\mathcal{N}]\right)_{\boldsymbol{i},\boldsymbol{j}}, \tag{C13}$$

where in the last step we use the fact that $\mathrm{Tr}[\mathrm{P_iP_j}] = \delta_{\mathrm{i,j}}$.

Lemma C1. Let ρ be a n-qubit quantum state. For arbitrary $\boldsymbol{i} \in \{0,1,2,3\}^n$, it holds that $|\langle\langle P_{\boldsymbol{i}}|\rho\rangle\rangle| \leqslant 1$.

Proof. By definition, $\langle\langle P_{\boldsymbol{i}}|\rho\rangle\rangle = \mathrm{Tr}\left[\mathrm{P_i}\rho\right]$. Notice that each Pauli operator $P_{\boldsymbol{i}}$ has eigenvalues ±1, this implies that $\mathrm{Tr}\left[\mathrm{P_i}\rho\right]$ is real. Assume the eigenstates of $P_{\boldsymbol{i}}$ is $\{|v_j\rangle\}_j$. We expand ρ in this basis: $\rho = \sum_{jk}\rho_{kj}|v_j\rangle\langle v_k|$. Then

$$\left|\mathrm{Tr}\left[\mathrm{P_i}\rho\right]\right| = \left|\sum_{\mathrm{jk}}\rho_{\mathrm{kj}}\langle\mathrm{v_k}|\,\mathrm{P_i}\,|\mathrm{v_j}\rangle\right| \leqslant \sum_{\mathrm{j}}\rho_{\mathrm{jj}} = \mathrm{Tr}[\rho] = 1, \tag{C14}$$

where the inequality follows from the fact that $P_{\boldsymbol{i}}$ has eigenvalues ±1.

Appendix D Proof of Theorem 1

Proof. Using Proposition 1, we have

$$\left|\mathrm{Tr}[\mathrm{O}\rho] - \sum_{\mathrm{k=1}}^{\mathrm{K+1}}\mathrm{c_K(k-1)E^{(k)}}\right| = \left|\langle\langle O|(I-[\mathcal{N}])^{K+1}|\rho\rangle\rangle\right|. \tag{D1}$$

We need to upper bound the tail $\left|\langle\langle O|(I-[\mathcal{N}])^{K+1}|\rho\rangle\rangle\right|$. Consider the following chain of inequalities:

$$\left|\langle\langle O|(I-[\mathcal{N}])^{K+1}|\rho\rangle\rangle\right| = \left|\langle\langle O\left(\sum_{\boldsymbol{i}}|P_{\boldsymbol{i}}\rangle\rangle\langle\langle P_{\boldsymbol{i}}|\right)(I-[\mathcal{N}])^{K+1}\left(\sum_{\boldsymbol{j}}|P_{\boldsymbol{j}}\rangle\rangle\langle\langle P_{\boldsymbol{j}}|\right)|\rho\rangle\rangle\right| \tag{D2}$$

$$= \left|\sum_{\boldsymbol{i},\boldsymbol{j}}\langle\langle O|P_{\boldsymbol{i}}\rangle\rangle\langle\langle P_{\boldsymbol{j}}|\rho\rangle\rangle\langle\langle P_{\boldsymbol{i}}|(I-[\mathcal{N}])^{K+1}|P_{\boldsymbol{j}}\rangle\rangle\right| \tag{D3}$$

$$\leqslant \sum_{\boldsymbol{i},\boldsymbol{j}}|\langle\langle O|P_{\boldsymbol{i}}\rangle\rangle|\cdot|\langle\langle P_{\boldsymbol{j}}|\rho\rangle\rangle|\cdot|\langle\langle P_{\boldsymbol{i}}|(I-[\mathcal{N}])^{K+1}|P_{\boldsymbol{j}}\rangle\rangle| \tag{D4}$$

$$\leqslant \sum_{\boldsymbol{i},\boldsymbol{j}}|\langle\langle O|P_{\boldsymbol{i}}\rangle\rangle|\cdot|\langle\langle P_{\boldsymbol{i}}|(I-[\mathcal{N}])^{K+1}|P_{\boldsymbol{j}}\rangle\rangle| \tag{D5}$$

$$= \sum_{\boldsymbol{i}}|\langle\langle O|P_{\boldsymbol{i}}\rangle\rangle|\left(\sum_{\boldsymbol{j}}|\langle\langle P_{\boldsymbol{i}}|(I-[\mathcal{N}])^{K+1}|P_{\boldsymbol{j}}\rangle\rangle|\right) \tag{D6}$$

$$\leqslant \sum_{\boldsymbol{i}}|\langle\langle O|P_{\boldsymbol{i}}\rangle\rangle|\left\|(I-[\mathcal{N}])^{K+1}\right\|_{\infty} \tag{D7}$$

$$= \|\langle\langle O|\|_{\infty}\left\|(I-[\mathcal{N}])^{K+1}\right\|_{\infty} \tag{D8}$$

$$\leqslant \|\langle\langle O|\|_{\infty}\|I-[\mathcal{N}]\|_{\infty}^{K+1}, \tag{D9}$$

where Eq. (D5) follows from Lemma C1, Eq. (D7) follows from the definition of the matrix ∞-norm, and Eq. (D9) follows from the submultiplicativity of the matrix ∞-norm.

Appendix E Proof of Eq. (10)

Proof. Let $\boldsymbol{s}^{m,k} \in \{0,1\}^n$ be the outcome in the m-th round when estimating $E_g^{(k)}$. By definition,

$$\overline{\eta} = \sum_{k=1}^{K+1}c_K(k-1)\eta^{(k)} \tag{E1}$$

$$= \frac{1}{M}\sum_{k=1}^{K+1}\sum_{m=1}^{M}c_K(k-1)O(\boldsymbol{s}^{m,k}) \tag{E2}$$

$$= \frac{1}{M(K+1)}\sum_{k=1}^{K+1}\sum_{m=1}^{M}(K+1)c_K(k-1)O(\boldsymbol{s}^{m,k}). \tag{E3}$$

Introducing the new random variables $X_{m,k} := (K+1)c_K(k-1)O(\boldsymbol{s}^{m,k})$, we have

$$\overline{\eta} = \frac{1}{M(K+1)}\sum_{k=1}^{K+1}\sum_{m=1}^{M}X_{m,k}. \tag{E4}$$

Intuitively, Eq. (E4) says that η can be viewed as the empirical mean value of the set of random variables

$$\{X_{m,k} : m = 1,\cdots,M; k = 1,\cdots,K+1\}. \tag{E5}$$

First, we show that the absolute value of each $X_{m,k}$ is upper bounded as

$$|X_{m,k}| = |(K+1)c_K(k-1)O(\boldsymbol{s}^{m,k})| \leqslant (K+1)|c_K(k-1)||O(\boldsymbol{s}^{m,k})| \leqslant (K+1)|c_K(k-1)|, \tag{E6}$$

where the second inequality follows from the assumption of O (recall that O is diagonal in the computational basis and $\|O\|_2 \leqslant 1$). Then, we show that $\overline{\eta}$ is an unbiased estimator of the quantity $\sum_{k=1}^{K+1} c_K(k-1)E^{(k)}$:

$$\begin{aligned}
\mathbb{E}[\overline{\eta}] &= \mathbb{E}\left[\frac{1}{M(K+1)}\sum_{k=1}^{K+1}\sum_{m=1}^{M} X_{m,k}\right] && \text{(E7a)}\\
&= \mathbb{E}\left[\frac{1}{M}\sum_{k=1}^{K+1}\sum_{m=1}^{M} c_K(k-1)O(\boldsymbol{s}^{m,k})\right] && \text{(E7b)}\\
&= \sum_{k=1}^{K+1} c_K(k-1)\left(\sum_{\boldsymbol{x}} O(\boldsymbol{x})\langle\boldsymbol{x}|\mathbb{E}_M\left[\frac{1}{M}\sum_{m=1}^{M}|\boldsymbol{s}^{m,k}\rangle\right]\right) && \text{(E7c)}\\
&= \sum_{k=1}^{K+1} c_K(k-1)\left(\sum_{\boldsymbol{x}} O(\boldsymbol{x})\langle\boldsymbol{x}|A^k \operatorname{vec}(\rho)\right) && \text{(E7d)}\\
&= \sum_{k=1}^{K+1} c_K(k-1)E^{(k)}, && \text{(E7e)}
\end{aligned}$$

where the last equality follows from (13). Eqs. (E6) and (E7) together guarantee that the prerequisites of the Hoeffding's inequality hold. By the Hoeffding's equality, we have

$$\begin{aligned}
\Pr\left\{\left|\overline{\eta} - \sum_{k=1}^{K+1} c_K(k-1)E^{(k)}\right| \geqslant \varepsilon\right\} &\leqslant 2\exp\left(-\frac{2M^2(K+1)^2\varepsilon^2}{4\sum_{k=1}^{K+1}\sum_{m=1}^{M}((K+1)c_K(k))^2}\right) && \text{(E8)}\\
&= 2\exp\left(-\frac{2M^2(K+1)^2\varepsilon^2}{4M(K+1)^3\left(\sum_{k=0}^{K}[c_K(k)]^2\right)}\right) && \text{(E9)}\\
&= 2\exp\left(-\frac{M\varepsilon^2}{2(K+1)\Delta}\right), && \text{(E10)}
\end{aligned}$$

where $\Delta := \sum_{k=1}^{K+1}[c_K(k)]^2 = \binom{2K+2}{K+1} - 1$. Solving

$$2\exp\left(-\frac{M\varepsilon^2}{2(K+1)\Delta}\right) \leqslant \delta \tag{E11}$$

gives

$$M \geqslant 2(K+1)\Delta\log(2/\delta)/\varepsilon^2. \tag{E12}$$

To summarize, choosing $K = \lceil\log\varepsilon/\log\xi - 1\rceil$ and $M = \lceil 2(K+1)\Delta\log(2/\delta)/\varepsilon^2\rceil$, we are able obtain the following two statements

$$\Pr\left\{\left|\overline{\eta} - \sum_{k=1}^{K+1} c_K(k-1)E^{(k)}\right| \geqslant \varepsilon\right\} \leqslant \delta, \tag{E13}$$

$$\left|\mathrm{Tr}[\mathrm{O}\rho] - \sum_{\mathrm{k}=1}^{\mathrm{K}+1} \mathrm{c}_{\mathrm{K}}(\mathrm{k}-1)\mathrm{E}^{(\mathrm{k})}\right| \leqslant \varepsilon, \tag{E14}$$

where the first one is shown above and the second one is proved in Theorem 1. Using the union bound and the triangle inequality, we conclude that $\overline{\eta}$ can estimate the ideal expectation value $\mathrm{Tr}[\mathrm{O}\rho]$ with error 2ε at a probability greater than $1-\delta$.

Appendix F Proof of Theorem 2

Proof. Using Proposition 1, we have

$$\left|\mathrm{Tr}[\mathrm{O}\rho] - \sum_{\mathrm{k}=1}^{\mathrm{K}+1} \mathrm{c}_{\mathrm{K}}(\mathrm{k}-1)\mathrm{E}^{(\mathrm{k})}\right| = \left|\sum_{\boldsymbol{x}\in\{0,1\}^n} O(\boldsymbol{x})\langle\boldsymbol{x}|(I-A)^{K+1}\operatorname{vec}(\rho)\right|. \tag{F1}$$

Now we show that the quantity in (F1) can be bounded from above. Define the matrix 1-norm of a $m \times n$ matrix B as

$$\|B\|_1 := \max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|B_{ij}| \equiv \max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|\langle i|B|j\rangle|, \tag{F2}$$

which is simply the maximum absolute column sum of the matrix. Let $\rho(\boldsymbol{y})$ is the $\boldsymbol{y}$-th diagonal element of the quantum state ρ. Consider the following chain of inequalities:

$$\left|\sum_{\boldsymbol{x}\in\{0,1\}^n} O(\boldsymbol{x})\langle\boldsymbol{x}|(I-A)^{K+1}\operatorname{vec}(\rho)\right| = \left|\sum_{\boldsymbol{x}\in\{0,1\}^n}\sum_{\boldsymbol{y}\in\{0,1\}^n} O(\boldsymbol{x})\rho(\boldsymbol{y})\langle\boldsymbol{x}|(I-A)^{K+1}|\boldsymbol{y}\rangle\right| \tag{F3a}$$

$$\leqslant \sum_{\boldsymbol{x}\in\{0,1\}^n} \sum_{\boldsymbol{y}\in\{0,1\}^n} |O(\boldsymbol{x})| \cdot \rho(\boldsymbol{y}) \cdot \left|\langle\boldsymbol{x}|(I-A)^{K+1}|\boldsymbol{y}\rangle\right| \tag{F3b}$$

$$\leqslant \sum_{\boldsymbol{x}\in\{0,1\}^n} \sum_{\boldsymbol{y}\in\{0,1\}^n} \rho(\boldsymbol{y}) \left|\langle\boldsymbol{x}|(I-A)^{K+1}|\boldsymbol{y}\rangle\right| \tag{F3c}$$

$$= \sum_{\boldsymbol{y}\in\{0,1\}^n} \rho(\boldsymbol{y}) \sum_{\boldsymbol{x}\in\{0,1\}^n} \left|\langle\boldsymbol{x}|(I-A)^{K+1}|\boldsymbol{y}\rangle\right| \tag{F3d}$$

$$\leqslant \sum_{\boldsymbol{y}\in\{0,1\}^n} \rho(\boldsymbol{y}) \|(I-A)^{K+1}\|_1 \tag{F3e}$$

$$= \|(I-A)^{K+1}\|_1 \tag{F3f}$$

$$\leqslant \|I-A\|_1^{K+1} \tag{F3g}$$

$$= \xi_m^{K+1}, \tag{F3h}$$

where (F3c) follows from the assumption that O is diagonalized in the computational basis and $\|O\|_2 \leqslant 1$, (F3e) follows from the definition of matrix 1-norm, (F3f) follows from the fact that ρ is a quantum state and thus $\sum_{\boldsymbol{y}} \rho(\boldsymbol{y}) = 1$, (F3g) follows from the submultiplicativity property of the matrix 1-norm, and (F3h) follows from Lemma F1 stated below. We are done.

Lemma F1. Let A be a column stochastic matrix of size $d \times d$. It holds that

$$\xi_m(A) = \|I-A\|_1, \tag{F4}$$

where $\xi_m(A)$ is defined in Eq. (11).

Proof. Since A is column stochastic, $I-A$ has non-negative diagonal elements and negative off-negative elements. Thus

$$\|I-A\|_1 = \max_{1\leqslant j\leqslant d} \left(1 - A_{jj} + \sum_{i\neq j} A_{ij}\right) \tag{F5}$$

$$= \max_{1\leqslant j\leqslant d} (1 - A_{jj} + 1 - A_{jj}) \tag{F6}$$

$$= 2\max_{1\leqslant j\leqslant d} (1 - A_{jj}) \tag{F7}$$

$$= 2 - 2\min_{1\leqslant j\leqslant d} A_{jj} \tag{F8}$$

$$=: \xi_m(A), \tag{F9}$$

where the second equality follows from the fact that A is column stochastic.

Appendix G Sequential measurements

In this Appendix, we prove that the classical noise model describing the sequential measurement repeating k times is effectively characterized by the stochastic matrix A^k. We begin with the simple case $k = 2$. Since the noise model is classical and linear in the input, it suffices to consider the computational basis states as inputs. As shown in Figure G1, we apply the noisy quantum measurement device two times sequentially on the input state $|\boldsymbol{x}\rangle\langle\boldsymbol{x}|$ in computational basis where $\boldsymbol{x} \in \{0,1\}^n$. Assume the measurement outcome of the first measurement is $\boldsymbol{y}$ and the measurement outcome of the second measurement is $\boldsymbol{z}$, where $\boldsymbol{y},\boldsymbol{z} \in \{0,1\}^n$. Assume that the error matrix associated with this sequential measurement is A'. That is, the probability of obtaining the outcome $\boldsymbol{z}$ provided the true outcome is $\boldsymbol{x}$ is given by $A'_{\boldsymbol{z}\boldsymbol{x}}$. Practically, we input $|\boldsymbol{x}\rangle\langle\boldsymbol{x}|$ to the first noisy measurement device and obtain the outcome $\boldsymbol{y}$. The probability of this event is $A_{\boldsymbol{y}\boldsymbol{x}}$, by the definition of the error matrix. Similarly, we input $|\boldsymbol{y}\rangle\langle\boldsymbol{y}|$ to the second noisy measurement device and obtain the outcome $\boldsymbol{z}$. The probability of this event is $A_{\boldsymbol{z}\boldsymbol{y}}$. Inspecting the chain $\boldsymbol{x} \to \boldsymbol{y} \to \boldsymbol{z}$, we have

$$A'_{\boldsymbol{z}\boldsymbol{x}} = \sum_{\boldsymbol{y}\in\{0,1\}^n} A_{\boldsymbol{y}\boldsymbol{x}} A_{\boldsymbol{z}\boldsymbol{y}} = A^2_{\boldsymbol{z}\boldsymbol{x}}. \tag{G1}$$

The above analysis justifies that the classical noise model describing the sequential measurement repeating 2 times is effectively characterized by the stochastic matrix A^2. The general case can be analyzed similarly.

Mathematically, quantum measurements can be modeled as quantum-classical quantum channels [3, Chapter 4.6.6] where they take a quantum system to a classical one. Experimentally, the implementation of quantum measurement is platform-dependent and has different characterizations. For example, the fabrication and control of quantum coherent superconducting circuits have enabled experiments that implement quantum measurement [4]. Based on the outcome data, experimental measurements are typically categorized into two types: those only output classical outcomes and those output both classical outcomes and quantum states. That is, besides the usually classical outcome sequences, the measurement device will also output a quantum state on the computational basis corresponding to the classical outcome. For the former type, we can implement the sequential measurement via the *qubit reset* [5–7] approach, by which we mean the ability to re-initialize the qubits into a known state, usually a state in the computational basis, during the course of the computation. Technically, when the i-th noisy measurement outputs an outcome sequence $\boldsymbol{s}^i \in \{0,1\}^n$, we use the qubit reset technique to prepare the computational basis state $|\boldsymbol{s}^i\rangle\langle\boldsymbol{s}^i|$ and feed it to the $(i+1)$-th noisy measurement (cf. Figure G1). In this case, the noisy measurement device can be reused. For the latter type, the sequential measurement can be implemented efficiently: when the i-th noisy measurement outputs a classical sequence and a quantum state on the computational basis, we feed the quantum state to the $(i+1)$-th noisy measurement.

Appendix H Demonstrative examples

In this section, we consider some common noise channels as demonstrative examples to investigate the experimental relevance of the proposed gate error mitigation method.

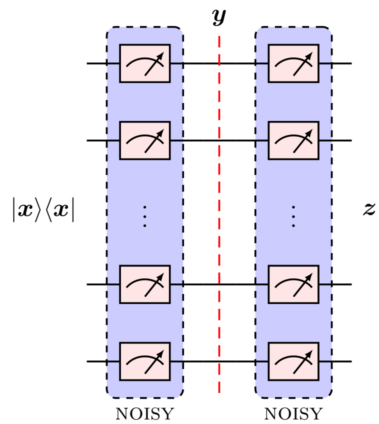

Figure G1 Apply the noisy quantum measurement device two times sequentially on the input state $|\boldsymbol{x}\rangle\langle\boldsymbol{x}|$ where $\boldsymbol{x} \in \{0,1\}^n$. The measurement outcome of the first measurement is $\boldsymbol{y}$ and the measurement outcome of the second measurement is $\boldsymbol{z}$.

Depolarizing noise. The qubit depolarization error replaces the input state with a completely mixed state with probability p, and does nothing with probability $1-p$. It describes a noise process where information is completely lost with probability p. For a single qubit, the depolarizing noise is defined as:

$$\Omega_p(\rho) := \left(1 - \frac{3p}{4}\right)\rho + \frac{p}{4}\left(X\rho X + Y\rho Y + Z\rho Z\right), \tag{H1}$$

where $p \in [0,1]$. The PTM representation of Ω_p is [2]

$$[\Omega_p] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-p & 0 & 0 \\ 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 1-p \end{pmatrix} \tag{H2}$$

and thus $\xi(\Omega_p) = p$. The same as dephasing noise, the noise resistance of the depolarizing noise implies that our method fails only when $p = 1$, for which the truncation error cannot be bounded any more.

Dephasing noise. The qubit dephasing error arises when the energy splitting of a qubit fluctuates as a function of time due to coupling to the environment. Charge noise affecting a transmon qubit is of this type. Dephasing is represented by a phase-flip channel, which describes the loss of phase information with probability p. This channel projects the state onto the Z-axis of the Bloch sphere with probability p, and does nothing with probability $1-p$:

$$\mathcal{D}_p(\rho) := \left(1 - \frac{p}{2}\right)\rho + \frac{p}{2} Z\rho Z, \tag{H3}$$

where $p \in [0,1]$. The PTM representation of $\mathcal{D}_p$ is [2]

$$[\mathcal{D}_p] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-p & 0 & 0 \\ 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{H4}$$

and thus $\xi(\mathcal{D}_p) = p$. The noise resistance implies that our method fails only when $p = 1$, for which the truncation error cannot be bounded any more.

Amplitude damping noise. The amplitude damping channel is defined as

$$\mathcal{A}_\gamma(\rho) := E_1 \rho E_1^\dagger + E_2 \rho E_2^\dagger, \tag{H5}$$

where $\gamma \in [0,1]$, $E_1 := |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|$, and $E_2 := \sqrt{\gamma}|0\rangle\langle 1|$. γ can be thought of as the probability of losing a photon. The PTM representation of $\mathcal{A}_\gamma$ is [2]

$$[\mathcal{A}_\gamma] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-\gamma} & 0 & 0 \\ 0 & 0 & \sqrt{1-\gamma} & 0 \\ \gamma & 0 & 0 & 1-\gamma \end{pmatrix} \tag{H6}$$

and thus $\xi(\mathcal{A}_\gamma) = 2\gamma$. Likewise, our method fails for the amplitude damping channel whose noise parameter γ exceeds $1/2$, for which the truncation error cannot be bounded.

Resource analysis for the common noise channels. Figure H1 shows the optimal truncated number K as a function of the noise parameter p or γ for the above common noise channels, after ignoring negligible extra terms. This figure provides practical guidance for implementing our error mitigation method on near-term quantum devices. We have fixed the error tolerance parameter as $\varepsilon = 0.01$. As seen from the figure, smaller noise parameters lead to smaller truncated numbers and reduced overhead. For example, for the depolarizing and dephasing noises with $p \in [0, 0.3]$, the optimal truncated number K is at most 3, implying an overhead of $4^3 = 64$ is sufficient to achieve the desired mitigation accuracy.

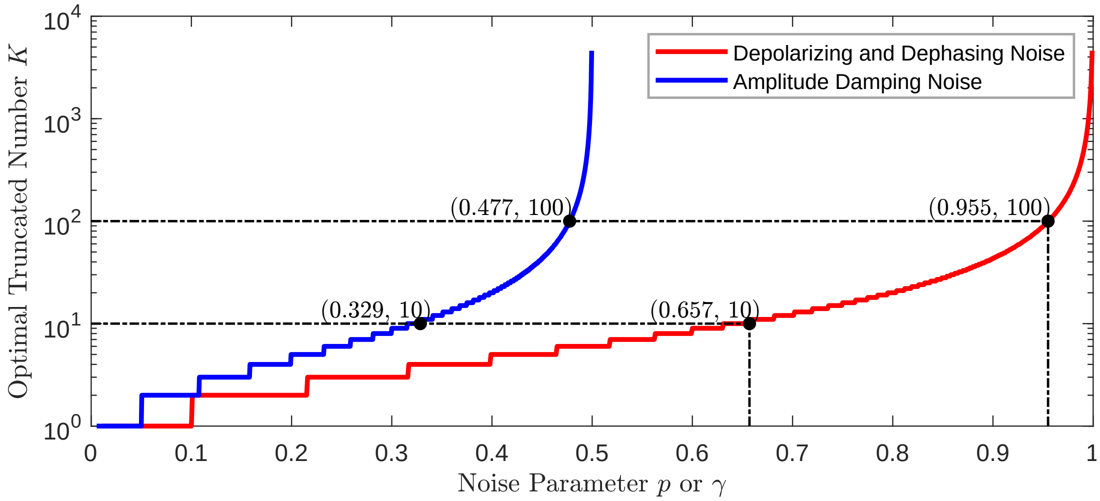

Figure H1 The (simplified) truncated number K as a function of the noise parameter p or γ of the single-qubit depolarizing, dephasing, and amplitude damping noises, where $\varepsilon = 0.01$.

References

1 Gilbert W Stewart. *Matrix Algorithms: Volume 1: Basic Decompositions.* SIAM, 1998.
2 Daniel Greenbaum. Introduction to quantum gate set tomography. *arXiv preprint arXiv:1509.02921*, 2015.
3 Mark M Wilde. *Quantum Information Theory 2nd Edition.* Cambridge University Press, 2016.
4 Mahdi Naghiloo. Introduction to experimental quantum measurement with superconducting qubits. *arXiv preprint arXiv:1904.09291*, 2019.
5 Daniel J Egger, Max Werninghaus, Marc Ganzhorn, Gian Salis, Andreas Fuhrer, Peter Mueller, and Stefan Filipp. Pulsed reset protocol for fixed-frequency superconducting qubits. *Physical Review Applied*, 10 (4): 044030, 2018.
6 Paul Magnard, Philipp Kurpiers, Baptiste Royer, Theo Walter, J-C Besse, Simone Gasparinetti, Marek Pechal, Johannes Heinsoo, Simon Storz, Alexandre Blais, et al. Fast and unconditional all-microwave reset of a superconducting qubit. *Physical Review Letters*, 121 (6): 060502, 2018.
7 Justin Yirka and Yiğit Subaşı. Qubit-efficient entanglement spectroscopy using qubit resets. *Quantum*, 5: 535, 2021.