Supplemental Information for "Circuit-Noise-Resilient Virtual Distillation"

Supplementary Note 1. Virtual Distillation

For an N-qubit noisy state ρ, the complete spectral decomposition of it can be expressed as

$$\rho = (1-\epsilon)|\psi\rangle\langle\psi| + \epsilon\sum_k \lambda|\psi_k\rangle\langle\psi_k|, \tag{S1}$$

where $0 < \epsilon < 1$ is the state preparation error rate and $|\psi\rangle\langle\psi|$ is the dominant eigenstate of ρ, along with the summation $\sum_k$ which satisfies $\sum_k \lambda_k = 1$ taking over all the noisy components of ρ. Given n identical copies of ρ, the error-mitigated expectation value of the Pauli observable O via virtual distillation (VD) is

$$\langle O\rangle_{\mathrm{VD}} = \frac{\mathrm{Tr}[\rho^n O]}{\mathrm{Tr}[\rho^n]} = \frac{(1-\epsilon)^n\langle\psi|O|\psi\rangle + \epsilon^n\sum_k \lambda_k^n\langle\psi_k|O|\psi_k\rangle}{(1-\epsilon)^n + \epsilon^n\sum_k\lambda_k^n} = \langle\psi|O|\psi\rangle + \mathcal{O}(\epsilon^n), \tag{S2}$$

which indicates that the approximation error $\mathcal{O}(\epsilon^n)$ decays exponentially with the number of copies n. To efficiently estimate $\mathrm{Tr}[\rho^n O]$ ($\mathrm{Tr}[\rho^n]$), we utilize the VD circuit which consists of the controlled derangment operator D_n and the controlled operator O. The n-order VD circuit involves a three-qubit CSWAP gate chain $\prod_{t=0}^{n-1}\prod_{i=0}^{N-1}\mathrm{CSWAP}_{i,t}$ where $\mathrm{CSWAP}_{i,t}$ acts on qubit Q_0, Q_{i+1+tN} and $Q_{i+1+(t+1)N}$, and a sequence of two-qubit controlled Pauli gates $\prod_{j=1}^{k}\mathrm{CO}_{i_j}$ where CO_{i_j} is applied on qubit Q_0 and Q_{i_j+1}. And we denote the set of index of qubits on the nontrivial Pauli terms of the k-weight Pauli observable O as $\mathcal{I}_k$.

We denote the probability of measuring the ancilla qubit Q_0 in the zero state as $p_0(\rho, O)$ and the $\mathrm{Tr}[\rho^n O]$ can be estimated by $2p_0(\rho, O) - 1$. Take the second-order VD ($n = 2$) for example, due to the linear properties of quantum computing, we can rewrite the probability $p_0(\rho, O)$ according to the spectral decomposition

$$p_0(\rho, O) = (1-\epsilon)^n p_0(|\psi\rangle^{\otimes 2}, O) + 2\epsilon(1-\epsilon)\sum_k \lambda_k p_0(|\psi\rangle|\psi_k\rangle, O) + \epsilon^2\sum_{k,k'}\lambda_k\lambda_{k'}p_0(|\psi_k\rangle|\psi_{k'}\rangle, O). \tag{S3}$$

In this way, we can focus on probability of the pure state in form of $p_0(|\psi_a\rangle|\psi_b\rangle, O)$. The input state of the VD circuit is $|+\rangle|\psi_a\rangle|\psi_b\rangle$, and we can obtain the final $(2N+1)$-qubit density matrix ρ^{out} just before measurement by

$$\begin{aligned}\rho^{\mathrm{fin}} =& \frac{1}{2}(|+\rangle|\psi_a\rangle|\psi_b\rangle\langle+|\langle\psi_a|\langle\psi_b| \;+\; |-\rangle O|\psi_b\rangle|\psi_a\rangle\langle-|\langle\psi_b|O\langle\psi_a| \;+\\ & |+\rangle|\psi_a\rangle|\psi_b\rangle\langle-|\langle\psi_b|O\langle\psi_a| \;+\; |-\rangle O|\psi_b\rangle|\psi_a\rangle\langle+|\langle\psi_a|\langle\psi_b|).\end{aligned} \tag{S4}$$

Then we calculate the probability $p_0(|\psi_a\rangle|\psi_b\rangle, O)$ via

$$p_0(|\psi_a\rangle|\psi_b\rangle, O) = \mathrm{Tr}[\rho^{\mathrm{fin}}\cdot(|0\rangle\langle 0|\otimes I_{2N})] = \begin{cases}\frac{1}{2}, \text{ if } a\neq b,\\ \\ \frac{1}{2}+\frac{1}{2}\mathrm{Tr}[|\psi_a\rangle\langle\psi_a|O], \text{ if } a=b,\end{cases}$$

where the derangement operator D_2 filters the asymmetric components of $\rho^{\otimes 2}$ to include only permutation-symmetric parts such as $|\psi_a\rangle|\psi_a\rangle$. Then we can arrive at

$$\begin{aligned}2p_0(\rho, O) - 1 =& (1-\epsilon)^2(1+\langle\psi|O|\psi\rangle) + 2\epsilon(1-\epsilon)\sum_k + \epsilon^2\sum_{k,k'}\lambda_k\lambda_{k'} + \epsilon^2\sum_k\langle|\psi_k|O|\psi_k\rangle - 1\\ =& (1-\epsilon)^2 - 1 + \mathrm{Tr}[\rho^2 O] + 2\sum_k\lambda_k\epsilon(1-\epsilon) + \epsilon^2\sum_{k,k'}\lambda_k\lambda_{k'}\\ =& \mathrm{Tr}[\rho^2 O].\end{aligned} \tag{S5}$$

Supplementary Note 2. Noisy Virtual Distillation under Pauli gate noise

In the maintext, we consider the stochastic Pauli gate noise $\mathcal{E}_m^{\mathrm{P}}(\rho) = (1-p_m)\rho + p_m\sum_{s=1}^{4^m-1}\delta_s^m P_s^m\rho P_s^m$, where $p_m \geq 0$ is the Pauli error rate and the coefficients $\{\delta_s^m\}_s$ of the m-qubit Pauli error terms $\{P_s^m\}_s$ satisfies $\delta_s^m \geq 0, \sum_s\delta_s^m = 1$.

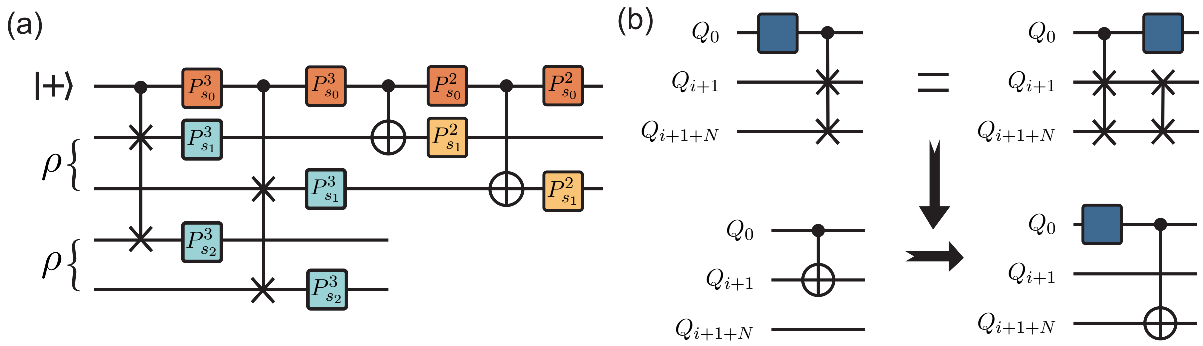

FIG. S1. **(a) The 2-qubit and 3-qubit Pauli gate noise in the second-order VD circuit. (b) The effect of the σ_x or σ_y Pauli error (blue square) acting on the controlled qubit Q_0 of the CSWAP gate where we denote the j-th qubit of the VD circuit as Q_j.** As $(\sigma_x(\sigma_y)\otimes I\otimes I)\cdot \mathrm{CSWAP} = \mathrm{CSWAP}\cdot(\sigma_x(\sigma_y)\otimes \mathrm{SWAP})$ which introduces a SWAP gate in the qubit Q_{i+1} and Q_{i+1+N}, it leads to the transformation of controlled Pauli gate CO_{i+1} to CO_{i+1+N}.

We denote the m-qubit $(m=1,2,3)$ Pauli error term as $P_s^m = \sum_{j=0}^{m-1} P_{s_j}^m$ where $P_{s_j}^m$ is from the single-qubit Pauli group $\{I,\sigma_x,\sigma_y,\sigma_z\}^{\otimes m}$. We now will provide the details of the noisy virtual distillation (VD) under Pauli gate noise in the VD circuit in this section.

We assume the local Pauli gate noise model, which means that the noise introduced by a gate only affects the qubits on which the gate is applied. For simplicity, we assume that the single-qubit Pauli gate noise can be absorbed into the nearest multi-qubit gate noise. As a result, only 2-qubit Pauli gate noise affects controlled Pauli gates, and 3-qubit Pauli gate noise affects CSWAP gates, respectively. We define the noisy CSWAP gate $\widetilde{\mathrm{CSWAP}}_{i,t}$ and controlled Pauli gate $\widetilde{\mathrm{CO}}_{i_j}$ as

$$\widetilde{\mathrm{CSWAP}}_{i,t} = \mathcal{E}_3^{\mathrm{P}}\cdot \mathrm{CSWAP}_{i,t},\quad \widetilde{\mathrm{CO}}_{i_j} = \mathcal{E}_2^{\mathrm{P}}\cdot \mathrm{CO}_{i_j}, \tag{S6}$$

as illustrated in the Fig. S1(a). As we solely measure the ancilla qubit Q_0, it is essential to consider how the Pauli gate noise in the VD circuit affects the measurement outcomes of Q_0.

The noise in the CSWAP gate chain— As the N CSWAP gates in the VD circuit only share their controlled qubit Q_0, only the first qubit of the Pauli error term $P_{s_0}^3$ can influence its following CSWAP gates. Then we can investigate the impact of Pauli terms in the form of $P_{s_0}^3\otimes I\otimes I$ before the CSWAP gate,

$$\mathrm{CSWAP}\cdot\left(P_{s_0}^3\otimes I\otimes I\right) = \begin{cases}\left(P_{s_0}^3\otimes I\otimes I\right)\cdot \mathrm{CSWAP},\ \text{if } P_{s_0}^3 = \sigma_z, I\\[2ex] \left(P_{s_0}^3\otimes \mathrm{SWAP}\right)\cdot \mathrm{CSWAP},\ \text{if } P_{s_0}^3 = \sigma_x,\sigma_y,\end{cases}$$

which means that the σ_x or σ_y Pauli error occurring on the ancilla qubit Q_0 can generate a undesired SWAP gate on the data qubits for each CSWAP gate behind it. Due to the light-cone principle, not all erroneous SWAP gates will affect the measurement outcome. When the SWAP gate acting on $(i+1, i+1+N)$ meets $i\in\mathcal{I}_k$, it will lead to the transformation of the corresponding controlled Pauli gate from CO_{i+1} to CO_{i+1+N}, as demonstrated in Fig. S1(b). This can further result in a portion of single-qubit Pauli terms of O, which originally act only on the register corresponding to the first copy of ρ, also affecting the second register. We define $O = O_1 + O_2$ and $i = 1,2$ refers to the index of the register which the operator O_i acts on. Then the final density matirx turns to

$$\begin{aligned}\tilde{\rho}^{\mathrm{fin}} =& \frac{1}{2}(|+\rangle|\psi_a\rangle|\psi_b\rangle\langle+|\langle\psi_a|\langle\psi_b| \ + \ |-\rangle O_1|\psi_b\rangle O_2|\psi_a\rangle\langle-|\langle\psi_b|O_1\langle\psi_a|O_2 \ + \\ & |+\rangle|\psi_a\rangle|\psi_b\rangle\langle-|\langle\psi_b|O_1\langle\psi_a|O_2 \ + \ |-\rangle O_1|\psi_b\rangle O_2|\psi_a\rangle\langle+|\langle\psi_a|\langle\psi_b|),\end{aligned} \tag{S7}$$

and the noisy probability $\tilde{p}_0(|\psi_a\rangle|\psi_a\rangle, O)$ becomes

$$\tilde{p}_0(|\psi_a\rangle|\psi_b\rangle, O) = \mathrm{Tr}[\tilde{\rho}^{\mathrm{fin}}\cdot(|0\rangle\langle 0|\otimes I_{2N})] = \begin{cases}\frac{1}{2},\ \text{if } a\neq b,\\[2ex] \frac{1}{2}+\frac{1}{2}\mathrm{Tr}[|\psi_a\rangle\langle\psi_a|O_1]\cdot\mathrm{Tr}[|\psi_a\rangle\langle\psi_a|O_2],\ \text{if } a=b.\end{cases}$$

Hence, the noisy estimation of $\mathrm{Tr}[\rho^2 O]$ contaminated by the undesired SWAP gate caused by the σ_x or σ_y Pauli error on the ancilla is

$$\begin{aligned}2\tilde{p}_0(\rho,O)-1=&(1-\epsilon)^2(1+\mathrm{Tr}[|\psi\rangle\langle\psi|O_1]\cdot\mathrm{Tr}[|\psi\rangle\langle\psi|O_2])+2\epsilon(1-\epsilon)\sum_k\lambda_k+\\&\epsilon^2\sum_{k,k'}\lambda_k\lambda_{k'}+\epsilon^2\sum_k\lambda_k\mathrm{Tr}[|\psi_k\rangle\langle\psi_k|O_1]\cdot\mathrm{Tr}[|\psi_k\rangle\langle\psi_k|O_2]-1\\=&(1-\epsilon)^2\mathrm{Tr}[|\psi\rangle\langle\psi|O_1]\cdot\mathrm{Tr}[|\psi\rangle\langle\psi|O_2]+\epsilon^2\sum_k\lambda_k\mathrm{Tr}[|\psi_k\rangle\langle\psi_k|O_1]\cdot\mathrm{Tr}[|\psi_k\rangle\langle\psi_k|O_2].\end{aligned}\tag{S8}$$

The noise in the controlled Pauli gate chain— We then turn our attention to the pauli gate noise in the controlled Pauli gates, along with the residual Pauli noise in the CSWAP gate chain which commutes with the CSWAP gate chain. As the controlled Pauli gate chain is Clifford which conserve the Pauli group, we can serve these Pauli errors as a $(2N+1)$-qubit stochastic Pauli noise channel which acts after the final density matrix ρ^{fin}. We consider the effect of the single-qubit Pauli error term P_{s_0}, it gives

$$\tilde{p}_0(|\psi_a\rangle|\psi_b\rangle,O)=\begin{cases}\frac{1}{2},\ \text{if } a\neq b,\ \text{or } a=b \text{ and } P_{s_0}=\sigma_y,\\[2ex]\frac{1}{2}+\frac{1}{2}\mathrm{Tr}[|\psi_a\rangle\langle\psi_a|O],\ \text{if } a=b \text{ and } P_{s_0}=I,\sigma_x,\\[2ex]\frac{1}{2}+\frac{1}{2}\mathrm{Tr}[|\psi_a\rangle\langle\psi_a|O],\ \text{if } a=b \text{ and } P_{s_0}=\sigma_z,\end{cases}$$

and then the noisy estimation is

$$2\tilde{p}_0(\rho,O)-1=\begin{cases}\mathrm{Tr}[\rho^2O],\ \text{if } P_{s_0}=I,\sigma_x,\\[2ex]0,\ \text{if } P_{s_0}=\sigma_y, & \equiv a\mathrm{Tr}[\rho^2O],\\[2ex]-\mathrm{Tr}[\rho^2O],\ \text{if } P_{s_0}=\sigma_z.\end{cases}$$

where the parameter a is ρ-independent and only determined by the structure of the VD circuit and the noise level of the Pauli gate noise in the VD circuit.

The outcome of noisy virtual distillation— From above discussions, it can be inferred that the undesired SWAP errors due to the σ_x,σ_y errors on the ancilla qubit Q_0 can lead to a partial shift of the Pauli observable O from register 1 to register 2, while the remaining Pauli errors result in multiplying the shifted expectation value by a parameter independent of ρ. When the erroneously generated SWAP gates separates O into O_1 and O_2, which reside in registers 1 and 2 respectively, we can obtain a similar expression for $\tilde{p}_0(|\psi_a\rangle|\psi_a\rangle,O)$ as

$$\tilde{p}_0(|\psi_a\rangle|\psi_a\rangle,O)=\begin{cases}\frac{1}{2}+\frac{1}{2}\langle\psi_a|O_1|\psi_a\rangle\cdot\langle\psi_a|O_2|\psi_a\rangle,\ \text{if } P_{s_0}=I,\sigma_x,\\[2ex]\frac{1}{2},\ \text{if } P_{s_0}=\sigma_y,\\[2ex]\frac{1}{2}-\frac{1}{2}\langle\psi_a|O_1|\psi_a\rangle\cdot\langle\psi_a|O_2|\psi_a\rangle,\ \text{if } P_{s_0}=\sigma_z,\end{cases}$$

However, the σ_x and σ_y errors on the ancilla qubit Q_0 also cancel each other out, making it difficult to infer the number of the unwanted SWAPs generated based on the occurrences of σ_x,σ_y errors. Therefore, we combine similar terms of expectation values resulting from different Pauli error terms based on the number of unwanted SWAPs in the VD circuit. Then the outcome obtained from the noisy VD circuit can be represented as a function that

$$2\tilde{p}_0(\rho,O)-1=\sum_{j=0}^{\lceil\frac{k}{2}\rceil}\sum_{t=0}^{\binom{k}{j}-1}a_{j,t}^k(p_1,p_2,p_3,O)A_j(t),\tag{S9}$$

where j represents the total number of SWAP errors that lead to nontrivial terms of O being transferred out of register 1, corresponding to $\binom{k}{j}$ situations. And t refers to the indices of these situations. It is worth noting that when k is even and $j=k/2$, the number of all the possible combinations of O_1 and O_2 is $\binom{k}{j}/2$. We do not make specific mention of this here in order to maintain the uniformity in the expression of $2\tilde{p}_0(\rho,O)-1$.

Supplementary Note 3. Circuit-Noise-Resilient virtual distillation under general Markovian noise model

A. The proof for the full resilience to noise in the VD circuit for the CNR-VD

In the maintext we have defined the effective noise $\mathcal{E}^{\text{eff}}$ of the VD circuit, then the noisy probability $\tilde{p}_0(\rho, O)$ of measuring zero state in the ancilla is

$$\tilde{p}_0(\rho, O) = \mathrm{Tr}[\mathcal{E}^{\text{eff}}\mathcal{U}(|+\rangle\langle+| \otimes \rho^{\otimes n})(|0\rangle\langle 0| \otimes I_{nN})], \tag{S10}$$

here the operator $\mathcal{U}$ refers to the controlled-D_n, controlled-O and the Hadamard gate H. We have defined the CNR-VD estimator as

$$\hat{O}_{\text{CNR-VD}}(\rho) = \frac{\hat{O}_{\text{VD}}(\rho)}{\hat{O}_{\text{VD}}(s)}, \tag{S11}$$

when the effective noise $\mathcal{E}^{\text{eff}}$ satisfies any of these two conditions:

- There exists a set of Kraus operators such that: each Kraus operator $K_i = K_i^{\text{R}} \otimes K_i^{\text{D}}$ of $\mathcal{E}^{\text{eff}}$ is separate into subsystems ρ^{R} and ρ^{D}, along with unitary K_i^{R} and K_i^{D}.
- $\mathcal{E}^{\text{eff}} = \mathcal{E}^{\text{eff,R}} \otimes \mathcal{E}^{\text{eff,D}}$, along with the single-qubit unital noise channel $\mathcal{E}^{\text{eff,R}}$ and the CPTP noise channel $\mathcal{E}^{\text{eff,D}}$,

then the CNR-VD estimator $\hat{O}_{\text{CNR-VD}}(\rho)$ can reduce the error in the expectation value indued by $\mathcal{E}^{\text{eff}}$.

Proof for Condition 1.—For simplicity, we suppose the state preparation error for s is negligible compared to $\mathcal{E}^{\text{eff}}$. And we take second-order VD and we assume that this two copies of noisy state are the same. Then the tensor product of the state $\rho^{\otimes 2}$ can be written as a linear combination of four types of pure state terms as shown in Eq. S1: $|\psi\rangle|\psi\rangle$, $|\psi\rangle|\psi_k\rangle$, $|\psi_k\rangle|\psi\rangle$ and $|\psi_k\rangle|\psi_{k'}\rangle$. Then we can decompose the $\tilde{p}_0(\rho, O)$ in Eq. S10 as $\tilde{p}_0(\rho, O) = (1-\epsilon)^2\tilde{p}_0(|\psi\rangle|\psi\rangle, O) + \epsilon(1-\epsilon)\sum_k \lambda_k[\tilde{p}_0(|\psi\rangle|\psi_k\rangle, O) + \tilde{p}_0(|\psi_k\rangle|\psi\rangle, O)] + \epsilon^2 \sum_{k,k'} \lambda_k\lambda_{k'}\tilde{p}_0(|\psi_k\rangle|\psi_{k'}\rangle, O)$. When the effective noise $\mathcal{E}^{\text{eff}}$ is unital noise channel and each Kraus operator conforms to $K_i = K_i^{\text{R}} \otimes K_i^{\text{D}}$ along with unitary K_i^{R} and K_i^{D}, then the expression of $\tilde{p}_0(|\psi\rangle\langle\psi|^{\otimes 2}, O)$ in Eq. S21 turns to

$$\begin{aligned}
\tilde{p}_0(|\psi\rangle|\psi\rangle, O) &= \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff}}\left(\mathcal{U}(|+\rangle|\psi\rangle|\psi\rangle) \cdot \mathcal{U}(|+\rangle|\psi\rangle|\psi\rangle)^{\dagger}\right) \cdot (|0\rangle\langle 0| \otimes I_{2N})] \\
&= \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff}}(|+\rangle\langle+| \otimes |\psi\rangle|\psi\rangle\langle\psi|\langle\psi|) \cdot (|0\rangle\langle 0| \otimes I_{2N})] + \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff}}(|-\rangle\langle-| \otimes |\psi\rangle O|\psi\rangle\langle\psi|O\langle\psi|) \\
&\cdot (|0\rangle\langle 0| \otimes I_{2N})] + \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff}}(|+\rangle\langle-| \otimes |\psi\rangle|\psi\rangle\langle\psi|O\langle\psi|) \cdot (|0\rangle\langle 0| \otimes I_{2N})] \\
&+ \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff}}(|-\rangle\langle+| \otimes |\psi\rangle O|\psi\rangle\langle\psi|\langle\psi|) \cdot (|0\rangle\langle 0| \otimes I_{2N})] \\
&= \frac{1}{2}\sum_i \mathrm{Tr}[(K_i^{\text{R}}|+\rangle\langle+|K_i^{\text{R}\dagger})|0\rangle\langle 0|] \cdot \mathrm{Tr}[K_i^{\text{D}}|\psi\rangle\langle\psi|^{\otimes 2}K_i^{\text{D}\dagger}] \\
&+ \frac{1}{2}\sum_i \mathrm{Tr}[(K_i^{\text{R}}|-\rangle\langle-|K_i^{\text{R}\dagger})|0\rangle\langle 0|] \cdot \mathrm{Tr}[K_i^{\text{D}}(O|\psi\rangle\langle\psi|O)^{\otimes 2}K_i^{\text{D}\dagger}] \\
&+ \frac{1}{2}\sum_i \mathrm{Tr}[(K_i^{\text{R}}|+\rangle\langle-|K_i^{\text{R}\dagger})|0\rangle\langle 0|] \cdot \mathrm{Tr}[K_i^{\text{D}}((|\psi\rangle\langle\psi|O) \otimes |\psi\rangle\langle\psi|)K_i^{\text{D}\dagger}] \\
&+ \frac{1}{2}\sum_i \mathrm{Tr}[(K_i^{\text{R}}|-\rangle\langle+|K_i^{\text{R}\dagger})|0\rangle\langle 0|] \cdot \mathrm{Tr}[K_i^{\text{D}}(|\psi\rangle\langle\psi| \otimes (|\psi\rangle\langle\psi|O))K_i^{\text{D}\dagger}] \\
&= \frac{1}{2}\sum_i \mathrm{Tr}[(K_i^{\text{R}}IK_i^{\text{R}\dagger})|0\rangle\langle 0|] + \frac{1}{2}\mathrm{Tr}[|\psi\rangle\langle\psi|O]\sum_i \mathrm{Tr}[(K_i^{\text{R}}ZK_i^{\text{R}\dagger})|0\rangle\langle 0|] \\
&:= \frac{1}{2} + b\mathrm{Tr}[|\psi\rangle\langle\psi|O],
\end{aligned} \tag{S12}$$

where we use the fact that the unitary Kraus operator meeting $K_i^{\text{R}}K_i^{\text{R}\dagger} = I$. And for other three pure state terms, the same approach can be employed for computation, yielding the following result:

$$\begin{aligned}
\tilde{p}_0(|\psi\rangle|\psi_k\rangle, O) &= a, \\
\tilde{p}_0(|\psi_k\rangle|\psi\rangle, O) &= a,
\end{aligned} \tag{S13}$$

$$\tilde{p}_0(|\psi_k\rangle|\psi_{k'}\rangle, O) = \begin{cases} a, \text{if } k \neq k' \\ a + b\mathrm{Tr}[|\psi_k\rangle\langle\psi_k|O], \text{ else.} \end{cases}$$

Thus we can obtain the expression of $\tilde{p}_0(\rho, O)$ as

$$\begin{aligned}\tilde{p}_0(\rho, O) &= (1-\epsilon)^2(\frac{1}{2} + b\mathrm{Tr}[|\psi\rangle\langle\psi|O]) + \epsilon(1-\epsilon) + \epsilon^2(\frac{1}{2} + b\sum_k \lambda_k^2 \mathrm{Tr}[|\psi_k\rangle\langle\psi_k|O]) \\ &= \frac{1}{2} + b\left((1-\epsilon)^2\mathrm{Tr}[|\psi\rangle\langle\psi|O] + \epsilon^2\sum_k \lambda_k^2\mathrm{Tr}[|\psi_k\rangle\langle\psi_k|O]\right) \\ &= \frac{1}{2} + b\mathrm{Tr}[\rho^2 O].\end{aligned} \tag{S14}$$

Then we can deduce that

$$\frac{2\tilde{p}_0(\rho, O) - 1}{2\tilde{p}_0(s, O) - 1} = \mathrm{Tr}[\rho^2 O]. \tag{S15}$$

Similarly, it can be obtained that

$$\frac{2\tilde{p}_0(\rho, I) - 1}{2\tilde{p}_0(s, I) - 1} = \mathrm{Tr}[\rho^2]. \tag{S16}$$

Then in this situation, the simplified CNR-VD estimator comply with

$$\hat{O}_{\text{CNR-VD}} = \frac{2\tilde{p}_0(\rho, O) - 1}{2\tilde{p}_0(\rho, I) - 1} \cdot \frac{2\tilde{p}_0(s, O) - 1}{2\tilde{p}_0(s, I) - 1} = \frac{\hat{O}_{\text{VD}}(s)}{\hat{O}_{\text{VD}}(\rho)} = \frac{\mathrm{Tr}[\rho^2 O]}{\mathrm{Tr}[\rho^2]}, \tag{S17}$$

which can also be easily extended to $n > 2$.

Proof for Condition 2.—When the effective noise $\mathcal{E}^{\text{eff}}$ is separate on two subsystems R and D that $\mathcal{E}^{\text{eff}} = \mathcal{E}^{\text{eff,R}} \otimes \mathcal{E}^{\text{eff,D}}$ along with the single-qubit unital noise channel $\mathcal{E}^{\text{eff,R}}$ and CPTP noise channel $\mathcal{E}^{\text{eff,D}}$, then this yields

$$\begin{aligned}\tilde{p}_0(|\psi\rangle|\psi\rangle, O) &= \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff,R}}(|+\rangle\langle+|)|0\rangle\langle 0|] \cdot \mathrm{Tr}[\mathcal{E}^{\text{eff,D}}(|\psi\rangle\langle\psi|^{\otimes 2})] \\ &+ \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff,R}}(|-\rangle\langle-|)|0\rangle\langle 0|] \cdot \mathrm{Tr}[\mathcal{E}^{\text{eff,D}}((O|\psi\rangle\langle\psi|O)^{\otimes 2})] \\ &+ \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff,R}}(|+\rangle\langle-|)|0\rangle\langle 0|] \cdot \mathrm{Tr}[\mathcal{E}^{\text{eff,D}}((|\psi\rangle\langle\psi|O) \otimes |\psi\rangle\langle\psi|)] \\ &+ \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff,R}}(|-\rangle\langle+|)|0\rangle\langle 0|] \cdot \mathrm{Tr}[\mathcal{E}^{\text{eff,D}}(|\psi\rangle\langle\psi| \otimes (|\psi\rangle\langle\psi|O))] \\ &= \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff,R}}(I)|0\rangle\langle 0|] + \frac{1}{2}\mathrm{Tr}[\mathcal{E}^{\text{eff,D}}(Z)|0\rangle\langle 0|] \cdot \mathrm{Tr}[|\psi\rangle\langle\psi|O] \\ &= \frac{1}{2} + b\mathrm{Tr}[|\psi\rangle\langle\psi|O],\end{aligned} \tag{S18}$$

where we use the fact $\mathrm{Tr}[\mathcal{E}^{\text{eff,D}}(\sigma)] = \mathrm{Tr}[\sigma]$ for the CPTP noise channel $\mathcal{E}^{\text{eff,D}}$ and any operator σ, and $\mathcal{E}^{\text{eff,R}}(I) = I$ for the unital noise channel $\mathcal{E}^{\text{eff,R}}$. Therefore, through a similar deduction process as condition 1, we can conclude that the Eq. S17 also holds under condition 2.

B. CNR-VD estimator with loosen requirement of the effective noise

We also establish other form of the CNR-VD estimator with relaxed requirement. For differentiation, we denote the s^+ and s^- as the eigenstate of Pauli O with eigenvalue $+1$ and -1, respectively. As for $O = I$ where only eigenvalue $+1$ exists, we replace the first copy of s^+ with g^0 ($g^{\frac{1}{2}}$) which satisfies $|\langle s^+|g^0\rangle|^2 = 0$ (and $|\langle s^+|g^{\frac{1}{2}}\rangle|^2 = 1/2$), and the noisy probability is referred to $\tilde{p}_0(g^0, I)$ ($\tilde{p}_0(g^{\frac{1}{2}}, I)$). We will utilize these noisy probabilities obtained from both noisy VD and the calibration stage to establish the CNR-VD estimator in the form of

$$\hat{O}_{\text{CNR-VD}}(\rho) = \left(\frac{2\tilde{p}_0(\rho, O) - \tilde{p}_0(s^+, O) - \tilde{p}_0(s^-, O)}{\tilde{p}_0(\rho, I) - \tilde{p}_0(g^0, I)}\right) \times \left(\frac{2(\tilde{p}_0(g^{\frac{1}{2}}, I) - \tilde{p}_0(g^0, I))}{\tilde{p}_0(s^+, O) - \tilde{p}_0(s^-, O)}\right). \tag{S19}$$

We now turn our attention to the precise condition when $\hat{O}_{\text{CNR-VD}}(\rho)$ exhibits full resilience to the effective noise. We can get the following condition.

Theorem 1. *If the effective noise $\mathcal{E}^{\text{eff}}$ satisfys*

$$Tr_{1:nN}[\mathcal{E}^{\text{eff}}(\rho^{RD})] = \sum_i c_i\, Tr[\rho^D]\rho_i^R, \quad \text{with } \rho_i^R = Tr_{1:nN}[K_i(\rho^R \otimes I_{nN})K_i^\dagger], \tag{S20}$$

where $Tr_{1:nN}$ denotes the partial trace over qubits from index 1 to nN and c_i is some constant which is exclusively associated with $\mathcal{E}^{\text{eff}}$, then the CNR-VD estimator $\hat{O}_{\text{CNR-VD}}(\rho)$ given in Eq. S19 can eliminate the undesired bias in the expectation value caused by $\mathcal{E}^{\text{eff}}$.

The criteria of $\mathcal{E}^{\text{eff}}$ set out in Eq. S20 means that the effective noise does not introduce entanglement between ρ^{R} and ρ^{D}.

Proof.— For the effective noise $\mathcal{E}^{\text{eff}}$ satisfying the conditions described in theorem 1, $\tilde{p}_0(|\psi\rangle|\psi\rangle, O)$ equals

$$\begin{aligned}
\tilde{p}_0(|\psi\rangle|\psi\rangle, O) &= \frac{1}{2}\text{Tr}[\mathcal{E}^{\text{eff}}\left(\mathcal{U}(|+\rangle|\psi\rangle|\psi\rangle)\cdot\mathcal{U}(|+\rangle|\psi\rangle|\psi\rangle)^\dagger\right)\cdot(|0\rangle\langle 0|\otimes I_{2N})] \\
&= \frac{1}{2}\text{Tr}[\mathcal{E}^{\text{eff}}(|+\rangle\langle +|\otimes|\psi\rangle|\psi\rangle\langle\psi|\langle\psi|)\cdot(|0\rangle\langle 0|\otimes I_{2N})] + \frac{1}{2}\text{Tr}[\mathcal{E}^{\text{eff}}(|-\rangle\langle -|\otimes|\psi\rangle O|\psi\rangle\langle\psi|O\langle\psi|) \\
&\cdot(|0\rangle\langle 0|\otimes I_{2N})] + \frac{1}{2}\text{Tr}[\mathcal{E}^{\text{eff}}(|+\rangle\langle -|\otimes|\psi\rangle|\psi\rangle\langle\psi|O\langle\psi|)\cdot(|0\rangle\langle 0|\otimes I_{2N})] \\
&+ \frac{1}{2}\text{Tr}[\mathcal{E}^{\text{eff}}(|-\rangle\langle +|\otimes|\psi\rangle O|\psi\rangle\langle\psi|\langle\psi|)\cdot(|0\rangle\langle 0|\otimes I_{2N})] \\
&= \frac{1}{2}\sum_i c_i\text{Tr}[K_i((|+\rangle\langle +| + |-\rangle\langle -|)\otimes I_{2N})K_i^\dagger\cdot(|0\rangle\langle 0|\otimes I_{2N})] \\
&+ \frac{1}{2}\text{Tr}[|\psi\rangle\langle\psi|O]\cdot\sum_i c_i\text{Tr}[K_i((|+\rangle\langle -| + |-\rangle\langle +|)\otimes I_{2N})K_i^\dagger\cdot(|0\rangle\langle 0|\otimes I_{2N})] \\
&= \frac{1}{2}\sum_i c_i\text{Tr}[K_iK_i^\dagger\cdot(|0\rangle\langle 0|\otimes I_{2N})] + \frac{1}{2}\text{Tr}[|\psi\rangle\langle\psi|O]\cdot\sum_i c_i\text{Tr}[K_i((Z\otimes I_{2N})K_i^\dagger\cdot(|0\rangle\langle 0|\otimes I_{2N})] \\
&:= a + b\text{Tr}[|\psi\rangle\langle\psi|O],
\end{aligned} \tag{S21}$$

where the noise parameters a and b are unrelated to the input state $|\psi\rangle\langle\psi|^{\otimes 2}$ since c_i is only exclusively relative with $\mathcal{E}^{\text{eff}}$. Thus we can obtain the expression of $\tilde{p}_0(\rho, O)$ in the similar way as in Eq. S14

$$\begin{aligned}
\tilde{p}_0(\rho, O) &= (1-\epsilon)^2(a + b\text{Tr}[|\psi\rangle\langle\psi|O]) + 2a\epsilon(1-\epsilon) + \epsilon^2(a + b\sum_k \lambda_k^2\text{Tr}[|\psi_\rangle\langle\psi_k|O]) \\
&= a + b\left((1-\epsilon)^2\text{Tr}[|\psi\rangle\langle\psi|O] + \epsilon^2\sum_k\lambda_k^2\text{Tr}[|\psi_\rangle\langle\psi_k|O]\right) \\
&= a + b\text{Tr}[\rho^2 O].
\end{aligned} \tag{S22}$$

When we input the eigenstate s^+ (s^-) of the Pauli observable O with eigenvalue 1 (-1), we can obtain $\tilde{p}_0(s^+, O) = a+b$ ($\tilde{p}_0(s^-, O) = a-b$). As a result, it follows that

$$\frac{2\tilde{p}_0(\rho, O) - \tilde{p}_0(s^+, O) - \tilde{p}_0(s^-, O)}{\tilde{p}_0(s^+, O) - \tilde{p}_0(s^-, O)} = \text{Tr}[\rho^2 O]. \tag{S23}$$

For $O = I$, the $\tilde{p}_0(\rho, I) = a' + b'\text{Tr}[\rho^2]$ and $\tilde{p}_0(s^+, I) = a' + b'$. When we replace the first copy of s^+ with g^0, since s^+ and g^0 are both pure states we can get

$$\tilde{p}_0(g^0, I) = a' + b'|\langle s^+|g^0\rangle|^2 = a', \tag{S24}$$

following Eq. S21. And the same way can be used to obatin $\tilde{p}_0(g^{\frac{1}{2}}, I) = a' + b'/2$. Then we get

$$\frac{\tilde{p}_0(\rho, I) - \tilde{p}_0(g^0, I)}{2(\tilde{p}_0(g^{\frac{1}{2}}, I) - \tilde{p}_0(g^0, I))} = \text{Tr}[\rho^2], \tag{S25}$$

which means the CNR-VD estimator defined in Theorem 1 satisfies

$$\begin{aligned}\hat{O}_{\text{CNR-VD}}(\rho) = &\frac{2\tilde{p}_0(\rho,O)-\tilde{p}_0(s^+,O)-\tilde{p}_0(s^-,O)}{\tilde{p}_0(\rho,I)-\tilde{p}_0(g^0,I)}\\ &\cdot\frac{2(\tilde{p}_0(g^{\frac{1}{2}},I)-\tilde{p}_0(g^0,I))}{\tilde{p}_0(s^+,O)-\tilde{p}_0(s^-,O)} = \frac{\text{Tr}[\rho^2 O]}{\text{Tr}[\rho^2]}.\end{aligned} \tag{S26}$$

And this can be easily extended to $n > 2$ by expanding ρ^n in terms of the binomial series.

C. The variance of the CNR-VD estimator

Since the CNR-VD estimator is in a form of quotient, we can employ the error propagation formula to calculate the variance of the CNR-VD estimator, incorporating the first-order Taylor expansion into the calculation. Consider a function $Y = A/B$, where A and B are random variables. When A and B are mutual independent, we can compute the variance of Y denoted as $\text{Var}[Y]$ by

$$\text{Var}[Y] \approx \left(\frac{\mathbb{E}[A]}{\mathbb{E}[B]}\right)^2 \cdot \left(\frac{\text{Var}[A]}{(\mathbb{E}[A])^2} + \frac{\text{Var}[B]}{(\mathbb{E}[B])^2}\right). \tag{S27}$$

Here we only calculate the variance of the CNR-VD estimator in Eq. S11 and we can compute the variance of other version of CNR-VD estimator in the similar way. The variance of the simpified $\hat{O}_{\text{CNR-VD}}(\rho)$ is

$$\begin{aligned}\text{Var}[\hat{O}_{\text{CNR-VD}}(\rho)] &= \text{Var}[\frac{\hat{O}_{\text{VD}}(\rho)}{\hat{O}_{\text{VD}}(s)}]\\ &= \left(\frac{\mathbb{E}[\hat{O}_{\text{VD}}(\rho)]}{\mathbb{E}[\hat{O}_{\text{VD}}(s)}\right)^2 \left(\frac{\text{Var}[\hat{O}_{\text{VD}}(\rho)]}{(\mathbb{E}[\hat{O}_{\text{VD}}(\rho)])^2} + \frac{\text{Var}[\hat{O}_{\text{VD}}(s)]}{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^2}\right).\end{aligned} \tag{S28}$$

If we assume $\hat{O}_{\text{VD}}(\rho)$ has the same variance for any input state under the same measurement shots number M, we can infer that the variance of $\hat{O}_{\text{CNR-VD}}(\rho)$ scales proportionally to that of $\hat{O}_{\text{VD}}(\rho)$ by

$$\begin{aligned}&\text{Var}[\hat{O}_{\text{CNR-VD}}(\rho)] =\\ &\left(\frac{1}{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^2} + \frac{(\mathbb{E}[\hat{O}_{\text{VD}}(\rho)])^2}{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^4}\right)\text{Var}[\hat{O}_{\text{VD}}(\rho)]\\ &\leq \left(\frac{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^2+1}{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^4}\right)\text{Var}[\hat{O}_{\text{VD}}(\rho)],\end{aligned} \tag{S29}$$

where the factor is totally determined by $\mathbb{E}[\hat{O}_{\text{VD}}(s)]$ which indicates the noise level in the VD circuit.

To investigate the relationship among $\mathbb{E}[\hat{O}_{\text{VD}}(s)]$, the qubit number N, and the order n more intuitively, we consider the simplest scenario where only local depolarizing gate noise with error rate γ occurs in the circuit. Since the depolarizing noise meets $\mathcal{E}_\gamma^{\text{depo}}(P) = (1-\frac{4}{3}\gamma)P$ for non-trivial Pauli terms X, Y and Z, we can get $\mathcal{E}[\hat{O}_{\text{VD}}(s)] = \frac{(1-\frac{4}{3}\gamma)^{(n-1)N+k}}{(1-\frac{4}{3}\gamma)^{(n-1)N}} = (1-\frac{4}{3}\gamma)^k$ and k is weight of observable O which is the count of non-trivial Pauli terms in O. Consequently, we deduce

$$\frac{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^2+1}{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^4} = \frac{(1-\frac{4}{3}\gamma)^{2k}+1}{(1-\frac{4}{3}\gamma)^{4k}} = \frac{1}{(1-\frac{4}{3}\gamma)^{2k}} + \frac{1}{(1-\frac{4}{3}\gamma)^{4k}}. \tag{S30}$$

We perform a Taylor expansion around $\gamma = 0$, truncating the series at $o(\gamma^2)$,

$$(1-\frac{4}{3}\gamma)^{-k} = 1 + \frac{4}{3}k\gamma + \frac{8k(k+1)}{9}\gamma^2. \tag{S31}$$

Thus, we arrive at

$$\frac{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^2+1}{(\mathbb{E}[\hat{O}_{\text{VD}}(s)])^4} \approx 2 + (8\gamma + \frac{16}{3}\gamma^2)k + \frac{160}{9}\gamma^2 k^2, \tag{S32}$$

which entails that the scaling factor of variance for the CNR-VD estimator remains uncorrelated with the order of VD, and when the level of gate noise is minimal, the scaling factor exhibits a polynomial growth with the weight k of the observable.

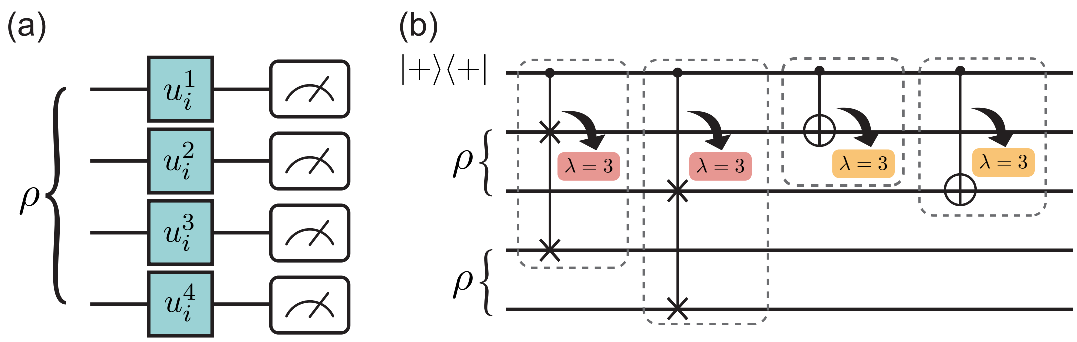

FIG. S2. **The visualization presented depicts the process of 4-qubit shadow distillation (SD) and 2-qubit VD integrated with zero-noise extrapolation (ZNE-VD) for observable $X^{\otimes 2}$.** .(a) Random single-qubit Clifford gate $u_i^{(j)}(j = 1, 2, 3, 4)$ is applied to each qubit of ρ, followed by the computation of $\frac{\mathrm{Tr}[\rho^2 X^{\otimes 2}]}{\mathrm{Tr}[\rho^2]}$ through shadow estimation utilizing the measurement outcomes of a set of random gates $\{U_i = \prod_{j=1}^4 u_i^{(j)}\}_i$. (b) The noise level is amplified three times for zero-noise extrapolation, with an amplification error 94%.

Supplementary Note 4. Details of numerical simulations

A. The implementation of shadow distillation and VD integrated with zero-noise extrapolation

In the numerical simulations conducted in this maintext, we compare the CNR-VD with two common VD-based methods which are designed against noise in the VD circuit, namely Shadow Distillation and VD integrated with Zero-Noise Extrapolation. In this section, we will provide a comprehensive description of the implementation process for these two methods, and we only use second-order VD in this simulations. The illustration of these two methods are given in Fig. S2

1. *Shadow Distillation*

Shadow Distillation (SD) is one of the variants of VD which use shadow estimation technique to estimate the unlinear function $\frac{\mathrm{Tr}[\rho^2 O]}{\mathrm{Tr}[\rho^2]} = \frac{\mathrm{Tr}[D_2(\rho\otimes(\rho O))]}{\mathrm{Tr}[D_2\rho^{\otimes 2}]}$ of ρ. In the SD, we treat the operator $D_2(I \otimes O)$ (D_2) as an observable. By randomly choosing N_u local Clifford gates $\{U_i\}_i$, we obtain the corresponding shadow $\bar{\rho}_i$ from N_s measuremeant outcomes for each U_i. Then we can use the estimator $\hat{O}_{\mathrm{SD}}(\rho) = \hat{O}^{\mathrm{num}}_{\mathrm{SD}}(\rho)/\hat{O}^{\mathrm{den}}_{\mathrm{SD}}(\rho)$ to compute the error-mitiagted expectation value where

$$\hat{O}^{\mathrm{num}}_{\mathrm{SD}}(\rho) = \frac{1}{N_u(N_u - 1)} \sum_{i\neq i'} \mathrm{Tr}[D_2\bar{\rho}_i \otimes (\bar{\rho}_{i'} O)], \tag{S33}$$

and we can get $\hat{O}^{\mathrm{den}}_{\mathrm{SD}}(\rho)$ by replacing O with I. In our simulations, we set $N_s = 10$ if the total measurement shots $M \leq 10^4$, otherwise $N_s = 50$.

2. *VD integrated with Zero-Noise Extrapolation*

Typical quantum error mitigation methods such as Zero-Noise Extrapolation (ZNE) can also be employed to minimize the noise in the VD circuit, and we denote this method as ZNE-VD. ZNE employs data gathered across varying error rates to develop a model of expectation values, which can be extrapolated to zero noise limit. We use the two-points exponential extrapolation [S1] to calculate $\mathrm{Tr}[\rho^2 O]$ by

$$\hat{O}^{\mathrm{num}}_{\mathrm{ZNE\text{-}VD}}(\rho) = \left(\frac{\hat{O}^{\lambda}_{\mathrm{VD}}(\rho, \epsilon)}{\hat{O}_{\mathrm{VD}}(\rho, \lambda\epsilon)}\right)^{\frac{1}{\lambda - 1}}, \tag{S34}$$

here we employ the denotation that $\hat{O}^{\mathrm{num}}_{\mathrm{VD}}(\rho, \lambda\epsilon)$ is the estimation of $\mathrm{Tr}[\rho^2 O]$ under error rate $\lambda\epsilon$ where $\lambda >= 1$ is the amplified coefficient for the original noise level ϵ. Then the error-mitiagted expectation value of ZNE-VD is computed via $\hat{O}_{\text{ZNE-VD}}(\rho) = \hat{O}^{\text{num}}_{\text{ZNE-VD}}(\rho)/\hat{O}^{\text{dun}}_{\text{ZNE-VD}}(\rho)$. The noise level of gate noise is amplified by $\lambda = 3$ with the amplification error 94% [S2]. In the numerical simulations, we also employ RC technique in the VD circuit to enhance the efficacy of ZNE. The implementation of RC is the same as illustrated in Fig. ??.

B. Noise model used in the numerical simulations

In the maintext, we assume that quantum circuits are mainly affected by the multi-qubit gate noise. And we also assume clean readout in the VD-based circuit since it only requires single-qubit measurement on the ancilla qubit. When we perform shadow distillation (SD) and unmitigated estimation, noisy multi-qubit measuremeants are still needed. Hence we also employ the multi-qubit readout error model in simulations. Moreover, we also configure the state preparation error for the noisy state applied to the VD circuit. In this section, we will provide details of the gate-noise model, multi-qubit readout error model and state preparation error in the following. All the numerical simulations are performed on Cirq [S3, S4].

1. *Gate noise model*

We employ two types of gate noise model which are the stochastic Pauli gate noise model and the composite gate noise model consisting of depolarizing and amplitude damping.

The stochastic Pauli gate noise.—We set ϵ to be the noise level which is the average error rate for 2-qubit gates. The benchmark of the noise (when $\epsilon = 1$) is from *Zuchongzhi* 2.1 [S5]. With the benchmark $c_1 = 1.6 \times 10^{-3}$ for single-qubit error and $c_2 = 6 \times 10^{-3}$ for 2-qubit error, we can calculate the Pauli error rates $p_1 = (\epsilon \cdot c_1/(1-1/2)) \cdot (1-1/4)$ and $p_2 = (\epsilon \cdot c_2/(1-1/4)) \cdot (1-1/16)$ [S6] with randomly chosen Pauli noise parameters $\{\delta_s^2\}_s$. We use the same stochastic Pauli noise within each independent experiment including noisy VD and the calibration stage. Taking into consideration that the experimental implementation of the CSWAP gate typically involves the use of 2-qubit gates for compilation, we have adopted the compilation method outlined in Ref [S2], where a minimum of six 2-qubit gates is required to compile the CSWAP gate. Consequently, according to Ref [S6], the process fidelity of the CSWAP gate is six times greater than that of 2-qubit gates under Pauli error rate p_2. As a result, we are able to calculate that the fidelity of the CSWAP gate is $\mathcal{F}(\text{CSWAP}) = (1-p_2)^6$ and derive the corresponding Pauli error rate $p_3 = 1 - \mathcal{F}(\text{CSWAP}) = 1 - (1-p_2)^6$.

The composite gate noise.—Given the noise level ϵ, we employ single-qubit depolarizing gate noise

$$\mathcal{E}_{\text{depo}}(\cdot) = (1 - \epsilon \cdot p_{\text{depo}})(\cdot) + \frac{1}{2}\epsilon \cdot p_{\text{depo}} I_2,$$

for each qubit of the gate and the p_{depo} is computed in the similar way as the Pauli gate noise [S6]. We also ultilize single-qubit amplitude damping $\mathcal{E}_{\text{ampl}}$ with the following definition

$$\mathcal{E}_{\text{ampl}}(\cdot) = E_0(\cdot)E_0^\dagger + E_1(\cdot)E_1^\dagger$$

where the Kraus operator $E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\epsilon \cdot p_{\text{ampl}}} \end{bmatrix}$, $E_1 = \begin{bmatrix} 0 & \sqrt{\epsilon \cdot p_{\text{ampl}}} \\ 0 & 0 \end{bmatrix}$. Here the p_{ampl} is calculated using $p_{\text{ampl}} = 1 - e^{-\frac{d_2}{d_1}}$ where $d_1 = 22.67 \times 10^{-6}$ and $d_2 = 20 \times 10^{-9}$ represent the average data for the T1 and T2 duration collected from *Zuchongzhi* 2.1 [S5, S7]. In our simulations, we set the noise level $\epsilon = 1$.

2. *Readout noise model*

It is common to assume the classical readout error model where the transfer matrix Λ is employed to discribe the transformation between ideal and noisy measuremeant probability distributions which are denoted as $\mathbf{p}^{\text{ideal}}$ and $\mathbf{p}^{\text{noisy}}$

$$\mathbf{p}^{\text{ideal}} = \Lambda^{-1}\mathbf{p}^{\text{noisy}}. \tag{S35}$$

The element $\Lambda_{\mathbf{x},\mathbf{y}}$ of the response matrix Λ is expressed as

$$\Lambda_{\mathbf{x},\mathbf{y}} = \langle \mathbf{x}|\rho|\mathbf{y}\rangle, \quad \mathbf{x}, \mathbf{y} \in \mathbb{Z}_2^N, \tag{S36}$$

where $\mathbf{x} \in \mathbb{Z}_2^N$ refers to the measuremeant outcome. We utilize the Iterative Bayesian Unfolding (IBU) method [S8] to calculate $\mathbf{p}^{\text{ideal}}$ in order to avoid the direct inversion in Eq. S35, which may produce unphysical probabilites and

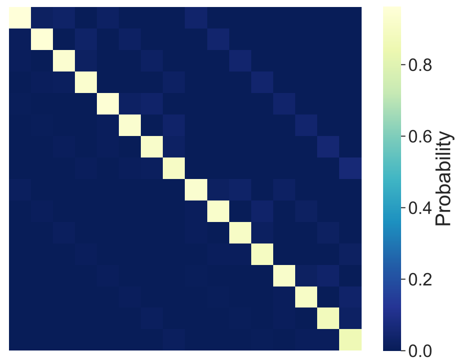

FIG. S3. **The 4-qubit transfer matrix obtained from *Zuchongzhi* 2.1.**.

amplify the statistical uncertainties in Λ. We use 100 IBU iterations and iteration formula for the $(t+1)$-th iteration is calculated as follows

$$m_i^{t+1} = \sum_j \frac{\Lambda_{ji} m_i^t}{\sum_k \Lambda_{jk} m_k^t} \times r_j, \tag{S37}$$

where m_i and r_j are the i-th and j-th terms in the probability estimated by IBU and the $\mathbf{p}^{\text{noisy}}$, respectively.

The full transfer matrix Λ we used as readout error is built on the calibration data in *Zuchongzhi* 2.1 [S5] as shown in Fig. S3 and the estimated transfer matrix $\bar{\Lambda}$ for IBU is obtained under the Tensor Product Noise [S9] (TPN) model that $\bar{\Lambda} = \bar{\Lambda}_1 \otimes \bar{\Lambda}_2 \otimes ... \otimes \bar{\Lambda}_N$. The sigle-qubit transfer matrix $\bar{\Lambda}_i$ on $(i-1)$-th qubit also relies on the calibration data in *Zuchongzhi* 2.1. We use the TPN model because full matrix calibration is exponentially costly. However, this assumption completely ignores the correlated readout error, which leads to additional correlated measurement noise in SD and unmitigated estimation.

3. *State preparation noise*

In our numerical simulations in the maintext, two types of noisy states are employed. The first is the random state $\rho_r = (1-\epsilon)|\psi\rangle\langle\psi| + \epsilon|\psi_e\rangle\langle\psi_e|$ with certain purity $(1-\epsilon)^2 + \epsilon^2$. We set the density matrix $\rho_{\rm r}$ as the initial state applied to the VD circuit without specific state preparation circuit. The secend is the N-qubit GHZ state $|\text{GHZ}\rangle_N = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes N} + |1\rangle^{\otimes N})$, we configure the N-qubit global depolarizing error for $|\text{GHZ}\rangle_N$ as

$$\mathcal{E}_{\text{depo}}^N(\cdot) = (1 - \epsilon \cdot p_{\text{depo}})(\cdot) + \epsilon \cdot p_{\text{depo}} I_N, \tag{S38}$$

where the error rate p_{depo} is set to be 10^{-1}.

C. Relative simulation results

1. *Simulation results for higher order VD*

We focus on the second-order VD in the maintext. Here, we supplement several simulations for 3-qubit random noisy state ρ_r concerning higher-order VD and explore the relationship between the accuracy of CNR-VD and the increase in the order of VD with order 2 and order 3, noise level in the VD circuit changing from 0.1 to 10.0 and the state preparation error rate varying from 0.05 to 0.30. Referring to the data illustrated in Fig. S4(a), CNR-VD demonstrates an impressive ability to recover the ideal behavior of VD, with noise level lower than 1. However, as the noise level increases, the incremental improvement in accuracy for CNR-VD from increasing the order diminishes gradually, reaching a point at 5 where the results for order 3 are actually inferior to those for order 2. We can also observe this from Fig. S4(b), where we define the decay rate as the logarithm of the ratio of the accuary of order 2 to

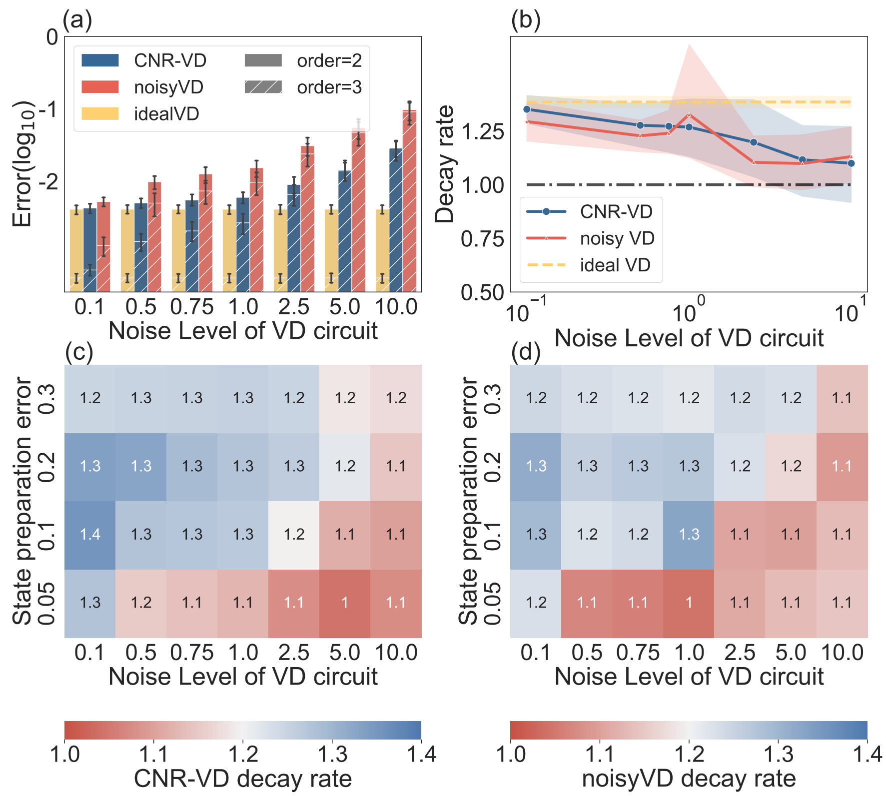

FIG. S4. **Performance on CNR-VD of random noisy states, with different orders of VD.**(a) The complete visualization of the variation in accuracy with order 2,3 and varying noise levels. (b) The decay rate of the expectation value estimation error varying with the noise levels. (c) The heat map of the decay rate for CNR-VD varying with state preparation error rates and noise levels of the VD circuit. (d) The heat map of the decay rate for noisy VD. The displayed data represents the average outcome derived from 50 separate experiments with different random states.

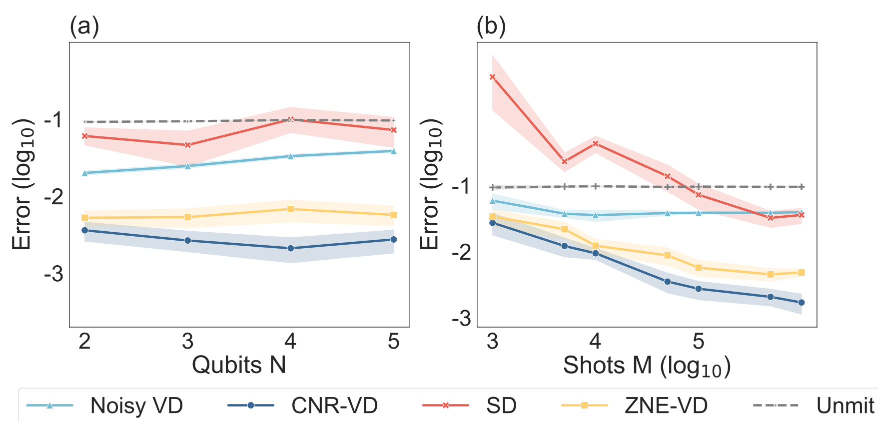

FIG. S5. **Accuracy of several VD-based methods for GHZ state, with increasing qubit number N and shot number M.** The displayed data represents the average outcome derived from 20 separate experiments.

that of order 3. It is obvious that the cost-effectiveness of increasing the order becomes diminished as the noise level increases. We further present a detailed demonstration of the decay rate of CNR-VD and noisy VD as they vary with the changes in noise levels of the VD circuit and state preparation error rates, as shown in Fig. S4(c) and Fig. S4(d).

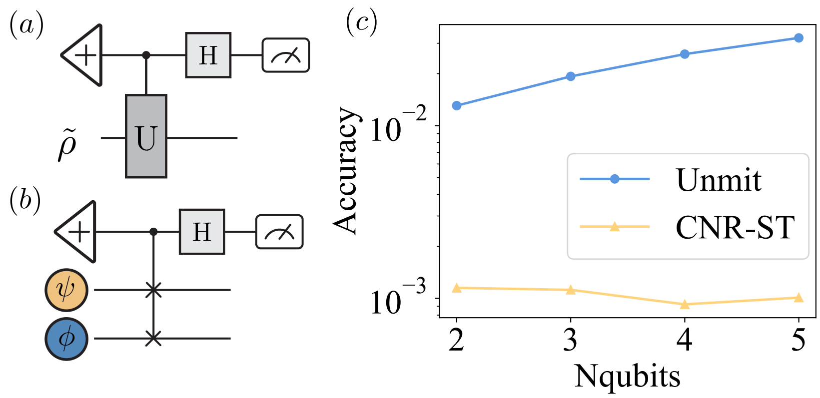

FIG. S6. **Coalesced presentation showcasing the circuit of Hadamard-Test and SWAP-Test and the simulation results pertaining to the SWAP-Test.** (a) The quantum circuit of Hadamard-Test. (b) The illustration of SWAP-Test to estimate the overlap of two quantum states ψ and ϕ (c) The simulation results of SWAP-Test with Unmit and CNR-ST. The displayed data represents the average outcome derived from 50 separate experiments, with the error bars indicating the corresponding standard deviation.

2. *Detailed simulation results for GHZ state*

In the maintext, we compare the performance on five estimators: CNR-VD, ZNE-VD, SD, noisy VD and unmitigated estimation (Unmit) for N-qubit GHZ state and Pauli observable $\sigma_x^{\otimes N}$, with varying qubit number N and total shot number M. The total shot number M contains all the procedures of the estimators. For instance, CNR-VD takes double circuit instances of noisy VD due to the calibration stage. Moreover, as we employ RC for CNR-VD with 4 random instances, CNR-VD takes 16 circuit instances and the measuremeant shot number of each circuit instance is $M/16$. Here, we illustrate the detailed variation curves of the accuracy of these estimators with respect to N and M, as shown in Fig. S5. As discussed in the maintext, the mitigated VD methods including CNR-VD, ZNE-VD and SD introduce additional sampling overhead to mitigate the noise in the VD circuit, leading to poorer results when the total shot number M is low, but CNR-VD and ZNE-VD are still superior to unmitigated estimations and noisy VD.

Supplementary Note 5. Extension for general applications

In order to demonstrate the generality of CNR-VD, we perform numerical simulations on a representative application of the Hadamard-Test, namely the SWAP-Test [S10, S11]. SWAP-Test is a fundamental protocol in quantum computing. By swapping the two states via performing CSWAP operators, SWAP-Test enables to estimate the state overlap which provides valuable insights into quantum state comparison and discrimination. We conducted numerical experiments utilizing the SWAP-Test on two randomly generated quantum states ψ and ϕ as illustrated in Fig. S6(b). We call the SWAP-Test using the CNR estimator as CNR-ST, and the input state s is $|0\rangle^{\otimes N}$ The accuary are still calculated using the absolute error.

To alleviate the impact of the extreme values of the ideal overlap on the estimation of absolute error, we generated the pair of random quantum states with a fixed overlap

$$|\psi\rangle = U_{\mathrm{R}}|0\rangle^{\otimes N}, |\phi\rangle = U_{\mathrm{R}}|+\rangle^{\otimes N}, \tag{S39}$$

where the random unitary operator U_{R} can maintain the inner product of ψ and ϕ as $\langle\psi|\phi\rangle = \langle 0|^{\otimes N} U_{\mathrm{R}}^{\dagger} U_{\mathrm{R}} |+\rangle^{\otimes N} = \langle 0|^{\otimes N} |+\rangle^{\otimes N} = \frac{1}{2}$. We conduct simulations on estimating the overlap of random states on qubits with number ranging from 2 to 5 and compare the accuracy between the unmitigated SWAP-Test (Unmit) and CNR-ST. The results are given in Fig. S6(c). This two approaches are implemented with the same number of total meaurement shots 10^5. There is clearly an improvement in the accuracy of the estimated overlap with an order of magnitude, and the accuracy does not decrease significantly with an increasing number of qubits as unmitigated SWAP-Test does.

[S1] S. Endo, S. C. Benjamin, and Y. Li, Phys. Rev. X **8**, 031027 (2018).

[S2] B. Koczor, Phys. Rev. X **11**, 031057 (2021).

[S3] T. C. Developers, "Cirq: A python framework for creating, editing, and invoking quantum circuits," `https://github.com/quantumlib/Cirq` (Accessed: 2023-10-12).

[S4] S. V. Isakov, D. Kafri, O. Martin, C. V. Heidweiller, W. Mruczkiewicz, M. P. Harrigan, N. C. Rubin, R. Thomson, M. Broughton, K. Kissell, E. Peters, E. Gustafson, A. C. Y. Li, H. Lamm, G. Perdue, A. K. Ho, D. Strain, and S. Boixo, "Simulations of quantum circuits with approximate noise using qsim and cirq," (2021), arXiv:2111.02396.

[S5] Q. Zhu, S. Cao, F. Chen, M.-C. Chen, X. Chen, T.-H. Chung, H. Deng, Y. Du, D. Fan, M. Gong, C. Guo, C. Guo, S. Guo, L. Han, L. Hong, H.-L. Huang, Y.-H. Huo, L. Li, N. Li, S. Li, Y. Li, F. Liang, C. Lin, J. Lin, H. Qian, D. Qiao, H. Rong, H. Su, L. Sun, L. Wang, S. Wang, D. Wu, Y. Wu, Y. Xu, K. Yan, W. Yang, Y. Yang, Y. Ye, J. Yin, C. Ying, J. Yu, C. Zha, C. Zhang, H. Zhang, K. Zhang, Y. Zhang, H. Zhao, Y. Zhao, L. Zhou, C.-Y. Lu, C.-Z. Peng, X. Zhu, and J.-W. Pan, Sci. Bull. **67**, 240 (2022).

[S6] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, M. Ho, Alanand Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, J. R. Mandrà, Salvatoreand McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, Nature **574**, 505 (2019).

[S7] C. Ding, X.-Y. Xu, Y.-F. Niu, S. Zhang, W.-S. Bao, and H.-L. Huang, Sci. China Phys. Mech. Astron. **66**, 230311 (2023).

[S8] B. Nachman, M. Urbanek, W. A. de Jong, and C. W. Bauer, Npj Quantum Inf. **6**, 84 (2020).

[S9] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, Phys. Rev. A **103**, 042605 (2021).

[S10] M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti, Phys. Rev. Lett. **124**, 060503 (2020).

[S11] J. C. Garcia-Escartin and P. Chamorro-Posada, Phys. Rev. A **87**, 052330 (2013).