Quantum error suppression with subgroup stabilisation projectors

Quantum Error Suppression with Subgroup Stabilisation Projectors

Bo Yang,[1,*] Elham Kashefi,[1,2,†] Dominik Leichtle,[1,‡] and Harold Ollivier[3,§]

[1] *Laboratorie d'Informatique de Paris 6, Centre National de la Recherche Scientifique, Sorbonne Université, 4 place Jussieu, 75005 Paris, France*

[2] *School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom*

[3] *Institut National de Recherche en Informatique et en Automatique, 2 Rue Simone Iff, 75012 Paris, France*

Quantum state purification is the functionality that, given multiple copies of an unknown state, outputs a state with increased purity. This is an essential building block for near- and middle-term quantum ecosystems before the availability of full fault tolerance, where one may want to suppress errors not only in expectation values but also in quantum states. We propose an effective state purification gadget with a moderate quantum overhead by projecting M noisy quantum inputs to their symmetric subspace defined by a set of projectors forming a symmetric subgroup with order M. Our method, applied in every short evolution over M redundant copies of noisy states, can suppress both coherent and stochastic errors by a factor of $1/M$. This reduces the circuit implementation cost M times smaller than the state projection to the full symmetric subspace proposed more than two decades ago by Barenco et al. We also show that our gadget purifies the depolarised inputs with probability p to asymptotically $O\left(p^2\right)$ with an optimal choice of M when p is small. Our method provides flexible choices of state purification depending on the hardware restrictions before fully fault-tolerant computing is available. Our method may also find its application in designing robust verification protocols for quantum outputs.

I. INTRODUCTION

The practical execution of quantum algorithms with notable advantages over classical computers [1–4] may require fault-tolerant quantum computation (FTQC), which can be achieved with quantum error correction (QEC) codes by encoding quantum information into redundant code space to detect and correct errors during the computation [5–10]. The FTQC procedure, including logical gate operations and error correction, is performed with $O\left(\text{poly}\log\left(N\right)\right)$ times of overhead to the polynomial time computation with problem size N and under the bounded failure probability in each physical gate operation. Although recent advances in quantum hardware have reached a juncture where the break-even fidelity of logical qubits encoded by certain QEC codes can be witnessed [11, 12], the realisation of practical FTQC remains elusive as the size of current physical devices is still small and physical operations on them are still too noisy.

Instead, quantum error mitigation (QEM) [13–18] attached to hybrid quantum-classical approaches [19–24] has become mainstream for harnessing the abilities of contemporary and near-term quantum devices. However, since the QEM methods restore noiseless measurement results (i.e. expectation values) through classical pre- and postprocessing, they are targeted for BQP computation, while the suppression of errors in the output of quantum computations for sampling problems may require quantum pre- and postprocessing. Besides, a wide class of QEM methods come with an exponential sampling cost in the desired estimation bias or the circuit depth [25–28], revealing their limitations in scalability.

The aforementioned issues highlight the demand for suppressing errors in quantum states, particularly to address sampling problems. In this line, the state purification methods by post-selection find a unique position as being capable of improving the quality of the quantum state before it is sampled. This scheme suppresses errors for quantum outputs consuming quantum resources in accord with QEM, which is advantageous for quantum computers with small scales before the arrival of full FTQC. The early works can be found over two decades ago [29–32], sharing the idea of taking noisy quantum states and returning a purified quantum state by state projection and post-selection.

Barenco et al. [32] proposed the stabilisation of multiple inputs to their symmetric subspace by the projectors forming a full symmetric group. The simplest example is the SWAP test circuit, which maps two inputs into their SWAP invariant subspace. They suppose applying this gadget many times per a short time interval over M redundant states each under the same computational task. Projecting M states to their permutation invariant subspace, this

* Bo.Yang@lip6.fr
† Elham.Kashefi@lip6.fr
‡ dominik.leichtle@lip6.fr
§ harold.ollivier@inria.fr

gadget reduces errors by a factor of $1/M$ in terms of the probability of the purified output being unwanted states. Particularly they demonstrated it for noisy inputs under either local coherent errors or first-order perturbed stochastic errors.

The perfect execution of the gadget by Barenco et al. can effectively stabilise errors growing exponentially to evolution time, such as decoherence. Supposing we execute a polynomial time computational task which gives the correct result at each step with probability $1-\epsilon$ with a constant ϵ, the success probability is at least $(1-\epsilon)^N \sim e^{-\epsilon N}$ after N steps. The gadget of Barenco et al. aims to amplify this into $e^{-\epsilon N/M}$ by stabilising M copies at each step. Thus the success probability can be kept at the level $1-\delta$ by taking $M = -\epsilon N/\log(1-\delta)$ which is polynomial in N.

However, the quantum circuit for this gadget requires $O(M\log(M))$ measurements and $O\left(M^2\log(M)\right)$ controlled-SWAP operations, which may not yet be feasible for applying it many times during the computation on the near- and middle-term devices. In this work, we propose a lighter implementation for suppressing errors by projecting M redundant states to their rotation invariant subspace using a linear combination of projectors forming the cyclic group C_M, which we name "Cyclic Group Gadget (CGG)". The implementation cost of our gadget is reduced to $O(\log(M))$ measurements and $O(M\log(M))$ controlled-SWAP operations since the stabilisation condition of rotation invariance is looser than that of permutation invariance by the symmetric group S_M. Yet our gadget still keeps the same purification rate as Barenco et al. when stabilising the errors in a short evolution time. We also provide another implementation projecting M inputs to a symmetric subspace associated with the group $(\mathbb{Z}/2\mathbb{Z})^{\log(M)}$, which we name "Generalised SWAP Gadget (GSG)". The group $(\mathbb{Z}/2\mathbb{Z})^{\log(M)}$ has the same order as the cyclic group C_M, and the circuit implementation cost and purification rate of GSG is the same as CGG.

Our gadget may also be used in a single round, assuming a fixed level of errors in inputs. The purification of single-qubit states is well studied by Cirac, Ekert, and Macchiavello [30] who found an optimal purification procedure under the depolarising noise. Keyl and Werner [31] also analysed the fidelity of purified states under different noise assumptions for single-qubit states. Recently, Childs et al. [33] have proved the optimality, in terms of sampling cost, of the recursive application of SWAP gadgets for purifying the depolarised states with higher dimensions. We also analyse the error suppression rate of single-round application of our method over M depolarised inputs with the depolarising rate p. We observe that there is an optimal number of copies M depending on p for both CGG and GSG that maximises the error suppression rate, which we have analysed both analytically and numerically. When p is small, the maximal purification rate can be approximated as $O\left(p^2\right)$ with $O\left(p^{-1}\right)$ sampling cost, showing the same order to p in purification rate and sampling cost as the recursive application of SWAP gadgets by Childs et al. [33].

Through the construction and analysis of our proposed method, we also comprehensively discuss the implementation overhead for purifying quantum states with multiple redundant noisy inputs. Besides the number of gates, we remark that our method is efficient in the number of measurements for the post-selection. While the projection to the full symmetric subspace uses $O(M\log(M))$ measurements and the recursive SWAP gadget uses $M-1$ measurements, our method requires only $O(\log(M))$. Therefore, our gadget would be more feasible on the hardware where the measurement error is the dominant noise factor.

The contents of this paper are organised as follows. First in Section II, we review the stabilisation by S_M [32] and recursive SWAP gadgets [33]. Based on [32], we introduce our proposed error suppression method in Section III, the cyclic group gadgets (CGG) and the generalised SWAP gadgets (GSG). Then, we analyse the purification rate of our proposed methods when applying them in a short time evolution under general noise (Section IV), and when taking the depolarised states (Section V). Finally in Section VI, we compare the implementation cost among different variations of state purification methods and discuss which gadget may fit better in which hardware and regime.

II. TWO STRATEGIES FOR STATE PURIFICATION

In this section, we review two different variations of purification gadgets: the recursive application of SWAP gadgets by Childs et al. [33] and the state projection with the symmetric group by Barenco et al. [32]. Both approaches aim to obtain a purified output from multiple copies of noisy inputs by projecting them to symmetric subspaces. The difference is that Childs et al. [33] minimises the required quantum memories while Barenco et al. [32] exploits the full symmetric group over multiple copies.

A. SWAP Gadget and Its Recursive Application

It is widely known that the outcome state of the SWAP test improves the purity of input quantum states by projecting them onto their symmetric subspace. We refer to this purification procedure as a SWAP gadget. Given two inputs ρ and σ, the SWAP gadget is described as the quantum circuit in Fig. 1, where the purified state is denoted

by $\tilde{\rho}$. After the controlled-SWAP operation to the whole system, the post-selection on $|0\rangle\langle 0|$ state in the ancillary qubit projects the input applies the projector

$$\frac{1}{2}(P_{12}+P_{21})=\frac{1}{2}(I+\mathrm{SWAP}) \tag{1}$$

to the input state $\rho\otimes\sigma$, where $P_{i_1\cdots i_M}$ is a permutation operation of quantum registers that maps the position k to i_k. Let $\mathrm{P}_{\vec{0}}$ denote the probability of post-selecting $|0\rangle\langle 0|$ state in the ancillary qubit. Ignoring the second register of

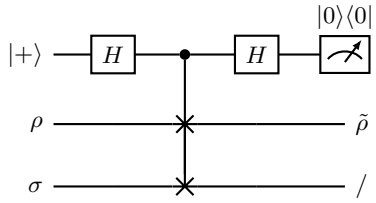


FIG. 1. The quantum circuit for the SWAP gadget.

the output in Fig. 1 yields the quantum state $\tilde{\rho}$ in the first register as

$$\begin{aligned}
\tilde{\rho}&=\frac{1}{4\,\mathrm{P}_{\vec{0}}}\,\mathrm{Tr}_2\left[(P_{12}+P_{21})\,(\rho\otimes\sigma)\,(P_{12}+P_{21})\right]\\
&=\frac{1}{4\,\mathrm{P}_{\vec{0}}}\,\mathrm{Tr}_2\left[\rho\otimes\sigma+(\rho\otimes\sigma)\,P_{21}+P_{21}\,(\rho\otimes\sigma)+\sigma\otimes\rho\right]\\
&=\frac{1}{4\,\mathrm{P}_{\vec{0}}}\left(\rho+\mathrm{Tr}_2\left[(\rho\otimes\sigma)\,P_{21}\right]+\mathrm{Tr}_2\left[(\rho\otimes\sigma)\,P_{21}\right]+\sigma\right)\\
&=\frac{1}{4\,\mathrm{P}_{\vec{0}}}\left(\rho+\rho\sigma+\sigma\rho+\sigma\right),\\
\mathrm{P}_{\vec{0}}&=\mathrm{Tr}\left[\frac{1}{4}\left(\rho+\rho\sigma+\sigma\rho+\sigma\right)\right]=\frac{1}{2}\left(1+\mathrm{Tr}\left[\rho\sigma\right]\right).
\end{aligned} \tag{2}$$

To obtain the last line for $\tilde{\rho}$, we use the identity $\mathrm{Tr}_2\left[(\rho\otimes\sigma)\,P_{21}\right]=\rho\sigma$, which can be verified as follows, assuming $\rho=\sum_i \lambda_i|\lambda_i\rangle\langle\lambda_i|$ and $\sigma=\sum_j \lambda_j^{'}|\lambda_j^{'}\rangle\langle\lambda_j^{'}|$.

$$\begin{aligned}
\mathrm{Tr}_2\left[(\rho\otimes\sigma)\,P_{21}\right]&=\mathrm{Tr}_2\left[\left(\sum_i \lambda_i\lambda_j^{'}|\lambda_i\rangle\langle\lambda_i|\otimes|\lambda_j^{'}\rangle\langle\lambda_j^{'}|\right)P_{21}\right]=\mathrm{Tr}_2\left[\sum_i \lambda_i\lambda_j^{'}|\lambda_i\rangle\langle\lambda_j^{'}|\otimes|\lambda_j^{'}\rangle\langle\lambda_i|\right]\\
&=\sum_i \lambda_i\lambda_j^{'}|\lambda_i\rangle\langle\lambda_j^{'}|\,\mathrm{Tr}_2\left[|\lambda_j^{'}\rangle\langle\lambda_i|\right]=\sum_i \lambda_i\lambda_j^{'}|\lambda_i\rangle\langle\lambda_j^{'}|\langle\lambda_i|\lambda_j^{'}\rangle=\sum_i \lambda_i\lambda_j^{'}|\lambda_i\rangle\langle\lambda_i|\lambda_j^{'}\rangle\langle\lambda_j^{'}|\\
&=\left(\sum_i \lambda_i|\lambda_i\rangle\langle\lambda_i|\right)\left(\sum_j \lambda_j^{'}|\lambda_j^{'}\rangle\langle\lambda_j^{'}|\right)=\rho\sigma
\end{aligned} \tag{3}$$

This identity can also be easily checked in a tensor network representation as shown in Fig. 2. Following the same deduction as Eq. (2), we also obtain the following rule.

$$\mathcal{E}_{\mathrm{SG}}\left(\rho^k\otimes\rho^l\right)\sim\mathrm{Tr}_2\left[\frac{1}{2}\left(P_{12}+P_{21}\right)\left(\rho^k\otimes\rho^l\right)\frac{1}{2}\left(P_{12}+P_{21}\right)\right]=\frac{1}{4}\left(\mathrm{Tr}_2\left[\rho^l\right]\rho^k+\mathrm{Tr}_2\left[\rho^k\right]\rho^l+2\rho^{k+l}\right) \tag{4}$$

When $\sigma=\rho=\sum_i \lambda_i|\lambda_i\rangle\langle\lambda_i|,\ \ (\lambda_0>\lambda_1>\lambda_2>\cdots)$, the reduced density matrix $\tilde{\rho}$ and the post-selection probability $\mathrm{P}_{\vec{0}}$ can be simplified as

$$\begin{aligned}
\tilde{\rho}&=\frac{1}{2\,\mathrm{P}_{\vec{0}}}\left(\rho+\rho^2\right)=\frac{\rho+\rho^2}{1+\mathrm{Tr}\left[\rho^2\right]}=\sum_i \frac{\lambda_i+\lambda_i^2}{1+\mathrm{Tr}\left[\rho^2\right]}|\lambda_i\rangle\langle\lambda_i|,\\
\mathrm{P}_{\vec{0}}&=\frac{1}{2}\left(1+\mathrm{Tr}\left[\rho^2\right]\right).
\end{aligned} \tag{5}$$

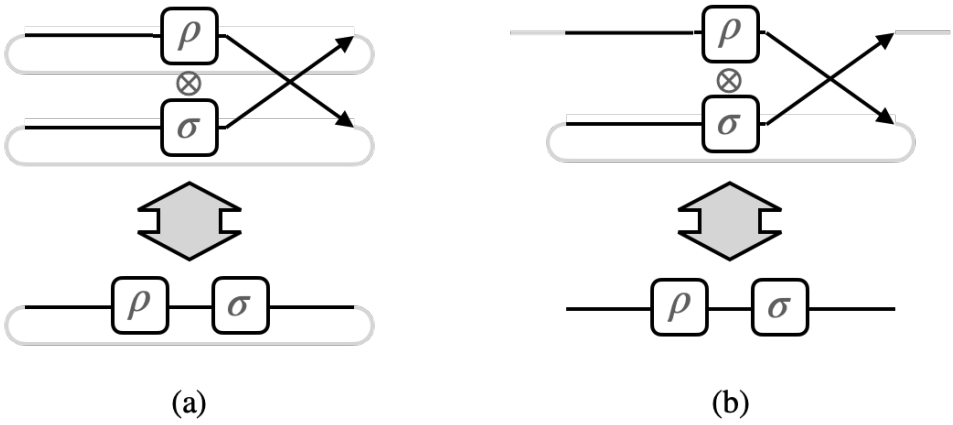


FIG. 2. The tensor network representation of (a) $\mathrm{Tr}\left[(\rho\otimes\sigma)\,P_{21}\right]=\mathrm{Tr}\left[\rho\sigma\right]$ and (b) $\mathrm{Tr}_2\left[(\rho\otimes\sigma)\,P_{21}\right]=\rho\sigma$. Taking the partial trace of a quantum register corresponds to drawing a loop around its tensor.

Since $\{\lambda_i\}_i$ is labelled in descending order, the contribution of larger eigenvalues in ρ is amplified in $\tilde{\rho}$.

On quantum hardware, we can assume the noise due to the system-environment interaction is stochastic noise with a small probability. In this scenario, the target state of interest in the noisy density matrix can be seen as the dominant eigenvector $|\lambda_0\rangle\langle\lambda_0|$ with a dominant probability (i.e. eigenvalue). Hence, according to Eq. (5), the SWAP gadget can "distil" the desired quantum state from two noisy inputs by selecting out unwanted states through post-selection.

The sequential application of the SWAP gadget will effectively purify multiple copies of noisy inputs. Childs et al. [33] show that the cascaded recursive application of the SWAP gadget can protect the quantum advantage of Simon's problem against the depolarising error with a constant depolarising rate occurring in the faulty oracle. We review their approach and purification performance.

Let us introduce the depolarised state ρ with depolarising rate p as

$$\rho=(1-p)\,\rho_0+p\frac{I}{d}, \tag{6}$$

where $\rho_0=|\psi\rangle\langle\psi|$, I is identity matrix, and d is the dimension of ρ, ρ_0 and I. Given two inputs of ρ, the SWAP gadget yields $\tilde{\rho}$ as

$$\begin{aligned}\tilde{\rho}&=\frac{1}{\mathrm{P}_{\vec{0}}}\left(\frac{1}{2}\left((1-p)\rho_0+p\frac{I}{d}\right)+\frac{1}{2}\left((1-p)^2\rho_0^2+\frac{2}{d}p(1-p)\rho_0+\frac{p^2}{d}\frac{I}{d}\right)\right)\\&=\frac{1}{2\,\mathrm{P}_{\vec{0}}}\left((1-p)\left(2-\left(1-\frac{2}{d}\right)p\right)\rho_0+p\left(1+\frac{p}{d}\right)\frac{I}{d}\right),\\\mathrm{P}_{\vec{0}}&=1-\left(1-\frac{1}{d}\right)p+\frac{1}{2}\left(1-\frac{1}{d}\right)p^2.\end{aligned} \tag{7}$$

According to [33], the recursive application of SWAP gadgets using $M=2^n$ inputs results in the finally purified state

$$\begin{aligned}\tilde{\rho}^{(n)}&=(1-\tilde{p})\rho_0+\tilde{p}\frac{I}{d},\\\text{where}\quad\tilde{p}&<\frac{p}{2^n(1-2p)+2p}<\frac{p}{2^n+1},\quad p<\frac{1}{2}.\end{aligned} \tag{8}$$

This means the purification rate scales linearly to the number of copies. For convenience, we refer to this recursive application of the SWAP gadget as "Recursive SWAP Gadget (RSG)". Childs et al. also showed that RSG is optimal in the sampling cost with respect to the expected number of copies required to suppress the depolarisation error in the inputs. To obtain one purified state with the arbitrary depolarising rate ϵ, RSG consumes $O\left(p/\epsilon\right)$ noisy inputs with the depolarising rate p. In terms of implementation cost, obtaining one quantum state through RSG from M noisy copies requires $M-1$ times of measurements and post-selection on $|0\rangle\langle 0|$ and $M-1$ controlled-SWAP gates in total.

B. Full symmetric group Gadget

We can also generalise the state purification by the SWAP gadget into the projection to the full symmetric group over M copies, which dates back to the work by Barenco et al. [32] during the early days of quantum error correction. For convenience, we call their method symmetric group Gadget (SGG). The projector of SGG over M copies includes all possible permutations in the symmetric group S_M. Hence the projector can be represented as

$$P_{S_M} = \frac{1}{M!} \sum_{i=0}^{M!-1} P_{\sigma_i}, \tag{9}$$

where P_{σ_i} is a permutation operator corresponding to the permutation σ_i. Following the procedure in [32], the five steps below complete the projection to the symmetric subspace by Eq. (9).

1. Introduce an ancillary Hilbert space with dimension $\log(M!)$ and initialise it to $|\vec{0}\rangle$.
2. Prepare the uniform superposition state $\frac{1}{\sqrt{M!}} \sum_{i=0}^{M!-1} |i\rangle$. For $M = 2^n$, the superposition is achieved by $H^{\otimes n}$, and for general M, this is achieved by QFT on the ancillary space.
3. Apply the i-th permutation operation P_{σ_i} in the symmetric group based on each state $|i\rangle$ in the ancillary space.
4. Apply the inverse operation of step 2 on the ancillary space.
5. Measure the ancillary space and post-select $|\vec{0}\rangle$ from the measurement result. For the post-selected copies, save the first quantum register by taking the partial trace on all other copies, i.e., throw them away. The purified quantum state $\tilde{\rho}$ is then held by the first register.

The circuit construction of SGG consumes $O(M \log(M))$ ancillary qubits in step 1. In step 2, creating superposition in the ancillary qubits requires at worse $O\left((M \log(M))^2\right)$ gates due to the use of QFT. Finally, in step 3, projecting M inputs to their full symmetric subspace requires $O\left(M^2 \log(M)\right)$ controlled-SWAP gates. In total, SGG requires $O\left((M \log(M))^2 + M^2 \log(M)\right)$ steps of non-local gate operations. The rough sketch of the quantum circuit for SGG is shown in Fig. 3.

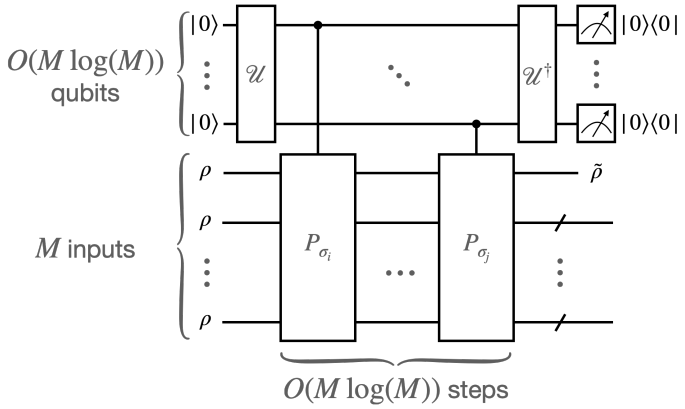


FIG. 3. The quantum circuit implementing SGG with M inputs. The unitary operation $\mathcal{U}$ prepares a uniform superposition state over $\log(M!) = O(M \log(M))$ qubits.

Barenco et al. showed that SGG suppresses the coherent error as the random independent unitary rotation into $1/M$ with M copies. To see this, we assume that the M copies of the input state are subject to independent drift of eigenvalues as $e^{iH_k t}$ on the k-th copy, where the Hamiltonian H_k is random and independent. Hence, the whole quantum state of M noisy inputs is assumed to be

$$|\psi\rangle = \bigotimes_{k=1}^{M} e^{iH_k t}|0\rangle. \tag{10}$$

Besides, the devices were intended for state storage, the rate of drift is suitably bounded as $|$ eigenvalues of $H_k| \leq \epsilon$ for $k = 1, \ldots, M$ with a small constant ϵ. With these assumptions, we can introduce the specific form of Hamiltonian H_k as

$$H_k = \begin{pmatrix} a_k & c_k^* \\ c_k & b_k \end{pmatrix}. \tag{11}$$

Finally, we assume that the purification gadget is applied in a short interval δt and performed instantaneously. This introduces the following quantum state for the approximation of $|\psi\rangle$.

$$|\psi(\delta t)\rangle \sim \bigotimes_{k=1}^{M} (I + iH_k\delta t)|0\rangle = \bigotimes_{k=1}^{M} ((1 + ia_k\delta t)|0\rangle + ic_k\delta t|1\rangle) \tag{12}$$

By keeping only the contribution up to $O(\delta t)$ terms,

$$|\psi(\delta t)\rangle = \left(1 + i\delta t \sum_{k=1}^{M} a_k\right)|0\rangle^{\otimes M} + i\delta t \sum_{k=1}^{M} c_k|0\ldots\underset{k\text{-th}}{010}\ldots0\rangle + O\left((\delta t)^2\right). \tag{13}$$

When applying the projector $P_{\mathrm{CGG}} = \dfrac{1}{M}\sum_{l=1}^{M} P_l$ in CGG to Eq. (13), each term $|0\ldots\underset{k\text{-th}}{010}\ldots0\rangle$ becomes

$$\frac{1}{M}\sum_{l=1}^{M} P_l|0\ldots\underset{k\text{-th}}{010}\ldots0\rangle = \frac{1}{\sqrt{M}}|e_1\rangle, \tag{14}$$

where we define $|e_1\rangle = \dfrac{1}{\sqrt{M}}(|10\ldots0\rangle + |010\ldots0\rangle + \cdots + |0\ldots01\rangle)$. Using $|e_1\rangle$, the whole projected state $P_{\mathrm{SGG}}|\psi(\delta t)\rangle$ becomes

$$P_{\mathrm{SGG}}|\psi(\delta t)\rangle = \left(1 + i\delta t \sum_{k=1}^{M} a_k\right)|0\rangle^{\otimes M} + \frac{i\delta t}{\sqrt{M}}\sum_{k=1}^{M} c_k|e_1\rangle + O\left((\delta t)^2\right). \tag{15}$$

As we have assumed $|$ eigenvalues of $H_k| \leq \epsilon$, we can estimate the scale of c_k by approximating it with $c_k \approx 2\epsilon e^{i\theta_k}$ with a random phase θ_k. This leads to the following estimation

$$\mathbb{E}\left[\left|\frac{i\delta t}{\sqrt{M}}\sum_{k=1}^{M} c_k\right|^2\right] = 4\epsilon^2(\delta t)^2. \tag{16}$$

The probability that the first register becomes $|1\rangle$ is

$${}_{M-1}C_{1-1}\left|\frac{1}{\sqrt{{}_MC_1}}\frac{i\delta t}{\sqrt{M}}\sum_{k=1}^{M} c_k\right|^2 = \frac{4\epsilon^2(\delta t)^2}{M}, \tag{17}$$

where we observe the suppression of coherent deviation linear to the number of copies M.

Barenco et al. also showed this gadget suppresses the level of stochastic error considering only the terms of the first order in the perturbations into $1/M$. Here, we review their performance analysis on these stochastic errors. Consider M copies of noisy inputs initially all prepared in pure state $\rho_0 = |\lambda_0\rangle\langle\lambda_0|$. The stochastic error is introduced as the independent interaction between each state and its environment such that after a short period δt the state of the i-th copy will have undergone an evolution $\rho^{(i)} = \rho_0 + \sigma_i$ for some Hermitian σ_i satisfying $\mathrm{Tr}[\sigma_i] = 0$. Retaining the terms of the first-order in the perturbations $\{\sigma_i\}_{i=1}^{M}$, the overall noisy inputs at time δt becomes

$$\begin{aligned} \rho^{(1\ldots M)} = \rho_0 \otimes \cdots \otimes \rho_0 &+ \sigma_1 \otimes \rho_0 \otimes \cdots \otimes \rho_0 \\ &+ \rho_0 \otimes \sigma_2 \otimes \cdots \otimes \rho_0 \\ &+ \cdots \\ &+ \rho_0 \otimes \rho_0 \otimes \cdots \otimes \sigma_M \\ &+ O(\sigma_i\sigma_j). \end{aligned} \tag{18}$$

Applying the full symmetric group projector P_{S_M} in Eq. (9) to $\rho^{(1\dots M)}$ and tracing out the quantum registers except for the first one, the output quantum state $\tilde{\rho}$ is described as

$$\begin{aligned}\tilde{\rho} = &\left(1-(M-1)\operatorname{Tr}\left[\rho_0\tilde{\sigma}\right]\right)\rho_0 + \frac{1}{M}\tilde{\sigma} \\ &+ (M-1)\left(A\rho_0\tilde{\sigma}\rho_0 + B\left(\rho_0\tilde{\sigma}+\tilde{\sigma}\rho_0\right) + C\rho_0\operatorname{Tr}\left[\rho_0\tilde{\sigma}\right]\right) + O\left(\sigma_i\sigma_j\right),\end{aligned} \tag{19}$$

where $A+2B+C=1$, and $\tilde{\sigma} = \dfrac{1}{M}\displaystyle\sum_{i=1}^{M}\sigma_i$.

The purification quality can be compared by the success probability of obtaining ρ_0 in $\tilde{\rho}$. The average success probability before SGG is

$$\frac{1}{M}\sum_i \operatorname{Tr}\left[\rho_0\left(\rho_0+\sigma_i\right)\right] = 1+\operatorname{Tr}\left[\rho_0\tilde{\sigma}\right], \tag{20}$$

and the average success probability after SGG is

$$\operatorname{Tr}\left[\rho_0\tilde{\rho}\right] = 1+\frac{1}{M}\operatorname{Tr}\left[\rho_0\tilde{\sigma}\right]. \tag{21}$$

Therefore, the probability of obtaining an erroneous state has been reduced by a factor of M. Note that this success probability can also be interpreted as state fidelity.

In addition, the average purity is also enhanced after SGG, as we can check by

$$\frac{1}{M}\sum_{i=1}^{M}\operatorname{Tr}\left[\left(\rho_0+\sigma_i\right)^2\right] = 1+2\operatorname{Tr}\left[\rho_0\tilde{\sigma}\right] \tag{22}$$

for the noisy inputs and

$$\operatorname{Tr}\left[\tilde{\rho}^2\right] = 1+\frac{2}{M}\operatorname{Tr}\left[\rho_0\tilde{\sigma}\right]. \tag{23}$$

for the purified output. Note that we keep ignoring the contribution $O\left(\sigma_i\sigma_j\right)$ during all calculations and $\operatorname{Tr}\left[\rho_0\tilde{\sigma}\right] < 0$.

III. EXPLOITING SYMMETRIC SUBGROUP PROJECTORS

As we have seen in the previous section, projecting multiple noisy states to their symmetric subspace followed by extracting one quantum register can amplify the purity of quantum states. Here we review our proposed state purification gadgets using the projectors by subgroups of the symmetric group, in particular, the cyclic group C_M and the group isomorphic to $(\mathbb{Z}/2\mathbb{Z})^{\log(M)}$.

A. Cyclic Group Gadget

First, we consider the post-selection projector as a linear combination of permutation operators in the cyclic group to amplify the contribution of higher power degrees of the density matrix. The projector is designed by a summation of all the projectors generated by every operator in the cyclic group. Introducing the cyclic group C_M of order M for M copies of inputs, the projected unnormalised quantum state $\mathrm{P}_{\vec{0}}\,\tilde{\rho}$ and the post-selection probability $\mathrm{P}_{\vec{0}}$ are formalised as

$$\begin{aligned}\mathrm{P}_{\vec{0}}\,\tilde{\rho} &= \operatorname{Tr}_{2\dots M}\left[\left(\frac{1}{M}\sum_{P_i\in C_M}P_i\right)\left(\rho^{\otimes M}\right)\left(\frac{1}{M}\sum_{P_i\in C_M}P_i\right)\right] = \frac{1}{M}\sum_{m|M}\varphi(m)\operatorname{Tr}\left[\rho^m\right]^{\frac{M}{m}-1}\rho^m, \\ \mathrm{P}_{\vec{0}} &= \frac{1}{M}\sum_{m|M}\varphi(m)\operatorname{Tr}\left[\rho^m\right]^{\frac{M}{m}},\end{aligned} \tag{24}$$

where $m|M$ means m divides M, and φ is Euler's totient function. Table I aligns the coefficients by φ in the polynomials of ρ of different M. Euler's totient function shows up in Eq. (24) because, for the cyclic group with order m, the number of its generators corresponds to $\varphi(m)$.

		power degree of ρ							
		1	2	3	4	5	6	7	8
M	1	1							
	2	1	1						
	3	1		2					
	4	1	$\mathrm{Tr}\left[\rho^2\right]$		2				
	5	1				4			
	6	1	$\mathrm{Tr}\left[\rho^2\right]^2$	$2\,\mathrm{Tr}\left[\rho^3\right]$			2		
	7	1						6	
	8	1	$\mathrm{Tr}\left[\rho^2\right]^3$		$2\,\mathrm{Tr}\left[\rho^4\right]$				4

TABLE I. The table of coefficients in each term of Eq. (24). Ignoring the factor of trace values of powers of ρ, the integer coefficients correspond to the values of Euler's totient function $\varphi(m)$ until they reach the maximum degree M, i.e. $\varphi(M)$.

When M is a prime number, (24) can be simplified as

$$\begin{aligned} \mathrm{P}_{\vec{0}}\,\tilde{\rho} &= \frac{1}{M}\sum_{m|M}\varphi(m)\,\mathrm{Tr}\left[\rho^m\right]^{\frac{M}{m}-1}\rho^m = \frac{1}{M}\rho + \frac{M-1}{M}\rho^M, \\ \mathrm{P}_{\vec{0}} &= \frac{1}{M} + \frac{M-1}{M}\,\mathrm{Tr}\left[\rho^M\right], \end{aligned} \tag{25}$$

since $\varphi(M) = M-1$ for a prime number M.

When M takes the power of an integer, i.e. $M = \alpha^n$, the output reduced density matrix becomes

$$\mathrm{P}_{\vec{0}}\,\tilde{\rho} = \frac{1}{\alpha^n}\rho + \frac{1}{\alpha^n}\sum_{i=1}^{n}\varphi(\alpha^i)\,\mathrm{Tr}\left[\rho^{\alpha^i}\right]^{\alpha^{n-i}-1}\rho^{\alpha^i} = \frac{1}{\alpha^n}\rho + \frac{1}{\alpha^n}\sum_{i=1}^{n}\alpha^{i-1}\,\mathrm{Tr}\left[\rho^{\alpha^i}\right]^{\alpha^{n-i}-1}\rho^{\alpha^i}, \tag{26}$$

in particular, $\alpha = 2$, the purified state $\tilde{\rho}$ becomes

$$\mathrm{P}_{\vec{0}}\,\tilde{\rho} = \frac{1}{2^n}\rho + \frac{1}{2^n}\sum_{i=1}^{n}\varphi(2^i)\,\mathrm{Tr}\left[\rho^{2^i}\right]^{2^{n-i}-1}\rho^{2^i} = \frac{1}{2^n}\rho + \frac{1}{2^n}\sum_{i=1}^{n}2^{i-1}\,\mathrm{Tr}\left[\rho^{2^i}\right]^{2^{n-i}-1}\rho^{2^i}, \tag{27}$$

where we use $\varphi(\alpha^n) = \alpha^{n-1}$ for two integers α and n. For $n = 3$, i.e. $M = 8$, the purified state in Eq. (27) becomes

$$\mathrm{P}_{\vec{0}}\,\tilde{\rho} = \frac{1}{8}\rho + \frac{1}{8}\sum_{i=1}^{3}2^{i-1}\,\mathrm{Tr}\left[\rho^{2^i}\right]^{2^{3-i}-1}\rho^{2^i} = \frac{1}{8}\left(\rho + \mathrm{Tr}\left[\rho^2\right]^3\rho^2 + 2\,\mathrm{Tr}\left[\rho^4\right]\rho^4 + 4\rho^8\right). \tag{28}$$

The coefficient of each term in Eq. (28) can be visually explained by Fig. 4.

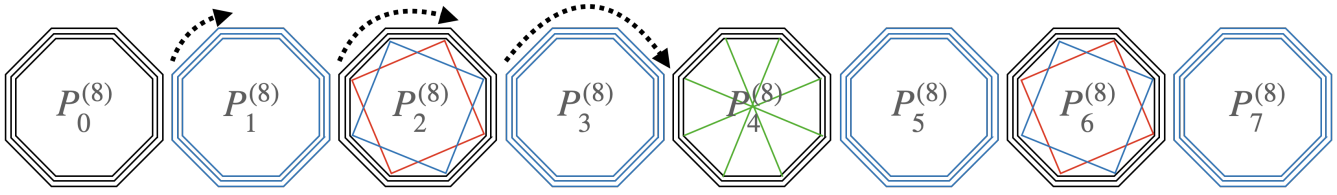


FIG. 4. The illustration of rotation operations in the cyclic group C_8 with order 8. $P_i^{(8)}$ denotes the rotation which maps teach element $k \in C_8$ to $k+i$, and $P_0^{(8)}$ represents the identity operation. In this figure, we observe that $P_1^{(8)}$, $P_3^{(8)}$, $P_5^{(8)}$, and $P_7^{(8)}$ make irreducible transition loops with length 8 since $1, 3, 5, 7$ are coprime to $8 = 2^3$. These operators contribute to the coefficient of term ρ^8 in Eq. (28). The operators $P_2^{(8)}$ and $P_6^{(8)}$ contribute to the coefficient of ρ^4 since the transition of elements by them results in two loops with length 4. Finally, $P_4^{(8)}$ makes four loops with length 2 as $P_4^{(8)}$ can be decomposed to four independent pair-wise SWAP operations.

Let us focus on Eq. (25) again. As Eq. (25) is a weighted sum of ρ and ρ^M, adding M will linearly amplify the contribution of ρ^M in which the ratio of the dominant eigenvector is amplified exponentially. This implies when the desired state ρ_0 is the dominant eigenvector of ρ, we can amplify the contribution of ρ_0 with CGG. However, we have to note that $\mathrm{Tr}\left[\rho^M\right]$ shrinks to 0 also in exponential speed with M when ρ is a mixed state. This implies that there

is an optimal number of copies M^* that maximises the purification rate of CGG and using more copies than M^* will project the copies closer to the initial noisy state. We analyse the optimal M and its purification rate for the depolarised inputs in Section V A.

To perform CGG in the circuit picture, one can adopt a circuit in Fig. 5(a), where $\mathcal{U}$ creates the equal superposition of M different states over $\log(M)$ ancillary qubits. Note that we can reduce the number of ancillary qubits by reusing them. Figure 5(b) shows the controlled-$P_1^{(4)}$ gate, which can be implemented by three (i.e. $4-1$) controlled-SWAP gates. In general, implementing a controlled-$P_i^{(M)}$ gate requires $M-1$ controlled-SWAP gates. Given the equal superposition over $\log(M)$ ancillary qubits, it is also enough to use $\log(M)$ steps of controlled-$P_i^{(M)}$ gates. Therefore, the total controlled-SWAP gates required in the projection is $O(M\log(M))$. Taking $M=8$ as an example, the whole quantum circuit of CGG becomes Fig. 5(c).

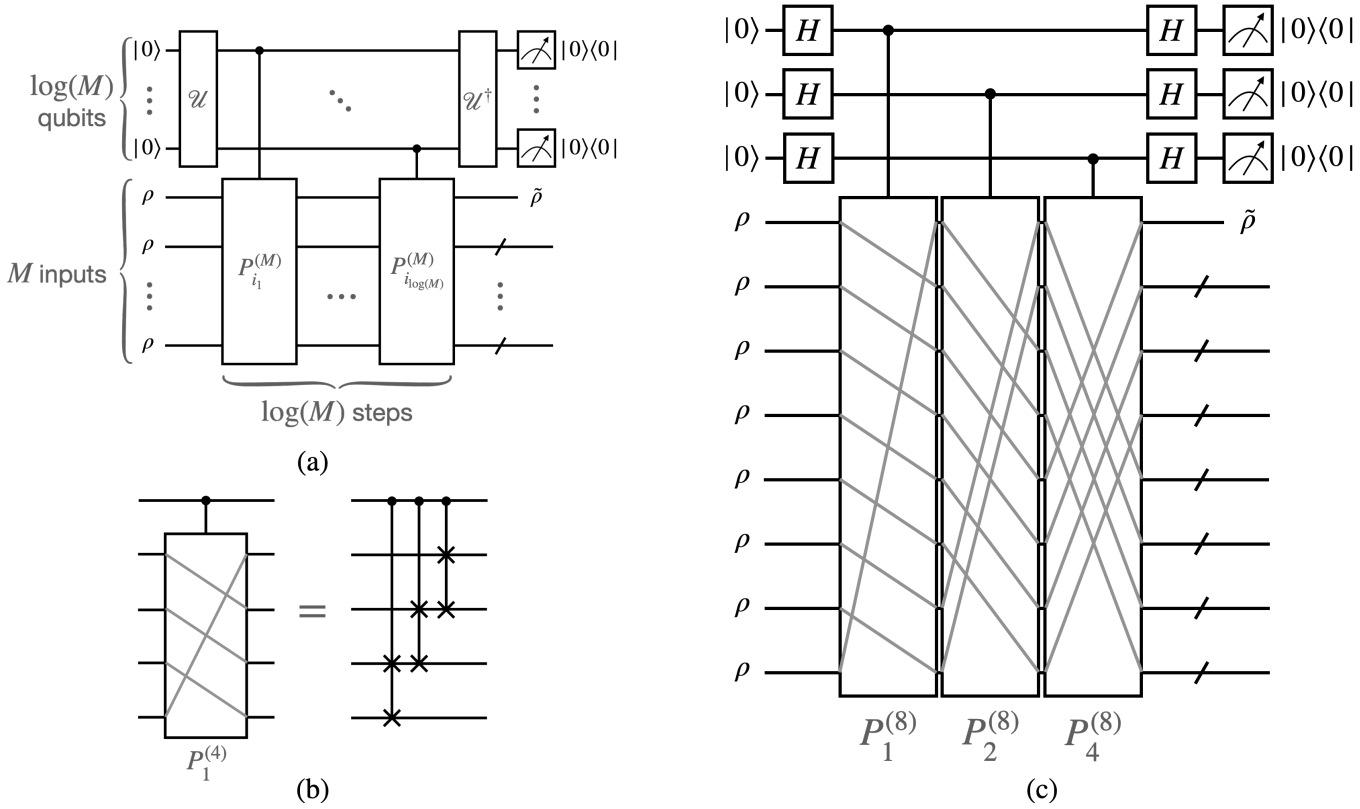


FIG. 5. (a) The quantum circuit implementing CGG taking M inputs in general. (b) The controlled-$P_1^{(4)}$ for $M=4$. The operator $P_1^{(4)}$ represents the cyclic shift to the next element, which can be decomposed into a sequential application of controlled-SWAP gates. The grey lines connect the original place and the destination of the quantum state moved by the operation. (c) The quantum circuit of CGG for $M=8$.

The overhead of preparing equal superposition on the ancillary qubits is exponentially smaller than the state projection part, which becomes more efficient than Barenco et al. [32]. For a general M, the equal superposition can also be prepared with QFT consuming $O\left((\log(M))^2\right)$ gate operations. Therefore, the circuit implementation of CGG has less quantum overhead than that of [32], requiring only $O(\log(M))$ ancillary qubits and $O(M\log(M))$ controlled-SWAP operations. Note the equal superposition on the ancillary qubits can be prepared with only a single step of local Hadamard gates when $M=2^n$.

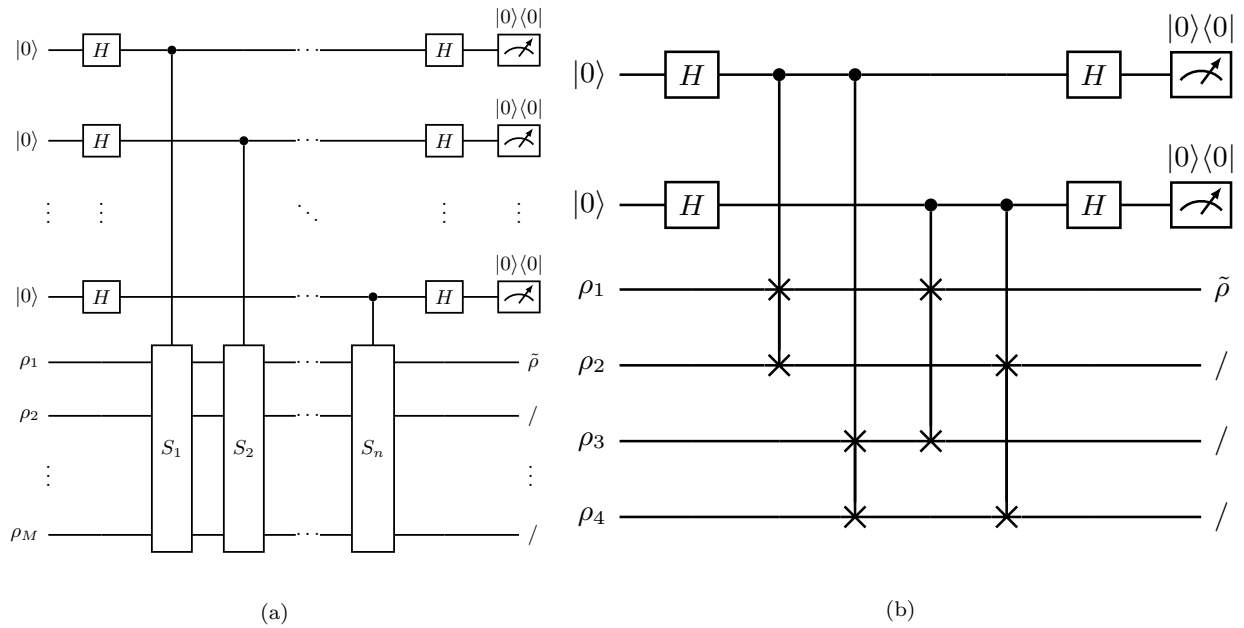
B. Generalised SWAP Gadget

Next, we propose our next purification gadget inspired by the generalised SWAP test proposed by Chabaud et al. [34]. The generalised SWAP test is to approximate any projective measurement optimally to a given one-sided error using M state inputs. The optimality is achieved by projecting an unknown quantum state σ with $M-1$ reference states ρ to their pairwise SWAP symmetric subspace using parallel SWAP operations. When $M=2^n$, the

parallel SWAP operations are defined for $k = 1, 2, \ldots, n$ as

$$S_k = \prod_{i\in[1,2^{k-1}]} \prod_{j\in[1,2^{n-k}]} \left((2j-2)2^{k-1}+i, (2j-1)2^{k-1}+i\right) \tag{29}$$

where the ordered pair (i, j) denotes the SWAP operation between symbols i and j. Note that the set $\{S_k\}_k$ is an Abelian group and forms a group isomorphic to $(\mathbb{Z}/2\mathbb{Z})^n$ as mentioned in [34], and the swapping operations can also be seen as the butterfly diagram appearing in the fast Fourier transformation (FFT).



(a) (b)

FIG. 6. (a) The quantum circuit of generalised SWAP gadget (GSG) for $M = 2^n$ copies. The difference between GSG and generalised SWAP gadgets is whether to keep the first quantum register or not. (b) The quantum circuit of GSG for $M = 2^n = 4, (n = 2)$. S_1, i.e. $k = 1$, has $i \in [1, 1]$, $j \in [1, 2]$, containing SWAP operations between $(1, 2)$ and $(3, 4)$. S_2, i.e. $k = 2$, has $i \in [1, 2]$, $j \in [1, 1]$, containing SWAP operations between $(1, 3)$ and $(2, 4)$.

Chabaud et al. [34] showed the generalised SWAP test optimally amplifies the state overlap in the post-selection probability by implementing all the SWAP operations in $\{S_k\}_k$ sequentially controlled by each superposition in the ancillary qubits shown in Fig. 6(a). The specific example for $M = 4$ is shown in Fig. 6(b), which implements the swapping operations in Klein's four-group. The gate operation of S_k is represented as

$$S_k = \bigotimes_{i\in[1,2^{k-1}]} \bigotimes_{j\in[1,2^{n-k}]} \text{SWAP}\left[(2j-2)2^{k-1}+i, (2j-1)2^{k-1}+i\right]. \tag{30}$$

The circuit implementation of [34] offers the same gate and qubit overhead as CGG, requiring $O(\log(M))$ ancillary qubits and $O(M \log(M))$ controlled-SWAP operations.

Given two pure states $|\phi\rangle$ as an unknown state and $M-1$ copies of $|\psi\rangle$ as the reference state, the projective measurement is calculated through the post-selection probability $\mathrm{P}_{\vec{0}}$ of $|\vec{0}\rangle\langle\vec{0}|$ in the generalised SWAP test

$$\mathrm{P}_{\vec{0}} = \frac{1}{M} + \frac{M-1}{M}|\langle\phi|\psi\rangle|^2. \tag{31}$$

Therefore, the generalised SWAP test with M inputs offers a binary test that outputs 0 with probability $\frac{1}{M} + \frac{M-1}{M}|\langle\phi \mid \psi\rangle|^2$ and 1 with probability $\left(\frac{M-1}{M}\right)\left(1 - |\langle\phi \mid \psi\rangle|^2\right)$. If the outcome 0 (resp. 1) is obtained, the test concludes that the states $|\phi\rangle$ and $|\psi\rangle$ were identical (resp. different).

This generalised SWAP test is also suitable for implementation on photonic devices using Hadamard interferometers, as discussed in [34]. When M is not the power of two, it is also possible to construct the quantum circuit by replacing

the $H^{\otimes n}$ part in Fig. 6(a) with the (inverse) quantum Fourier transform (QFT) to prepare the M uniform superposition in the ancillary qubits. This is also the case for the optical implementation.

Now, let us consider purifying the multiple copies of the input state using the generalised SWAP gadget. If all the inputs are identical in Fig 6(a), i.e. $\rho_k = \rho$ for all $k = 1, \ldots, M$, the output state $\tilde{\rho}$ in the first quantum register can be described as

$$\begin{aligned} \tilde{\rho} &= \frac{1}{\mathrm{P}_{\vec{0}}} \left(\frac{1}{M}\rho + \frac{M-1}{M} \mathrm{Tr}\left[\rho^2\right]^{\frac{M}{2}-1} \rho^2 \right), \\ \mathrm{P}_{\vec{0}} &= \frac{1}{M} + \frac{M-1}{M} \mathrm{Tr}\left[\rho^2\right]^{\frac{M}{2}}. \end{aligned} \tag{32}$$

Compared to the output state of CGG shown in Eq. (24), the eigenvectors in the output state of GSG are less concentrated in the dominant one.

IV. PURIFYING INPUTS UNDER SMALL COHERENT OR STOCHASTIC ERRORS

We first consider applying our gadgets over M redundant copies many times per every small time evolution δt. The noise assumption can be described in the same way as that introduced in Barenco et al. [32]. Particularly, we analyse the noisy inputs either under the coherent noise drifted by bounded local Hamiltonian or stochastic noise generated by the system-environment interaction.

A. Coherent errors in a short interval

We introduce several assumptions on the coherent errors in input states as assumed in [32]. Suppose that the M copies of the input state are subject to independent drift of eigenvalues as $e^{iH_k t}$ on the k-th copy, where the Hamiltonian H_k is random and independent. Hence, the whole quantum state of M noisy inputs is assumed to be

$$|\psi\rangle = \bigotimes_{k=1}^{M} e^{iH_k \delta t} |\lambda_0\rangle \tag{33}$$

We can introduce the specific form of Hamiltonian H_k as $d \times d$ matrix

$$H_k = \begin{pmatrix} a_k & \vec{c}_k^{\dagger} \\ \vec{c}_k & B_k \end{pmatrix}, \tag{34}$$

assuming the dimension of each quantum state in each quantum register is d. Finally, the purification gadget is applied in a short interval δt and performed instantaneously. This introduces the following quantum state for the approximation of $|\psi\rangle$.

$$|\psi(\delta t)\rangle \sim \bigotimes_{k=1}^{M} (I + iH_k \delta t) |\lambda_0\rangle = \bigotimes_{k=1}^{M} \left((1 + ia_k \delta t) |\lambda_0\rangle + i\delta t |\lambda_{\perp,k}\rangle \right), \tag{35}$$

where $|\lambda_{\perp,k}\rangle = \sum_{j>0} c_{kj} |\lambda_j\rangle$. By keeping only the contribution up to $O(\delta t)$ terms,

$$|\psi(\delta t)\rangle = \left(1 + i\delta t \sum_{k=1}^{M} a_k\right) |\lambda_0\rangle^{\otimes M} + i\delta t \sum_{k=1}^{M} |\lambda_0\rangle \cdots \underset{k\text{-th}}{|\lambda_{\perp,k}\rangle} \cdots |\lambda_0\rangle + O\left((\delta t)^2\right). \tag{36}$$

When applying the projector $P_{\mathrm{CGG}} = \frac{1}{M} \sum_{l=1}^{M} P_l$ in CGG to Eq. (36), each term $|\lambda_0\rangle \cdots \underset{k\text{-th}}{|\lambda_{\perp,k}\rangle} \cdots |\lambda_0\rangle$ in Eq. (36) becomes

$$\frac{1}{M} \sum_{l=1}^{M} P_l |\lambda_0\rangle \cdots \underset{k\text{-th}}{|\lambda_{\perp,k}\rangle} \cdots |\lambda_0\rangle = \frac{1}{\sqrt{M}} |E_{1,k}\rangle, \tag{37}$$

where we define $|E_{1,k}\rangle$ as

$$|E_{1,k}\rangle = \frac{1}{\sqrt{M}} \left(|\lambda_{\perp,k}\rangle|\lambda_0\rangle \cdots |\lambda_0\rangle + |\lambda_0\rangle|\lambda_{\perp,k}\rangle \cdots |\lambda_0\rangle + \cdots + |\lambda_0\rangle \cdots |\lambda_0\rangle|\lambda_{\perp,k}\rangle \right). \tag{38}$$

Using $|E_{1,k}\rangle$, the whole projected state $P_{\mathrm{CGG}}|\psi(\delta t)\rangle$ becomes

$$P_{\mathrm{CGG}}|\psi(\delta t)\rangle = \left(1 + i\delta t \sum_{k=1}^{M} a_k\right) |\lambda_0\rangle^{\otimes M} + \frac{i\delta t}{\sqrt{M}} \sum_{k=1}^{M} |E_{1,k}\rangle + O\left((\delta t)^2\right). \tag{39}$$

Here we compute the unnormalised probability of quantum states other than $|\lambda_0\rangle$ from $P_{\mathrm{CGG}}|\psi\rangle$ being in the first quantum register after CGG.

$$\begin{aligned}
&\left\| \left((I - |\lambda_0\rangle\langle\lambda_0|) \otimes I^{\otimes(M-1)} \right) P_{\mathrm{CGG}}|\psi\rangle \right\|_2^2 \\
&= \left\| \left((I - |\lambda_0\rangle\langle\lambda_0|) \otimes I^{\otimes(M-1)} \right) \frac{i\delta t}{\sqrt{M}} \sum_{k=1}^{M} \sum_{l=1}^{M} |\lambda_0\rangle \cdots |\lambda_0\rangle \underset{l\text{-th}}{|\lambda_{\perp,k}\rangle} |\lambda_0\rangle \cdots |\lambda_0\rangle \right\|_2^2 \\
&= \left(\frac{\delta t}{M}\right)^2 \left\| \sum_{k=1}^{M} \left(\sum_{l=1}^{M} |\lambda_0\rangle \cdots \underset{l\text{-th}}{|\lambda_0\rangle|\lambda_{\perp,k}\rangle} |\lambda_0\rangle \cdots |\lambda_0\rangle - \sum_{l=2}^{M} |\lambda_0\rangle \cdots \underset{l\text{-th}}{|\lambda_0\rangle|\lambda_{\perp,k}\rangle} |\lambda_0\rangle \cdots |\lambda_0\rangle \right) \right\|_2^2 \\
&= \left(\frac{\delta t}{M}\right)^2 \left\| \sum_{k=1}^{M} |\lambda_{\perp,k}\rangle \right\|^2 = \frac{(\delta t)^2}{M} \left\| \frac{1}{\sqrt{M}} \sum_{k=1}^{M} |\lambda_{\perp,k}\rangle \right\|_2^2 = \frac{(\delta t)^2}{M} \langle \tilde{\lambda}_\perp | \tilde{\lambda}_\perp \rangle.
\end{aligned} \tag{40}$$

where $|\tilde{\lambda}_\perp\rangle$ is defined as the average deviated state $|\tilde{\lambda}_\perp\rangle = \frac{1}{\sqrt{M}} \sum_{k=1}^{M} |\lambda_{\perp,k}\rangle$.

On the other hand, the unnormalised probability of quantum states other than $|\lambda_0\rangle$ from $P_{\mathrm{CGG}}|\psi\rangle$ being in the first quantum register without the subspace projection by CGG is

$$\begin{aligned}
&\left\| \left((I - |\lambda_0\rangle\langle\lambda_0|) \otimes I^{\otimes(M-1)} \right) |\psi\rangle \right\|_2^2 \\
&= (\delta t)^2 \left\| \left((I - |\lambda_0\rangle\langle\lambda_0|) \otimes I^{\otimes(M-1)} \right) \sum_{k=1}^{M} |\lambda_0\rangle \cdots \underset{k\text{-th}}{|\lambda_{\perp,k}\rangle} |\lambda_0\rangle \cdots |\lambda_0\rangle \right\|_2^2 \\
&= (\delta t)^2 \left\| \sum_{k=1}^{M} |\lambda_0\rangle \cdots \underset{k\text{-th}}{|\lambda_{\perp,k}\rangle} |\lambda_0\rangle \cdots |\lambda_0\rangle - \sum_{k=2}^{M} |\lambda_0\rangle \cdots \underset{k\text{-th}}{|\lambda_{\perp,k}\rangle} |\lambda_0\rangle \cdots |\lambda_0\rangle \right\|_2^2 \\
&= (\delta t)^2 \langle \lambda_{\perp,1} | \lambda_{\perp,1} \rangle.
\end{aligned} \tag{41}$$

When the devices are intended for state storage, the rate of drift is suitably bounded as |eigenvalues of $H_k| \leq \epsilon$ for $k = 1, \ldots, M$ with a small constant ϵ. With this assumption, both $\langle \lambda_{\perp,1} | \lambda_{\perp,1} \rangle$ and $\langle \tilde{\lambda}_\perp | \tilde{\lambda}_\perp \rangle$ can be bounded by the Frobenius norm $\|H_k\|_F \leq d\epsilon^2$ and $\|\tilde{H}\|_F = \left\| \frac{1}{M} \sum_{k=1}^{M} H_k \right\|_F \leq d\epsilon^2$ as

$$\begin{aligned}
\langle \lambda_{\perp,1} | \lambda_{\perp,1} \rangle &= \|\vec{c}_1\|_2^2 \leq \|H_k\|_F \leq d\epsilon^2, \\
\langle \tilde{\lambda}_\perp | \tilde{\lambda}_\perp \rangle &= \|\tilde{c}\|_2^2 \leq \|\tilde{H}\|_F \leq d\epsilon^2,
\end{aligned} \tag{42}$$

where $\tilde{c}$ is the average vector defined by $\tilde{c} = \frac{1}{M} \sum_{k=1}^{M} \vec{c}_k$. Therefore, we observe that the probability of obtaining deviated states after CGG from $|\lambda_0\rangle$ is suppressed into $1/M$ of that of the noisy state without CGG.

B. Stochastic errors up to the first-order perturbation

Recall the noise setting introduced in Section II B, the stochastic error is introduced as the independent interaction between each state and its environment such that after a short period of time δt, the state of the i-th copy will have

undergone an evolution $\rho^{(i)} = \rho_0 + \sigma_i$ for some Hermitian σ_i satisfying $\mathrm{Tr}\left[\sigma_i\right] = 0$. Retaining only the terms of the first-order in the perturbations $\{\sigma_i\}_{i=1}^{M}$, the overall noisy inputs at time δt becomes

$$\begin{aligned}\rho^{(1\ldots M)} = \rho_0 \otimes \cdots \otimes \rho_0 &+ \sigma_1 \otimes \rho_0 \otimes \cdots \otimes \rho_0 \\ &+ \rho_0 \otimes \sigma_2 \otimes \cdots \otimes \rho_0 \\ &+ \cdots \\ &+ \rho_0 \otimes \rho_0 \otimes \cdots \otimes \sigma_M \\ &+ O\left(\sigma_i \sigma_j\right).\end{aligned} \tag{43}$$

We are going to apply each CGG and GSG to this state, respectively, keeping only the first-order perturbation of $\{\sigma_i\}_{i=1}^{M}$ during the calculation.

First, we calculate the output state from CGG. Following the discussion in [32], we consider the application of the projection operator $\frac{1}{M}\sum_{i=0}^{M-1} P_i^{(M)}$ to each of the M terms of $\rho_0 \otimes \cdots \otimes \sigma_i \otimes \cdots \otimes \rho_0$ in Eq. (43) of the form

$$\frac{1}{M^2} P_l^{(M)} \left(\rho_0 \otimes \cdots \otimes \sigma_i \otimes \cdots \otimes \rho_0\right) P_r^{(M)} \tag{44}$$

where $l, r \in \{0, \ldots, M-1\}$. For each σ_i, we observe that the M^2 terms taking the form of (44) reduce to the following cases.

1. One term equal to $\frac{1}{M^2}\sigma_i$, when $P_l = P_{M-i-1}^{(M)}$ and $P_r = P_i^{(M)}$.

2. $M-1$ terms equal to 0, when $P_l = P_{M-j-1}^{(M)}$ and $P_r = P_j^{(M)}$ for $j \neq i$.

3. $M-1$ terms equal to $\frac{1}{M^2}\rho_0\sigma_i$, when $P_l = P_{M-i-1}^{(M)}$ and $P_r = P_j^{(M)}$ for $j \neq i$.

4. $M-1$ terms equal to $\frac{1}{M^2}\sigma_i\rho_0$, when $P_l = P_{M-j-1}^{(M)}$ for $j \neq i$ and $P_r = P_i^{(M)}$.

5. $(M-1)(M-2)$ terms equal to $\frac{1}{M^2}\rho_0\sigma_i\rho_0$, when $P_l = P_{M-j-1}^{(M)}$ for $j \neq i$ and $P_r = P_k^{(M)}$ for $k \neq i$ and $P_l^\dagger \neq P_r$.

These five cases are visualised in Fig. 7.

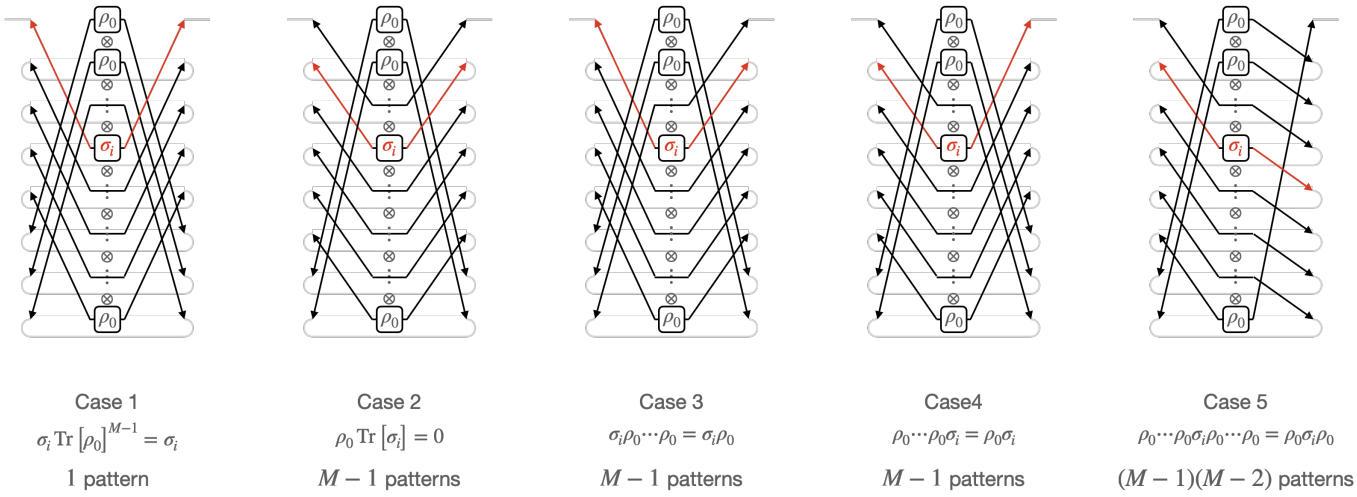


FIG. 7. The five cases of combinations of cyclic projectors resulting in different quantum states in tensor network representation. The effect of the cyclic projectors is shown by the arrows. The partial trace operations taken from the second quantum register to the last quantum register are shown by grey loops.

Thus, taking the average $\tilde{\sigma} = \frac{1}{M}\sum_{i=1}^{M}\sigma_i$ of the local noise terms $\{\sigma_i\}_{i=1}^{M}$, the resulting density matrix $\tilde{\rho}'$ before normalisation becomes

$$\begin{aligned}\tilde{\rho}'_{\mathrm{CGG}} &= \rho_0 + \sum_{i=1}^{M}\left(\frac{1}{M^2}\left(\sigma_i + (M-1)\left(\rho_0\sigma_i + \sigma_i\rho_0\right) + (M-1)(M-2)\rho_0\sigma_i\rho_0\right)\right) \\ &= \rho_0 + \frac{1}{M}\cdot\frac{1}{M}\sum_{i=1}^{M}\sigma_i + \frac{M-1}{M}\left(\rho_0\frac{1}{M}\sum_{i=1}^{M}\sigma_i + \frac{1}{M}\sum_{i=1}^{M}\sigma_i\rho_0\right) + \frac{(M-1)(M-2)}{M}\rho_0\frac{1}{M}\sum_{i=1}^{M}\sigma_i\rho_0 \\ &= \rho_0 + \frac{1}{M}\tilde{\sigma} + \frac{M-1}{M}\left(\rho_0\tilde{\sigma} + \tilde{\sigma}\rho_0 + (M-2)\rho_0\tilde{\sigma}\rho_0\right).\end{aligned} \tag{45}$$

This density matrix has a trace given by

$$\mathrm{Tr}\left[\tilde{\rho}'_{\mathrm{CGG}}\right] = 1 + \frac{M-1}{M}\cdot(1+1+M-2)\,\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right] = 1 + (M-1)\,\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]. \tag{46}$$

Eventually, we obtain the purified quantum state $\tilde{\rho}$ as

$$\begin{aligned}\tilde{\rho}_{\mathrm{CGG}} = \frac{\tilde{\rho}'}{\mathrm{Tr}\left[\tilde{\rho}'\right]} &= \rho_0 - (M-1)\,\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]\rho_0 + \frac{1}{M}\tilde{\sigma} + \frac{M-1}{M}\left(\rho_0\tilde{\sigma} + \tilde{\sigma}\rho_0 + (M-2)\rho_0\tilde{\sigma}\rho_0\right) \\ &= \left(1 - (M-1)\,\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]\right)\rho_0 + \frac{1}{M}\tilde{\sigma} + \frac{M-1}{M}\left(\rho_0\tilde{\sigma} + \tilde{\sigma}\rho_0 + (M-2)\rho_0\tilde{\sigma}\rho_0\right).\end{aligned} \tag{47}$$

GSG follows a similar discussion to CGG and gives the same output as CGG since the difference between CGG and GSG only lies in the way to shuffle the input quantum registers, which results in slightly different proportion of ρ_0, $\tilde{\sigma}$, $\rho_0\tilde{\sigma}$, $\tilde{\sigma}\rho_0$, $\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]\rho_0$ and $\rho_0\tilde{\sigma}\rho_0$, as shown in Fig. 8. The form of the purified state from GSG is described as

$$\begin{aligned}\tilde{\rho}_{\mathrm{GSG}} = \frac{\tilde{\rho}'_{\mathrm{GSG}}}{\mathrm{Tr}\left[\tilde{\rho}'_{\mathrm{GSG}}\right]} &= \left(1 - \mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]\right)\rho_0 - (M-1)\,\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]\rho_0 + \frac{1}{M}\tilde{\sigma} + \frac{M-1}{M}\left(\rho_0\tilde{\sigma} + \tilde{\sigma}\rho_0\right) \\ &= \left(1 - 2\frac{M-1}{M}\,\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]\right)\rho_0 + \frac{1}{M}\tilde{\sigma} + \frac{M-1}{M}\left(\rho_0\tilde{\sigma} + \tilde{\sigma}\rho_0\right).\end{aligned} \tag{48}$$

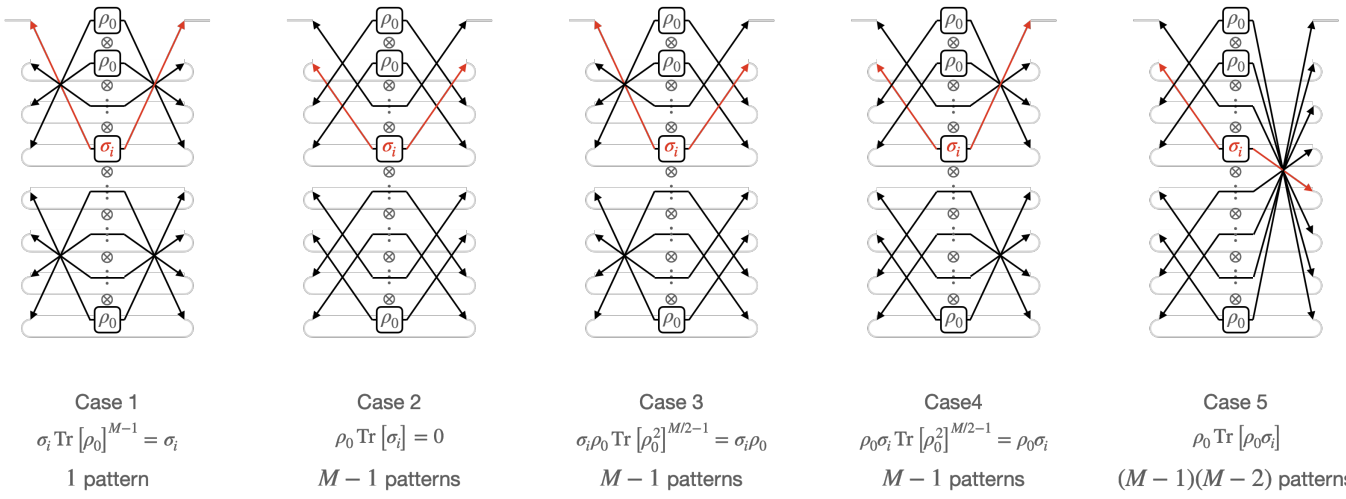


FIG. 8. The five cases of combinations of parallel SWAP projectors resulting in different quantum states in tensor network representation. The effect of the parallel SWAP projectors is shown by the arrows. Grey loops show the partial trace operations taken from the second quantum register to the last quantum register.

Here we find that the success probability of obtaining ρ_0 in $\tilde{\rho}$ after either CGG or GSG is the same as that of SGG under the first-order perturbation of $\{\sigma_i\}_{i=0}^{M-1}$. In fact, the average success probability after CGG and GSG is

$$\mathrm{Tr}\left[\rho_0\tilde{\rho}\right] = 1 + \frac{1}{M}\,\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]. \tag{49}$$

Recall the average success probability among the noisy inputs is represented as

$$\frac{1}{M}\sum_i \mathrm{Tr}\left[\rho_0\left(\rho_0+\sigma_i\right)\right] = 1+\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right], \tag{50}$$

CGG and GSG linearly suppress the probability of accepting unwanted states, which can also be seen as increasing the state fidelity to the desired state by a factor of M.

In addition, the average purity is also enhanced to the same extent as SGG. The purity of $\tilde{\rho}$ is

$$\mathrm{Tr}\left[\tilde{\rho}^2\right] = 1+\frac{2}{M}\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right] \tag{51}$$

which is also enhanced by a factor of M from the purity of the noisy state

$$\frac{1}{M}\sum_{i=1}^{M}\mathrm{Tr}\left[\left(\rho_0+\sigma_i\right)^2\right] = 1+2\,\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right]. \tag{52}$$

Note that we assume $\mathrm{Tr}\left[\rho_0\tilde{\sigma}\right] < 0$ in our calculation as well.

In conclusion, under local interaction with an external environment when the coherence length of the reservoir is less than the spatial separation assumed in [32], it is enough to use CGG and GSG to achieve linear scaling of errors suppression to the number of copies M, instead of exploiting the symmetry by the full symmetric group. This reduces the implementation cost of quantum circuits M times lighter regarding the number of ancillary qubits and the number of controlled-SWAP gates from SGG.

V. PURIFYING DEPOLARISED INPUTS

It is also possible to apply our gadget only once over M redundant copies under a bounded error rate. Here we analyse the purification rate of CGG and GSG given depolarised inputs ρ with a depolarising rate $0 < p < 1$ and dimension d described as

$$\rho = (1-p)\,\rho_0 + p\frac{I}{d}. \tag{53}$$

We see that both CGG and GSG suppress the input depolarising rate p to $O\left(p^2\right)$ when p is small.

For later use, let β and γ respectively be

$$\beta = 1-p+\frac{p}{d} = 1-\left(1-\frac{1}{d}\right)p, \quad \gamma = \frac{p}{d}. \tag{54}$$

Note that $0 < \gamma < \beta < 1$. Using β and γ, ρ can be rewritten as

$$\begin{aligned} \rho &= (\beta-\gamma)\rho_0 + \gamma I, \\ \mathrm{Tr}\left[\rho^k\right] &= \mathrm{Tr}\left[\sum_{i=0}^{k} {}_kC_i(\beta-\gamma)^{k-i}\gamma^i\rho_0^{k-i}I^i\right] = \sum_{i=0}^{k} {}_kC_i(\beta-\gamma)^{k-i}\gamma^i + \left(d^k-1\right)\gamma^k = \beta^k + (1-\beta)\,\gamma^{k-1}. \end{aligned} \tag{55}$$

A. Cyclic Group Gadget with Depolarised Inputs

For CGG, we analyse the purification performance of CGG for prime M and $M = 2^n$ under depolarising noise and general stochastic noise. We focus on these two cases because the prime M gives a simple form of purified states which is easy to analyse, and the ancillary qubits for $M = 2^n$ can be easily implemented with local Hadamard gates.

1. When M is prime

For prime M, we expand Eq. (25) using β and γ into

$$\begin{aligned}
\mathrm{P}_{\vec{0}}\,\tilde{\rho} &= \frac{1}{M}\left(1-p\right)\rho_0 + \frac{1}{M}p\frac{I}{d} + \frac{M-1}{M}\left(\left(1-p\right)\rho_0 + p\frac{I}{d}\right)^M \\
&= \frac{1}{M}\left(1-p+(M-1)\left(\left(1-p+\frac{p}{d}\right)^M - \left(\frac{p}{d}\right)^M\right)\right)\rho_0 + \frac{1}{M}\left(1+(M-1)\left(\frac{p}{d}\right)^{M-1}\right)p\frac{I}{d} \\
&= \frac{1}{M}\left(\beta-\gamma+(M-1)\left(\beta^M-\gamma^M\right)\right)\rho_0 + \frac{1}{M}\left(1+(M-1)\,\gamma^{M-1}\right)p\frac{I}{d}, \\
\mathrm{P}_{\vec{0}} &= \frac{1}{M}\left(1+(M-1)\left(\left(1-p+\frac{p}{d}\right)^M + \left(\frac{p}{d}\right)^M(d-1)\right)\right) \\
&= \frac{1}{M}\left(1+(M-1)\left(\beta^M+\gamma^{M-1}\left(1-\beta\right)\right)\right).
\end{aligned} \tag{56}$$

Therefore, the output state $\tilde{\rho}$ also takes the depolarised state $\tilde{\rho} = (1-\tilde{p})\,\rho_0 + \tilde{p}\frac{I}{d}$ with depolarising rate $\tilde{p}$ as

$$\tilde{p} = \frac{1+(M-1)\left(\frac{p}{d}\right)^{M-1}}{1+(M-1)\left(\left(1-p+\frac{p}{d}\right)^M + \left(\frac{p}{d}\right)^M(d-1)\right)}p = \frac{1+(M-1)\,\gamma^{M-1}}{1+(M-1)\left(\beta^M+\gamma^{M-1}\left(1-\beta\right)\right)}p, \tag{57}$$

It is clear from Eq. (57) that $\tilde{p} < p$ always holds for $M \geq 2, d \geq 2$. When d goes to infinity,

$$\tilde{p} \overset{d\to\infty}{\to} \frac{1}{1+(M-1)\left(1-p\right)^M}p =: \tilde{p}_{d=\infty} \tag{58}$$

Note that $\tilde{p} < \tilde{p}_{d=\infty}$ holds for $d > 1$, which can be easily checked since $\tilde{p} < \tilde{p}_{d=\infty}$ is equivalent to the conversion to the following inequalities.

$$\begin{aligned}
&\left(1+(M-1)\left(1-p\right)^M\right)\left(1+(M-1)\left(\frac{p}{d}\right)^M\right) < 1+(M-1)\left(\left(1-p+\frac{p}{d}\right)^M + \left(\frac{p}{d}\right)^M(d-1)\right), \\
&\left(\frac{p}{d}\right)^M + (1-p)^M + (1-p)\left(\frac{p}{d}\right)^{M-1} + (M-1)\left(1-p\right)^M\left(\frac{p}{d}\right)^{M-1} < \left(1-p+\frac{p}{d}\right)^M, \\
&(1-p)\left(\frac{p}{d}\right)^{M-1} + (M-1)\left(1-p\right)^M\left(\frac{p}{d}\right)^{M-1} < \sum_{i=1}^{M-1} {}_MC_i\left(1-p\right)^{M-i}\left(\frac{p}{d}\right)^i.
\end{aligned} \tag{59}$$

The last line clearly holds because $(1-p)\left(\frac{p}{d}\right)^{M-1} < {}_MC_{M-1}\left(1-p\right)\left(\frac{p}{d}\right)^{M-1}$ and $(M-1)\left(1-p\right)^M\left(\frac{p}{d}\right)^{M-1} < {}_MC_1\left(1-p\right)^{M-1}\left(\frac{p}{d}\right)$.

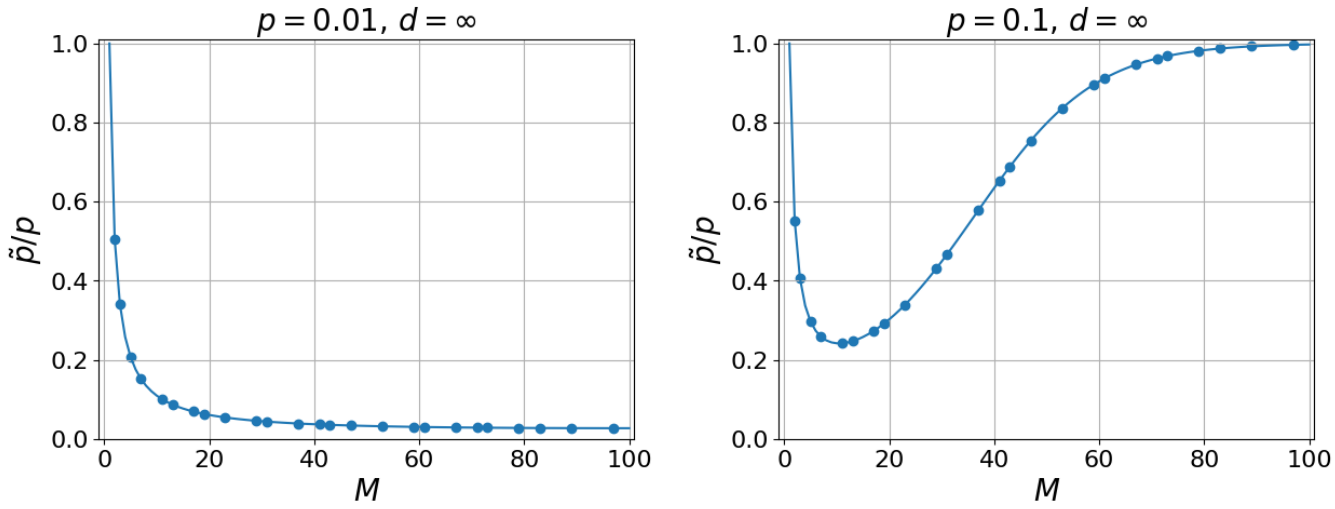
We observe from Fig. 9 that $\tilde{p}/p$ is not suppressed monotonically as M increases. As mentioned in Section. III A, an intuitive interpretation comes from the fact that $\mathrm{Tr}\left[\rho^M\right]$ shrinks to 0 exponentially with M. The monotonicity of $\tilde{p}$ to M can be analysed by checking the sign of $\frac{d}{dM}\tilde{p}$, which can be expanded as

$$\frac{d}{dM}\tilde{p} = \frac{\beta\gamma^M - \beta^M\gamma + (M-1)\left(\log\left(\gamma\right)\beta\gamma^M - \log\left(\beta\right)\beta^M\gamma\right) + (M-1)^2\left(\log\left(\beta\right) - \log\left(\gamma\right)\right)\beta^M\gamma^M}{\gamma\left(1+(M-1)\left(\beta^M+\gamma^{M-1}\left(1-\beta\right)\right)\right)^2}. \tag{60}$$

The condition $\tilde{p}$ decreases monotonically to M is $\frac{d}{dM}\tilde{p} < 0$, which can be described by β and γ as

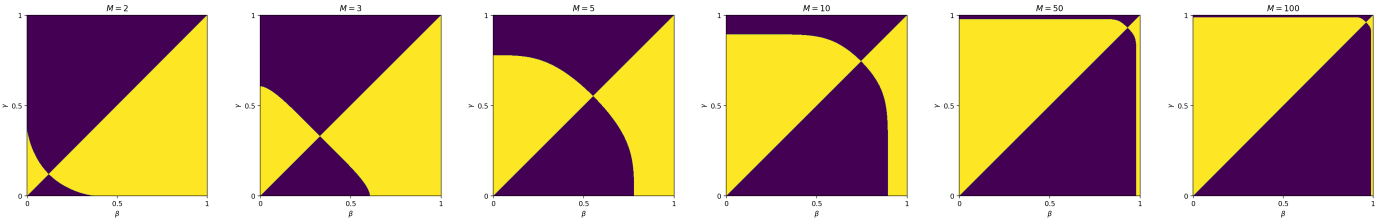
$$f(\beta,\gamma) - f(\gamma,\beta) > 0, \tag{61}$$

by introducing a function $f(x,y) = x^{M-1}\left(1+(M-1)\left(1+(M-1)y^{M-1}\right)\log(x)\right)$. The areas representing $f(\beta,\gamma) - f(\gamma,\beta) > 0$ under different M are visualised in Fig. 10.



(a) (b)

FIG. 9. The purification ratio $\tilde{p}/p$ in CGG when taking M copies of depolarised states with infinite dimensions. Scattered plots show the ratio when M is prime numbers. (a) $p = 0.1$. (b) $p = 0.01$.



(a) $M = 2$ (b) $M = 3$ (c) $M = 5$ (d) $M = 10$ (e) $M = 50$ (f) $M = 100$

FIG. 10. The heatmaps that show the sign of $f(\beta,\gamma) - f(\gamma,\beta)$ for $0 < \gamma < \beta < 1$. In each figure, the areas representing $f(\beta,\gamma) - f(\gamma,\beta) > 0$ are coloured yellow. The variable β takes from 0 to 1 on the x-axis, and the variable γ takes from 0 to 1 on the y-axis.

We show the condition for M which satisfies $\frac{d}{dM}\tilde{p} < 0$ for $\gamma = 0$ also holds for $0 < \gamma < 1$. First when $\gamma = 0$, the inequality $f(\beta,0) - f(0,\beta) > 0$ can be simplified into

$$\begin{aligned} & 1 + (M-1)\log\left(1 - \left(1 - \frac{1}{d}\right)p\right) > 0 \\ \text{(solving the first line for } p) \qquad & p < \frac{d}{d-1}\left(1 - e^{\frac{1}{1-M}}\right) \\ \text{(solving the first line for } M) \qquad & M < 1 - \frac{1}{\log\left(1 - \left(1 - \frac{1}{d}\right)p\right)} = 1 - \frac{1}{\log(\beta)} \end{aligned} \tag{62}$$

When $\gamma \ll 1$ and inequalities in Eq. (62) hold, we can also easily check $\tilde{p}/p$ converges to 0 as follows:

$$\frac{\tilde{p}}{p} = \frac{1 + (M-1)\gamma^{M-1}}{1 + (M-1)\left(\beta^M + \gamma^{M-1}(1-\beta)\right)} \simeq \frac{1}{1 + (M-1)\beta^M} < \frac{1}{1 + (M-1)e^{-1}} \overset{M\to\infty}{\longrightarrow} 0. \tag{63}$$

Next, we check $\frac{d}{dM}\tilde{p} < 0$ also holds for $0 < \gamma < 1$ under the condition Eq. (62). On the monotonicity of $\tilde{p}$, we

prove $f(\beta,\gamma)-f(\gamma,\beta)>0$. We first bound $f(\beta,\gamma)-f(\gamma,\beta)$ by using Eq. (62) and $0<\gamma<\beta<1$ into

$$\begin{aligned} f(\beta,\gamma)-f(\gamma,\beta) &= \beta^{M-1}\left(1+(M-1)\log(\beta)+(M-1)^2\gamma^{M-1}\left(\log(\beta)-\log(\gamma)\right)\right)-\gamma^{M-1}\left(1+(M-1)\log(\gamma)\right) \\ &> \frac{1}{e}\left(1+(M-1)\frac{1}{1-M}+(M-1)^2\gamma^{M-1}\cdot 0\right)-\gamma^{M-1}\left(1+(M-1)\log(\gamma)\right) \\ &= \left(\frac{p}{d}\right)^{M-1}\left((M-1)\left(\log(d)-\log(p)\right)-1\right). \end{aligned} \tag{64}$$

Thus it is sufficient to check $(M-1)\left(\log(d)-\log(p)\right)>1$, which is equivalent to $p<de^{\frac{1}{M-1}}$. In fact, this holds for $d>1$ and $M>1$ since $de^{\frac{1}{M-1}}$ is always larger than $\frac{d}{d-1}\left(1-e^{\frac{1}{1-M}}\right)$ by checking the following inequality:

$$de^{\frac{1}{M-1}}-\frac{d}{d-1}\left(1-e^{\frac{1}{1-M}}\right)>0, \tag{65}$$

which is equivalent to $M>1+\dfrac{1}{\log(d)}$ and thus always holds.

On the destination of the convergence, we also check that $\tilde{p}/p$ converges to 0 when M goes infinity under condition Eq. (62). As $\tilde{p}/p$ can be introduced as

$$\frac{\tilde{p}}{p}=\frac{1}{1+(M-1)\left(\beta^M+\gamma^{M-1}(1-\beta)\right)}+\frac{(M-1)\gamma^{M-1}}{1+(M-1)\left(\beta^M+\gamma^{M-1}(1-\beta)\right)}, \tag{66}$$

the convergence of the first term and the inverse of the second term respectively follow

$$\begin{aligned} \frac{1}{1+(M-1)\left(\beta^M+\gamma^{M-1}(1-\beta)\right)} &< \frac{1}{1+(M-1)e^{\frac{1-M}{M}}}=\frac{1}{1+(M-1)e^{\frac{1}{M}-1}} \overset{M\to\infty}{\longrightarrow} 0 \\ \frac{1+(M-1)\left(\beta^M+\gamma^{M-1}(1-\beta)\right)}{(M-1)\gamma^{M-1}} &= \frac{1}{(M-1)\gamma^{M-1}}+\frac{\beta^M}{\gamma^{M-1}}+1-\beta \overset{M\to\infty}{\longrightarrow} \infty, \end{aligned} \tag{67}$$

where we applied $0<\gamma<\beta<1$ and Eq. (62) again. As a result, we obtain $\tilde{p}/p \overset{M\to\infty}{\longrightarrow} 0$.

This concludes for a prime number M, the condition Eq. (62) is a sufficient condition for the monotonicity of $\tilde{p}$. Figure. 11(a) visually implies $\tilde{p}$ is linearly suppressed by M until the optimal number of M when the dimension d of the input state ρ goes to infinity. The condition $d=\infty$ is the worst case in the purification performance, which can be numerically checked in Fig. 11(b) as well. We also observe from Fig. 11(b) that increasing d does not significantly influence the optimal purification rate.

Here, we compute the purified depolarising rate for the optimal case, $M=1-\dfrac{1}{\log(\beta)}$, especially in the limit of $d\to\infty$. We let M^* denote this optimal M. Assigning this condition to Eq. (57), we obtain the optimal purified depolarising rate $\tilde{p}^*$ as

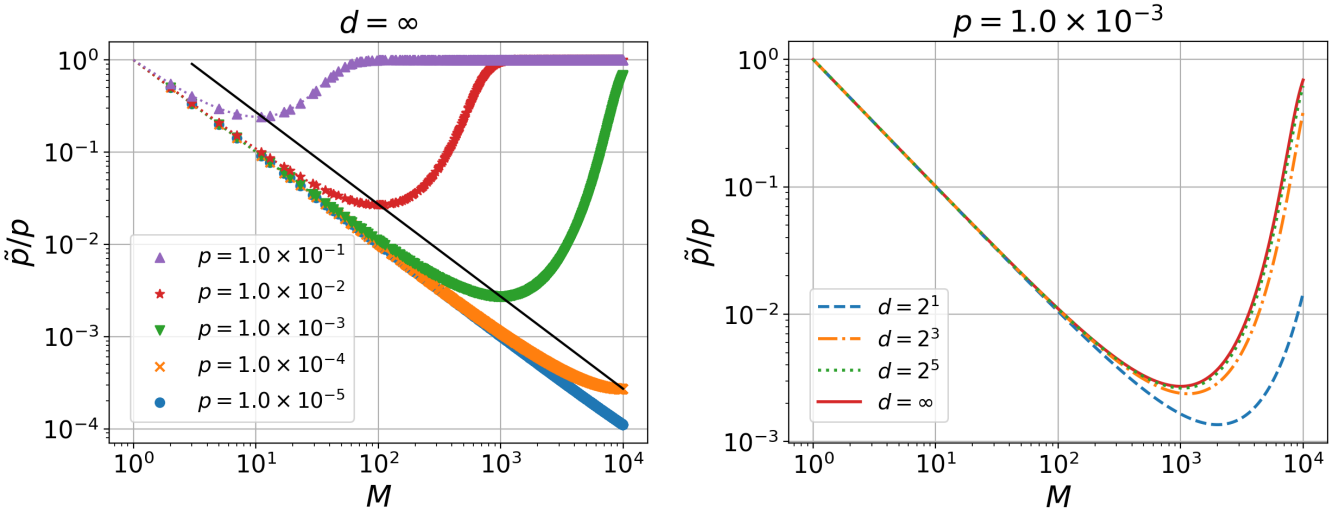
$$\tilde{p}^*=\frac{(M^*-1)\,d\left(\frac{1}{d-1}\left(1-e^{\frac{1}{1-M^*}}\right)\right)^{M^*}+\frac{1}{d-1}\left(1-e^{\frac{1}{1-M^*}}\right)}{1+(M^*-1)\left(e^{\frac{M^*}{1-M^*}}+(d-1)\left(\frac{1}{d-1}\left(1-e^{\frac{1}{1-M^*}}\right)\right)^{M^*}\right)} \overset{d\to\infty}{\longrightarrow} \frac{1-e^{\frac{1}{1-M^*}}}{1+(M^*-1)e^{\frac{M^*}{1-M^*}}}. \tag{68}$$

Since $p=\dfrac{d}{d-1}\left(1-e^{\frac{1}{1-M^*}}\right)\overset{d\to\infty}{\longrightarrow}1-e^{\frac{1}{1-M^*}}$, the purification ratio of $\tilde{p}^*$ to p becomes

$$\frac{\tilde{p}^*}{p}\overset{d\to\infty}{\longrightarrow}\frac{1}{1+(M^*-1)e^{\frac{M^*}{1-M^*}}}. \tag{69}$$

We further bound this ratio by observing $e^{-2}<e^{\frac{M}{1-M}}<e^{-1}$ for $M>1$ and $e^{\frac{M}{1-M}}\overset{M\to\infty}{\longrightarrow}e^{-1}$, which results in the following bound

$$\frac{e}{M^*+(e-1)}<\frac{\tilde{p}^*}{p}<\frac{e^2}{M^*+(e^2-1)}. \tag{70}$$



(a) (b)

FIG. 11. The ratio $\tilde{p}/p$ between output depolarising rate $\tilde{p}$ and p with M copies CGG when M is a prime number. Scattered plots show the ratio of prime numbers. (a) The purification ratio with different input depolarising probability p when the state dimension is infinity. The coloured points, with symbols for different p, show the purification ratio for each prime number M. The dotted curves fit the plots for each input depolarising probability p. The solid black line approximates the valley of the curves according to Eq. (71). (b) The purification ratio with different state dimension d when the input depolarising probability is set to 1.0×10^{-3}.

Hence the error suppression rate scales in $O\left(\frac{1}{M}\right)$ to the number of copies M.

Taking the logarithm of Eq. (68) for $d = \infty$, we can deduce the fitting function of the valley point in Fig. 11(a).

$$\log\left(\frac{\tilde{p}^*}{p}\right) = -\log\left(1 + (M^* - 1)\, e^{\frac{M^*}{1-M^*}}\right) < -\log\left(M^* e^{\frac{M^*}{1-M^*}}\right) = -\log(M^*) + \frac{1}{1 - \frac{1}{M^*}} < -\log(M^*) + 1 \tag{71}$$

Therefore there exists $M < 1 - \frac{1}{\log(\beta)}$ that achieves the purification rate in logarithmic scale: $\log\left(\frac{\tilde{p}}{p}\right) = -\log(M) + 1$. This is displayed as the solid black line in Fig. 11(a).

As M^* depends on p, the optimally purified depolarising rate $\tilde{p}^*$ in Eq. (68) can be alternatively described with p as

$$\tilde{p}^* = \frac{\log\left(1 - p + \frac{p}{d}\right) - \left(\frac{p}{d}\right)^{-\frac{1}{\log\left(1-p+\frac{p}{d}\right)}}}{\log\left(1 - p + \frac{p}{d}\right) - \frac{1}{e}\left(1 - p + \frac{p}{d}\right) + \left(\frac{p}{d}\right)^{-\frac{1}{\log\left(1-p+\frac{p}{d}\right)}} p\left(1 - \frac{1}{d}\right)} p \xrightarrow{d\to\infty} \frac{ep\log(1-p)}{e\log(1-p) + p - 1}. \tag{72}$$

We observe that the Taylor expansion of $\tilde{p}^*$ around $p = 0$ is $ep^2 + O\left(p^3\right)$.

When p is small, our gadget outputs the same purification rate as RSG [33]. To see this, we check the purification rate of RSG taking $M^* = 1 - \frac{1}{\log(1-p)}$ copies of depolarised inputs. Recalling that the purified depolarising rate $\tilde{p}_{RSG}$ is described as Eq. (8) for $p < \frac{1}{2}$, we obtain

$$\tilde{p}_{RSG} < \frac{1}{\left(1 - \frac{1}{\log(1-p)}\right)(1-2p) + 2p} p = \frac{p\log(1-p)}{\log(1-p) + 2p - 1}, \tag{73}$$

by replacing 2^n with M^*. Taking up to the second order of p around $p = 0$ in the Taylor expansion, Eq. (73) becomes $p^2 + O\left(p^3\right)$. This implies the purification rate of our method differs from that of RSG only on a constant scale, with the same number of copies.

We analytically compare the sampling cost of CGG and that of RSG for prime M as well. The acceptance probability of CGG $\mathrm{P}_{\vec{0}}$ with M copies when $d=\infty$ is

$$\mathrm{P}_{\vec{0}}=\frac{1}{M}\left(1+(M-1)\left(\left(1-p+\frac{p}{d}\right)^{M}+\left(\frac{p}{d}\right)^{M}(d-1)\right)\right)\overset{d\to\infty}{\to}\frac{1}{M}\left(1+(M-1)(1-p)^{M}\right). \tag{74}$$

Thus the expected number of copies required to post-select the all-zero state in the ancillary qubits for the optimal M is

$$\frac{M^{*}}{\mathrm{P}_{\vec{0}}}=\frac{(\log(1-p)-1)^{2}}{\left(\log(1-p)-\frac{1}{e}(1-p)\right)\log(1-p)}. \tag{75}$$

The Laurent series of Eq. (75) up to the first order of p around 0 is

$$\frac{e}{p}+\left(\frac{5}{2}-e\right)e+\left(\frac{41e}{12}-4e^{2}+e^{3}\right)p+O\left(p^{2}\right). \tag{76}$$

To achieve the same purification quality with RSG, the sampling cost becomes

$$C(n,d)\leq\frac{2p}{\tilde{p}^{*}(1-2p)^{2}}=\frac{2(e\log(1-p)+p-1)}{e(1-2p)^{2}\log(1-p)}. \tag{77}$$

The Laurent series of Eq. (77) up to the first order of p around 0 is

$$\frac{2}{ep}+\left(2+\frac{5}{e}\right)+\left(8+\frac{77}{6e}\right)p+O\left(p^{2}\right). \tag{78}$$

This implies our method achieves the same order in p up to the constant difference as the RSG with fewer measurement operations when p is small enough.

2. *When M is power of two*

For $M=2^{n}$, noting that we have defined $\beta=1-p+\frac{p}{d}$ and $\gamma=\frac{p}{d}$, the unnormalised output quantum state is

$$\begin{aligned}
MP_{\vec{0}}\tilde{\rho}&=\rho+\sum_{i=1}^{n}\alpha^{i-1}\operatorname{Tr}\left[\rho^{\alpha^{i}}\right]^{\alpha^{n-i}-1}\rho^{\alpha^{i}}\\
&=(1-p)\rho_{0}+p\frac{I}{d}+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^{i}}+\gamma^{\alpha^{i}-1}(1-\beta)\right)^{\alpha^{n-i}-1}\left(\beta^{\alpha^{i}}\rho_{0}+\gamma^{\alpha^{i}}(I-\rho_{0})\right)\\
&=(1-p)\rho_{0}+p\frac{I}{d}+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^{i}}+\gamma^{\alpha^{i}-1}(1-\beta)\right)^{\alpha^{n-i}-1}\left(\left(\beta^{\alpha^{i}}-\gamma^{\alpha^{i}}\right)\rho_{0}+\gamma^{\alpha^{i}}I\right)\\
&=\left((1-p)+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^{i}}+\gamma^{\alpha^{i}-1}(1-\beta)\right)^{\alpha^{n-i}-1}\left(\beta^{\alpha^{i}}-\gamma^{\alpha^{i}}\right)\right)\rho_{0}\\
&\quad+\left(p+d\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^{i}}+\gamma^{\alpha^{i}-1}(1-\beta)\right)^{\alpha^{n-1}-1}\gamma^{\alpha^{i}}\right)\frac{I}{d},
\end{aligned} \tag{79}$$

with the post-selection probability $P_{\vec{0}}$ as

$$\begin{aligned}
MP_{\vec{0}} &= 1-p+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-i}-1}\left(\beta^{\alpha^i}-\gamma^{\alpha^i}\right)+p+d\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-1}-1}\gamma^{\alpha^i} \\
&= 1+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-i}-1}\left(\beta^{\alpha^i}-\gamma^{\alpha^i}\right)+\alpha\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-i}-1}\gamma^{\alpha^i} \\
&= 1+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-i}-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right) \\
&= 1+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-i}}.
\end{aligned} \tag{80}$$

Therefore, the normalised output state of CGG becomes

$$\begin{aligned}
\tilde{\rho} &= (1-\tilde{p})\,\rho_0+\tilde{p}\frac{I}{d} \\
\tilde{p} &= \frac{p+d\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-1}-1}\gamma^{\alpha^i}}{1+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-i}}}=\frac{1+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-1}-1}\gamma^{\alpha^i-1}}{1+\sum_{i=1}^{n}\alpha^{i-1}\left(\beta^{\alpha^i}+\gamma^{\alpha^i-1}(1-\beta)\right)^{\alpha^{n-i}}}p.
\end{aligned} \tag{81}$$

When $\alpha=2$,

$$\frac{\tilde{p}}{p}=\frac{1+\sum_{i=1}^{n}2^{i-1}\left(\beta^{2^i}+\gamma^{2^i-1}(1-\beta)\right)^{2^{n-1}-1}\gamma^{2^i-1}}{1+\sum_{i=1}^{n}2^{i-1}\left(\beta^{2^i}+\gamma^{2^i-1}(1-\beta)\right)^{2^{n-i}}} \tag{82}$$

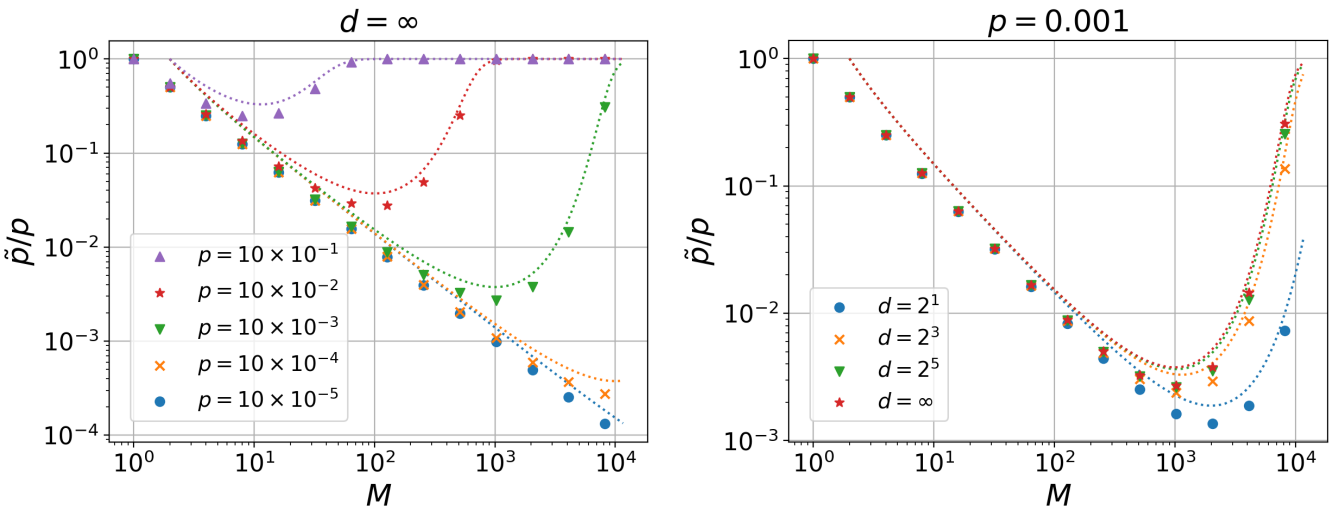
Taking $M=2^n$ copies, the uniform superposition in n ancillary qubits can be trivially prepared by applying local Hadamard gates to $|0\rangle$ in each ancillary qubit.

To see the behaviour of $\tilde{p}$, we numerically plot $\tilde{p}/p$ for different p in Fig. 12. This figure also implies that CGG suppresses the error linearly small to the number of copies M until the balance point.

B. Generalised SWAP Gadget with Depolarised Inputs

Taking the depolarised inputs as described by Eq. (53), the generalised SWAP gadget outputs the following quantum state:

$$\begin{aligned}
\tilde{\rho} &= \frac{1}{\mathrm{P}_{\vec{0}}}\left(\frac{1}{M}\left((1-p)\rho_0+p\frac{I}{d}\right)+\frac{M-1}{M}\left((1-p)^2\,\mathrm{Tr}\left[\rho_0^2\right]+\frac{1}{d}p(2-p)\right)^{\frac{M}{2}-1}\left((1-p)^2\rho_0^2+\frac{2}{d}p(1-p)\rho_0+\frac{p^2}{d}\frac{I}{d}\right)\right), \\
&= \frac{1}{\mathrm{P}_{\vec{0}}}\left(\frac{M-1}{M}\left((1-p)^2\,\mathrm{Tr}\left[\rho_0^2\right]+\frac{1}{d}p(2-p)\right)^{\frac{M}{2}-1}(1-p)^2\,\mathrm{Tr}\left[\rho_0^2\right]\frac{\rho_0^2}{\mathrm{Tr}\left[\rho_0^2\right]}\right. \\
&\quad+\left(\frac{1}{M}(1-p)+\frac{M-1}{M}\left((1-p)^2\,\mathrm{Tr}\left[\rho_0^2\right]+\frac{1}{d}p(2-p)\right)^{\frac{M}{2}-1}\frac{2}{d}p(1-p)\right)\rho_0 \\
&\quad\left.+\left(\frac{1}{M}p+\frac{M-1}{M}\left((1-p)^2\,\mathrm{Tr}\left[\rho_0^2\right]+\frac{1}{d}p(2-p)\right)^{\frac{M}{2}-1}\frac{p^2}{d}\right)\frac{I}{d}\right) \\
\mathrm{P}_{\vec{0}} &= \frac{1}{M}+\frac{M-1}{M}\left((1-p)^2\,\mathrm{Tr}\left[\rho_0^2\right]+\frac{1}{d}p(2-p)\right)^{\frac{M}{2}},
\end{aligned} \tag{83}$$



(a) (b)

FIG. 12. The purification performance $\tilde{p}/p$ with $M = 2^n$ copies in the cyclic group test. (a) Changing the input depolarising probability p, (b) Changing the dimension d. The fitting function used in both (a) and (b) derives from the upper bound of the continuitisation of Eq. (81) to aid the visibility of plots.

where we used

$$\begin{aligned} \rho^2 &= (1-p)^2 \operatorname{Tr}\left[\rho_0^2\right] \frac{\rho_0^2}{\operatorname{Tr}\left[\rho_0^2\right]} + \frac{2}{d} p(1-p)\rho_0 + \frac{p^2}{d}\frac{I}{d}, \\ \operatorname{Tr}\left[\rho^2\right] &= (1-p)^2 \operatorname{Tr}\left[\rho_0^2\right] + \frac{1}{d} p\,(2-p)\,. \end{aligned} \tag{84}$$

From (83), when ρ_0 is a pure state,

$$\begin{aligned} \tilde{\rho} &= \frac{1}{\mathrm{P}_{\vec{0}}} \left(\left(\frac{1}{M}(1-p) + \frac{M-1}{M} \left((1-p)^2 + \frac{1}{d} p(2-p) \right)^{\frac{M}{2}-1} \left((1-p)^2 + \frac{2}{d} p(1-p) \right) \right) \rho_0 \right. \\ &\quad \left. + \left(\frac{1}{M} p + \frac{M-1}{M} \left((1-p)^2 + \frac{1}{d} p(2-p) \right)^{\frac{M}{2}-1} \frac{p^2}{d} \right) \frac{I}{d} \right) \\ &= \frac{1}{\mathrm{P}_{\vec{0}}} \left(\left(\frac{1}{M}(1-p) + \frac{M-1}{M} \left(1 - 2\left(1-\frac{1}{d}\right) p + \left(1-\frac{1}{d}\right) p^2 \right)^{\frac{M}{2}-1} \left(1 - 2\left(1-\frac{1}{d}\right) p + \left(1-\frac{2}{d}\right) p^2 \right) \right) \rho_0 \right. \\ &\quad \left. + \left(\frac{1}{M} p + \frac{M-1}{M} \left(1 - 2\left(1-\frac{1}{d}\right) p + \left(1-\frac{1}{d}\right) p^2 \right)^{\frac{M}{2}-1} \frac{p^2}{d} \right) \frac{I}{d} \right) \\ \mathrm{P}_{\vec{0}}, &= \frac{1}{M} + \frac{M-1}{M} \left(1 - 2\left(1-\frac{1}{d}\right) p + \left(1-\frac{1}{d}\right) p^2 \right)^{\frac{M}{2}} . \end{aligned} \tag{85}$$

It is clear to see Eq. (85) corresponds to Eq. (8) when $M = 2$.

When the dimension d goes to infinity in Eq. (85), $\tilde{\rho}$ and $\mathrm{P}_{\vec{0}}$ become

$$\begin{aligned} \tilde{\rho} &\xrightarrow{d\to\infty} \frac{1}{\mathrm{P}_{\vec{0}}} \left(\frac{1-p}{M} \left(1 + (M-1)\,(1-p)^{M-1} \right) \rho_0 + \frac{p}{M}\frac{I}{d} \right), \\ \mathrm{P}_{\vec{0}} &\xrightarrow{d\to\infty} \frac{1}{M} + \frac{M-1}{M} (1-p)^M , \end{aligned} \tag{86}$$

which correspond to the outputs of CGG shown in Eq. (58) and Eq. (74). Therefore, the purification rate of GSG is equivalent to CGG as far as the difference in the way of applying permutation in quantum circuits.

VI. DISCUSSION

In this work, we have proposed the cyclic group gadget (CGG) and generalised SWAP gadget (GSG) to obtain a purified quantum state from multiple noisy inputs. Here we discuss how and where our gadgets can be used. As we have seen in Section IV, CGG and GSG suppress the first-order approximation of coherent errors and stochastic errors by the factor of $1/M$ to the number of copies M. This assumption can be justified when we apply the gadget over M copies for every small evolution time δt, which is the same strategy as SGG [32]. Compared to SGG, our proposed methods, CGG and GSG, improve the circuit implementation cost M times smaller than SGG, particularly from $O(M\log(M))$ measurements to $O(\log(M))$, and from $O\left(M^2\log(M)\right)$ controlled-SWAP operations to $O(M\log(M))$ controlled-SWAP operations. The reduction of the measurement operations also means the reduction of the cost for preparing a uniform superposition with dimension M from $O\left((M\log(M))^2\right)$ to $O\left((\log(M))^2\right)$, which is smaller than $O(M\log(M))$. Compared to the purification-based QEM methods [15, 16] which consume M nearly identical noisy inputs as well, the implementation cost of our gadgets is larger only $O(\log(M))$ times with respect to the number of measurements and controlled-SWAP gates (see Appendix A for their implementation cost). Therefore, the implementation cost of our proposed method is still in line with that of QEM schemes, which would be feasible in the near- and middle-term quantum devices.

When it comes to the single application of our gadgets over M copies, our analysis of the depolarised inputs demonstrates that the purified depolarising rate $\tilde{p}$ is also suppressed by a factor of $1/M$ up to the optimal number of M associated with the input depolarising rate p. When p is small, we have seen that our gadget achieves the same purification rate and sampling cost as RSG whose optimality of sampling cost is proven to be optimal for purifying depolarised inputs [33]. Taking the progress of quantum devices into account, and since both RSG and our gadgets are supposed to be performed perfectly, the assumption that p is small is likely to fit the actual situation for applying these purification gadgets. The difference between our gadgets and RSG in the circuit implementation is that our gadget uses $O(\log(M))$ ancillary qubits for measurements $O(M\log(M))$ controlled-SWAP operations, while RSG consumes $O(M)$ ancillary qubits for measurements $O(M)$ controlled-SWAP operations, which is in a trade-off between the amount of measurement and controlled-SWAP operations. On devices where the measurement process is noisy, our method may display its advantage over others. Therefore, for a single application of the purification gadget, there is room for choosing which gadget to use depending on the hardware restriction.

There are many other open questions to be explored in the future. One interesting direction of future research would be the analysis of lighter implementations and the recursive application of our methods. In our gadgets, we are paying $O(M\log(M))$ controlled-SWAP operations so that all of the M outputs are stabilised equally, i.e. all M outputs are purified. This allows us to stabilise the errors over M copies of the same computation, while in the case where we want to apply it once and retrieve only one of the purified states, it might be possible to reduce the number of controlled-SWAP gates or recursively apply the gadget by reusing all of the outputs. When it comes to the recursive application as adopted in RSG [33], we may adopt some lighter implementations of our methods to pass only one output quantum state to the next recursion, or effectively utilise all M purified outputs to save the required number of copies in the recursion.

The purification rate of our gadgets under more general errors also remains to be examined analytically and numerically. With the first-order approximation with respect to the evolution time, we show that our gadgets suppress the coherent and stochastic errors linearly smaller to the number of copies. The performance of CGG and GSG with inputs affected by more general noise without approximation, such as general stochastic errors or coherent errors, also remains to be examined analytically and numerically. In Appendix B, we give a preliminary analytical observation on the fidelity of output state taking the inputs under the general stochastic noise $\rho = (1-p)\rho_0 + p\sigma$ with any density matrix σ. This may come into question when we consider combining our purification gadgets with verification protocols as discussed below.

Analysing the purification rate against general errors is also of interest from the motivation to apply these purification gadgets in verification protocols. Quantum verification protocols [35–39] are designed to guarantee the correctness of results obtained from computations offloaded to untrusted devices. They are typically described in a client-and-server model, in which a client with very limited quantum abilities delegates a computation to a powerful server, a setting likely to be encountered in a future quantum ecosystem. While very efficient verification protocols exist for computations in BQP, any protocol that avoids prohibitive hardware overheads has only achieved weak security guarantees for computations with nondeterministic classical or quantum output. One could hope that purification gadgets could be used to improve the quality of outputs produced by such protocols. This requires a rigorous analysis of the structure of the errors that may affect the outputs of these verification protocols in the presence of a malicious server and to match these errors with the classes of noise for which purification gadgets guaranteedly perform well.

ACKNOWLEDGMENTS

BY acknowledges insightful discussions with Suguru Endo from NTT. This work received funding from the ANR research grant ANR-21-CE47-0014 (SecNISQ), and from the European Union's Horizon 2020 research and innovation program through the FET project PHOQUSING ("PHOtonic Quantum SamplING machine" – Grant Agreement No. 899544). This work was supported by the Quantum Advantage Pathfinder (QAP) research programme within the UK's National Quantum Computing Centre (NQCC), and by the Quantum Computing and Simulation (QCS) Hub.

[1] P. Shor, in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, SFCS-94 (IEEE Comput. Soc. Press, 1994).
[2] P. W. Shor, SIAM review **41**, 303 (1999).
[3] L. K. Grover, in *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing* (1996) pp. 212–219.
[4] A. W. Harrow, A. Hassidim, and S. Lloyd, Physical review letters **103**, 150502 (2009).
[5] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
[6] D. A. Lidar and T. A. Brun, *Quantum error correction* (Cambridge university press, 2013).
[7] J. Roffe, Contemporary Physics **60**, 226 (2019).
[8] D. Horsman, A. G. Fowler, S. Devitt, and R. Van Meter, New Journal of Physics **14**, 123011 (2012).
[9] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, Physical Review A **86**, 032324 (2012).
[10] A. G. Fowler and C. Gidney, arXiv preprint arXiv:1808.06709 (2018).
[11] Z. Ni, S. Li, X. Deng, Y. Cai, L. Zhang, W. Wang, Z.-B. Yang, H. Yu, F. Yan, S. Liu, *et al.*, Nature **616**, 56 (2023).
[12] R. S. Gupta, N. Sundaresan, T. Alexander, C. J. Wood, S. T. Merkel, M. B. Healy, M. Hillenbrand, T. Jochym-O'Connor, J. R. Wootton, T. J. Yoder, *et al.*, Nature **625**, 259 (2024).
[13] K. Temme, S. Bravyi, and J. M. Gambetta, Physical review letters **119**, 180509 (2017).
[14] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Reviews of Modern Physics **95**, 045005 (2023).
[15] B. Koczor, Physical Review X **11**, 031057 (2021).
[16] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Physical Review X **11**, 041036 (2021).
[17] B. Yang, R. Raymond, and S. Uno, Physical Review A **106**, 012423 (2022).
[18] B. Yang, N. Yoshioka, H. Harada, S. Hakkaku, Y. Tokunaga, H. Hakoshima, K. Yamamoto, and S. Endo, arXiv preprint arXiv:2309.14171 (2023).
[19] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, *et al.*, Nature Reviews Physics **3**, 625 (2021).
[20] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Journal of the Physical Society of Japan **90**, 032001 (2021).
[21] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, Physical review letters **125**, 150504 (2020).
[22] X. Yuan, J. Sun, J. Liu, Q. Zhao, and Y. Zhou, Physical Review Letters **127**, 040501 (2021).
[23] H. Harada, Y. Suzuki, B. Yang, Y. Tokunaga, and S. Endo, arXiv preprint arXiv:2309.15761 (2023).
[24] A. Eddins, M. Motta, T. P. Gujarati, S. Bravyi, A. Mezzacapo, C. Hadfield, and S. Sheldon, PRX Quantum **3**, 010309 (2022).
[25] R. Takagi, S. Endo, S. Minagawa, and M. Gu, npj Quantum Information **8**, 114 (2022).
[26] R. Takagi, H. Tajima, and M. Gu, Physical Review Letters **131**, 210602 (2023).
[27] K. Tsubouchi, T. Sagawa, and N. Yoshioka, Physical Review Letters **131**, 210601 (2023).
[28] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, arXiv preprint arXiv:2210.11505 (2022).
[29] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, Physical review letters **76**, 722 (1996).
[30] J. I. Cirac, A. Ekert, and C. Macchiavello, Physical review letters **82**, 4344 (1999).
[31] M. Keyl and R. F. Werner, in *Annales Henri Poincare*, Vol. 2 (Springer, 2001) pp. 1–26.
[32] A. Barenco, A. Berthiaume, D. Deutsch, A. Ekert, R. Jozsa, and C. Macchiavello, SIAM Journal on Computing **26**, 1541 (1997).
[33] A. M. Childs, H. Fu, D. Leung, Z. Li, M. Ozols, and V. Vyas, arXiv preprint arXiv:2309.16387 (2023).
[34] U. Chabaud, E. Diamanti, D. Markham, E. Kashefi, and A. Joux, Physical Review A **98**, 062318 (2018).
[35] A. Gheorghiu, T. Kapourniotis, and E. Kashefi, Theory of computing systems **63**, 715 (2019).
[36] M. Hayashi and T. Morimae, Physical review letters **115**, 220502 (2015).
[37] U. Mahadev, in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2018) pp. 259–267.
[38] J. F. Fitzsimons and E. Kashefi, Physical Review A **96**, 012303 (2017).
[39] D. Leichtle, L. Music, E. Kashefi, and H. Ollivier, PRX Quantum **2**, 040302 (2021).

Appendix A: Purification-based QEM Gadget

Recently, with the advent of near-term quantum devices, many quantum error mitigation (QEM) techniques have been proposed to suppress the bias in expectation values of measurement results through classical pre- and postprocessing [14–16, 18, 20]. The purification-based QEM methods, such as virtual distillation (VD) [16] and exponential suppression by derangement (ESD) [15], adopt the same idea as projecting the multiple copies of noisy states into their symmetric subspace. As those methods aim to improve the expectation value instead of the quantum state itself, they can make their quantum circuits much lighter.

Here we review the exponential error suppression by derangement (ESD) method [15] and design a state purification gadget from it. The ESD circuit uses M copies of noisy quantum states to suppress the stochastic errors in magnitude to M by applying one derangement operation to M copies. Then the mitigated expectation value for an observable O becomes

$$\langle O\rangle_{\mathrm{ESD}} = \frac{\mathrm{Tr}\left[\rho^M O\right]}{\mathrm{Tr}\left[\rho^M\right]}. \tag{A1}$$

When ρ is given as a classical mixture of the target state $|\lambda_0\rangle\langle\lambda_0|$ with a dominant probability and unwanted biased states $\{|\lambda_i\rangle\langle\lambda_i|\}_{i>1}$ with small probabilities, ρ can be expanded as $\rho = \sum_k \lambda_k |\lambda_k\rangle\langle\lambda_k|$ whose eigenvalues are aligned in descending order. The M-th power of ρ becomes

$$\rho^M = \sum_k \lambda_k^M |\lambda_k\rangle\langle\lambda_k| = \lambda_0^M \left(|\lambda_0\rangle\langle\lambda_0| + \sum_{k>0} \left(\frac{\lambda_k}{\lambda_0}\right)^M |\lambda_k\rangle\langle\lambda_k| \right). \tag{A2}$$

Then, the mitigated expectation value $\langle O\rangle_{\mathrm{ESD}}$ can be expanded as

$$\begin{aligned}\langle O\rangle_{\mathrm{ESD}} &= \frac{\lambda_0^M \left(\mathrm{Tr}\left[|\lambda_0\rangle\langle\lambda_0|O\right] + \sum_{k>0} \left(\frac{\lambda_k}{\lambda_0}\right)^M \mathrm{Tr}\left[|\lambda_k\rangle\langle\lambda_k|O\right] \right)}{\lambda_0^M \left(\mathrm{Tr}\left[|\lambda_0\rangle\langle\lambda_0|\right] + \sum_{k>0} \left(\frac{\lambda_k}{\lambda_0}\right)^M \mathrm{Tr}\left[|\lambda_k\rangle\langle\lambda_k|\right] \right)} \\ &= \mathrm{Tr}\left[|\lambda_0\rangle\langle\lambda_0|O\right] + \sum_{k>0} \left(\frac{\lambda_k}{\lambda_0}\right)^M \left(\mathrm{Tr}\left[|\lambda_k\rangle\langle\lambda_k|O\right] - \mathrm{Tr}\left[|\lambda_0\rangle\langle\lambda_0|O\right]\right) + O\left(x^2\right)\end{aligned} \tag{A3}$$

where $x = \sum_{k>0} \left(\frac{\lambda_k}{\lambda_0}\right)^M$. Since $\lambda_k < \lambda_0$, the factors of unwanted states and the contribution of $O\left(x^2\right)$ are suppressed exponentially to M in Eq. (A3).

In ESD, $\mathrm{Tr}\left[\rho^M O\right]$ and $\mathrm{Tr}\left[\rho^M\right]$ are obtained from the quantum circuits with a derangement operation that performs a permutation such that no state appears in its original position. We can pick up cyclic shift $P_{23\cdots M1}$ as an example of the derangement operator. The quantum circuit for $\mathrm{Tr}\left[\rho^M O\right]$ and $\mathrm{Tr}\left[\rho^M\right]$ contains only one ancillary qubit and one derangement operation as shown in Fig. 13(a). Since the post-selection on $|0\rangle\langle 0|$ state in the ancillary qubit gives the post-selection probability $\mathrm{P}_0 = \frac{1}{2}\left(1 + \mathrm{Tr}\left[\rho^M O\right]\right)$ for $\rho_1 = \cdots = \rho_M = \rho$, $\mathrm{Tr}\left[\rho^M O\right]$ can be computed as

$$\mathrm{Tr}\left[\rho^M O\right] = 1 - 2\,\mathrm{P}_0\,. \tag{A4}$$

While the ESD circuit is simpler than that of the generalised SWAP test and that of the SGG, the exponential error suppression to the number of state copies M is available only for mitigating the expectation value. In fact, the state purification performance of the purification gadget inspired by the ESD circuit in Fig. 13(b) does not amplify the contribution of $\mathrm{Tr}\left[\rho^M\right]$ scaling to M, which can be immediately checked by computing $\tilde{\rho}$ and P_0:

$$\begin{aligned}\tilde{\rho} &= \frac{1}{2\,\mathrm{P}_{\vec{0}}}\left(\rho + \rho^M\right), \\ \mathrm{P}_0 &= \frac{1}{2}\left(1 + \mathrm{Tr}\left[\rho^M\right]\right).\end{aligned} \tag{A5}$$

Since $\mathrm{Tr}\left[\rho^M\right]$ shrinks in magnitude to M, the contribution of ρ^M vanishes quickly in $\tilde{\rho}$.

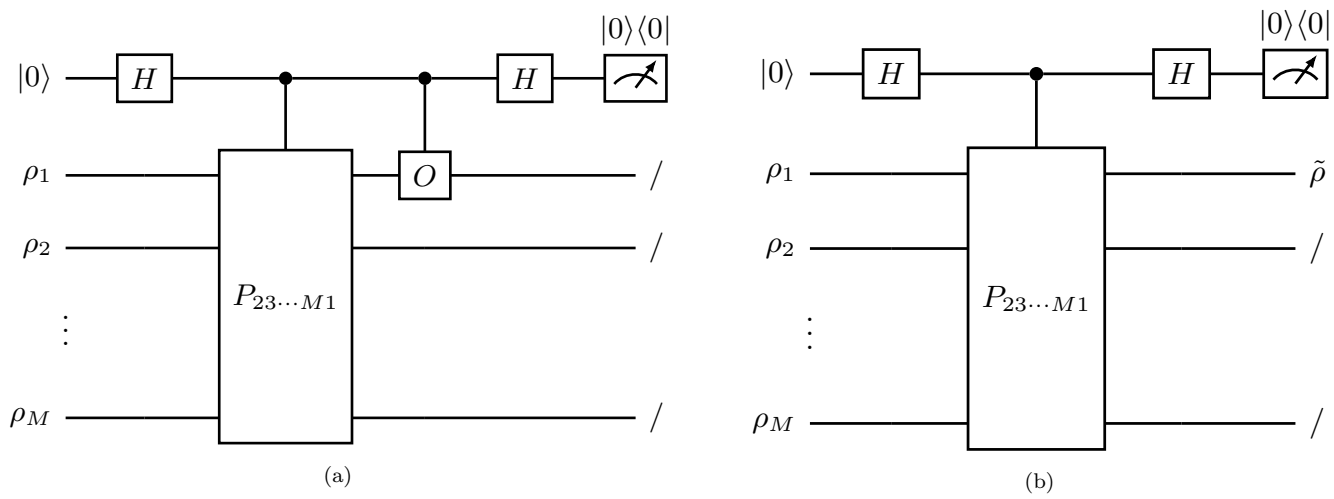


FIG. 13. (a) The quantum circuit for computing $\mathrm{Tr}\left[\rho_1\rho_2\cdots\rho_M O\right]$ in ESD. The cyclic shift $P_{23\cdots M1}$ over M state copies is one of the derangement operators. The expectation value $\mathrm{Tr}\left[\rho_1\rho_2\cdots\rho_M\right]$ is obtained by removing the controlled-O operator. (b) The quantum circuit which post-selects the first copy after the ESD process.

We remark that this does not surpass the purification rate of the SWAP gadget with two depolarised inputs. Taking depolarised inputs $\rho=(1-p)\,\rho_0+p\dfrac{I}{d}$, the output depolarising rate $\tilde{p}$ becomes

$$\tilde{p}=\frac{1+\left(\frac{p}{d}\right)^{M-1}}{1+\left(1-p+\frac{p}{d}\right)^{M}+\left(\frac{p}{d}\right)^{M}(d-1)}p\overset{d\to\infty}{\longrightarrow}\frac{1}{1+(1-p)^{M}}p<\frac{1}{2-p}p. \tag{A6}$$

Therefore there is no improvement in the purification rate when increasing the number of copies with this gadget.

Appendix B: Fidelity of Output State with Inputs Under Arbitrary Stochastic Noise

In this section, we generalise the noise in input states from the depolarising noise to the general form of stochastic noise described as

$$\rho\left(p\right)=\left(1-p\right)\rho_0+p\sigma. \tag{B1}$$

The fidelity $F=\mathrm{Tr}\left[\rho_0\rho\right]$ between the target pure state $\rho_0=|\psi\rangle\langle\psi|$ and the noisy input state ρ becomes

$$F=\mathrm{Tr}\left[\rho_0\rho\right]=\mathrm{Tr}\left[\rho_0\left(\left(1-p\right)\rho_0+p\sigma\right)\right]=1-p+pF_1=1-p\left(1-F_1\right)=1-pI_1, \tag{B2}$$

where $F_1=\mathrm{Tr}\left[\rho_0\sigma\right]$ is the fidelity between ρ_0 and σ and $I_1=1-F_1$ is infidelity.

When we apply CGG over M copies of ρ, the fidelity between ρ_0 and the output state $\tilde{\rho}$ for prime M is

$$\begin{aligned}\tilde{F}&=F(\rho_0,\tilde{\rho})\\&=\frac{1}{\mathrm{P}_{\vec{0}}}\mathrm{Tr}\left[\rho_0\left(\frac{1}{M}\rho+\frac{M-1}{M}\rho^M\right)\right]\\&=\frac{1}{\mathrm{P}_{\vec{0}}}\left(\frac{1}{M}\mathrm{Tr}\left[\rho_0\rho\right]+\frac{M-1}{M}\mathrm{Tr}\left[\rho_0\rho^M\right]\right)\\&=\frac{1}{\mathrm{P}_{\vec{0}}}\left(\frac{1}{M}F_1+\frac{M-1}{M}\mathrm{Tr}\left[\rho_0\rho^M\right]\right).\end{aligned} \tag{B3}$$

Note that $\mathrm{P}_{\vec{0}}$ in the denominator is still the same as Eq. (25), $\mathrm{P}_{\vec{0}}=\dfrac{1}{M}+\dfrac{M-1}{M}\mathrm{Tr}\left[\rho^M\right]$. Now, we analyse $\mathrm{Tr}\left[\rho_0\rho^M\right]$

in Eq. (B3) by expanding it to

$$\begin{aligned}
\mathrm{Tr}\left[\rho_0\rho^M\right] &= \mathrm{Tr}\left[\rho_0\left(\left(1-p\right)\rho_0+p\sigma\right)^M\right] \\
&= \left(1-p\right)^M + \sum_{i=1}^{M}\left(1-p\right)^{M-i}p^i \sum_{l=1}^{\min\{i,M-i+1\}} \sum_{\substack{j\in\text{partition}(i,l)\\ 1^{j_1}2^{j_2}\ldots i^{j_i}}} \prod_{k=1}^{i} {}_{(M-i+1)-(\sum_{k'=1}^{k-1} j_{k'})}C_{j_k} F_k^{j_k} \\
&= \left(1-p\right)^M + \sum_{i=1}^{M}\left(1-p\right)^{M-i}p^i \sum_{l=1}^{\min\{i,M-i+1\}} {}_{M-i+1}C_l \sum_{\substack{j\in\text{partition}(i,l)\\ 1^{j_1}2^{j_2}\ldots i^{j_i}}} \frac{l!}{j_1!j_2!\cdots j_i!} F_k^{j_k} \\
&= \left(1-p\right)^M + \sum_{i=1}^{M}\left(1-p\right)^{M-i}p^i \sum_{l=1}^{i} {}_{M-i+1}C_l \sum_{\substack{j\in\text{partition}(i,l)\\ 1^{j_1}2^{j_2}\ldots i^{j_i}}} \frac{l!}{j_1!j_2!\cdots j_i!} F_k^{j_k},
\end{aligned} \tag{B4}$$

where $F_k = \mathrm{Tr}\left[\rho_0\sigma^k\right]$. We obtain the last line of Eq. (B4) by defining ${}_nC_r = 0$ for $n < r$.

Each coefficient of $F_k^{j_k}$ is associated with each Young diagram. The case for $M = 5$ is shown in Fig. 14. The coefficient $\prod_{k=1}^{i} {}_{(M-i+1)-(\sum_{k'=1}^{k-1} j_{k'})}C_{j_k}$ comes from sequentially choosing j_k identical sequences of multiplication of k σ from all possible remaining $(M-i+1)-(\sum_{k'=1}^{k-1} j_{k'})$ positions for the sequence of multiplication of σ, following the stars-and-bars reasoning, so that the sequences of multiplication of σ will not be next to each other. In the third line of Eq. (B4), we can also interpret the coefficient as first choosing the position of locating l sequences of multiplication of σ, and then consider the permutation of l sequences by identifying sequences consisting of the same number of σ.

We might be able to simplify Eq. (B4) by introducing several different assumptions on the noisy state σ.

1. σ is pure: $\sigma^2 = \sigma$,

2. ρ_0 and σ are commutative: $[\rho_0, \sigma] = 0$.

These conditions are realistic as condition (1) covers the dephasing error, and condition (2) covers the depolarising error and general orthogonal errors.

1. When σ is pure: $\sigma^2 = \sigma$

First, when the condition (1) holds, the fidelity can be simplified as

$$F_k = \mathrm{Tr}\left[\rho_0\sigma^k\right] = \mathrm{Tr}\left[\rho_0\sigma\right] = F_1. \tag{B5}$$

Therefore, Eq. (B4) can be simplified into

$$\begin{aligned}
\mathrm{Tr}\left[\rho_0\rho^M\right] &= \left(1-p\right)^M + \sum_{i=1}^{M}\left(1-p\right)^{M-i}p^i \sum_{l=1}^{\min\{i,M-i+1\}} {}_{i-1}C_{l-1} \times {}_{l+1}C_{M-i-(l-1)} F_1^l \\
&= \left(1-p\right)^M + \sum_{i=1}^{M}\left(1-p\right)^{M-i}p^i \sum_{l=1}^{i} {}_{i-1}C_{l-1} \times {}_{l+1}C_{M-i-(l-1)} F_1^l,
\end{aligned} \tag{B6}$$

if we again define ${}_nC_r = 0$ for $n < r$. The number of combinations ${}_{i-1}C_{l-1} \times {}_{l+1}C_{M-i-(l-1)}$ for each i and l is obtained by considering the following rule. Since l represents the number of separated sequences of multiplication of σ, and each sequence should contain at least one σ, we have to choose which place to insert ρ_0 as separators. There are ${}_{i-1}C_{l-1}$ ways to choose the positions for the separators. Next, we still have to consider different arrangements of remaining ρ_0 and separated sequences of multiplication of σ. Note that in this step, we can identify the sequences of multiplication with different numbers of σ. Therefore, the number of the arrangements can be computed as ${}_{l+1}C_{M-i-(l-1)}$, since we are choosing $M-i-(l-1)$ remaining ρ_0 from $l+1$ positions following the stars-and-bars reasoning.

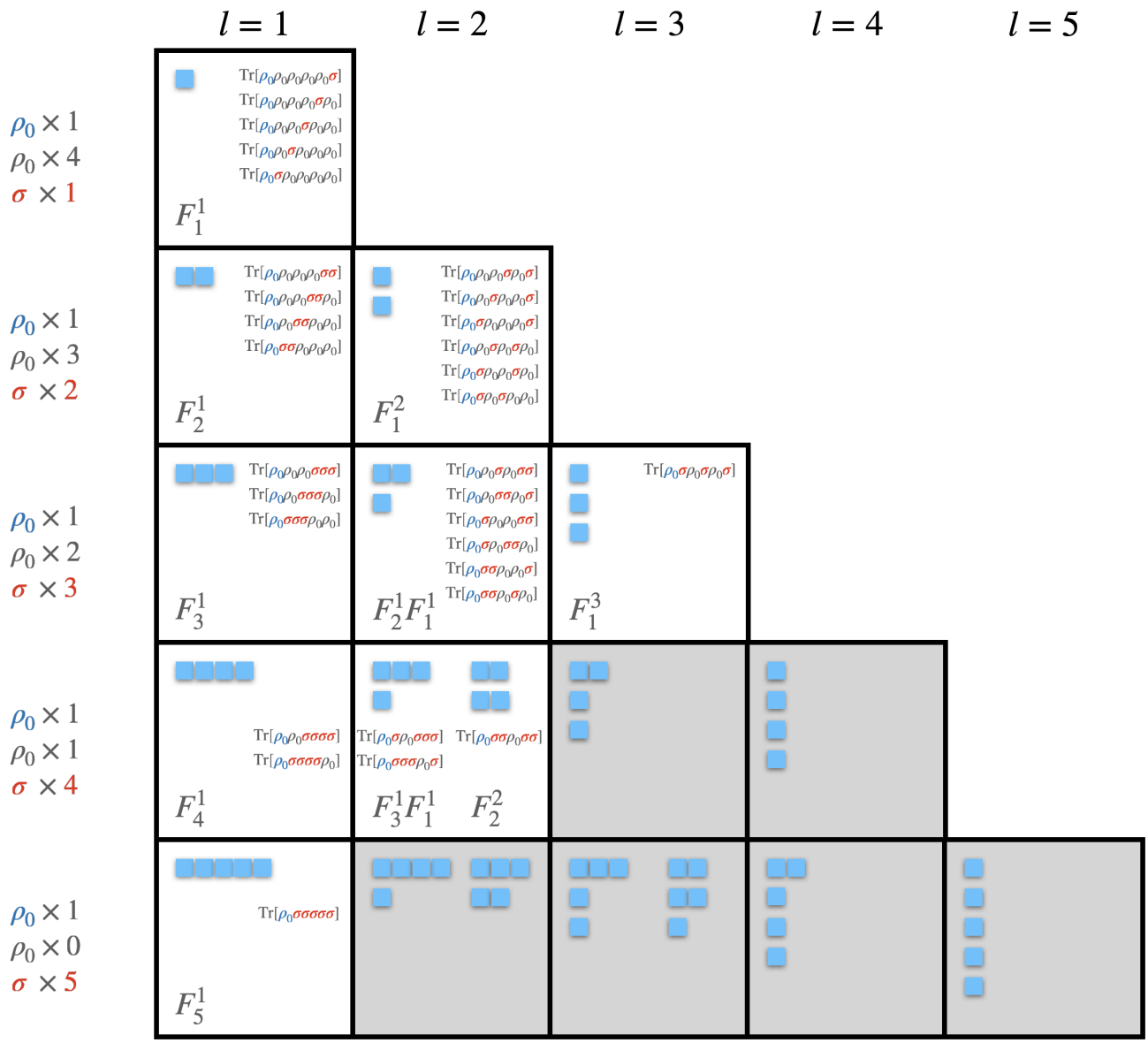


FIG. 14. The table of Young diagrams for $M = 5$. The grey cells are not counted (or counted as 0) in the summation due to the restriction $\#(\rho_0) + \#(\sigma) = M$ (or to the definition ${}_nC_r = 0$ for $n < r$).

2. When ρ_0 and σ are commutative: $[\rho_0, \sigma] = 0$

Next, when the condition (2) holds, the fidelity can be simplified as

$$\prod_{k=1}^{i} F_k^{j_k} = \mathrm{Tr}\left[\rho_0^{M-i+1}\sigma^i\right] = \mathrm{Tr}\left[\rho_0\sigma^i\right] = F_i = \mathrm{Tr}\left[\rho_0\sigma\rho_0\sigma\cdots\rho_0\sigma\right] = F_1^i, \tag{B7}$$

which means $F_i = F_1^i$. Since $[\rho_0, \sigma] = 0$, ρ_0 and σ can be simultaneously diagonalised. Thus we can further put $\rho_0 = |\psi\rangle\langle\psi| = |\lambda_0\rangle\langle\lambda_0|$ and $\sigma = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$, which gives $F_1 = \lambda_0$. These settings simplify Eq. (B4) to

$$\begin{aligned}
\mathrm{Tr}\left[\rho_0 \rho^M\right] &= \sum_{i=0}^{M} (1-p)^{M-i} p^i {}_M C_i F_i \\
&= \sum_{i=0}^{M} (1-p)^{M-i} \left(pF_1\right)^i {}_M C_i \\
&= (1 - p + pF_1)^M \\
&= (1 - p(1-F_1))^M = (1 - p(1-\lambda_0))^M \\
&= (1 - pI_1)^M ,
\end{aligned} \tag{B8}$$

where we also defined the infidelity between ρ_0 and σ as $I_1 = 1 - F_1$. Note that

$$\begin{aligned}
\mathrm{Tr}\left[\rho^M\right] &= \mathrm{Tr}\left[\left((1-p)\rho_0 + p\sigma\right)^M\right] \\
&= \mathrm{Tr}\left[\sum_{i=0}^{M} {}_M C_i (1-p)^{M-i} p^i \rho_0^{M-i} \sigma^i\right] \\
&= \sum_{i=0}^{M} {}_M C_i (1-p)^{M-i} p^i \,\mathrm{Tr}\left[\rho_0 \sigma^i\right] + p^M \left(\mathrm{Tr}\left[\sigma^M\right] - \mathrm{Tr}\left[\rho_0 \sigma^M\right]\right) \\
&= (1 - p(1-F_1))^M + p^M \left(\mathrm{Tr}\left[\sigma^M\right] - F_1^M\right) \\
&= (1 - p(1-\lambda_0))^M + p^M \left(\sum_i \lambda_i^M - \lambda_0^M\right) \\
&= (1 - p(1-\lambda_0))^M + \sum_{i>0} (p\lambda_i)^M .
\end{aligned} \tag{B9}$$

Hence, replacing $\mathrm{Tr}\left[\rho_0 \rho^M\right]$ in Eq. (B3) with Eq. (B8) gives

$$\begin{aligned}
\tilde{F} &= \frac{F_1 + (M-1)\,\mathrm{Tr}\left[\rho_0 \rho^M\right]}{1 + (M-1)\,\mathrm{Tr}\left[\rho^M\right]} \\
&= \frac{\lambda_0 + (M-1)\left(1 - p(1-\lambda_0)\right)^M}{1 + (M-1)\left(1 - p(1-\lambda_0)\right)^M + (M-1)\sum_{i>0} (p\lambda_i)^M} \\
&= 1 - \frac{1 - \lambda_0 + (M-1)\sum_{i>0} (p\lambda_i)^M}{1 + (M-1)\left(1 - p(1-\lambda_0)\right)^M + (M-1)\sum_{i>0} (p\lambda_i)^M} \\
&= 1 - \tilde{I},
\end{aligned} \tag{B10}$$

where we have introduced $\tilde{I} = 1 - \tilde{F}$ as the state infidelity between ρ_0 and $\tilde{\rho}$.