# Quantum Error Mitigation Relying on Permutation Filtering

Yifeng Xiong, Soon Xin Ng, *Senior Member, IEEE*, and Lajos Hanzo, *Life Fellow, IEEE*

***Abstract*— Quantum error mitigation (QEM) is a class of promising techniques capable of reducing the computational error of variational quantum algorithms tailored for current noisy intermediate-scale quantum computers. The recently proposed permutation-based methods are practically attractive, since they do not rely on any *a priori* information concerning the quantum channels. In this treatise, we propose a general framework termed as permutation filters, which includes the existing permutation-based methods as special cases. In particular, we show that the proposed filter design algorithm always converge to the global optimum, and that the optimal filters can provide substantial improvements over the existing permutation-based methods in the presence of narrowband quantum noise, corresponding to large-depth, high-error-rate quantum circuits.**

***Index Terms*— Quantum error mitigation, permutation filtering, permutation symmetry, variational quantum algorithms.**

## NOTATIONS

- Scalars, vectors and matrices are represented by $x$, $\boldsymbol{x}$, and $\boldsymbol{X}$, respectively. Sets and operators are denoted as $\mathscr{X}$ and $\mathcal{X}$, respectively.
- The notations $\mathbf{1}_n$, $\mathbf{0}_n$, $\mathbf{0}_{m\times n}$, and $\boldsymbol{I}_k$, represent the $n$-dimensional all-one vector, the $n$-dimensional all-zero vector, the $m \times n$ dimensional all-zero matrix, and the $k \times k$ identity matrix, respectively.
- The notation $\|\boldsymbol{x}\|_p$ represents the $\ell_p$-norm of vector $\boldsymbol{x}$, and the subscript may be omitted when $p = 2$. For matrices, $\|\boldsymbol{A}\|_p$ denotes the matrix norm induced by the corresponding $\ell_p$ vector norm.
- The notation $[\boldsymbol{A}]_{i,j}$ denotes the $(i, j)$-th entry of matrix $\boldsymbol{A}$. For a vector $\boldsymbol{x}$, $[\boldsymbol{x}]_i$ denotes its $i$-th element. The submatrix obtained by extracting the $i_1$-th to $i_2$-th rows and the $j_1$-th to $j_2$-th columns from $\boldsymbol{A}$ is denoted as $[\boldsymbol{A}]_{i_1:i_2,j_1:j_2}$. The notation $[\boldsymbol{A}]_{:,i}$ represents the $i$-th column of $\boldsymbol{A}$, and $[\boldsymbol{A}]_{i,:}$ denotes the $i$-th row, respectively.
- The trace of matrix $\boldsymbol{A}$ is denoted as $\mathrm{Tr}\{\boldsymbol{A}\}$.
- The notation $\boldsymbol{A} \otimes \boldsymbol{B}$ represents the Kronecker product between matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.
- Pure states are denoted by "kets" $|\psi\rangle$, and their dual vectors are denoted by "bras" $\langle\psi|$.

Manuscript received May 21, 2021; revised September 28, 2021 and November 29, 2021; accepted December 1, 2021. Date of publication December 6, 2021; date of current version March 17, 2022. Lajos Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/P034284/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028). This work is also supported in part by China Scholarship Council (CSC). The associate editor coordinating the review of this article and approving it for publication was M. Safari. *(Corresponding author: Lajos Hanzo.)*

The authors are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3132914.

Digital Object Identifier 10.1109/TCOMM.2021.3132914

## I. INTRODUCTION

QUANTUM technologies have entered the era of noisy intermediate-scale quantum (NISQ) computation [1]. These computers typically rely on dozens to a few hundreds of qubits. Remarkably, NISQ computers based on both superconductive [2] and photonic technologies [3], have shown quantum advantage in computing certain tasks.

However, NISQ computers may not afford fully fault-tolerant operations [4] enabled by quantum error correction codes [5]–[9], since the qubit overhead is still prohibitive for state-of-the-art devices. Consequently, quantum algorithms requiring long coherence time, such as the quantum phase estimation algorithm [10] and the quantum amplitude amplification [11], [12], may not be practical for quantum computers available at the time of writing. Notably, these algorithms are often used as subroutines of more sophisticated quantum algorithms relying on the assumption of fault-tolerance, including Shor's factoring algorithm [13] and Grover's search algorithm [14]–[16]. This suggests that a paradigm shift both for algorithm design and for error control techniques might be necessary for NISQ computers.

As proposed in [17], variational quantum algorithms [17]–[21] constitute one of the new algorithm design paradigms harnessing the computational power of NISQ computers without relying on quantum error correction techniques, including the celebrated variational quantum eigensolver [17] and the quantum approximate optimization algorithm (QAOA) [19]. Specifically, the eigenvalue evaluation subroutine, which is typically realized using the quantum phase estimation algorithm in "traditional" quantum algorithms, is implemented in variational quantum algorithms by directly measuring the corresponding quantum observables [22]. The workflow of a typical variational quantum algorithm is portrayed in Fig. 1. To elaborate further, these algorithms aim for designing parametric state preparation circuits using an iterative, hybrid quantum-classical optimization procedure, that output (approximate) eigenstates of the Hamiltonian encoding the computational task. The eigenvalues can then be estimated by directly measuring the observables. By contrast, in the quantum phase algorithm,

0090-6778 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

TABLE I
COMPARISONS BETWEEN DIFFERENT QEM METHODS

| | Main overhead | Prior knowledge required | Remark |
|---|---|---|---|
| Zero-noise Extrapolation [24]–[26] | Sampling overhead | No | Requires pulse-level control |
| Channel inversion [24], [27], [28] | Sampling overhead | Channel estimation (gate set tomography) | Has error floor due to imperfect channel estimation |
| Learning-based [29]–[31] | Sampling overhead | Pre-training on certain circuit sets | |
| Symmetry verification [32], [33] | Sampling overhead, qubit overhead | Type of symmetries in the computational task | **Symmetry-based** |
| Virtual distillation [34], [35] | Sampling overhead, qubit overhead | No | **Symmetry-based**; Has error floor due to mismatch in the dominant eigenvector [36] |
| **This treatise** | Sampling overhead, qubit overhead | No | **Improves the accuracy of VD at a similar overhead** |

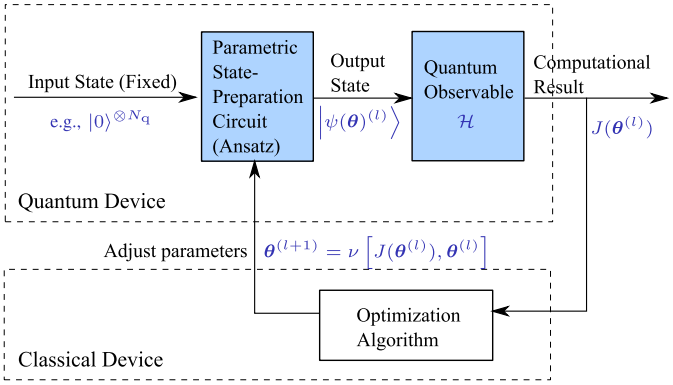



Fig. 1. The workflow of a typical variational quantum algorithm.

the Hamiltonian simulation [23] subroutine is executed for $O(1/\epsilon)$ times, where $\epsilon$ denotes the required accuracy, hence the coherence time requirements of physical qubits are more strict than that of variational algorithms.

Despite that the parametric state preparation circuits in variational quantum algorithms have relatively short depth (compared to that of the quantum phase estimation algorithm), they can still be so deep that the imperfections of the circuits accumulate to an amount that lead to significant computational errors. This calls for effective error control methods that do not rely on the qubit-demanding fault-tolerant scheme. One of the most representative error control strategy conceived for NISQ computers is quantum error mitigation (QEM) [24] tailored for variational quantum algorithms. Typically, QEM methods mitigate the error with the aid of classical post-processing. This reduces both the additional errors introduced by error control quantum operations as well as the qubit overhead represented by the number of ancillas used in quantum error correction.

Broadly speaking, there have been four types of QEM methods. One of them collects the computational results produced by circuits having different error rates, and then extrapolates the results to the point where the error rate tends to zero [24]–[26]. Another idea is to construct a set of probabilistic quantum circuits effectively implementing the inverse of the error operator (also known as the quantum channel) [24], [27], [28]. There have also been learning-based methods that mitigate the error of practical sophisticated circuits using statistical models that pre-trained on Clifford circuits, which have known efficient simulation algorithms on classical computers [30], [31]. The fourth concept exploits the symmetry (redundancy) of the quantum states or the computational task itself for mitigating the error rate, by preventing the states that do not satisfy certain symmetry conditions from contributing to the computational result [32], [33]. The characteristics of the QEM methods are summarized in Table I. In general, these methods are not mutually exclusive in practical applications. Instead, potentially beneficial combinations have been conceived [34]. For a comprehensive comparison between these methods, interested readers may refer to [37].

Recently, a new class of symmetry-aided QEM methods, namely the virtual distillation (VD) [34], [35], has been proposed, which relies on the permutation symmetry of quantum states. To elaborate, they prepare multiple copies of the same quantum state, and filter out the components in the states that are not identical across all copies, as shown in Fig. 2. The observables are then measured on one of the copies. Compared to previous QEM methods, the advantage of these techniques is that they do not require *a priori* knowledge about the quantum channels, and that the symmetry of the states can be easily manipulated by adjusting the number of copies.

From the spectral analysis perspective of quantum states, when the noise is not extremely strong, the dominant eigenvector of the output state serves as a good approximation of the ideal noise-free output state [34]. In this sense, the permutation-based QEM methods may be viewed as high-pass filters in the spectral domain. In this treatise, we generalize this idea by proposing a general framework for designing optimal filters in the spectral domain of quantum states. These filters assume a similar form as the finite impulse response (FIR) filters widely used in classical signal processing tasks, by computing a weighted average over the outputs of multiple virtual distillation circuits of different orders, as shown in Fig. 3. Our novel contributions are summarized below.

- We propose a general permutation filter design framework, including the functional form of the filters and the performance metric to be optimized. We will show that existing permutation-based QEM methods may be viewed as specific cases of permutation filters.

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

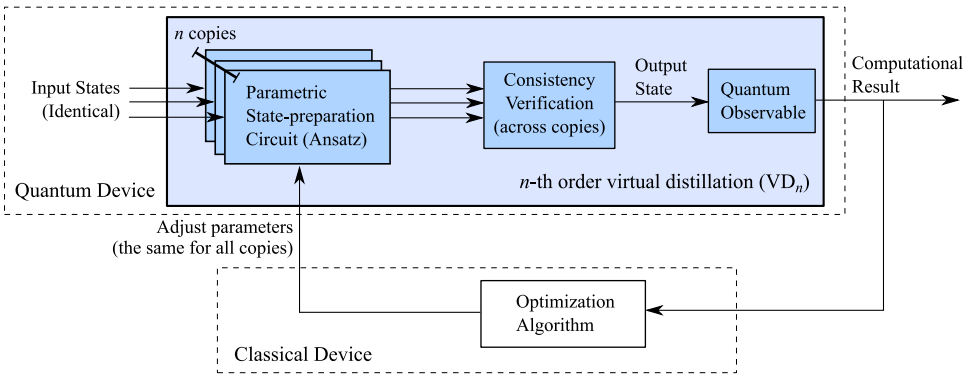



Fig. 2. An $n$-th order virtual distillation method (relying on $n$ copies of the parametric state-preparation circuits) applied to a variational quantum algorithm.

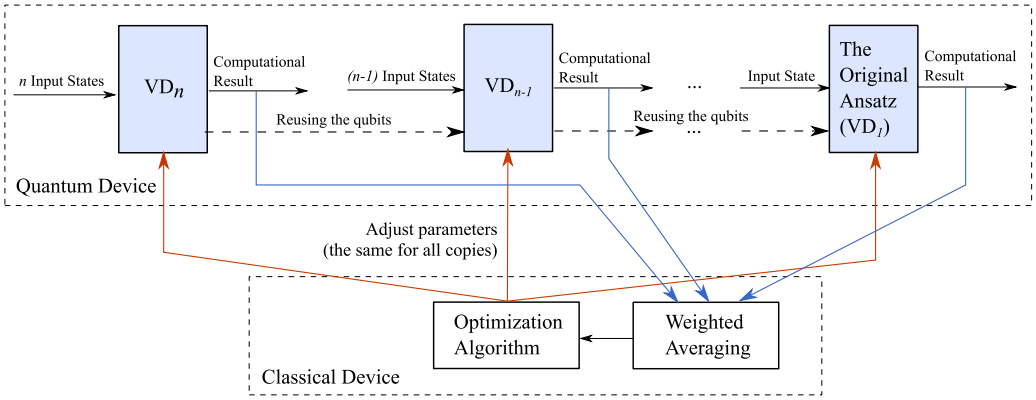



Fig. 3. An $n$-th order permutation filter proposed in this treatise applied to a variational quantum algorithm.

- We propose an algorithm for optimal permutation filter design. In particular, we show that the local optimum of the optimization problem is unique, hence the global optimal solution is attainable by the proposed algorithm.
- We show that permutation filters are particularly efficient in combating narrowband noise. Specifically, they are capable of providing an error-reduction improvement scaling polynomially with respect to the noise bandwidth, compared to the existing permutation-based QEM methods.
- We also show that the noise bandwidth decreases exponentially with the depth of the quantum circuit. This suggests that the proposed permutation filters can be used for supporting the employment of quantum circuits having an increased depth without degrading their fidelity.

The rest of this treatise is organized as follows. In Section II we provide a brief introduction to variational quantum algorithms and permutation-based QEM methods. In Section III, we describe the permutation filter as well as its design algorithm. Then, in Section IV we analyze the error-reduction performance of permutation filters. The results are further illustrated using numerical results in Section V. Finally, we conclude the paper in Section VI.

## II. Preliminaries

### A. Variational Quantum Algorithms

Variational quantum algorithms constitute a class of hybrid quantum-classical algorithms [38] tailored for NISQ computers, which aim for solving optimization problems of the following form

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}), \quad J(\boldsymbol{\theta}) = \langle \psi(\boldsymbol{\theta}|) \mathcal{H} | \psi(\boldsymbol{\theta}) \rangle, \tag{1}$$

where $\mathcal{H}$ is the Hamiltonian encoding the optimization cost function, and the mapping from $\boldsymbol{\theta}$ to the quantum state $|\psi(\boldsymbol{\theta})\rangle$ is implemented by a parametric state preparation circuit, also known as the *ansatz* [39].

When we work on qubits, it is often convenient to decompose the Hamiltonian into a weighted sum of Pauli operators (so-called "Pauli-strings" defined in [40]). In particular, a Hamiltonian acting upon $N_{\mathrm{q}}$ qubits may be expressed as

$$\mathcal{H} = \sum_{i=1}^{4^{N_{\mathrm{q}}}} w_i \mathcal{S}_i^{(N_{\mathrm{q}})}, \tag{2}$$

where $\mathcal{S}_i^{(N_{\mathrm{q}})}$ denotes the $i$-th Pauli string acting upon $N_{\mathrm{q}}$ qubits, given by

$$\mathcal{S}_i^{(N_{\mathrm{q}})} = \bigotimes_{j=1}^{N_{\mathrm{q}}} \mathcal{S}^{(1)}_{\mathrm{digit}(i,j)+1}, \tag{3}$$

where $\mathrm{digit}(i,j)$ represents the $j$-th digit of $i$ when treated as a base-4 number. The single-qubit Pauli operators $\mathcal{S}_k^{(1)}$,

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

$k = 1, 2, 3, 4$, are given by

$$\mathcal{S}_1^{(1)} = \boldsymbol{S}_{\mathcal{I}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathcal{S}_2^{(1)} = \boldsymbol{S}_{\mathcal{X}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$
$$\mathcal{S}_3^{(1)} = \boldsymbol{S}_{\mathcal{Y}} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \mathcal{S}_4^{(1)} = \boldsymbol{S}_{\mathcal{Z}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The number of $t_i$ values satisfying $t_i \neq 1$ is called the weight $\omega(\mathcal{S}_{p(\boldsymbol{t})})$ of the Pauli string $\mathcal{S}_{p(\boldsymbol{t})}$, and in general we have $1 < \omega(\mathcal{S}_{p(\boldsymbol{t})}) \leq N_{\mathrm{q}}$.

In variational quantum algorithms, the observation of the complicated Hamiltonian $\mathcal{H}$ is implemented by a set of observations of the corresponding Pauli strings, as follows [22]:

$$\langle \psi(\boldsymbol{\theta}|)\mathcal{H}|\psi(\boldsymbol{\theta}\rangle) = \sum_{i=1}^{4^{N_{\mathrm{q}}}} w_i \langle \psi(\boldsymbol{\theta}|)\mathcal{S}_i^{(N_{\mathrm{q}})}|\psi(\boldsymbol{\theta}\rangle). \tag{4}$$

To take full advantage of the computational power of both classical and quantum devices, the variational quantum algorithms solve the optimization problem in an iterative fashion as follows (also shown in Fig. 1):

$$J(\boldsymbol{\theta}^{(l)}) = \sum_{i=1}^{4^{N_{\mathrm{q}}}} w_i \langle \psi(\boldsymbol{\theta}|^{(l)})\mathcal{S}_i^{(N_{\mathrm{q}})}|\psi(\boldsymbol{\theta}\rangle^{(l)}), \tag{5a}$$
$$\boldsymbol{\theta}^{(l+1)} = \nu\left[J(\boldsymbol{\theta}^{(l)}), \boldsymbol{\theta}^{(l)}\right], \tag{5b}$$

where $\nu\left[J(\boldsymbol{\theta}^{(l)}), \boldsymbol{\theta}^{(l)}\right]$ is an update rule for the parameters defined by the specific algorithm. This hybrid quantum-classical optimization procedure aims for finding the optimal eigenvalue using short-depth circuits, thus avoiding the strict coherence time requirements of the quantum phase estimation algorithm.

In practice, the state preparation circuit outputs are contaminated by decoherence, which turns the output states into a mixed form. Hence, the practical version of (5a) is given by

$$\tilde{J}_l(\boldsymbol{\theta}^{(l)}) = \sum_{i=1}^{4^{N_{\mathrm{q}}}} w_i \mathrm{Tr}\left\{\rho(\boldsymbol{\theta}^{(l)})\mathcal{S}_i^{(N_{\mathrm{q}})}\right\}, \tag{6}$$

where $\rho(\boldsymbol{\theta}^{(l)})$ is a mixed state, as opposed to the pure state $|\psi(\boldsymbol{\theta}\rangle)$ of the previous discussion. Apparently, the noisy cost function $\tilde{J}_l(\cdot)$ would be different from the ideal cost function $J(\cdot)$, and hence their values at the specific parameter $\boldsymbol{\theta}^{(l)}$ would also be different. The difference will become more significant when the state preparation circuit is more complex (i.e., either deep or involves a large number of qubits). This necessitates the employment of quantum error mitigation, which aims for "purifying" the mixed state $\rho(\boldsymbol{\theta}^{(l)})$, in order to mitigate the contamination of the computed cost function values.

### *B. Permutation-Based Quantum Error Mitigation*

The permutation-based quantum error mitigation philosophy is inspired by the concept of permutation tests, which constitute generalizations of the swap test [41]. As portrayed in Fig. 4a, the swap test is implemented by controlled-SWAP gates. It is widely employed for evaluating the overlap between a pair of quantum states $\rho$ and $\sigma$, since the expected value of the measurement outcome is given by $\mathrm{Tr}\{\rho\sigma\}$. Naturally, when we have two copies of the same state $\rho$, we may compute $\mathrm{Tr}\{\rho^2\}$ using the swap test.

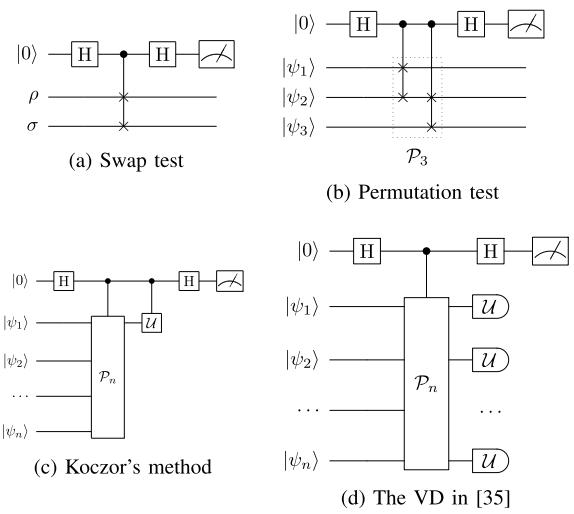



Fig. 4. Schematics of the swap test, the permutation test, and two circuit implementations of the virtual distillation method.

The permutation tests, exemplified by the cyclic-shift test [42], may be implemented using quantum circuits taking the form shown in Fig. 4b. As a generalization of the swap gate, an $n$-th order cyclic-shift circuit $\mathcal{P}_n$ taking an input of $n$ pure states $|\psi_1, \psi_2, \ldots, \psi_n\rangle$ would output a shifted state $|\psi_2, \psi_3, \ldots, \psi_n, \psi_1\rangle$. Note that the swap gate may be viewed as a specific case of cyclic-shift circuit, since it is equivalent to $\mathcal{P}_2$. Similar to the swap test, one may show that the expectation value of the outcome in an $n$-th order cyclic-shift test is given by $\mathrm{Tr}\{\rho^n\}$ [34], when the inputs are represented by $n$ copies of the same mixed state $\rho$.

Typically, when quantum circuits are contaminated by decoherence, the output state would approximately take the following form

$$\rho = \lambda_1 |\psi\rangle\langle\psi| + \sum_{i=2}^{2^{N_{\mathrm{q}}}} \lambda_i |\psi_i\rangle\langle\psi_i|, \tag{7}$$

where $|\psi_i\rangle$ denotes the eigenvector associated with the $i$-th largest eigenvalue of $\rho$, and $|\psi\rangle = |\psi_1\rangle$ is the dominant eigenvector, which approximates the noise-free output state [34], [36]. Inspired by these observations, Koczor [34] proposed the permutation-based quantum error mitigation concept (which has later been generalized to the concept of VD [35]), as portrayed in Fig. 4c. Compared to the permutation test shown in Fig. 4b, it may be observed that the output of the VD circuit for a given unitary observable $\mathcal{U}$ is given by

$$\tilde{y}_{\mathrm{VD}}^{(n)} = \mathrm{Tr}\left\{\rho^n \mathcal{U}\right\}, \tag{8}$$

where $n$ is the order of the circuit $\mathcal{P}_n$, and we will also refer to it as the order of VD. Another implementation yielding the same result as in (8) is proposed in [35], as shown in Fig. 4d. This implementation enables simultaneous measurement of multiple compatible observables, and thus reduces the total number of circuit repetitions. Note that all Pauli strings are unitary observables, hence they can be nicely fit into

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

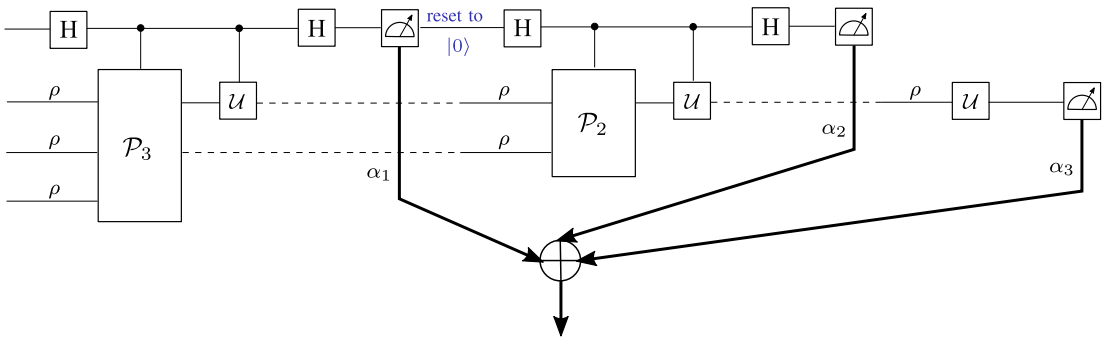



Fig. 5. Schematic of a third-order permutation filter $\mathrm{Tr}\{(\alpha_1\rho^3+\alpha_2\rho^2+\alpha_3\rho)\mathcal{U}\}$.

this framework. Next, upon replacing the observable $\mathcal{U}$ by the identity operator $\mathcal{I}$ (i.e., the original $n$-th order permutation test), one may also compute $\mathrm{Tr}\{\rho^n\}$, and obtain the final result[1]

$$y_{\mathrm{VD}}^{(n)}=\frac{\tilde{y}_{\mathrm{VD}}^{(n)}}{\mathrm{Tr}\{\rho^n\}}=\frac{\mathrm{Tr}\{\rho^n\mathcal{U}\}}{\mathrm{Tr}\{\rho^n\}}. \tag{9}$$

Note that

$$\tilde{y}_{\mathrm{VD}}^{(n)}=\lambda_1^n\langle\psi|\mathcal{U}|\psi\rangle+(1-\lambda_1)^n\sum_{i=2}^{2^{N_\mathrm{q}}}p_i^n\langle\psi_i|\mathcal{U}|\psi_i\rangle, \tag{10}$$

where $p_i=\lambda_i(1-\lambda_1)^{-1}$ satisfies $\sum_{i=2}^{N_\mathrm{q}}p_i=1$. When $\lambda_1$ is far larger than the other eigenvalues, it becomes clear from (10) that the term $(1-\lambda_1)^n$ decreases much more rapidly with $n$ than $\lambda_1^n$. Hence the contribution of the undesired components $|\psi_i\rangle,\ i>1$ to the final computation result is substantially reduced by VD.

## III. Permutation Filters

In this section, we propose a generalized version of virtual distillation, which will be referred to as "permutation filters". A third-order permutation filter is portrayed in Fig. 5. As it may be observed from the figure, the third-order filter consists of the third-order and the second-order VD circuits. In general, an $n$-th order permutation filter would contain all the $m$-th order VD circuits, where $m=2,3,\ldots,n$. Note that these circuits can be activated one after the other by reusing the same qubit resources, since the post-processing stage only involves a weighted averaging of the measured outcomes, which are classical quantities.

Formally, an $N$-th order permutation filter may be expressed as an $N$-th order polynomial of the input state $\rho$ formulated as

$$\mathcal{F}_{\boldsymbol{\alpha}}(\rho)=\sum_{n=1}^{N}\alpha_{N-n+1}\rho^n, \tag{11}$$

where $\boldsymbol{\alpha}=[\alpha_1\ \alpha_2\ \ldots\ \alpha_N]^{\mathrm{T}}\in\mathbb{R}^N$. Correspondingly, the eigenvalues of the output state are thus given by

$$h_{\boldsymbol{\alpha}}(\lambda)=\sum_{n=1}^{N}\alpha_{N-n+1}\lambda^n. \tag{12}$$

Observe that the function $h_{\boldsymbol{\alpha}}(\lambda)$ may be viewed as the "spectral response" of the filter, resembling the frequency response of conventional filters used in classical signal processing tasks. The final computational result with respect to an observable $\mathcal{U}$ is given by

$$y_{\mathrm{filter}}^{(N)}(\mathcal{U})=\frac{\mathrm{Tr}\{\mathcal{F}_{\boldsymbol{\alpha}}(\rho)\mathcal{U}\}}{\mathrm{Tr}\{\mathcal{F}_{\boldsymbol{\alpha}}(\rho)\}}. \tag{13}$$

The reason that we do not include the constant term $\alpha_{N+1}$ in (11) is that it does not contribute to the final computational results in (13) for most practical applications. To elaborate, consider the Pauli string decomposition (2) of observables used in variational quantum algorithms. Since the single-qubit Pauli operators except for the identity have a trace of zero, we have $\mathrm{Tr}\{\mathcal{U}\}=0$ for every Pauli string $\mathcal{U}$. Therefore, even if we include the constant coefficient $\alpha_{N+1}$ in our filter, it will not contribute to the final result, since we have:

$$\alpha_{N+1}\mathrm{Tr}\{\rho^0\mathcal{U}\}=0. \tag{14}$$

As for the term involving the identity operator, we could simply account for it by adding a constant to the final computational result, since $\mathrm{Tr}\{\rho\}=1$ always holds.

It is often convenient to design filters under an alternative parametrization, namely the pole-zero representation widely used in classical signal processing theory.[2] When considering "FIR-like" filters taking the form (11) (since there is no denominator in this formula), there are only zeros but no poles. Observe from (11) that the first zero is at $\beta=0$ due to the lack of the constant term. Upon denoting the remaining zeros by $\boldsymbol{\beta}=[\beta_1\ \ldots\ \beta_{N-1}]^{\mathrm{T}}$, we have

$$\mathcal{F}_{\boldsymbol{\beta}}(\rho)=\rho\prod_{n=1}^{N-1}(\rho-\beta_n\boldsymbol{I}), \tag{15}$$

[1]The accuracy of this normalization procedure may be further improved by replacing $\mathrm{Tr}\{\rho^n\}$ with $\lambda_1^n$. However, $\lambda_1$ is typically not known prior to the computation, and is also difficult to be computed exactly from the observations. By contrast, $\mathrm{Tr}\{\rho^n\}$ is readily obtainable by observing the identity operator.

[2]In classical signal processing theory, filters are represented by a ratio between two polynomials in the complex frequency domain. "Poles" refers to the roots of the denominator polynomial, while "zeros" refer to the roots of the numerator polynomial.

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

and

$$h_{\boldsymbol{\beta}}(\lambda) = \lambda \prod_{n=1}^{N-1} (\lambda - \beta_n). \tag{16}$$

The relationship between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is

$$\boldsymbol{\alpha} = \bigstar_{n=1}^{N-1} [1, \ -\beta_n]^{\mathrm{T}}, \tag{17}$$

where we define $\bigstar_{n=1}^{K} \boldsymbol{v}_n := \boldsymbol{v}_1 \star \boldsymbol{v}_2 \star \ldots \star \boldsymbol{v}_K$, and $\star$ denotes the discrete convolution given by

$$[\boldsymbol{x} \star \boldsymbol{y}]_n = \sum_{i=\max\{1,k+1-n\}}^{\min\{k,m\}} x_i y_{k-i+1},$$

where $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{y} \in \mathbb{R}^n$, and $\boldsymbol{x} \star \boldsymbol{y} \in \mathbb{R}^{m+n-1}$. Without loss of generality, we assume that

$$\beta_1 \leq \beta_2 \leq \ldots \leq \beta_{N-1}. \tag{18}$$

### A. The Performance Metric of Permutation Filter Design

For a given observable $\mathcal{U}$, we would hope to minimize the estimation error

$$\begin{aligned} \epsilon_{\mathcal{U}}(\boldsymbol{\beta}) &= \left| y_{\mathrm{filter}}^{(N)}(\mathcal{U}) - \langle\psi|\mathcal{U}|\psi\rangle \right| \\ &= \left| \frac{\frac{1}{h_{\boldsymbol{\beta}}(\lambda_1)} \sum_{i=2}^{2^{N_\mathrm{q}}} h_{\boldsymbol{\beta}}(\lambda_i) \left( \langle\psi_i|\mathcal{U}|\psi_i\rangle - \langle\psi|\mathcal{U}|\psi\rangle \right)}{1 + [h_{\boldsymbol{\beta}}(\lambda_1)]^{-1} \sum_{i=2}^{2^{N_\mathrm{q}}} h_{\boldsymbol{\beta}}(\lambda_i)} \right|. \end{aligned} \tag{19}$$

However, in a typical variational quantum algorithm, a large number of unitary observables $\mathcal{U}_1, \ldots, \mathcal{U}_{N_{\mathrm{ob}}}$ would have to be evaluated. In light of this, we consider the minimization of the following upper bound

$$\begin{aligned} \epsilon_{\mathcal{U}}(\boldsymbol{\beta}) &\leq \epsilon(\boldsymbol{\beta}) \\ &= \frac{2}{h_{\boldsymbol{\beta}}(\lambda_1)} \left\| \boldsymbol{h}_{\boldsymbol{\beta}}(\tilde{\boldsymbol{\lambda}}) \right\|_1, \end{aligned} \tag{20}$$

where $\tilde{\boldsymbol{\lambda}} = [\boldsymbol{\lambda}]_{2:2^{N_\mathrm{q}}}$, and $\boldsymbol{\lambda} = [\lambda_1 \ \ldots \ \lambda_{2^{N_\mathrm{q}}}]^{\mathrm{T}}$.

If we know *a priori* the distribution of $\tilde{\boldsymbol{\lambda}}$, or in other words, the spectral density of $\rho$ (excluding the dominant eigenvalue), we may directly minimize the cost function $\epsilon(\boldsymbol{\beta})$ as follows:

$$\begin{aligned} &\min_{\boldsymbol{\beta}} \ \epsilon(\boldsymbol{\beta}), \\ &\text{s.t. } \boldsymbol{\beta} \in \mathscr{B}, \ (18), \end{aligned} \tag{21}$$

where $\epsilon(\boldsymbol{\beta})$ can be rewritten as

$$\epsilon(\boldsymbol{\beta}) = \frac{1}{\lambda_1 \prod_{n=1}^{N-1} (\lambda_1 - \beta_n)} \int_{\lambda_\mathrm{m}}^{1} \left| \lambda \prod_{n=1}^{N-1} (\lambda - \beta_n) \right| f(\lambda) \mathrm{d}\lambda,$$

$\lambda_\mathrm{m} > 0$ denotes the minimum value of $\lambda$, and $f(\lambda)$ denotes the spectral density. The feasible region $\mathscr{B}$ is given by

$$\mathscr{B} = \{\boldsymbol{\beta} | \boldsymbol{\beta} \succeq \mathbf{0}, \beta_1 \leq \beta_2 \leq \ldots \leq \beta_{N-1}\}.$$

For most practical scenarios, we have $\beta_i \ll \lambda_1$, hence $\epsilon(\boldsymbol{\beta})$ may be approximated as

$$\epsilon(\boldsymbol{\beta}) \approx \tilde{\epsilon}(\boldsymbol{\beta}) = \int_{\lambda_\mathrm{m}}^{1} \left| \lambda \prod_{n=1}^{N-1} (\lambda - \beta_n) \right| f(\lambda) \mathrm{d}\lambda, \tag{22}$$

since the optimal solution is hardly affected by the denominator. Note that $h_{\boldsymbol{\beta}}(1)$ is always positive, hence we may further simplify the approximated objective function as follows:

$$\begin{aligned} \tilde{\epsilon}(\boldsymbol{\beta}) &= \int_{\lambda_\mathrm{m}}^{1} |G_{\boldsymbol{\beta}}(\lambda)| \, \mathrm{d}\lambda \\ &= \sum_{i=0}^{N-1} (-1)^i \int_{\beta_{N-i-1}}^{\beta_{N-i}} G_{\boldsymbol{\beta}}(\lambda) \mathrm{d}\lambda, \end{aligned} \tag{23}$$

where $G_{\boldsymbol{\beta}}(\lambda) = f(\lambda)\lambda \prod_{n=1}^{N-1}(\lambda - \beta_n)$, and additionally we define $\beta_N = 1$ and $\beta_0 = \lambda_\mathrm{m}$.

### B. Practical Permutation Filter Design Algorithms

When $f(\lambda)$ is known exactly, we may directly solve the optimization problem discussed in the previous subsection. However, for practical applications, $f(\lambda)$ is never known precisely; it has to be estimated from observations. In this treatise, we fit Pareto distribution [43], [44] to $f(\lambda)$ which is formulated as:

$$f(\lambda) = k\lambda_\mathrm{m}^k \lambda^{-(k+1)}, \tag{24}$$

where $k > 2$ is a shape parameter.

The reason for using the Pareto distribution is two-fold. First of all, it approximates our empirical observations concerning the output spectra of noisy quantum circuits quite closely. Secondly, it fits nicely with the polynomial form of the permutation filter, making the design problem more tractable. Specifically, under the parametrization of the Pareto distribution, the indefinite integral of $G_{\boldsymbol{\beta}}(\lambda)$ can be explicitly calculated as follows:

$$\begin{aligned} \tilde{G}_{\boldsymbol{\alpha}}(\lambda) &= \frac{1}{k\lambda_\mathrm{m}^k} \int G_{\boldsymbol{\beta}}(\lambda) \mathrm{d}\lambda \\ &= \int \lambda^{-k} \prod_{n=1}^{N-1} (\lambda - \beta_n) \mathrm{d}\lambda \\ &= \sum_{n=1}^{N} \frac{\alpha_{N-n+1}}{n-k} \cdot \lambda^{n-k}. \end{aligned} \tag{25}$$

The definite integrals in (23) can then be obtained as

$$\int_{\beta_i}^{\beta_{i+1}} G_{\boldsymbol{\beta}}(\lambda) \mathrm{d}\lambda = k\lambda_\mathrm{m}^k \left( \tilde{G}_{\boldsymbol{\alpha}}(\beta_{i+1}) - \tilde{G}_{\boldsymbol{\alpha}}(\beta_i) \right). \tag{26}$$

Note that for an $N$-th order permutation filter, we may obtain $N-1$ observations $\boldsymbol{m} = [m_1 \ \ldots \ m_{N-1}]^{\mathrm{T}}$ where $m_i = \mathrm{Tr}\left\{\rho^{i+1}\right\}$. These observations can be used to fit the Pareto distribution to $f(\lambda)$ using the method of moments [45]. For example, when $N = 3$, the equations of moments are given by

$$\begin{aligned} \frac{1 - \hat{\lambda}_1(\boldsymbol{m})}{2^{N_\mathrm{q}} - 1} &= \frac{k\lambda_\mathrm{m}}{k-1}, \\ \frac{m_1 - \hat{\lambda}_1(\boldsymbol{m})^2}{2^{N_\mathrm{q}} - 1} &= \frac{k\lambda_\mathrm{m}^2}{k-2}, \end{aligned} \tag{27}$$

where $\hat{\lambda}_1(\boldsymbol{m})$ is an estimate of $\lambda_1$. Here, the quantities $\frac{1-\hat{\lambda}_1(\boldsymbol{m})}{2^{N_\mathrm{q}}-1}$ and $\frac{m_1-\hat{\lambda}_1(\boldsymbol{m})^2}{2^{N_\mathrm{q}}-1}$ are estimates of the mean value and the variance of the spectrum, respectively. We do not

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

use the conventional sample mean and variance, because the eigenvalues cannot be sampled directly. A natural choice of $\hat{\lambda}_1(\boldsymbol{m})$ for an $N$-th order filter is

$$\hat{\lambda}_1(\boldsymbol{m}) = \|\boldsymbol{\lambda}\|_N = (m_{N-1})^{\frac{1}{N}}, \tag{28}$$

which is asymptotically exact as $N \to \infty$, since $\lambda_1 = \|\boldsymbol{\lambda}\|_\infty$.

Using the equations of moments in (27), we may then estimate the unknown parameters $k$ and $\lambda_{\mathrm{m}}$. However, for the $N = 2$ case, the method of moments would encounter an identifiability problem, since the number of observations (one) is less than the number of parameters (two). Fortunately, we may obtain the closed-form solution of $\beta_1$ as follows:

$$\beta_1 = \lambda_{\mathrm{m}} \left[2(1 + \lambda_{\mathrm{m}}^{k-1})^{-1}\right]^{\frac{1}{k-1}}, \tag{29}$$

which is obtained by taking the derivative of $\tilde{\epsilon}(\boldsymbol{\beta})$ with respect to $\beta_1$ and setting it to zero. For $k \geq 2$, $\beta_1$ can be closely approximated by

$$\beta_1 \approx \mu = k\lambda_{\mathrm{m}}(k-1)^{-1}, \tag{30}$$

where $\mu$ is the mean value of the Pareto distribution. This may be seen by neglecting the term $\lambda_{\mathrm{m}}^{k-1}$ (since typically $\lambda_{\mathrm{m}}^{k-1} \ll 1$ when $k \geq 2$), and noticing that the ratio $\mu/\beta_1$ is then approximately (approximately because of neglecting $\lambda_{\mathrm{m}}^{k-1}$) bounded by

$$1 \lessapprox \mu/\beta_1 \lessapprox \frac{2^{-\frac{1-\ln 2}{\ln 2}}}{\ln 2} \approx 1.062,$$

where the lower bound is attained at $k = 2$ and the upper bound is attained at $k = (1 - \ln 2)^{-1}$. The mean value $\mu$ may then be estimated by

$$\hat{\mu} = \frac{1 - \hat{\lambda}_1(\boldsymbol{m})}{2^{N_{\mathrm{q}}} - 1}. \tag{31}$$

For the $N > 2$ case, it is difficult to obtain closed-form solutions of $\boldsymbol{\beta}$. Furthermore, in general, the optimization problem with respect to $\boldsymbol{\beta}$ may no longer be convex. Fortunately, in the following proposition we show that $\tilde{\epsilon}(\boldsymbol{\beta})$ satisfies a generalized convexity property, which guarantees that the global optimum is always attainable.

*Proposition 1 Invexity of the Permutation Filter Design Problem:* The cost function $\tilde{\epsilon}(\boldsymbol{\beta})$ in (21) is an invex[3] function of $\boldsymbol{\beta}$ in the convex feasible region $\mathscr{B}$. In other words, every stationary point of $\tilde{\epsilon}(\boldsymbol{\beta})$ in $\mathscr{B}$ is a global minimum.

*Proof:* Please refer to Appendix I. ∎

Proposition 1 implies that the following simple projected gradient descent iteration rule

$$\begin{aligned} \tilde{\boldsymbol{\beta}}^{(\ell+1)} &= \boldsymbol{\beta}^{(\ell)} - \delta^{(\ell)} \cdot \left.\frac{\partial \tilde{\epsilon}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}^{(\ell)}}, \\ \boldsymbol{\beta}^{(\ell+1)} &= \mathcal{T}_{\mathscr{B}}\left[\tilde{\boldsymbol{\beta}}^{(\ell+1)}\right], \end{aligned} \tag{32}$$

may be used to solve the problem in (21), despite that $\tilde{\epsilon}(\boldsymbol{\beta})$ may not be convex with respect to $\boldsymbol{\beta}$. The operator $\mathcal{T}_{\mathscr{B}}(\cdot)$ projects its argument onto the convex feasible region $\mathscr{B}$, which can be implemented by simply sorting the entries of $\boldsymbol{\beta}$ after each iteration. The step size parameter $\delta^{(l)}$ can be determined using classic line search methods [47]. More sophisticated methods, such as modified Newton's method specifically tailored for invex optimization [48], may also be applied to accelerate the convergence.

[3]Invexity is a generalization of convexity, ensuring that the global optimal solutions can be found by using the Karush-Kuhn-Tucker conditions [46].

**Algorithm 1** Type-2 Permutation Filter Design

**Input:** Spectral density parameters $k$ and $\lambda_{\mathrm{m}}$
**Output:** The filter weight vector $\boldsymbol{\alpha}$

```
ℓ = 0; Initialize β^(0);
repeat
    Compute (∂ε̃(β)/∂β)|_{β^(ℓ)} using (25), (26), (33) and (34);
    Determine δ^(ℓ) using line search methods;
    Update β^(ℓ+1) = β^(ℓ) − δ^(ℓ) · (∂ε̃(β)/∂β)|_{β^(ℓ)};
    Sort the entries in β^(ℓ+1) in the ascending order;
    ℓ = ℓ + 1;
until convergence conditions are met
Compute α = φ(β^(ℓ)) using (17);
return α
```

According to our discussion in Appendix I, the cost function $\xi(\boldsymbol{\alpha})$ is a convex function of $\boldsymbol{\alpha}$. The reason that we do not solve directly this convex problem is that it is a challenge to differentiate the cost function $\xi(\boldsymbol{\alpha})$. By contrast, it is relatively simple to compute the gradient $\frac{\partial}{\partial \boldsymbol{\beta}}\tilde{\epsilon}(\boldsymbol{\beta})$, as follows:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}}\tilde{\epsilon}(\boldsymbol{\beta}) &= \sum_{i=0}^{N-1}(-1)^i \frac{\partial}{\partial \boldsymbol{\beta}} \int_{\beta_{N-i-1}}^{\beta_{N-i}} G_{\boldsymbol{\beta}}(\lambda)\mathrm{d}\lambda \\ &= \sum_{i=0}^{N-1}(-1)^i \int_{\beta_{N-i-1}}^{\beta_{N-i}} \frac{\partial}{\partial \boldsymbol{\beta}} G_{\boldsymbol{\beta}}(\lambda)\mathrm{d}\lambda \\ &= \sum_{i=0}^{N-1}(-1)^{i+1} \int_{\beta_{N-i-1}}^{\beta_{N-i}} \boldsymbol{g}_{\boldsymbol{\beta}}(\lambda)\mathrm{d}\lambda, \end{aligned} \tag{33}$$

where $[\boldsymbol{g}_{\boldsymbol{\beta}}(\lambda)]_i = \lambda^{-k}\prod_{\substack{n=1 \\ n\neq i}}^{N-1}(\lambda - \beta_n)$. The order between the integration and the differentiation is interchangeable, since $G_{\boldsymbol{\beta}}(\lambda) = 0$ for $\lambda = \beta_i,\ \forall i = 1, 2, \ldots, N-1$. The integrals can be computed using (25) and (26), but for $[\boldsymbol{g}_{\boldsymbol{\beta}}(\lambda)]_i$ the vector $\boldsymbol{\alpha}$ should be replaced by

$$\tilde{\boldsymbol{\alpha}}_i = \bigstar_{n=1, n\neq i}^{N-1} [1 \ -\beta_n]^{\mathrm{T}}. \tag{34}$$

When low-complexity methods are preferred, a simple heuristic alternative, which will be referred to as the "Type-1 permutation filter", is to set

$$\beta_1 = \beta_2 = \ldots = \beta_{N-1} = \mu. \tag{35}$$

Correspondingly, we refer to the aforementioned optimization-based method, summarized in Algorithm 1, as the "Type-2 permutation filter". In Section IV we will show that, even though the Type-1 filters rely on a heuristic method, they are capable of outperforming VD.

To conclude, the complete workflow of an $N$-th order permutation filter for a given observable $\mathcal{U}$ consists of the following steps:

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

1) Execute the original circuit and obtain the estimate of $\mathrm{Tr}\{\rho\mathcal{U}\}$;
2) Execute all $n$-th order virtual distillation circuits ($2 \leq n \leq N$), and obtain the estimates of $\mathrm{Tr}\{\rho^n\mathcal{U}\}$ as well as additional observations $m_{n-1} = \mathrm{Tr}\{\rho^n\}$;
3) Fit the spectral density model using the observations $\boldsymbol{m} = [m_1 \ \ldots \ m_{N-1}]^{\mathrm{T}}$, and determine the filter parameters $\boldsymbol{\alpha}$;
4) Obtain the final filtered result by classical post-processing.

### C. The Computational Overhead of Permutation Filters

In terms of the number of gates, the computational overhead of permutation filters is the same as virtual distillation. The number of gates in order to implement the permutation operation $\mathcal{P}_n$ (which is the additional gate cost of the protocol compared to the unprotected circuit) has been discussed in [34]. Specifically, if the original unprotected circuit acts on $N_{\mathrm{q}}$ qubits, implementing $\mathcal{P}_n$ would require $N_{\mathrm{q}}(n-1)$ controlled-SWAP gates (i.e. the Fredkin gate), which is on the order of $O(N_{\mathrm{q}})$. Hence we may conclude that the method would be beneficial when the algorithm circuit has an increasing depth with respected to $N_{\mathrm{q}}$.

As for the sampling overhead, permutation filters are slightly different from virtual distillation due to the weighted averaging process. For virtual distillation, an approximate expression for the variance of a given observable $\mathcal{U}$ has been presented in [35]. Using similar arguments, we may also obtain an expression for permutation filters as (36), as shown at the bottom of the page. The variance of the entire Hamiltonian $\mathcal{H}$ can then be calculated by a weighted summation over the Pauli observables. In light of this, the sampling overhead factor of permutation filters may be defined as the ratio between the variance of the Hamiltonian estimator based on the permutation filter and that based on the unprotected circuit. We will evaluate the sampling overhead of permutation filters applied to practical variational quantum algorithms using this metric in Section V-C.

## IV. The Error Reduction Performance of Permutation Filters

In this section, we quantify the error reduction of permutation filters compared to VD of the same order using the following performance metric.

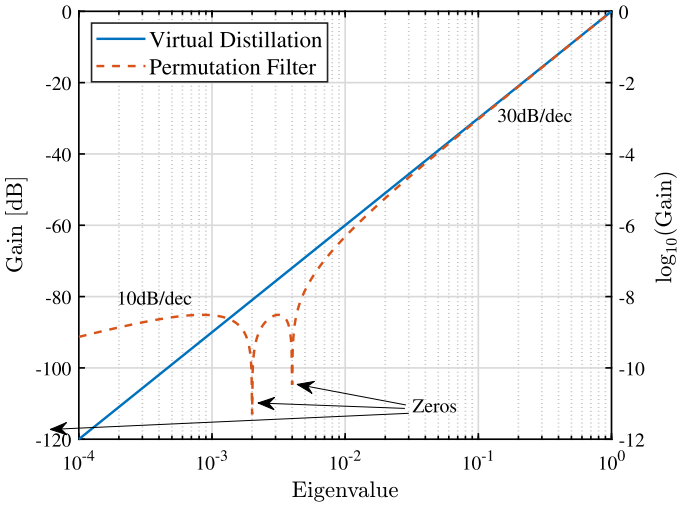



Fig. 6. The spectral response of a third-order permutation filter, compared to that of the third-order VD.

*Definition 1 (Error Ratio):* We define the error ratio between an $N$-th order permutation filter $\mathcal{F}_{\boldsymbol{\beta}}(\cdot)$ and its corresponding $N$-th order counterpart based on VD as follows:

$$R(\boldsymbol{\beta}) := \frac{\tilde{\epsilon}(\boldsymbol{\beta})}{\tilde{\epsilon}(\mathbf{0})}. \tag{37}$$

Note that VD is equivalent to a permutation filter that satisfies $\boldsymbol{\beta} = \mathbf{0}$.

Intuitively, the permutation filters are narrowband notch filters, hence they should perform better when the "bandwidth" of the undesired spectral components is lower. To see this more clearly, we consider the spectral response of a third-order permutation filter, as portrayed in Fig. 6. Observe that every zero contributes 10 dB per decade to the slope of the filter gain.[4] For both third-order permutation filters and for VD, the slope will be 30 dB per decade beyond the largest zero. In light of this, the only region where permutation filters have smaller gain is the narrowband range around the two largest zeros. Therefore, permutation filters perform the best when the noise components are concentrated in this region.

[4]For readers do not familiar with classical signal processing theory, please refer to Appendix II for further explanation.

$$\begin{aligned}\mathrm{Var}\{y^{(N)}_{\mathrm{filter}}(\mathcal{U})\} \approx\ & \frac{1-\sum_{n=1}^{N}\alpha_{N-n+1}^2\mathrm{Tr}\{\rho^n\mathcal{U}\}^2}{(\alpha_N+\sum_{n=2}^{N}\alpha_{N-n+1}\mathrm{Tr}\{\rho^n\})^2} \\ & -\frac{2(\sum_{n=1}^{N}\alpha_{N-n+1}\mathrm{Tr}\{\rho^n\mathcal{U}\})}{(\alpha_N+\sum_{n=2}^{N}\alpha_{N-n+1}\mathrm{Tr}\{\rho^n\})^3} \\ & \times\sum_{n=2}^{N}\alpha_{N-n+1}^2\left(\mathrm{Tr}\{\rho\mathcal{U}\}-\mathrm{Tr}\{\rho^n\mathcal{U}\}\,\mathrm{Tr}\{\rho^n\}\right) \\ & +\frac{(\sum_{n=1}^{N}\alpha_{N-n+1}\mathrm{Tr}\{\rho^n\mathcal{U}\})^2(1-\sum_{n=2}^{N}\alpha_{N-n+1}^2\mathrm{Tr}\{\rho^n\}^2)}{(\alpha_N+\sum_{n=2}^{N}\alpha_{N-n+1}\mathrm{Tr}\{\rho^n\})^4}.\end{aligned} \tag{36}$$

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

To make our aforementioned intuitions more rigorous, we define the following quantities to characterize the bandwidth.

*Definition 2 (Noise Bandwidth):* We define the bandwidth of the noise (i.e., the undesired spectral components $\tilde{\boldsymbol{\lambda}}$ in a mixed state $\rho$) as follows:

$$B(\tilde{\boldsymbol{\lambda}}) := \sqrt{\mathbb{E}\{|\lambda - \mu|^2\}}, \tag{38}$$

where

$$\mathbb{E}\{g(\lambda)\} := \int_{\lambda_{\mathrm{m}}}^{1} g(\lambda) f(\lambda) \mathrm{d}\lambda, \tag{39}$$

denotes the expectation operation, and $\mu = \mathbb{E}\{\lambda\}$ denotes the mean value of noise components. We also define the *relative noise bandwidth* as

$$b(\tilde{\boldsymbol{\lambda}}) := \mu^{-1} B(\tilde{\boldsymbol{\lambda}}). \tag{40}$$

Given the previous definitions, we are now prepared to state the following result concerning the error ratio of Type-1 permutation filters.

*Proposition 2 Generic Error Ratio Scaling Behaviour of Type-1 Permutation Filters:* The error ratio $R(\boldsymbol{\beta})$ of an $N$-th order Type-1 permutation filter, as a function of the relative noise bandwidth $b(\tilde{\boldsymbol{\lambda}})$, can be bounded by

$$R(\boldsymbol{\beta}) \leq \frac{1}{\mu} \left[ b(\tilde{\boldsymbol{\lambda}}) \sqrt{2^{N_{\mathrm{q}}} - 1} \right]^{N-1}, \tag{41}$$

as $b(\tilde{\boldsymbol{\lambda}}) \to 0$.

*Proof:* The term $\tilde{\epsilon}(\mathbf{0})$ can thus be written explicitly as

$$\tilde{\epsilon}(\mathbf{0}) = \mathbb{E}\{\lambda^N\}. \tag{42}$$

Using Jensen's inequality [49], we have

$$\tilde{\epsilon}(\mathbf{0}) \geq [\mathbb{E}\{\lambda\}]^N = \mu^N. \tag{43}$$

Therefore, from (37) we obtain

$$\begin{aligned} R(\boldsymbol{\beta}) &\leq \tilde{\epsilon}(\boldsymbol{\beta}) \mu^{-N} \\ &= \mathbb{E}\{|\lambda(\lambda - \mu)^{N-1}|\} \mu^{-N} \\ &\leq \frac{1}{\mu} \cdot \mathbb{E}\left\{ \left| (\lambda - \mu) \mu^{-1} \right|^{N-1} \right\}, \end{aligned} \tag{44}$$

where the last line follows from the fact that $\lambda \leq 1$ holds for all eigenvalues. Furthermore, assume that we have access to the actual values of $\tilde{\boldsymbol{\lambda}}$ (which will only be used for calculating intermediate results), we have

$$\begin{aligned} \mathbb{E}\left\{ \left| \frac{\lambda - \mu}{\mu} \right|^{N-1} \right\} &= \left( \frac{\mu^{-1} \left\| \tilde{\boldsymbol{\lambda}} - \mu \mathbf{1} \right\|_{N-1}}{(2^{N_{\mathrm{q}}} - 1)^{\frac{1}{N-1}}} \right)^{N-1} \\ &\leq \left( \mu^{-1} \left\| \tilde{\boldsymbol{\lambda}} - \mu \mathbf{1} \right\|_{\infty} \right)^{N-1} \\ &\leq \left( \mu^{-1} \left\| \tilde{\boldsymbol{\lambda}} - \mu \mathbf{1} \right\|_{2} \right)^{N-1} \\ &= \left( b(\tilde{\boldsymbol{\lambda}}) \sqrt{2^{N_{\mathrm{q}}} - 1} \right)^{N-1}. \end{aligned} \tag{45}$$

Hence the proof is completed. ∎

Proposition 2 supports our intuition that the error ratio decreases, as the noise bandwidth becomes smaller. However, the constant $\sqrt{2^{N_{\mathrm{q}}} - 1}$ in (41) can be extremely large for large $N_{\mathrm{q}}$, when the bound becomes of limited practical significance. In the following result we show that for spectral densities satisfying Pareto distributions, the dependence of the bound on $N_{\mathrm{q}}$ can be eliminated.

*Proposition 3 Type-1 Filters Applied to Pareto-Distributed States:* Assume that $f(\lambda)$ corresponds to a Pareto distribution, and that $b(\tilde{\boldsymbol{\lambda}}) < (N-1)^{-1}$. The error ratio of an $N$-th order Type-1 permutation filter can be bounded by

$$\begin{aligned} R(\boldsymbol{\beta}) &\leq \frac{(N-1)![1 + b(\tilde{\boldsymbol{\lambda}})]^N}{e \prod_{n=1}^{N-2} [1 - n b(\tilde{\boldsymbol{\lambda}})]} \cdot [b(\tilde{\boldsymbol{\lambda}})]^{N-1} \\ &= O\left\{ [b(\tilde{\boldsymbol{\lambda}})]^{N-1} \right\}. \end{aligned} \tag{46}$$

*Proof:* Please refer to Appendix III. ∎

Both Proposition 2 and 3 show that, the error ratio of Type-1 filters decreases exponentially with the filter order $N$. For Type-2 filters, this may be viewed as an upper bound of the error ratio, since their parameter vectors $\boldsymbol{\beta}$ are obtained via optimization. By contrast, the parameter vectors of Type-1 filters are determined using only the mean value of noise components, hence are suboptimal.

A natural question that arises is: under what practical conditions do the undesired spectral components have small relative bandwidth? In the following proposition, we show that the relative noise bandwidth decreases with the depth of quantum circuits, as well as with the error rate of the gates in the circuits.

*Proposition 4 Exponential Spectral Concentration of Deep Quantum Circuits:* Assume that each qubit is acted upon by at least $L$ gates, and that each of the gates is contaminated by quantum channels containing Pauli noise, which have matrix representations under the Pauli basis given in (71). We assume furthermore that the probability of each type of Pauli error (i.e., X error, Y error or Z error) on each qubit is lower bounded by $\epsilon_{\mathrm{l}}$. Under these assumptions, the relative noise bandwidth can be upper bounded by

$$\begin{aligned} b(\tilde{\boldsymbol{\lambda}}) &\leq \frac{1 + \sqrt{2^{N_{\mathrm{q}}} - 1}}{1 - 2^{-N_{\mathrm{q}}} - \exp(-4\epsilon_{\mathrm{l}} L)} \cdot \exp(-4\epsilon_{\mathrm{l}} L) \\ &= O\left\{ \exp(-4\epsilon_{\mathrm{l}} L) \right\}. \end{aligned} \tag{47}$$

*Proof:* Please refer to Appendix IV. ∎

From Proposition 4 we observe that the relative noise bandwidth decreases exponentially with the product of $\epsilon_{\mathrm{l}}$ and $L$. This implies that the proposed permutation filters would provide more significant performance improvements when the circuits are relatively deep, or the gates therein are noisy.

## V. Numerical Results

In this section, we further illustrate the results discussed in the previous sections using numerical simulations. In all simulations, we consider a class of parametric state preparation circuit consisting of different number of stages, for which a single stage is portrayed in Fig. 7. For illustration we drawn a four-qubit circuit, but in the actual simulations we set $N_{\mathrm{q}} = 10$. As observed from Fig. 7,[5] each stage of the

[5]The notation Rx, Ry, and Rz denote X-, Y-, and Z-rotation gates, respectively.

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

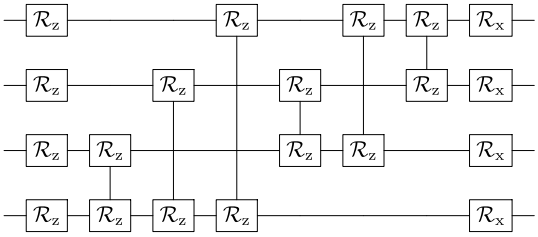


Fig. 7. Schematic of a stage in the parametric state preparation circuit used in the simulations. Here we set $N_{\mathrm{q}} = 4$ only for illustration.

circuit is constructed by two-qubit ZZ-rotation gates acting upon each pair of qubits, and single-qubit X- and Y-rotation gates acting upon each qubit. The rotation angle of each gate is a parameter to be determined. In the simulations, we choose the parameters by independent sampling from uniform distributions over $[-\pi, \pi]$, and the simulation results are averaged over 100 random instances of the circuits. The gates are inflicted by depolarizing errors occurring at varying probabilities, but we always set the depolarizing probabilities of two-qubit gates 10 times higher than that of single-qubit gates.

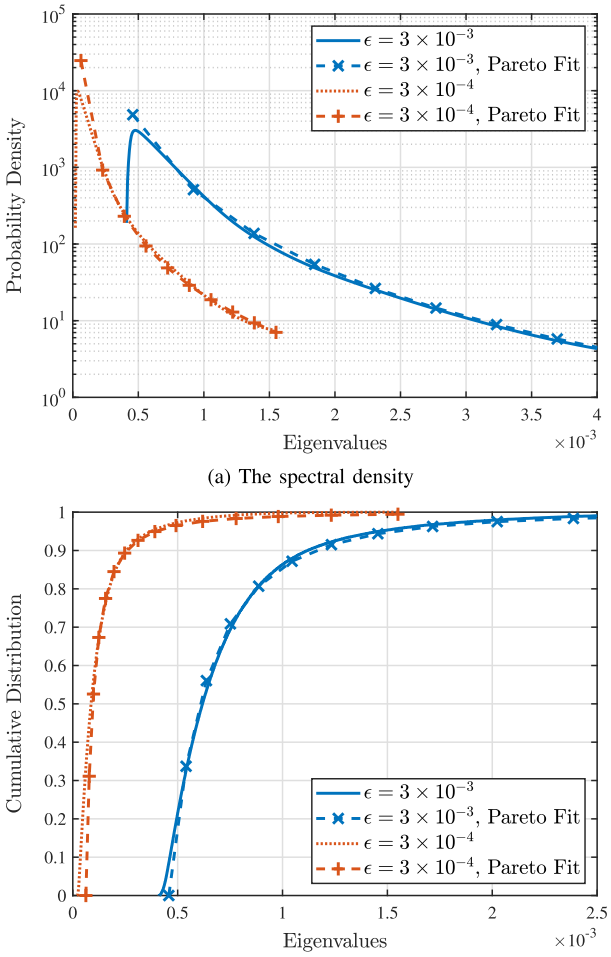
### A. Spectral Properties of the Output States

We first demonstrate the spectral densities of the output states. In particular, we consider parametric state preparation circuits having 10 stages acting on $N_{\mathrm{q}} = 10$ qubits. The spectral densities and the corresponding cumulative density functions for $\epsilon = 3 \times 10^{-4}$ and $\epsilon = 3 \times 10^{-3}$ are portrayed in Fig. 8, where $\epsilon$ denotes the depolarizing probability of each two-qubit gate. The Pareto fit are also plotted for comparison. We see that the Pareto distributions provide good approximations to the eigenvalue spectra, except for very small eigenvalues. This also suggests that the Pareto fit may become less accurate when the noise bandwidth is very narrow, for which the approximation error becomes more significant.

### B. The Filter Design Metric $\tilde{\epsilon}(\boldsymbol{\beta})$

Next, we investigate the values of the cost function $\tilde{\epsilon}(\boldsymbol{\beta})$ for filter design under different scenarios, which may be used for evaluating the performance of the filters irrespective of the specific choices of observables.

In Fig. 9a, we compare the values of $\tilde{\epsilon}(\boldsymbol{\beta})$ obtained both by our permutation filters and by VD, as functions of the number of stages in the state preparation circuits. The depolarizing probability of two-qubit gates is $1.25 \times 10^{-3}$. In this figure, the curve "Closed-form, 2nd order" corresponds to the second-order permutation filter designed based on the closed-form solution in (29)–(31). We observe from the figure that permutation filters significantly outperform VD, when the number of stages is large, for both the second-order case and the third-order case. In particular, in the second-order case, both the Type-1 and Type-2 permutation filters have the same parameters $\boldsymbol{\beta}$, and we see that their performance is very close to that of the optimal solution, which is obtained by directly solving (21) relying on the full *a priori* knowledge of the spectral density $f(\lambda)$.




(a) The spectral density

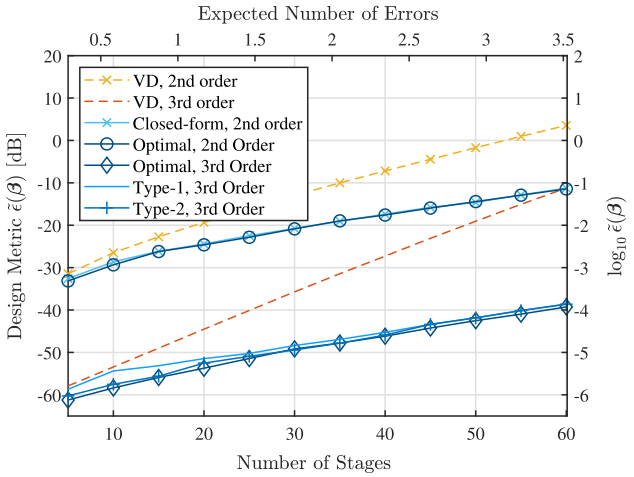
(b) The cumulative distribution function

Fig. 8. The spectra and the corresponding Pareto fits of the output states of parametric state preparation circuits having different depolarizing probability $\epsilon$.

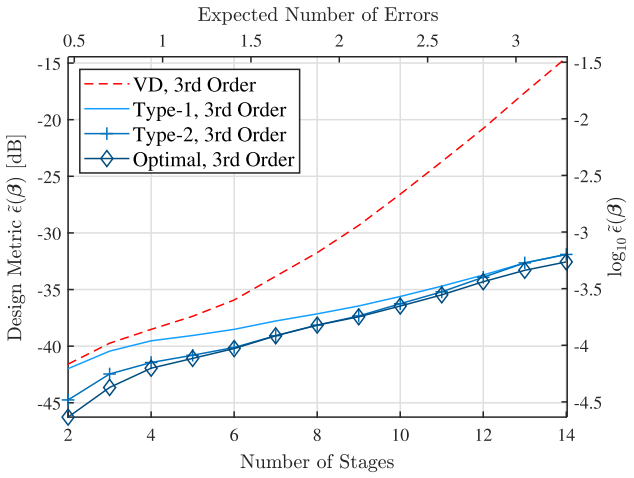
For the third-order case, we see that the Type-2 filter slightly outperforms the Type-1 filter, when the number of stages is relatively small. Intuitively, by adjusting the two zeros of the third-order filters, it is indeed possible to achieve a better error-reduction performance than that of simply placing the zeros at the same point. However, the effect of adjusting the positions of zeros would be less significant when the noise bandwidth is smaller, corresponding to the case where the number of stages is large. Closer scrutiny reveals that the performance of both the Type-1 and Type-2 third-order filters become similar when the number of stages is large, especially when it is larger than 45. By contrast, the performance of the Type-2 filter is near-optimal when the number of stages is moderate (around 25-40). This trend may prevail, because the Pareto fit becomes more accurate, when the noise bandwidth is moderate.

In Fig. 9b, we consider the case where the two-qubit depolarizing probability is $5 \times 10^{-3}$, which is four times that of Fig. 9a. The trends of the curves are similar to those of the lower depolarizing probability scenario. It may now be seen more clearly that the Type-2 permutation filter substantially outperforms its Type-1 counterpart, when the number of stages is small.

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.




(a) Two-qubit depolarizing probability $1.25 \times 10^{-3}$




(b) Two-qubit depolarizing probability $5 \times 10^{-3}$

Fig. 9. The value of the design metric $\tilde{\epsilon}(\boldsymbol{\beta})$ in (23) for both VD and for the proposed methods, as functions of the number of stages.

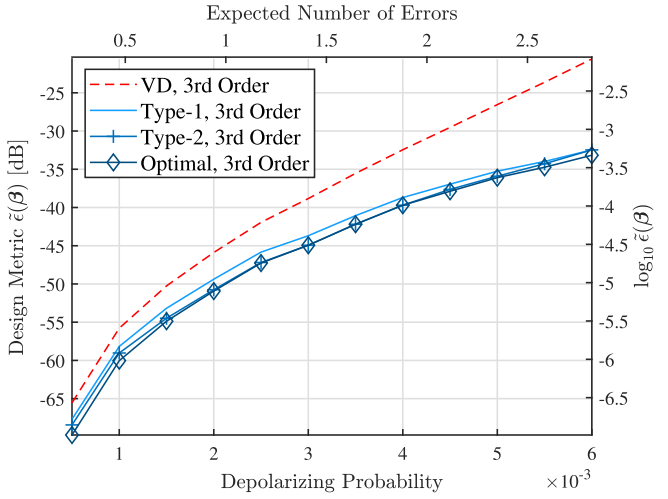



Fig. 10. The value of the design metric $\tilde{\epsilon}(\boldsymbol{\beta})$ in (23) for both VD and for the proposed methods, as functions of the depolarizing probability.

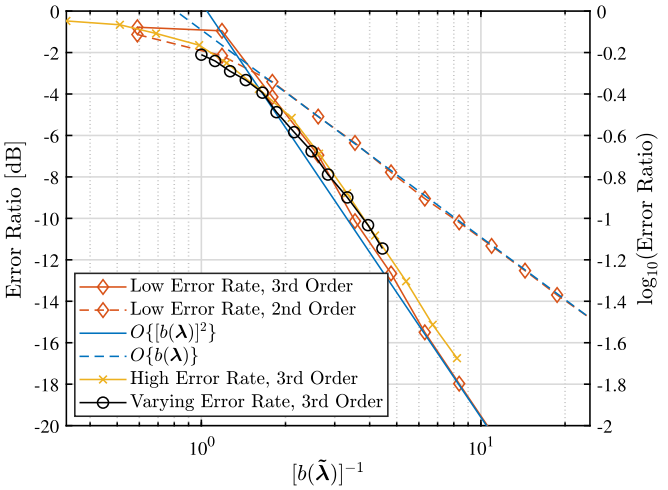



Fig. 11. The error ratio $R(\boldsymbol{\beta})$ in (37) between Type-1 permutation filters and VD vs. the reciprocal of the relative noise bandwidth $b(\tilde{\boldsymbol{\lambda}})$.

Next, in Fig. 10, we consider circuits having varying depolarizing probabilities. The number of stages is fixed to 10. We observe a similar increasing gap between the permutation filters and VD. In addition, the Type-2 permutation filter also exhibits better performance for moderate depolarizing probabilities.

In Fig. 11, we illustrate the scaling behaviour of the error ratio between the type-1 permutation filters and VD, which has been discussed in Section IV. In particular, we plot the error ratios computed using the data presented in Figures 9a, 9b and 10. We observe that when the relative noise bandwidth $b(\tilde{\boldsymbol{\lambda}})$ is small (less than around 0.5), all error ratios are reduced roughly polynomially with $[b(\tilde{\boldsymbol{\lambda}})]^{-1}$. Furthermore, the slopes of the curves are almost equal to the asymptotes scaling quadratically and linearly with $b(\tilde{\boldsymbol{\lambda}})$, respectively for third-order and second-order filters. These observations corroborate Propositions 2 and 3.

Finally, in Fig. 12, we demonstrate that the commonly used metric of noisiness, namely the expected number of errors, does not determine the relative noise bandwidth on its own, and hence does not solely determine the error ratio between permutation filters and VD. To this end, we fixed the number of expected errors, and change the number of stages and the depolarizing probability accordingly. As it can be seen from Fig. 12a, the relative bandwidth shrinks with the number of stages, even when the number of expected errors is fixed. Similarly, we observe from Fig. 12 that the error ratio decreases with decreasing depolarizing probability (or increasing number of stages).

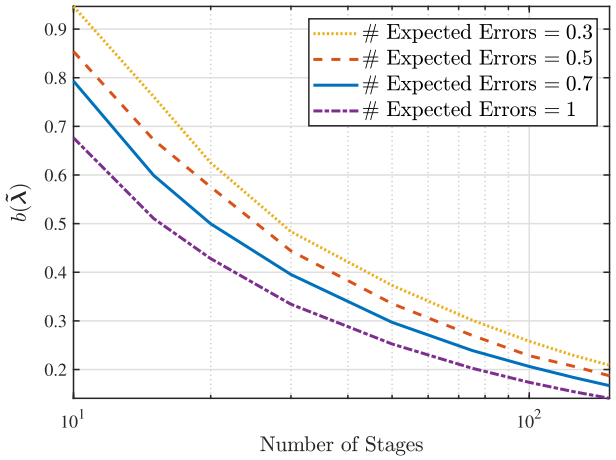
From the discussions in this subsection, we may conclude that the benefit of the permutation filter method is more significant when the circuit is rather noisy, or it is deep but is constituted by gates having relatively small error probabilities.

### *C. Case Study: QAOA-Aided Multi-User Detection*

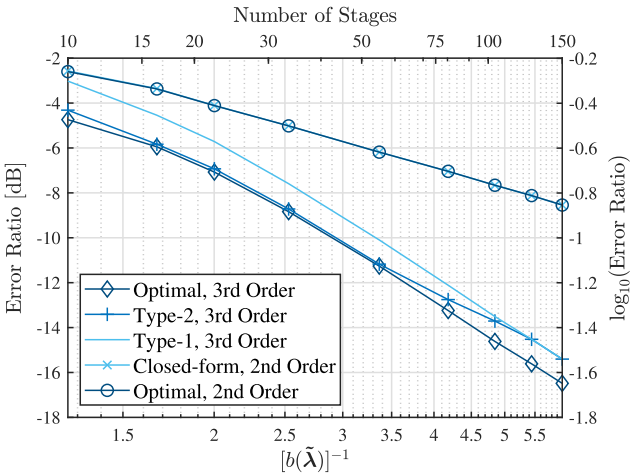
In this subsection we demonstrate the performance of permutation filters when applied to a practical variational quantum algorithm, namely the QAOA. The parametric state-preparation circuits of QAOA are multi-stage circuits having an alternating structure, which take a plus state $|+\rangle^{\otimes N_{\mathrm{q}}}$ as the input and produce the following output

$$|\psi\rangle_{\mathrm{out}} = e^{-\imath b_{N_{\mathrm{L}}}\mathcal{H}_{\mathrm{M}}} e^{-\imath c_{N_{\mathrm{L}}}\mathcal{H}_{\mathrm{P}}} \ldots e^{-\imath b_1\mathcal{H}_{\mathrm{M}}} e^{-\imath c_1\mathcal{H}_{\mathrm{P}}} |+\rangle^{\otimes N_{\mathrm{q}}}, \tag{48}$$

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

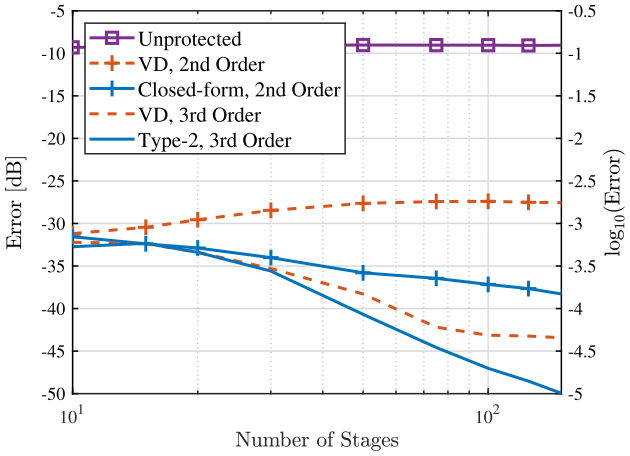



(a) The relative noise bandwidth

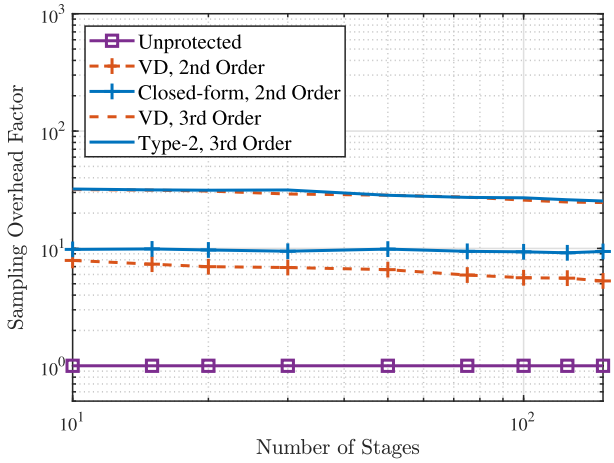



(b) The error ratio (The number of expected errors is 0.7)

Fig. 12. The relative noise bandwidth $b(\tilde{\boldsymbol{\lambda}})$, and the error ratio $R(\boldsymbol{\beta})$ in (37) between permutation filters and VD.




(a) The computational error




(b) The sampling overhead factor

Fig. 13. The computational error and the sampling overhead factor of permutation filters applied to QAOA-aided multi-user detection vs. the number of stages, where the number of expected errors is fixed at 0.7.

where $N_{\mathrm{L}}$ denotes the number of stages, $\mathcal{H}_{\mathrm{M}}$ denotes the mixing Hamiltonian defined as $\mathcal{H}_{\mathrm{M}} := \sum_{n=1}^{N_{\mathrm{q}}} \mathcal{X}_i$ ($\mathcal{X}_i$ denotes the Pauli-X operator acting on the $i$-th qubit), and $\mathcal{H}_{\mathrm{P}}$ denotes the phase Hamiltonian that encodes the problem to be solved. The parameters $\boldsymbol{b} = [b_1, \ldots, b_{N_{\mathrm{L}}}]^{\mathrm{T}}$ and $\boldsymbol{c} = [c_1, \ldots, c_{N_{\mathrm{L}}}]^{\mathrm{T}}$ control the dynamic of the algorithm, and are typically determined by an iterative optimization procedure [19]. Since we focus on the performance evaluation for error mitigation methods, here we consider a suboptimal linear scheduling [50] instead of optimizing for the parameters, given by $c_\ell = \ell/N_{\mathrm{L}}$ and $b_\ell = 1 - \ell/N_{\mathrm{L}}$.

In particular, we construct the phase Hamiltonian corresponding to the multi-user detection problem [51] for wireless communication systems.[6] For an $m \times n$ multiple-input mutliple-output (MIMO) system, the received signal may be modelled as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{\omega},$$

where $\boldsymbol{H}$ denotes the MIMO channel, $\boldsymbol{x}$ represents the transmitted signal, and $\boldsymbol{\omega}$ denotes the noise. For simplicity of the illustration, we assume that the noise is i.i.d. Gaussian on each receiving antenna, and that the modulation scheme is binary phase-shift keying (BPSK), hence $\boldsymbol{x} \in \{-1, 1\}^n$ and $\boldsymbol{H} \in \mathbb{R}^{m \times n}$. The phase Hamiltonian corresponding to the maximum likelihood estimator of $\boldsymbol{x}$ is thus given by

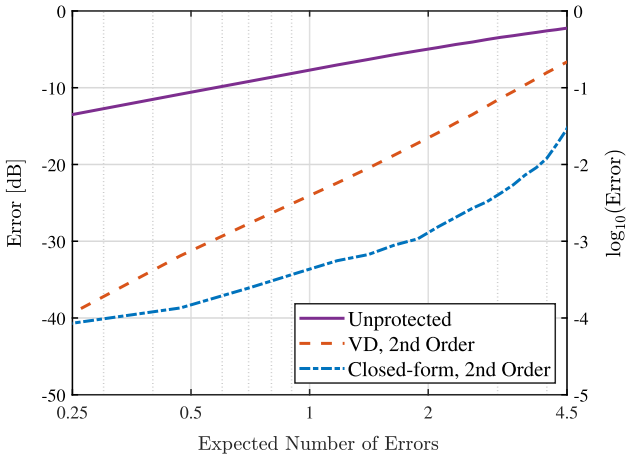
$$\sum_{k=1}^{n} [\boldsymbol{H}^{\mathrm{T}}\boldsymbol{y}]_i \mathcal{Z}_i - \sum_{i=1}^{n-1} \sum_{j>i} [\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}]_{i,j} \mathcal{Z}_i \mathcal{Z}_j. \tag{49}$$

We consider the following scenario for the numerical simulation: $N_{\mathrm{q}} = m = n = 10$, the channel $\boldsymbol{H}$ has i.i.d Gaussian entries with zero mean and a variance of $1/m = 0.1$, and the signal-to-noise ratio is 13dB, implying that $[\boldsymbol{\omega}]_i \sim \mathcal{N}(0, 0.05)$.

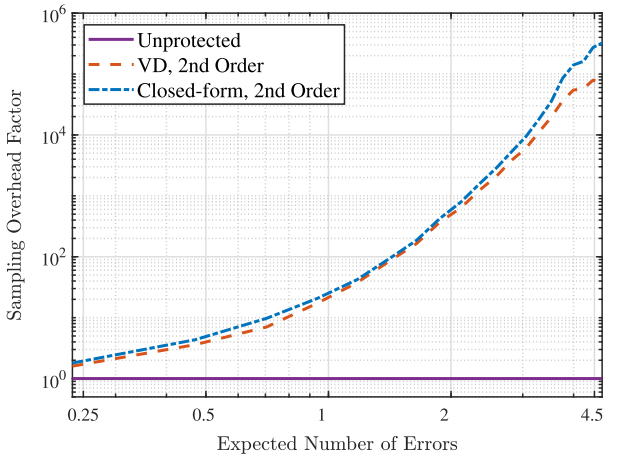
We first fix the number of expected errors at 0.7 and investigate the dependency of the computational error (the absolute difference between the error-free result and the result computed relying on noisy circuits based on the entire Hamiltonian) on the number of stages. As it may be seen from Fig. 13a, the permutation filters are more beneficial when the circuit is deep, as have been discussed in Section V-B. We may also observe from Fig. 13b that the sampling overhead is nearly constant with the number of stages, suggesting that the

[6]For readers not familiar with wireless communication, just note that it is a quadratic unconstrained binary optimization problem.

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.




(a) The computational error




(b) The sampling overhead factor

Fig. 14. The computational error and the sampling overhead factor of permutation filters applied to QAOA-aided multi-user detection vs. the number of expected errors, where the number of stages is fixed at 50.

number of expected errors might be the principal determining factor of the overhead.

Next, we present the relationship between the computational error and the number of expected error, with a fixed number of stages $N_{\mathrm{L}} = 50$, in Fig. 14a. It is seen from the figure that the permutation filter improves the error mitigation performance significantly when the number of expected errors is large. However, it should also be noted that the sampling overhead increases dramatically when the number of expected errors is larger than 1, as shown in Fig. 14b. Extra care should be taken for this issue, since a high sampling overhead may render the error mitigation method unfavorable in practice.

## VI. Conclusion

In this treatise, we have proposed a general framework for designing FIR-like permutation filters for mitigating the computational errors of variational quantum algorithms. In particular, the filter design problem is an invex problem, hence the algorithm is guaranteed to converge to the global optimum. For narrowband noise scenarios, we have also shown a polynomial error reduction compared to VD. This implies that permutation filters improve the error-reduction performance more substantially for quantum circuits having large depth or higher gate error rate.

The performance metric we used for filter design is an upper bound of the error magnitude across all unitary observables. A possible future research direction is to find other metrics better suited to specific classes of practical observables.

## Appendix I
## Proof of Proposition 1

*Proof:* Consider the transform from $\boldsymbol{\beta}$ to $\boldsymbol{\alpha}$, which helps us to reformulate (21) (where the cost function is approximated as in (22)) in the form of:

$$\min_{\boldsymbol{\alpha}} \ \xi(\boldsymbol{\alpha}), \quad \text{s.t. } [\boldsymbol{\alpha}]_1 = 1, \tag{50}$$

where

$$\xi(\boldsymbol{\alpha}) := \tilde{\epsilon}[\boldsymbol{\varphi}(\boldsymbol{\alpha})] = \int_{\lambda_{\mathrm{m}}}^{1} f(\lambda)\sqrt{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}(\lambda)\boldsymbol{\alpha}}\ \mathrm{d}\lambda, \tag{51}$$

$\boldsymbol{\varphi}(\cdot)$ is the mapping from $\boldsymbol{\beta}$ to $\boldsymbol{\alpha}$, $\boldsymbol{A}(\lambda)$ is defined by $\boldsymbol{A}(\lambda) := \boldsymbol{a}(\lambda)[\boldsymbol{a}(\lambda)]^{\mathrm{T}}$, and $\boldsymbol{a}(\lambda) := [\lambda^N \ \lambda^{N-1} \ \ldots \ \lambda]^{\mathrm{T}}$. Note that the term $\sqrt{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}(\lambda)\boldsymbol{\alpha}}$ is actually the Mahalanobis norm [52] of $\boldsymbol{\alpha}$ with respect to a positive semi-definite symmetric matrix $\boldsymbol{A}(\lambda)$, hence it is a convex function of $\boldsymbol{\alpha}$. Thus the objective function itself is also convex with respect to $\boldsymbol{\alpha}$, since the integration (weighted by a non-negative function $f(\lambda)$) preserves convexity.

Next, we observe that $\boldsymbol{\varphi}(\cdot)$ can be computed via (17), and its inverse may be obtained using the factorization of polynomials [53]. Since $\boldsymbol{\beta}$ satisfies the ordering (18), when $\boldsymbol{\alpha}$ is further constrained to be the coefficients of polynomials having only non-negative real-valued roots, it is clear that $\boldsymbol{\varphi}(\cdot)$ is a bijection, and hence the Jacobian $\boldsymbol{J}_{\boldsymbol{\beta}}$ that is given by

$$\boldsymbol{J}_{\boldsymbol{\beta}} = \left[\frac{\partial\boldsymbol{\varphi}(\boldsymbol{\beta})}{\partial\beta_1} \ \frac{\partial\boldsymbol{\varphi}(\boldsymbol{\beta})}{\partial\beta_2} \ \cdots \ \frac{\partial\boldsymbol{\varphi}(\boldsymbol{\beta})}{\partial\beta_{N-1}}\right]^{\mathrm{T}},$$

is invertible for every $\boldsymbol{\beta} \in \mathscr{B}$. This implies that $\boldsymbol{\varphi}(\cdot)$ is a diffeomorphism from $\boldsymbol{\beta}$ to $\boldsymbol{\alpha}$, and hence $\tilde{\epsilon}(\boldsymbol{\beta})$ is an invex function of $\boldsymbol{\beta}$ [46], [54], [55]. To elaborate further, we see that

$$\left.\frac{\partial\tilde{\epsilon}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right|_{\boldsymbol{\beta}_0} = \boldsymbol{J}_{\boldsymbol{\beta}_0}^{-1}\left.\frac{\partial\xi(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\varphi}(\boldsymbol{\beta}_0)} = \boldsymbol{0} \Leftrightarrow \left.\frac{\partial\xi(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right|_{\boldsymbol{\varphi}(\boldsymbol{\beta}_0)} = \boldsymbol{0}$$

holds for $\boldsymbol{\beta}_0 \in \mathscr{B}$, implying that $\boldsymbol{\beta}_0 \in \mathscr{B}$ is a stationary point of $\tilde{\epsilon}(\boldsymbol{\beta})$ if and only if $\boldsymbol{\varphi}(\boldsymbol{\beta}_0)$ is also a stationary point of $\xi(\boldsymbol{\alpha})$, which in turn is one of the global minima of $\xi(\boldsymbol{\alpha})$.

Our remaining task is to show that $\xi(\boldsymbol{\alpha})$ attains its global minimum when $\boldsymbol{\beta} = \boldsymbol{\varphi}^{-1}(\boldsymbol{\alpha})$ belongs to the feasible region $\mathscr{B}$. This may be proved using the method of contradiction. Assume by contrast that the minimum of $\xi(\boldsymbol{\alpha})$ is attained at $\boldsymbol{\alpha}_0 \notin \mathscr{B}$. Then the polynomial $\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{a}(\lambda)$ has either real negative roots or complex roots. For the former case, it is plausible that $|\boldsymbol{\alpha}_0^{\mathrm{T}}\boldsymbol{a}(\lambda)| > \lambda^N$ for all $\lambda > 0$, hence $\boldsymbol{\alpha}_0$ is not the optimum. For the latter case, we specifically consider a pair of conjugate complex roots $x \pm iy$. It is clear that

$$\begin{aligned}|(\lambda - x - iy)(\lambda - x + iy)| &= \lambda^2 - 2x\lambda + \sqrt{x^2 + y^2} \\ &\geq \lambda^2 - 2x\lambda + x^2 = (\lambda - x)^2,\end{aligned}$$

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

implying that the cost function value can be reduced by replacing the complex roots with real roots. Hence the proof is completed. ∎

## APPENDIX II
## NOTES ON THE SPECTRAL RESPONSE OF PERMUTATION FILTERS

Let us consider a third-order permutation filter as an example, which has the following spectral response:

$$h_{\boldsymbol{\beta}}(\lambda) = \lambda(\lambda-\beta_1)(\lambda-\beta_2), \tag{52}$$

where $\beta_1 \le \beta_2$. By taking the limit $\lambda \to \infty$, we see that $h_{\boldsymbol{\beta}}(\lambda) \sim \lambda^3$, implying that the spectral response can be well approximated by $\lambda^3$ when $\lambda \gg \beta_2$. Since the cubic function $\lambda^3$ satisfies $\lambda_1^3 = 10^3 \cdot \lambda_2^3$ when $\lambda_1 = 10\lambda_2$, we say that it "has a slope of 30dB per decade" (note that 10dB corresponds to $10\log_{10}(10) = 10$ times). Here, the "slope" refers to that of the spectral response curve on a log-log scale, which appears to be linear for power functions. Furthermore, if $\beta_1$ and $\beta_2$ is well separated, we see that $h_{\boldsymbol{\beta}}(\lambda) \sim \lambda^2$ when $\beta_1 \ll \lambda \ll \beta_2$, and hence "has a slope of 20dB per decade". In general, when the eigenvalue $\lambda$ is in the region $\beta_n \ll \lambda \ll \beta_{n+1}$, we see that the slope is (approximately) $10(n+1)$ dB per decade. Since the first zero is $\beta_0 = 0$, we may conclude that each zero $\beta_i \ll \lambda_0$ contributes 10dB/decade to the slope at the point $\lambda = \lambda_0$.

## APPENDIX III
## PROOF OF PROPOSITION 3

*Proof:* The term $\tilde{\epsilon}(\mathbf{0})$ may be viewed as the $N$-th moment of the Pareto distribution. Upon denoting the shape parameter and the minimum value of the Pareto distribution as $k$ and $\lambda_{\rm m}$, we have

$$\tilde{\epsilon}(\mathbf{0}) = \frac{k}{k-N} \cdot \lambda_{\rm m}^N. \tag{53}$$

From (23) we obtain

$$R(\boldsymbol{\beta}) = \frac{\tilde{\epsilon}(\boldsymbol{\beta})}{\tilde{\epsilon}(\mathbf{0})} = \frac{k-N}{k\lambda_{\rm m}^N} \sum_{i=0}^{N-1} (-1)^i \int_{\beta_{N-i-1}}^{\beta_{N-i}} G_{\boldsymbol{\beta}}(\lambda) {\rm d}\lambda. \tag{54}$$

Note that for Type-1 permutation filters, we have $\boldsymbol{\beta} = \frac{k\lambda_{\rm m}}{k-1}\mathbf{1}$. Hence (54) can be bounded as

$$\begin{aligned} R(\boldsymbol{\beta}) &= \frac{k-N}{k\lambda_{\rm m}^N} \left( \int_{\frac{k\lambda_{\rm m}}{k-1}}^{\infty} |G_{\boldsymbol{\beta}}(\lambda)| {\rm d}\lambda + \int_{\lambda_{\rm m}}^{\frac{k\lambda_{\rm m}}{k-1}} |G_{\boldsymbol{\beta}}(\lambda)| {\rm d}\lambda \right) \\ &= \left| G(\lambda_{\rm m}) - G\left(\frac{k\lambda_{\rm m}}{k-1}\right) \right| + \left| G\left(\frac{k\lambda_{\rm m}}{k-1}\right) \right| \\ &\le 2 \left| G\left(\frac{k\lambda_{\rm m}}{k-1}\right) \right|, \end{aligned} \tag{55}$$

where for simplicity of notations we have defined $G(\lambda) = \frac{k-N}{\lambda_{\rm m}^{N-k}} \tilde{G}_{\boldsymbol{\alpha}}(\lambda)$. The last line of (55) comes from the fact that

$$\int_{\lambda_{\rm m}}^{\frac{k\lambda_{\rm m}}{k-1}} |G_{\boldsymbol{\beta}}(\lambda)| {\rm d}\lambda \ge 0.$$

Furthermore, from (17) we have

$$\alpha_i = \tbinom{N-1}{i-1} \left( -\frac{k\lambda_{\rm m}}{k-1} \right)^{i-1}. \tag{56}$$

Thus we obtain

$$\begin{aligned} G(\lambda) &= \sum_{n=1}^{N} \alpha_{N-n+1} \frac{k-N}{n-k} \cdot \frac{\lambda^{n-k}}{\lambda_{\rm m}^{N-k}} \\ &= (k-N) \sum_{n=1}^{N} \frac{\binom{N-1}{N-n}}{n-k} \left( \frac{-k}{k-1} \right)^{N-n} \left( \frac{\lambda}{\lambda_{\rm m}} \right)^{n-k}. \end{aligned} \tag{57}$$

This implies that

$$G\left( \frac{k\lambda_{\rm m}}{k-1} \right) = \frac{k-N}{\left(\frac{k}{k-1}\right)^{k-N}} \sum_{n=1}^{N} \frac{\binom{N-1}{n-1}}{n-k} (-1)^{N-n}. \tag{58}$$

Next, we denote

$$\sum_{n=1}^{N} (-1)^{N-n} \tbinom{N-1}{n-1} \eta(n,k) = \boldsymbol{a}_{N-1}^{\rm T} \boldsymbol{\eta}, \tag{59}$$

where $[\boldsymbol{a}_{N-1}]_i = (-1)^{N-i}\binom{N-1}{i-1}$, $[\boldsymbol{\eta}]_i = \eta(i,k)$, and $\eta(n,k)$ denotes an arbitrary function of $n$ and $k$. Furthermore, we have

$$\boldsymbol{a}_{N-1}^{\rm T} \boldsymbol{\eta} = \mathbf{1}^{\rm T} \boldsymbol{A}_{N-1} \boldsymbol{\eta}, \tag{60}$$

where $\boldsymbol{A}_{N-1}$ is defined recursively by

$$\boldsymbol{A}_n = \begin{bmatrix} \boldsymbol{A}_{n-1} & \mathbf{0}_{2^{n-2}\times 1} \\ \mathbf{0}_{2^{n-2}\times 1} & -\boldsymbol{A}_{n-1} \end{bmatrix}, \tag{61}$$

and $\boldsymbol{A}_1 := [1 \ -1]$. Thus we have the following recursion

$$\mathbf{1}^{\rm T} \boldsymbol{A}_n \boldsymbol{x} = \mathbf{1}^{\rm T} \boldsymbol{A}_{n-1} \left( [\boldsymbol{x}]_{1:L-1} - [\boldsymbol{x}]_{2:L} \right)$$

for $\boldsymbol{x} \in \mathbb{R}^L$. From (58) we may now write $\boldsymbol{\eta}$ explicitly as

$$\boldsymbol{\eta} = \begin{bmatrix} \frac{1}{1-k} & \frac{1}{2-k} & \cdots & \frac{1}{N-k} \end{bmatrix}^{\rm T}. \tag{62}$$

When $N = 2$, we have

$$\begin{aligned} \mathbf{1}^{\rm T} \boldsymbol{A}_1 \boldsymbol{\eta} &= \frac{1}{1-k} - \frac{1}{2-k} \\ &= \frac{\Gamma(-k)}{\Gamma(1-k)} - \frac{\Gamma(1-k)}{\Gamma(2-k)}, \end{aligned}$$

where $\Gamma(\cdot)$ denotes the Gamma function [56]. Note that

$$\frac{\Gamma(-k)}{\Gamma(m-k)} - \frac{\Gamma(1-k)}{\Gamma(m-k+1)} = \frac{m\Gamma(-k)}{\Gamma(m+1-k)}. \tag{63}$$

Hence in general we have

$$\begin{aligned} \mathbf{1}^{\rm T} \boldsymbol{A}_{N-1} \boldsymbol{\eta} &= \frac{(N-1)!\Gamma(-k)}{\Gamma(N-k)} \\ &= (-1)^N (N-1)! \cdot \frac{\Gamma(k-N-1)}{\Gamma(k)}. \end{aligned}$$

This implies that

$$\begin{aligned} \left| G\left( \frac{k\lambda_{\rm m}}{k-1} \right) \right| &= \frac{(k-N)(N-1)!}{\left(\frac{k}{k-1}\right)^{k-N}} \cdot \frac{\Gamma(k-N-1)}{\Gamma(k)} \\ &= \frac{\left(1+\frac{1}{k-1}\right)^{N-k} (N-1)!}{\prod_{n=1}^{N-1}(k-n)}, \end{aligned} \tag{64}$$

as a function of $k$.

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

Finally, since we have assumed that the spectral density obeys a Pareto distribution, we may compute the relative noise bandwidth explicitly as follows:

$$b(\tilde{\boldsymbol{\lambda}}) = \sqrt{\frac{k}{(k-1)^2(k-2)}} \geq (k-1)^{-1}. \tag{65}$$

Combining (64) and (65), we obtain the desired scaling law in (46). ∎

## Appendix IV
## Proof of Proposition 4

*Proof:* To simplify the discussion, we will use the Pauli basis. Under the Pauli basis, a quantum channel $\mathcal{C}$ may be represented in a matrix form as

$$[\boldsymbol{C}]_{i,j} = \frac{1}{2^{N_\mathrm{q}}} \mathrm{Tr}\left\{\mathcal{S}_i \mathcal{C}(\mathcal{S}_j)\right\}, \tag{66}$$

where $\mathcal{S}_i$ denotes the $i$-th Pauli string acting upon $N_\mathrm{q}$ qubits. Correspondingly, a quantum state $\rho$ may be represented as a vector $[\boldsymbol{x}_\rho]_i = \frac{1}{\sqrt{2^{N_\mathrm{q}}}} \mathrm{Tr}\left\{\mathcal{S}_i \rho\right\}$. Since the Pauli operators are unitary and mutually orthogonal, both the transform from the conventional computation basis to the Pauli basis, as well as the inverse transform, are also unitary. This implies that

$$\|\boldsymbol{\lambda}_\rho\|_2 = \|\boldsymbol{x}_\rho\|_\mathrm{F} = \|\boldsymbol{x}_\rho\|_2, \tag{67}$$

due to the unitary invariance of the Frobenius norm [57], where $\boldsymbol{\lambda}_\rho$ denotes the vector containing all eigenvalues of $\rho$ sorted in descending order. Without loss of generality, we assume that the first Pauli operator is the identity operator $\mathcal{I}^{\otimes N_\mathrm{q}}$. In light of this, we have $\boldsymbol{x}_\rho = [2^{-N_\mathrm{q}/2}\ \tilde{\boldsymbol{x}}_\rho^\mathrm{T}]^\mathrm{T}$, since all quantum states satisfy $\mathrm{Tr}\left\{\rho\right\} = 1$.

We say that "a layer of gates" is activated if each qubit has been act upon by at least one gate. From our assumption we see that the circuit consists of at least $L$ layers. After the $l$-th layer, the output state $\boldsymbol{x}_{\rho_l}$ may be expressed as

$$\boldsymbol{x}_{\rho_l} = \widetilde{\boldsymbol{G}}_l \boldsymbol{x}_{\rho_{l-1}} = \boldsymbol{C}_l \boldsymbol{G}_l \boldsymbol{x}_{\rho_{l-1}}, \tag{68}$$

where $\boldsymbol{G}_l$ denotes the ideal noiseless operation corresponding to the $l$-th layer, and $\boldsymbol{C}_l$ denotes the associated quantum channel characterizing the noise. A perfect layer of gates $\boldsymbol{G}_i$, and the corresponding Pauli channel $\boldsymbol{C}_i$, can be expressed as

$$\boldsymbol{G}_i = \begin{bmatrix} 1 & \boldsymbol{0}^\mathrm{T} \\ \boldsymbol{0} & \boldsymbol{U}_i \end{bmatrix}, \quad \boldsymbol{C}_i = \begin{bmatrix} 1 & \boldsymbol{0}^\mathrm{T} \\ \boldsymbol{0} & \boldsymbol{D}_i \end{bmatrix}, \tag{69}$$

respectively, where $\boldsymbol{U}_i \in \mathbb{R}^{(4^{N_\mathrm{q}}-1)\times(4^{N_\mathrm{q}}-1)}$ is a unitary matrix, and $\boldsymbol{D}_i$ is a diagonal matrix, whose diagonal entries take values in the interval $[0, 1]$. We now see that the maximum singular value of $\widetilde{\boldsymbol{G}}_l$ is 1, while its second largest singular value $\sigma_2(\widetilde{\boldsymbol{G}}_l)$ is given by

$$\sigma_2(\widetilde{\boldsymbol{G}}_l) = \|\boldsymbol{D}_l\|_2. \tag{70}$$

Since the probability of each single-qubit Pauli error is at least $\epsilon_1$, we see that for a single-qubit channel $\mathcal{C}$ characterized by the error probabilities of $p_\mathrm{X}$, $p_\mathrm{Y}$ and $p_\mathrm{Z}$ corresponding to the X, Y and Z errors, respectively, the following holds:

$$\begin{aligned} \boldsymbol{C} &= \mathrm{diag}\left\{\widetilde{\boldsymbol{H}}[1-p_\mathrm{X}-p_\mathrm{Y}-p_\mathrm{Z} \quad p_\mathrm{X} \quad p_\mathrm{Y} \quad p_\mathrm{Z}]^\mathrm{T}\right\} \\ &= \boldsymbol{I} - 2\mathrm{diag}\left\{[p_\mathrm{X}+p_\mathrm{Z} \ p_\mathrm{Y}+p_\mathrm{Z} \ \ p_\mathrm{X}+p_\mathrm{Y}]\right\} \\ &\preceq (1-4\epsilon_1)\boldsymbol{I}, \end{aligned} \tag{71}$$

where $\widetilde{\boldsymbol{H}}$ denotes the inverse Hadamard transform over $N_\mathrm{q}$ qubits. Therefore, we obtain

$$\begin{aligned} \|\tilde{\boldsymbol{x}}_{\rho_L}\|_2 &\leq \sigma_2\left(\prod_{l=1}^{L} \widetilde{\boldsymbol{G}}_{L-l+1}\right) \\ &\leq \prod_{l=1}^{L} \|\boldsymbol{D}_l\|_2 \\ &\leq (1-4\epsilon_1)^L \\ &\leq \exp\left(-4\epsilon_1 L\right), \end{aligned} \tag{72}$$

where the last line follows from the fact that $\ln(1-x) \leq -x$ holds for all $x > 0$. This implies that

$$\|\boldsymbol{x}_{\rho_L} - [2^{N_\mathrm{q}/2}\ \boldsymbol{0}^\mathrm{T}]^\mathrm{T}\|_2 \leq \exp\left(-4\epsilon_1 L\right). \tag{73}$$

Note that $[2^{-N_\mathrm{q}/2}\ \boldsymbol{0}^\mathrm{T}]^\mathrm{T}$ corresponds to the completely mixed state $2^{-N_\mathrm{q}}\boldsymbol{I}$, hence from (67) we have

$$\begin{aligned} \|\boldsymbol{\lambda}_{\rho_L} - 2^{-N_\mathrm{q}}\boldsymbol{1}\|_2 &= \|\rho_L - 2^{-N_\mathrm{q}}\boldsymbol{I}\|_\mathrm{F} \\ &\leq \exp\left(-4\epsilon_1 L\right). \end{aligned} \tag{74}$$

The relative noise bandwidth is given by

$$b(\tilde{\boldsymbol{\lambda}}) = \mu^{-1}(2^{N_\mathrm{q}}-1)^{-\frac{1}{2}}\|\tilde{\boldsymbol{\lambda}} - \mu\boldsymbol{1}\|_2, \tag{75}$$

where $\mu = \frac{1-[\boldsymbol{\lambda}_{\rho_L}]_1}{2^{N_\mathrm{q}}-1}$ and $\tilde{\boldsymbol{\lambda}} = [\boldsymbol{\lambda}_{\rho_L}]_{2:2^{N_\mathrm{q}}}$. The term $\|\tilde{\boldsymbol{\lambda}} - (2^{N_\mathrm{q}}-1)^{-1}\boldsymbol{1}|_2$ can be bounded by

$$\begin{aligned} \|\tilde{\boldsymbol{\lambda}} - \mu\boldsymbol{1}\|_2 &\leq \|\tilde{\boldsymbol{\lambda}} - 2^{-N_\mathrm{q}}\boldsymbol{1}\|_2 + \frac{\left\|[\boldsymbol{\lambda}_{\rho_L}]_1 - 2^{-N_\mathrm{q}}\boldsymbol{1}\right\|_2}{2^{N_\mathrm{q}}-1} \\ &\leq \left(1 + (2^{N_\mathrm{q}}-1)^{-1/2}\right) e^{-4\epsilon_1 L}. \end{aligned} \tag{76}$$

In addition, we have

$$\begin{aligned} \mu &= \frac{1 - 2^{-N_\mathrm{q}} - |[\boldsymbol{\lambda}_{\rho_L}]_1 - 2^{-N_\mathrm{q}}|}{2^{N_\mathrm{q}}-1} \\ &\geq \frac{1 - 2^{-N_\mathrm{q}} - e^{-4\epsilon_1 L}}{2^{N_\mathrm{q}}-1}. \end{aligned} \tag{77}$$

Substituting (76) and (77) into (75), we obtain (47). Thus the proof is completed. ∎

## Acknowledgment

The insightful comments of Balint Koczor are gratefully acknowledged by the authors.

## References

[1] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, pp. 1–21, Aug. 2018.
[2] F. Arute *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.
[3] H.-S. Zhong *et al.*, "Quantum computational advantage using photons," *Science*, vol. 370, no. 6523, pp. 1460–1463, 2020.
[4] D. Aharonov and M. Ben-Or, "Fault-tolerant quantum computation with constant error rate," *SIAM J. Comput.*, vol. 38, no. 4, pp. 1207–1282, 2008.

Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.

[5] A. R. Calderbank, E. M. Rains, P. M. Shor, and N. J. A. Sloane, "Quantum error correction via codes over GF(4)," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1369–1387, Jul. 1998.

[6] C. H. Bennett and P. W. Shor, "Quantum information theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2724–2742, Oct. 1998.

[7] Z. Babar, D. Chandra, H. V. Nguyen, P. Botsinis, D. Alanis, S. X. Ng, and L. Hanzo, "Duality of quantum and classical error correction codes: Design principles and examples," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 970–1010, 1st Quart., 2018.

[8] D. Chandra, Z. Babar, H. V. Nguyen, D. Alanis, P. Botsinis, S. X. Ng, and L. Hanzo, "Quantum topological error correction codes: The classical-to-quantum isomorphism perspective," *IEEE Access*, vol. 6, pp. 13729–13757, 2017.

[9] Z. Babar, S. X. Ng, and L. Hanzo, "Near-capacity code design for entanglement-assisted classical communication over quantum depolarizing channels," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4801–4807, Dec. 2013.

[10] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 2011.

[11] L. K. Grover, "Quantum computers can search rapidly by using almost any transformation," *Phys. Rev. Lett.*, vol. 80, no. 19, p. 4329, 1998.

[12] G. Brassard and P. Hoyer, "An exact quantum polynomial-time algorithm for Simon's problem," in *Proc. 5th Israeli Symp. Theory Comput. Syst.*, Jun. 1997, pp. 12–23.

[13] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, Nov. 1994, pp. 124–134.

[14] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proc. 28th Annu. ACM Symp. Theory Comput.*, May 1996, pp. 212–219.

[15] P. Botsinis, S. X. Ng, and L. Hanzo, "Fixed-complexity quantum-assisted multi-user detection for CDMA and SDMA," *IEEE Trans. Commun.*, vol. 62, no. 3, pp. 990–1000, Mar. 2014.

[16] P. Botsinis, D. Alanis, Z. Babar, S. X. Ng, and L. Hanzo, "Iterative quantum-assisted multi-user detection for multi-carrier interleave division multiple access systems," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3713–3727, Oct. 2015.

[17] A. Peruzzo *et al.*, "A variational eigenvalue solver on a photonic quantum processor," *Nature Commun.*, vol. 5, no. 1, pp. 1–7, Jul. 2014.

[18] N. Moll, P. Barkoutsos, and L. S. Bishop, "Quantum optimization using variational algorithms on near-term quantum devices," *Quantum Sci. Technol.*, vol. 3, pp. 1–17, Jun. 2018.

[19] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014, *arXiv:1411.4028*.

[20] G. E. Crooks, "Performance of the quantum approximate optimization algorithm on the maximum cut problem," 2018, *arXiv:1811.08419*.

[21] X. Xu, J. Sun, S. Endo, Y. Li, S. C. Benjamin, and X. Yuan, "Variational algorithms for linear algebra," 2019, *arXiv:1909.03898*.

[22] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," *New J. Phys.*, vol. 18, pp. 1–22, Feb. 2016.

[23] G. H. Low and I. L. Chuang, "Optimal Hamiltonian simulation by quantum signal processing," *Phys. Rev. Lett.*, vol. 118, no. 1, Jan. 2017, Art. no. 010501.

[24] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits," *Phys. Rev. Lett.*, vol. 119, no. 18, pp. 1–5, Nov. 2017.

[25] S. Endo, S. C. Benjamin, and Y. Li, "Practical quantum error mitigation for near-future applications," *Phys. Rev. X*, vol. 8, no. 3, pp. 1–21, Jul. 2018.

[26] Z. Cai, "Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications," 2020, *arXiv:2007.01265*.

[27] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, "Quantum computation with universal error mitigation on a superconducting quantum processor," *Sci. Adv.*, vol. 5, no. 9, Sep. 2019, Art. no. eaaw5686.

[28] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, "Quantum error mitigation as a universal error-minimization technique: Applications from NISQ to FTQC eras," 2020, *arXiv:2010.03887*.

[29] A. Strikis, D. Qin, Y. Chen, S. C. Benjamin, and Y. Li, "Learning-based quantum error mitigation," *PRX Quantum*, vol. 2, no. 4, Nov. 2021, Art. no. 040330, doi: 10.1103/PRXQuantum.2.040330.

[30] P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Error mitigation with Clifford quantum-circuit data," 2020, *arXiv:2005.10189*.

[31] A. Lowe, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Unified approach to data-driven quantum error mitigation," *Phys. Rev. Res.*, vol. 3, no. 3, Jul. 2021, Art. no. 033098.

[32] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, "Low-cost error mitigation by symmetry verification," *Phys. Rev. A, Gen. Phys.*, vol. 98, no. 6, Dec. 2018, Art. no. 062339.

[33] R. Sagastizabal *et al.*, "Experimental error mitigation via symmetry verification in a variational quantum eigensolver," *Phys. Rev. A, Gen. Phys.*, vol. 100, no. 1, Jul. 2019, Art. no. 010302.

[34] B. Koczor, "Exponential error suppression for near-term quantum devices," *Phys. Rev. X*, vol. 11, no. 3, Sep. 2021, Art. no. 031057.

[35] W. J. Huggins *et al.*, "Virtual distillation for quantum error mitigation," 2020, *arXiv:2011.07064*.

[36] B. Koczor, "The dominant eigenvector of a noisy quantum state," *New J. Phys.*, to be published.

[37] D. Bultrini, M. H. Gordon, P. Czarnik, A. Arrasmith, P. J. Coles, and L. Cincio, "Unifying and benchmarking state-of-the-art quantum error mitigation techniques," 2021, *arXiv:2107.13470*.

[38] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, "Hybrid quantum-classical algorithms and quantum error mitigation," *J. Phys. Soc. Jpn.*, vol. 90, no. 3, Mar. 2021, Art. no. 032001.

[39] S. Hadfield, Z. Wang, B. O'Gorman, E. Rieffel, D. Venturelli, and R. Biswas, "From the quantum approximate optimization algorithm to a quantum alternating operator ansatz," *Algorithms*, vol. 12, pp. 1–45, Feb. 2019.

[40] A. V. Uvarov and J. D. Biamonte, "On barren plateaus and cost function locality in variational quantum algorithms," *J. Phys. A, Math. Theor.*, vol. 54, no. 24, Jun. 2021, Art. no. 245301.

[41] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf, "Quantum fingerprinting," *Phys. Rev. Lett.*, vol. 87, no. 16, Sep. 2001, Art. no. 167902.

[42] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, "Direct estimations of linear and nonlinear functionals of a quantum state," *Phys. Rev. Lett.*, vol. 88, no. 21, May 2002, Art. no. 217901.

[43] M. Wildemeersch, T. Q. S. Quek, M. Kountouris, A. Rabbachin, and C. H. Slump, "Successive interference cancellation in heterogeneous networks," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4440–4453, Dec. 2014.

[44] C.-F. Liu, M. Bennis, M. Debbah, and H. V. Poor, "Dynamic task offloading and resource allocation for ultra-reliable low-latency edge computing," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4132–4150, Jun. 2019.

[45] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, 1st ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[46] M. A. Hanson, "Invexity and the Kuhn–Tucker theorem," *J. Math. Anal. Appl.*, vol. 236, no. 2, pp. 594–604, Aug. 1999.

[47] J. J. Moré and D. J. Thuente, "Line search algorithms with guaranteed sufficient decrease," *ACM Trans. Math. Softw.*, vol. 20, no. 3, pp. 286–307, Sep. 1994.

[48] F. Roosta, Y. Liu, P. Xu, and M. W. Mahoney, "Newton-MR: Inexact Newton method with minimum residual sub-problem solver," 2018, *arXiv:1810.00303*.

[49] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[50] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, "A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem," *Science*, vol. 292, no. 5516, pp. 472–475, 2001.

[51] S. Verdu *et al.*, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[52] D. Pastor and Q.-T. Nguyen, "Random distortion testing and optimality of thresholding tests," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4161–4171, Aug. 2013.

[53] M. van Hoeij, "Factoring polynomials and the knapsack problem," *J. Number Theory*, vol. 95, no. 2, pp. 167–189, Aug. 2002.

[54] M. A. Hanson, "On sufficiency of the Kuhn–Tucker conditions," *J. Math. Anal. Appl.*, vol. 80, no. 2, pp. 545–550, Apr. 1981.

[55] M. Syed, P. Pardalos, and J. Principe, "Invexity of the minimum error entropy criterion," *IEEE Signal Process. Lett.*, vol. 20, no. 12, pp. 1159–1162, Dec. 2013.

[56] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York, NY, USA: Academic, 2014.

[57] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2012.


Authorized licensed use limited to: ULAKBIM UASL - DOKUZ EYLUL UNIVERSITESI. Downloaded on October 28,2022 at 12:40:10 UTC from IEEE Xplore. Restrictions apply.