Virtual distillation with noise dilution

Yong Siah Teo,* Seongwook Shin, Hyukgun Kwon, Seok-Hyung Lee, and Hyunseok Jeong†
Department of Physics and Astronomy, Seoul National University, Seoul 08826, South Korea

(Received 2 November 2022; accepted 18 January 2023; published 8 February 2023)

Virtual distillation is an error-mitigation technique that reduces quantum-computation errors without assuming the noise type. In scenarios where the user of a quantum circuit is required to additionally employ peripherals, such as delay lines, that introduce excess noise, we find that the error-mitigation performance can be improved if the peripheral, whenever possible, is split across the entire circuit, that is, when the noise channel is uniformly distributed in layers within the circuit. We show that under the multiqubit loss and Pauli noise channels, respectively, for a given overall error rate, the average mitigation performance improves monotonically as the noisy peripheral is split (diluted) into more layers, with each layer sandwiched between subcircuits that are sufficiently deep to behave as 2-designs. For both channels, analytical and numerical evidence show that second-order distillation is generally sufficient for (near-)optimal mitigation. We propose an application of these findings in designing a quantum-computing cluster that houses realistic noisy intermediate-scale quantum circuits that may be shallow in depth, where measurement detectors are limited and delay lines are necessary to queue output qubits from multiple circuits.

DOI: 10.1103/PhysRevA.107.022608

I. INTRODUCTION

In principle, owing to the existence of universal quantum gate sets [1–5] for computation, full-fledged quantum computers [6–9] that obey the laws of quantum mechanics are promising devices that permit universal fault-tolerant quantum computation [10–16], with the possibility of surpassing the performances of presently known classical algorithms. In practice, however, numerous challenges remain to be resolved before such a quantum vision can be realized. These include the qualities of qubit sources, gates and detectors [17–19], and the exponentially large number of components necessary to construct arbitrary quantum circuits [20].

At present, one only has access to noisy intermediate-scale quantum (NISQ) devices [21] that are capable of manipulating less than 1000 qubits using noisy unitary gates and measurements. These limitations motivated the development of several kinds of NISQ algorithms [22–30]. Most notably, variational quantum algorithms (VQAs) [31–36] perform computations using both classical and NISQ devices in a hybrid fashion, which find applications in variational quantum eigensolvers designed for quantum-chemistry [37–39] and combinatorial problems [40,41], and quantum machine learning [42–49].

Owing to noise accompanying any NISQ device, the output state of its circuit will deviate from the target state of interest to the computation procedure. Without resorting to quantum tomography [50–54] to certify the circuit output quality, the techniques of error mitigation may be considered as alternatives to directly reduce errors due to noise channels. The names of specific error-mitigation procedures have grown rapidly in recent years, which makes it implausible to cite them all. While interested readers may refer to [23,55] for a comprehensive survey, we will highlight some representative techniques. Methods such as Richardson or zero-noise extrapolation and probabilistic error cancellation [56–58] for estimating the circuit output in the near-noiseless regime require knowledge about the noise model and precise control of the quantum-circuit parameters, the former of which could be replaced by gate-set tomography [59–62], which should be carried out prior to the quantum computation. Other methods such as the subspace reduction [63,64] method demand additional ancillary qubits appended to the main NISQ circuit, which can be very large especially when the noise model is unknown. Iterative power-series-based or perturbative methods hold promise to mitigate errors under noise-agnostic situations. Nevertheless, such methods may require a large number of measurements that may be reduced by employing manipulative tricks for certain target measurement observables [65] and in other cases may only work on invertible noise channels without bias [66].

Virtual distillation [67–69] is yet another technique that can mitigate noise of small error rates in a model-agnostic manner. Moreover, mitigation happens on the fly either with an external correcting circuit [68] or with efficient data post-processing using shadow tomography [70] that requires only randomized single-qubit unitary gates. Consequently, this technique can cope with noise drifts, which is an attractive feature in addition to its technical simplicity that permits accessible analysis.

While it is true that noise can influence a NISQ device anywhere in an uncontrollable way, in this work we focus on scenarios where, other than errors originating from ambient environment, certain external peripherals employed in addition to the main quantum-computation circuit could

*yong.siah.teo@gmail.com
†h.jeong37@gmail.com

result in excess noise on the entire device. Examples of such peripherals could be delay lines and active switches. In these cases, one has the freedom of arranging peripherals within the circuit, thereby altering the effective noise channel acting on the device. This work investigates both the multiqubit loss and Pauli noise channels, which describe photon losses [71–80] and polarization disturbances [81–84], respectively, in optical components. For small error rates, we analytically show that virtual distillation can exponentially mitigate errors due to the loss channel with increasing distillation order, whereas mitigation improvement stagnates beyond the second distillation order for the multiqubit Pauli channel. Furthermore, virtual distillation can, on average, better mitigate errors when the peripherals are split into more layers across the quantum circuit. The former is equivalent to diluting the noise channel into multiple layers within the circuit, where one such noise layer (except for the last one) is sandwiched between two subcircuits. All analysis is carried out under the assumption that every sandwiching subcircuit is approximately a 2-design [85].

Finally, we give a practical application to this main result by studying the performance of virtual distillation on outputs of a quantum-computing cluster that contains several hardware-efficient circuits and limited detectors. In this situation, delay lines are necessary to queue the output qubits coming from concurrent circuits. According to the main result, under the same total delay time, it is evidently better to transmit qubits of uniformly delayed unitary operations to the detectors instead of delaying the transmission of qubits after all unitary operations are performed quickly.

II. VIRTUAL DISTILLATION

Suppose that an effective noise channel Φ_μ acts on a pure state $|\ \rangle\langle\ |$ of Hilbert-space dimension d and is parametrized by $0 \leqslant \mu \leqslant 1$, which quantifies the strength of Φ. Then the resulting noisy state

$$\rho' = \Phi_\mu[\rho] = |\lambda_0(\mu)\rangle\lambda_0(\mu)\langle\lambda_0(\mu)| + \sum_{k=1}^{d-1} |\lambda_k(\mu)\rangle\lambda_k(\mu)\langle\lambda_k(\mu)| \tag{1}$$

possesses a spectral decomposition with eigenvalues ordered according to $\lambda_0(\mu) > \lambda_1(\mu) \geqslant \lambda_2(\mu) \geqslant \cdots \geqslant \lambda_{d-1}(\mu)$, where for sufficiently small μ the eigenstate $|\lambda_0(\mu)\rangle\langle\lambda_0(\mu)|$ is in general close to the target pure state $|\ \rangle\langle\ |$. On the other hand, ρ' deviates significantly from $|\ \rangle\langle\ |$ at maximal channel noise ($\mu = 1$). Since Φ_μ is completely positive and trace preserving, it may also be expressed as

$$\rho' = \sum_m K_{m,\mu}|\ \rangle\langle\ |K_{m,\mu}^\dagger \tag{2}$$

using a set of Kraus operators $\{K_{0,\mu}, K_{1,\mu}, K_{2,\mu}, \ldots\}$ of the general property $K_{m,\mu=0} = \delta_{m,0}1$. Using a Hermitian operator basis $\{\Gamma_l\}_{l=0}^{d^2-1}$ such that $\Gamma_0 = 1/\sqrt{d}$, $\mathrm{tr}\{\Gamma_{l>0}\} = 0$, and $\mathrm{tr}\{\Gamma_l\Gamma_{l'}\} = \delta_{l,l'}$, we have the expansion

$$K_{m,\mu} = \frac{\gamma_0^{(m,\mu)}}{\sqrt{d}}1 + \sum_{l=1}^{d^2-1} \gamma_l^{(m,\mu)}\Gamma_l \tag{3}$$

in terms of complex coefficients $\gamma_l^{(m,\mu)}$, where $\gamma_l^{(m,\mu=0)} = \sqrt{d}\delta_{l,0}\delta_{m,0}$. We may now group all nondominant terms together and give an equivalently exact representation for ρ',

$$\rho' = |\ \rangle[1-\epsilon(\mu)]\langle\ | + \underbrace{\epsilon(\mu)\tilde{\rho}_{\mathrm{err}}(\mu, |\ \rangle\langle\ |)}_{\text{nondominant}}, \tag{4}$$

with $\epsilon(\mu) = 1 - |\gamma_0^{(0,\mu)}|^2/d$, so that $\epsilon(\mu \to 0) \to 0$. The state $\tilde{\rho}_{\mathrm{err}}(\mu, |\ \rangle\langle\ |)$ is the error component owing to $\Phi_\mu = \Phi_\epsilon$ of an error rate $\epsilon = \epsilon(\mu)$. Very generally, $\tilde{\rho}_{\mathrm{err}}(\mu, |\ \rangle\langle\ |)$ can be a nonlinear operator function of μ and $[\tilde{\rho}_{\mathrm{err}}(\mu, |\ \rangle\langle\ |), |\ \rangle\langle\ |] \neq 0$.

A special case of (4) arises in the regime $\mu, \epsilon \ll 1$. In this case, it can be deduced (see Appendix A) that

$$\rho' = |\ \rangle(1-\epsilon)\langle\ | + \epsilon\rho_{\mathrm{err}}(|\ \rangle\langle\ |), \tag{5}$$

where ρ_{err} is an ϵ-independent error component, but could still depend on $|\ \rangle\langle\ |$. Apart from this approximation, Eq. (5) is exact with *any* ϵ for many commonly known noise channels.

There exists a simple error-mitigation procedure that utilizes a basic linear-algebraic principle. To illustrate such a procedure, let us first recall, which we have earlier assumed with a sleight of hand, that the dominant eigenvalue $\lambda_0(\mu)$ in Eq. (1) is nondegenerate, which is the typical case for noisy environments, as the situation of coincidentally having another dominant eigenstate of the same eigenvalue is unlikely for small ϵ. Then, because $\lambda_0(\mu) > \lambda_{k>0}(\mu)$, it is clear that

$$\lim_{M\to\infty} \frac{\rho'^M}{\mathrm{tr}\{\rho'^M\}} = |\lambda_0(\mu)\rangle\langle\lambda_0(\mu)|, \tag{6}$$

that is, raising ρ' to a very large power M (the distillation order) and normalizing the answer amplifies the dominant eigenvalue $\lambda_0(\mu)$, thereby asymptotically leading to the singly dominant eigenstate. This method may be traced back to von Mises and Pollaczek-Geiringer [86] and is a common numerical technique for finding the largest eigenvalue of a matrix. Hence, for a sufficiently small μ or ϵ, this dominant pure state is close to the target state $|\ \rangle\langle\ |$ under such a simple purification or distillation scheme.

In a VQA setting, one is generally interested in measuring the expectation value $\langle O\rangle$ of a Hermitian observable O with respect to some target $|\ \rangle\langle\ |$. The use of such a distillation scheme would therefore entail the corresponding measurement of $\mathrm{tr}\{\rho'^M O\}/\mathrm{tr}\{\rho'^M\}$. Recipes to implement such a measurement from multiple copies of ρ' using two-qubit entangling gates have been proposed [67,68,87–92]. More recently, a separate idea of using shadow tomography [70,93–96] as an efficient postprocessing protocol for estimating $\mathrm{tr}\{\rho'^M O\}/\mathrm{tr}\{\rho'^M\}$ with only randomized single-qubit unitary rotations in addition to the VQA circuit further enhances the feasibility of this scheme. The name virtual distillation is fitting, since all practical implementations never directly generate the distilled state $\rho'^M/\mathrm{tr}\{\rho'^M\}$, but only estimate the corresponding observable measurements.

III. NOISE MODELS

A. Multiqubit loss channel

Although virtual distillation is agnostic to any particular noise model, for the purposes of analysis and discussion, we will investigate two particular classes of noise channels. The first class consists of the multiqubit loss channels. Let us start with the simplest case, that is, the single-qubit loss channel defined by the completely positive and trace-preserving (CPTP) map

$$\Phi_\epsilon^{\text{loss}}[\rho_{\text{qubit}}] = (1-\epsilon)\rho_{\text{qubit}} + |\text{vac}\rangle\epsilon\langle\text{vac}| \tag{7}$$

for any state ρ_{qubit}, where $|\text{vac}\rangle\langle\text{vac}|$ is the vacuum state. This map may be derived by solving for the response of the polarization degree of freedom with respect to the master equation [97]

$$\frac{\partial\rho'_{\text{qubit}}}{\partial t} = \gamma\sum_{j=0}^{1}\left(a_j\rho'_{\text{qubit}}a_j^\dagger - \frac{1}{2}a_j^\dagger a_j\rho'_{\text{qubit}} - \frac{1}{2}\rho'_{\text{qubit}}a_j^\dagger a_j\right), \tag{8}$$

where a_j is the annihilation operator on the polarization ket $|j\rangle$ and $\epsilon = 1-e^{-\gamma t}$ relates to the evolution time period t and decay rate γ (see Appendix B). The linearity of $\Phi_\epsilon^{\text{loss}}$, or any CPTP map for that matter, implies the following operator actions:

$$\begin{aligned}
\Phi_\epsilon^{\text{loss}}[|0\rangle\langle 0|] &= |0\rangle(1-\epsilon)\langle 0| + |\text{vac}\rangle\epsilon\langle\text{vac}|,\\
\Phi_\epsilon^{\text{loss}}[|1\rangle\langle 1|] &= |1\rangle(1-\epsilon)\langle 1| + |\text{vac}\rangle\epsilon\langle\text{vac}|,\\
\Phi_\epsilon^{\text{loss}}[|0\rangle\langle 1|] &= |0\rangle(1-\epsilon)\langle 1|,\\
\Phi_\epsilon^{\text{loss}}[|\text{vac}\rangle\langle\text{vac}|] &= |\text{vac}\rangle\langle\text{vac}|.
\end{aligned} \tag{9}$$

In the multiqubit product basis, an n-qubit noiseless state ρ may be written as

$$\rho = \sum_{l_1,\ldots,l_n=0}^{1}\sum_{l'_1,\ldots,l'_n=0}^{1}|l_1,\ldots,l_n\rangle\rho_{l_1,\ldots,l_n;l'_1,\ldots,l'_n}\langle l'_1,\ldots,l'_n|, \tag{10}$$

where each qubit basis ket is now extended to the three-dimensional space inasmuch as

$$|0\rangle \mathrel{\widehat{=}} \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad |1\rangle \mathrel{\widehat{=}} \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad |\text{vac}\rangle \mathrel{\widehat{=}} \begin{pmatrix}0\\0\\1\end{pmatrix}. \tag{11}$$

After subjecting each qubit to the loss channel in Eq. (7) under a common small error rate ϵ, we arrive at the noisy state

$$\begin{aligned}
\rho' &= (1-\epsilon)^n\rho + \epsilon(1-\epsilon)^{n-1}\sum_{l,l'}\rho_{l;l'}(|\text{vac}\rangle\delta_{l_1,l'_1}\langle\text{vac}|\otimes|l_2\rangle\langle l'_2|\otimes\cdots\otimes|l_n\rangle\langle l'_n|\\
&\quad + |l_1\rangle\langle l'_1|\otimes|\text{vac}\rangle\delta_{l_2,l'_2}\langle\text{vac}|\otimes\cdots\otimes|l_n\rangle\langle l'_n| + \cdots\\
&\quad + |l_1\rangle\langle l'_1|\otimes\cdots\otimes|l_{n-1}\rangle\langle l'_{n-1}|\otimes|\text{vac}\rangle\delta_{l_n,l'_n}\langle\text{vac}|)\\
&\quad + \{\text{vacuum-related terms of higher } \epsilon \text{ orders}\}\\
&\cong (1-n\epsilon)\rho + \epsilon(|\text{vac}\rangle\langle\text{vac}|\otimes\text{tr}_1\{\rho\} + \cdots + \text{tr}_n\{\rho\}\otimes|\text{vac}\rangle\langle\text{vac}|),
\end{aligned} \tag{12}$$

where we see that a photon-loss action on the jth qubit is equivalent to a partial trace $\text{tr}_j\{\cdot\}$ applied to that qubit followed by a vacuum-state substitution. The set of summation variables $\{l_1,\ldots,l_n\}$ and the primed ones are concisely represented by $\boldsymbol{l}$ and $\boldsymbol{l}'$, respectively. We emphasize here that all vacuum-related terms are mutually orthogonal with each other and the noiseless state ρ. Such mutual orthogonality will prove advantageous in subsequent analysis for this channel.

B. Multiqubit Pauli channel

The second class of multiqubit Pauli channels not only generates enormous interest in the fields of error correction and quantum computing [98–108], but is also relevant in the discussion of noise models generated from common quantum-optical peripherals such as polarization disturbances in optical-fiber-based delay lines [81–83].

For a single qubit, the Pauli channel is defined by the four Kraus operators $K_0 = \sqrt{1-\epsilon}1$, $K_1 = \sqrt{\epsilon_1}X$, $K_2 = \sqrt{\epsilon_2}Y$, and $K_3 = \sqrt{\epsilon_3}Z$, where $1 \geqslant \epsilon = \epsilon_1+\epsilon_2+\epsilon_3$, resulting in the single-qubit CPTP map

$$\Phi_\epsilon^{\text{Pauli}}[\rho_{\text{qubit}}] = (1-\epsilon)\rho_{\text{qubit}} + \epsilon_1 X\rho_{\text{qubit}}X + \epsilon_2 Y\rho_{\text{qubit}}Y + \epsilon_3 Z\rho_{\text{qubit}}Z. \tag{13}$$

The Pauli channel is isotropically depolarizing when $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon/3$.

The multiqubit version naturally involves Kraus operators that are tensor products. Given a pure target n-qubit state ρ, the corresponding noisy state ρ' as a result of the multiqubit Pauli noise channel with identical qubit error rates, up to first order in ϵ_j, is given by

$$\rho' \cong (1-n\epsilon)\rho + \epsilon_1\sum_{j=1}^{n}X_j\rho X_j + \epsilon_2\sum_{j=1}^{n}Y_j\rho Y_j + \epsilon_3\sum_{j=1}^{n}Z_j\rho Z_j, \tag{14}$$

where X_j, for instance, refers to the X operator for the jth qubit. For this channel, the corresponding error component $\rho_{\text{err}} \propto \epsilon_1\sum_{j=1}^{n}X_j\rho X_j + \epsilon_2\sum_{j=1}^{n}Y_j\rho Y_j + \epsilon_3\sum_{j=1}^{n}Z_j\rho Z_j$ does not commute with ρ in general.

IV. NOISE DILUTION

Equations (12) and (14) are representatives of independent and identically distributed (i.i.d.) noise models, namely, that all qubit suffers from the same kind of noise of equal error

rate ϵ, where the resulting noisy n-qubit state ρ' takes the form

$$\rho_1' = \Phi_\epsilon^{\text{i.i.d.}}[\rho] = (1-\epsilon)^n \rho + [1-(1-\epsilon)^n]\rho_{\text{err}} \tag{15}$$

for any $0 \leqslant \epsilon \leqslant 1$ and some error component ρ_{err}. That $\Phi_\epsilon^{\text{i.i.d.}}$ is the independent and identically distributed noise channel due to a given quantum-circuit peripheral invites the concept of noise dilution. Suppose that this peripheral is used in conjunction with a quantum circuit that is described by the unitary operator U and can be split into several smaller peripherals. The corresponding noiseless output state $\rho = U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger$, where $|\mathbf{0}\rangle\langle\mathbf{0}|$ is some fixed initialized pure state. Given the possible decomposition $U = W_1 W_2 \cdots W_{L_{\text{err}}}$ in terms of the unitary operators $W_1, W_2, \ldots, W_{L_{\text{err}}}$, the user may choose to distribute the peripheral across U so that the output state of the circuit is given by

$$\begin{aligned}\rho_{L_{\text{err}}}' &= \Phi_{\epsilon/L_{\text{err}}}^{\text{i.i.d.}}\left[W_{L_{\text{err}}} \cdots \Phi_{\epsilon/L_{\text{err}}}^{\text{i.i.d.}}\left[W_2 \Phi_{\epsilon/L_{\text{err}}}^{\text{i.i.d.}}[W_1|\mathbf{0}\rangle\langle\mathbf{0}|W_1^\dagger]W_2^\dagger\right]\cdots W_{L_{\text{err}}}^\dagger\right] \\ &= \left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{L_{\text{err}}n}\rho + \left[1-\left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{L_{\text{err}}n}\right]\rho_{\text{err}}^{(L_{\text{err}})},\end{aligned} \tag{16}$$

where each of the L_{err} independent and identically distributed channels now acts with an error rate of ϵ/L_{err}. An example situation that is relevant for such a choice is the arrangement of delay lines that are necessary in many circumstances for the practical implementation of the NISQ device. In this situation, the user may choose to either use delay lines after the computation with U [Eq. (15)] or distribute them uniformly across U [Eq. (16)].

For small error rates ϵ,

$$\begin{aligned}\rho_1' &\cong (1-n\epsilon)\rho + n\epsilon\rho_{\text{err}}, \\ \rho_{L_{\text{err}}}' &\cong (1-n\epsilon)\rho + n\epsilon\rho_{\text{err}}^{(L_{\text{err}})},\end{aligned} \tag{17}$$

where we see that the noise level for both choices, measured as the weight attributed to ρ, is the same and only the error components differ [$\rho_{\text{err}}^{(L_{\text{err}}=1)} = \rho_{\text{err}}$]. Figure 1 illustrates (17) for $L_{\text{err}} = 1, 2, 4$ with a four-qubit quantum circuit, where W_1, W_2, W_3, and W_4 are unitary operators of subcircuits that make up U. We will compare the performance of virtual distillation on noisy output states for various values of L_{err}.

There is a technical exception to (15)–(17), namely, with the independent and identically distributed loss channels as described in Sec. III A. For such a channel, it turns out that

$$\rho_{L_{\text{err}}}' \cong \left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{L_{\text{err}}n}\rho + \frac{n\epsilon}{L_{\text{err}}}\left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{L_{\text{err}}n-1}\rho_{\text{err}}^{(L_{\text{err}})} \tag{18}$$

for a small loss error rate ϵ [see also either (B15) or (D1)], with a trace that is unpreserved even up to first order in ϵ whenever $L_{\text{err}} > 1$, unlike the second equality in (16). Upon a trace renormalization,

$$\rho_{L_{\text{err}}}' \cong \left(1-\frac{n\epsilon}{L_{\text{err}}}\right)\rho + \frac{n\epsilon}{L_{\text{err}}}\rho_{\text{err}}^{(L_{\text{err}})}. \tag{19}$$

The reason for the trace-lossy form in (18) is that while the trace of ρ' in (16) is preserved throughout the noise dilution procedure under typical noise channels [even up to first-order approximation in ϵ as in (17)], this is not the case for the independent and identically distributed loss channel. If after the measurement phase data corresponding to detector clicks with missing qubits are to be discarded anyway, then the effective action of the subcircuit W_l on a lossy state at every dilution layer amounts to losing information about the error component, unless $L_{\text{err}} = 1$. In this sense, the vacuum state is invisible to circuit operations in practice.

Physically, the noise dilution strategy outlined here is equivalent to a redistribution of noisy peripherals such that certain aspects of the peripherals are conserved. As a working example that will be the main theme of this article, consider a realistic physical situation where the lossy peripheral is a delay line of a certain decay rate γ for which the error rate $\epsilon = 1 - e^{-\gamma\tau}$ after some delay time period τ. For a small τ, we find that $\epsilon \cong \gamma\tau$, so the L_{err}-layered noise dilution scheme outlined here is equivalent to splitting the delay line into L_{err} equal delay times τ/L_{err} and distributing them evenly throughout the quantum circuit while preserving the total delay time τ.

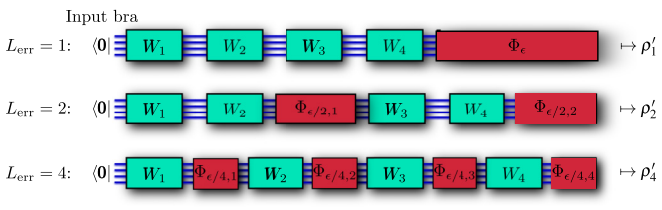

FIG. 1. Dilution of a peripheral independent and identically distributed noise channel $\Phi_\epsilon^{\text{i.i.d.}}$ in a four-qubit ($d = 2^4 = 16$) quantum circuit that accepts the initialized input state $|0\rangle\langle 0|$. Here the number of noise layers L_{err} is chosen to be 1 (no dilution), 2, and 4.

V. FIGURE OF MERIT AND CIRCUIT AVERAGING

To compare the mitigative power of virtual distillation in noise-diluted scenarios of various L_{err}, we take the figure of merit to be the Hilbert-Schmidt distance or mean-square error (MSE) [109] between a target pure state $\rho = |\ \rangle\langle\ | = U|\mathbf{0}\rangle\langle\mathbf{0}|U^\dagger$ and the mitigated state with respect to some noise-channel map $\Phi_\epsilon^{\text{i.i.d.}}$ of error rate ϵ. This is defined as

$$\mathcal{D} = \left\langle \text{tr}\left\{\left(\rho - \frac{\rho'^M}{\text{tr}\{\rho'^M\}}\right)^2\right\}\right\rangle, \tag{20}$$

where the average $\langle\cdot\rangle$ is taken over all possible independent circuit unitary operators. For instance, in Fig. 1, based on that particular decomposition of circuit unitary U, the average is taken over all possible W_1, W_2, W_3, and W_4. Such a

circuit-averaged figure of merit quantifies the average accuracy over all possible randomly chosen circuit parameters that define U.

To obtain analytical formulas, we will assume that all unitary operators W_j are 2-designs [110], that is, their first- and second-moment averages, such as $\langle W_j O_1 W_j^\dagger\rangle$ and $\langle W_j^{\otimes 2} O_2 W_j^{\dagger\otimes 2}\rangle$ with any operators O_1 and O_2 of appropriate dimensions, are those of the Haar measure over the unitary group [111,112]. Useful identities that apply for any d-dimensional 2-design unitary V, d-dimensional observable O_1, and d^2-dimensional observables O_2, A, B, and C include [36,113]

$$\langle V O_1 V^\dagger\rangle = \frac{\mathrm{tr}\{O_1\}}{d}, \tag{21}$$

$$\langle V^{\otimes 2} O_2 V^{\dagger\otimes 2}\rangle = \left(\frac{\mathrm{tr}\{O_2\}}{d^2-1} - \frac{\mathrm{tr}\{O_2\tau\}}{d(d^2-1)}\right)1 + \left(\frac{\mathrm{tr}\{O_2\tau\}}{d^2-1} - \frac{\mathrm{tr}\{O_2\}}{d(d^2-1)}\right)\tau, \tag{22}$$

$$\langle VAV^\dagger BVCV^\dagger\rangle = \frac{\mathrm{tr}\{A\}\mathrm{tr}\{C\}B + \mathrm{tr}\{B\}\mathrm{tr}\{AC\}1}{d^2-1} - \frac{\mathrm{tr}\{A\}\mathrm{tr}\{B\}\mathrm{tr}\{C\}1 + \mathrm{tr}\{AC\}B}{d(d^2-1)}, \tag{23}$$

where τ is the $d\times d$ bipartite SWAP operator. A popular 2-design distribution that we will adopt for more general simulation runs is the Haar measure itself. According to [114], one may generate random unitary operators (U_{Haar}) of dimension d that are distributed according to this measure from the following numerical recipe.

(i) Generate a random $d\times d$ matrix $\mathbf{A}$ with entries independent and identically distributed standard Gaussian distribution.

(ii) Compute the matrices $\mathbf{Q}$ and $\mathbf{R}$ from the QR decomposition $\mathbf{A} = \mathbf{QR}$.

(iii) Define $\mathbf{R}_{\mathrm{diag}} = \mathrm{diag}\{\mathbf{R}\}$.

(iv) Define $\mathbf{L} = \mathbf{R}_{\mathrm{diag}} \oslash |\mathbf{R}_{\mathrm{diag}}|$ ($\oslash$ refers to the Hadamard division).

(v) Define $U_{\mathrm{Haar}} \widehat{=} \mathbf{QL}$.

VI. RESULTS

A. 2-design networks

1. Independent and identically distributed loss channel

Because the vacuum state $|\mathrm{vac}\rangle\langle\mathrm{vac}|$ resides in the Hilbert-space sector that is orthogonal to that of ρ, for small error rate ϵ, it is possible to obtain the complete expression for the MSE between the target pure state $\rho = |\ \rangle\langle\ |$ and the mitigated state ρ' using virtual distillation, which reads

$$\mathcal{D}^{\mathrm{i.i.d.\ loss}}_{M,L_{\mathrm{err}}} = \left(\frac{\epsilon}{L_{\mathrm{err}}}\right)^{2M}\left[n\langle \mathrm{tr}\{\mathrm{tr}_1\{|\ \rangle\langle\ |\}^{2M}\}\rangle + \left\langle\left(\sum_{j=1}^{n}\mathrm{tr}\{\mathrm{tr}_j\{|\ \rangle\langle\ |\}^{M}\}\right)^2\right\rangle\right], \tag{24}$$

where the two averages $\langle\mathrm{tr}\{\mathrm{tr}_1\{|\ \rangle\langle\ |\}^{2M}\}\rangle$ and $\langle\{\sum_{j=1}^{n}\mathrm{tr}\{\mathrm{tr}_j\{|\ \rangle\langle\ |\}^{M}\}\}^2\rangle$ would depend specifically on the kind of 2-design circuit ansatz employed in the application. For $M = 1$, it can be shown that

$$\mathcal{D}^{\mathrm{i.i.d.\ loss}}_{M,L_{\mathrm{err}}} = \left(\frac{\epsilon}{L_{\mathrm{err}}}\right)^2\left[\frac{n(d+4)}{2(d+1)} + n^2\right], \tag{25}$$

which universally holds for all 2-design circuits (derived in Appendix B).

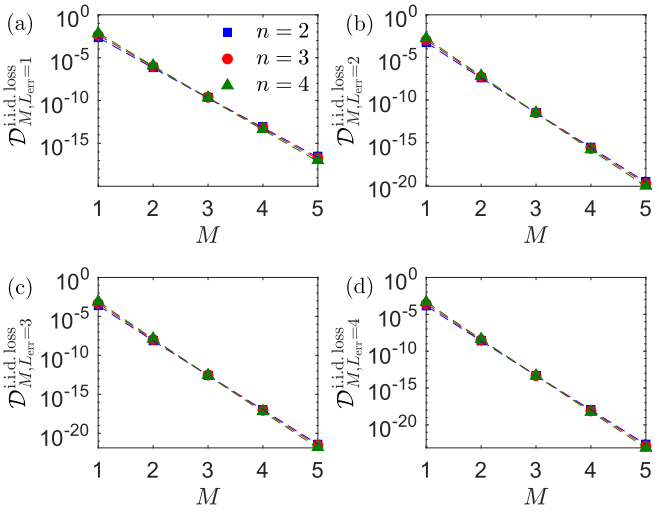

FIG. 2. Comparison between Monte Carlo simulations (markers) and theoretical predictions (dashed curves) of $\mathcal{D}^{\mathrm{i.i.d.\ loss}}_{M,L_{\mathrm{err}}}$ for $\epsilon = 0.02$, various numbers of qubits n and virtual-distillation order M, and (a) $L_{\mathrm{err}} = 1$, (b) $L_{\mathrm{err}} = 2$, (c) $L_{\mathrm{err}} = 3$, and (d) $L_{\mathrm{err}} = 4$. For each value of L_{err}, 100 sets of Haar unitary operators $\{W_l\}_{l=1}^{L_{\mathrm{err}}}$ are used to average every plot marker. All graphs are plotted with a vertical logarithmic scale.

We may now extract some key behaviors concerning virtual distillation with peripheral noise dilution. The first observation is that the MSE scales exponentially with M in the error rate ϵ for a fixed L_{err}. If the magnitude of ϵ is about 0.01, say, then a second-order virtual distillation results in an MSE that is about 10^{-8} in orders of magnitude. We therefore see that for such error rates, $M = 2$ is typically sufficient for practical purposes.

The next important finding is that for fixed ϵ and order M, the MSE decreases with increasing L_{err} according to a power law. This verifies the intuition that buffering the peripheral noise channel using the existing circuit components indeed results in higher mitigative power at least with virtual distillation. Figures 2–4 graphically showcase the precision of Eq. (24) relative to Monte Carlo simulations. In all simulations, the circuit unitary operators are distributed according to the Haar measure.

Moreover, for the independent and identically distributed loss channel, both Eq. (24) and Fig. 2 tell us that virtual distillation results in mitigated states that are asymptotically unbiased. That is, in the limit of large M, $\mathcal{D}^{\mathrm{i.i.d.\ loss}}_{M\to\infty,L_{\mathrm{err}}} \to 0$. One can understand this property by inspecting the structure of $\rho'^M/\mathrm{tr}\{\rho'^M\}$ in (B17), which possesses only the target-only term and the ϵ^M-dependent term, with no other cross terms of lower ϵ orders. These cross terms are responsible for a nonzero bias in $\lim_{M\to\infty}\rho'^M/\mathrm{tr}\{\rho'^M\}$, which is miss-

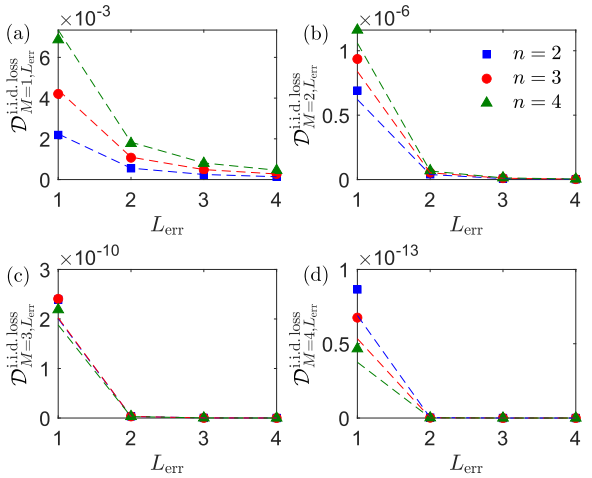

FIG. 3. Comparison between Monte Carlo simulations (markers) and theoretical predictions (dashed curves) of $\mathcal{D}^{\text{i.i.d. loss}}_{M,L_{\rm err}}$ for $\epsilon = 0.02$, various numbers of qubits n and diluted layers $L_{\rm err}$, and (a) $M = 1$, (b) $M = 2$, (c) $M = 3$, and (d) $M = 4$. For each value of M, 100 sets of Haar unitary operators $\{W_l\}_{l=1}^{L_{\rm err}}$ are used to average every plot marker.

ing for this channel, a consequence of the orthogonality between $|\text{vac}\rangle$ and the circuit Hilbert-space sector.

2. *Independent and identically distributed Pauli channel*

The independent and identically distributed Pauli channel takes on a very different structure than the independent and identically distributed loss channel. Most notably, its action described by (14) entails an error component that typically does not commute with the target state ρ. Furthermore, the nonorthogonality of this error component makes it persistent regardless of the virtual-distillation order [see Eq. (C9)], that

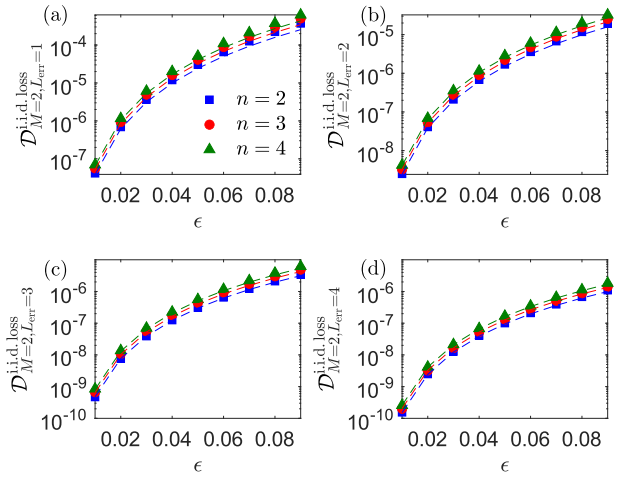

FIG. 4. Comparison between Monte Carlo simulations (markers) and theoretical predictions (dashed curves) of $\mathcal{D}^{\text{i.i.d. loss}}_{M,L_{\rm err}}$ for a virtual-distillation order of $M = 2$, various numbers of qubits n and error rate ϵ, and (a) $L_{\rm err} = 1$, (b) $L_{\rm err} = 2$, (c) $L_{\rm err} = 3$, and (d) $L_{\rm err} = 4$. For each value of $L_{\rm err}$, 100 sets of Haar unitary operators $\{W_l\}_{l=1}^{L_{\rm err}}$ are used to average every plot marker. All graphs are plotted with a vertical logarithmic scale.

is, the error component of the distilled noisy state $\rho'^M/\text{tr}\{\rho'^M\}$ carries a coefficient that is always of the leading order $O(\epsilon_1, \epsilon_2, \epsilon_3)$.

More specifically, the respective MSE expressions of distillation orders $M = 1$ and $M \geqslant 2$ for small ϵ and arbitrary $L_{\rm err}$ are

$$\mathcal{D}^{\text{i.i.d. Pauli}}_{M=1,L_{\rm err}} = (n\epsilon)^2\left[\frac{d^3}{(d+1)(d^2-1)} - \frac{d}{L_{\rm err}(d+1)(d^2-1)}\right] + \frac{nd}{L_{\rm err}(d+1)}\sum_{l=1}^{3}\epsilon_l^2, \tag{26}$$

$$\mathcal{D}^{\text{i.i.d. Pauli}}_{M\geqslant 2,L_{\rm err}} = 2\sum_{l,l'=1}^{3}\frac{\epsilon_l\epsilon_{l'}}{L_{\rm err}^2}\sum_{j,j'=1}^{n}\Big(\langle\text{tr}\{\rho T_j^{(l)}T_{j'}^{(l')}\}\rangle - \langle\text{tr}\{\rho T_j^{(l)}\}\text{tr}\{\rho T_{j'}^{(l')}\}\rangle\Big), \tag{27}$$

$$T_j^{(l)} = P_j^{(l)}\rho P_j^{(l)} + W_{L_{\rm err}}P_j^{(l)}\rho_{L_{\rm err}-1}P_j^{(l)}W_{L_{\rm err}}^\dagger + \cdots + W_{L_{\rm err}}W_{L_{\rm err}-1}\cdots W_2P_j^{(l)}\rho_1P_j^{(l)}W_2^\dagger\cdots W_{L_{\rm err}-1}^\dagger W_{L_{\rm err}}^\dagger, \tag{28}$$

where $P_j^{(l)}$ is a single-qubit Pauli operator (see Appendix C). For the same n and $L_{\rm err}$, the MSE therefore explicitly depends on the asymmetry of the Pauli channel in the loss rates ϵ_j [115]. A significant departure from the independent and identically distributed loss channel is that the MSE for $M \geqslant 2$ is independent of M. This may appear counterintuitive at first, but a closer inspection establishes consistency with (C9), that is, perfect error mitigation is not possible in the presence of a persistent error component for any M, since it is always accompanied by error coefficients of unit leading order regardless of the value of M. Without loss of generality, we will subsequently take the independent and identically distributed Pauli channel to be depolarizing, or $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon/3$.

Another observation we can make is that for fixed $\epsilon \ll 1/n$ and very large circuits ($n, d \gg 2$), $\mathcal{D}^{\text{i.i.d. Pauli}}_{M=1,L_{\rm err}} \to n\sum_{l=1}^{3}\epsilon_l^2/L_{\rm err}$, that is, the unmitigated distance increases with n and asymptotically approaches a finite limit. On the other hand, Fig. 5 shows that the mitigated MSE for $M \geqslant 2$ drops gradually with increasing n (excluding $n = 2$). As with all simulations up to this section, all circuit unitary operators are assumed to possess a Haar distribution. Figure 6 confirms the mitigation improvement as $L_{\rm err}$ increases. Figure 7 supplies the MSE graphs with respect to different error rates while fixing M and $L_{\rm err}$.

B. Eigenvalue distribution of the error component

To understand matters properly, let us for the moment consider the special case [67] of a noise-channel map

$$\rho' = |\lambda_0\rangle(1-\epsilon_0)\langle\lambda_0| + \epsilon_0\sum_{k=1}^{d-1}|\lambda_k\rangle p_k\langle\lambda_k| \tag{29}$$

that brings the ideal target pure state $|\ \rangle\langle\ | \equiv |\lambda_0\rangle\langle\lambda_0|$ to a noisy mixed state ρ', where the error component

$$\rho_{\rm err} = \sum_{k=1}^{d-1}|\lambda_k\rangle p_k\langle\lambda_k| \tag{30}$$

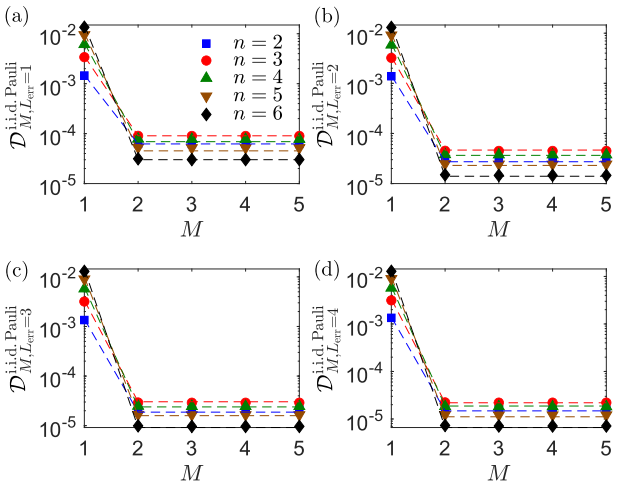

FIG. 5. Comparison between Monte Carlo simulations (markers) and theoretical predictions (dashed curves) of $\mathcal{D}^{\text{i.i.d. Pauli}}_{M,L_{\text{err}}}$ for $\epsilon = 0.02$, various numbers of qubits n and virtual-distillation order M, and (a) $L_{\text{err}} = 1$, (b) $L_{\text{err}} = 2$, (c) $L_{\text{err}} = 3$, and (d) $L_{\text{err}} = 4$. The channel is depolarizing with $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon/3$. For each value of L_{err}, 1000 sets of Haar unitary operators $\{W_l\}_{l=1}^{L_{\text{err}}}$ are used to average every plot marker. All graphs are plotted with a vertical logarithmic scale.

resides in the orthogonal subspace of the target ($\sum_{k=1}^{d-1} p_k = 1$ and $\sum_{k=0}^{d-1} |\lambda_k\rangle\langle\lambda_k| = 1$). Under a fixed error rate ϵ_0 and virtual-distillation order M, it is interesting to search for the optimal distribution of $\boldsymbol{p} = (p_1 p_2 \cdots p_{d-1})$ that minimizes the MSE defined in (20) (without circuit averaging for this situation). This finding would at least serve as a guide towards the optimal noise-coping strategy under this special case.

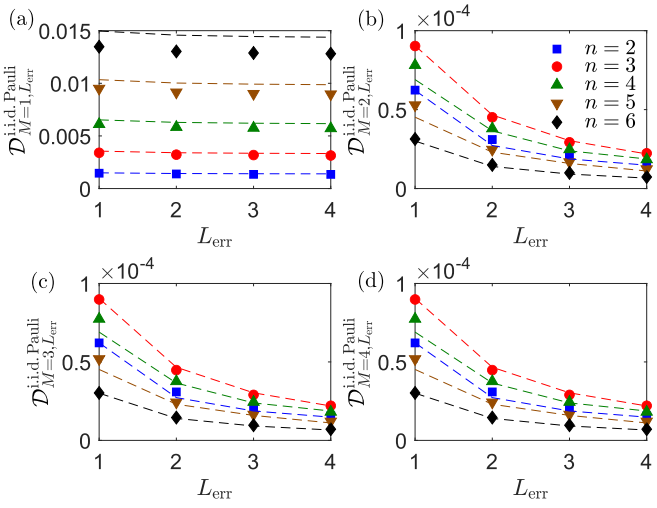

FIG. 6. Comparison between Monte Carlo simulations (markers) and theoretical predictions (dashed curves) of $\mathcal{D}^{\text{i.i.d. Pauli}}_{M,L_{\text{err}}}$ for $\epsilon = 0.02$, various qubit numbers n and L_{err}, and (a) $M = 1$, (b) $M = 2$, (c) $M = 3$, and (d) $M = 4$. The noise channel is depolarizing. For each M, 1000 sets of Haar unitary $\{W_l\}_{l=1}^{L_{\text{err}}}$ are used to average every plot marker.

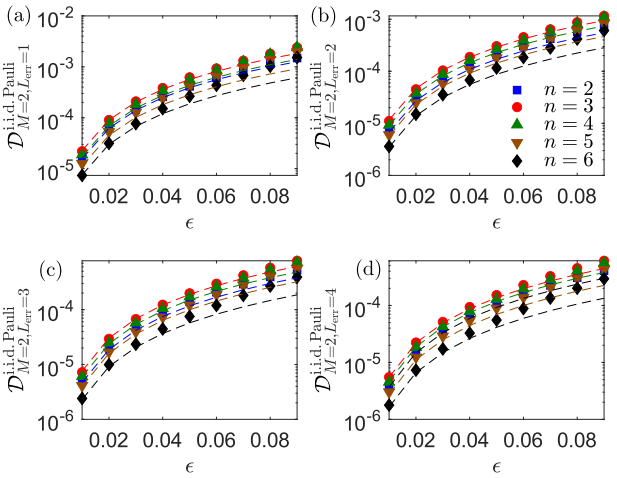

FIG. 7. Comparison between Monte Carlo simulations (markers) and theoretical predictions (dashed curves) of $\mathcal{D}^{\text{i.i.d. Pauli}}_{M,L_{\text{err}}}$ for a distillation order of $M = 2$, various n and ϵ, and (a) $L_{\text{err}} = 1$, (b) $L_{\text{err}} = 2$, (c) $L_{\text{err}} = 3$, and (d) $L_{\text{err}} = 4$. The noise channel is depolarizing. For each L_{err}, 1000 sets of Haar unitary $\{W_l\}_{l=1}^{L_{\text{err}}}$ are used to average every plot marker.

After some simple manipulation, upon defining $\mathcal{N}_{M,\epsilon_0} = (1-\epsilon_0)^M + \epsilon_0^M \sum_{k=1}^{d-1} p_k^M$, we find that for any $0 \leqslant \epsilon_0 \leqslant 1$,

$$\mathcal{D}_M = \text{tr}\left\{\left(\frac{\rho'^M}{\text{tr}\{\rho'^M\}} - |\ \rangle\langle\ |\right)^2\right\} = \left[1 - \frac{(1-\epsilon_0)^M}{\mathcal{N}_{M,\epsilon_0}}\right]^2 + \frac{\epsilon_0^{2M}}{\mathcal{N}^2_{M,\epsilon_0}} \sum_{k=1}^{d-1} p_k^{2M}. \tag{31}$$

Since $\{p_k\}_{k=1}^{d-1}$ is a discrete set of normalized probabilities, it may be parametrized as $p_k = a_k^2 / \sum_{k'=1}^{d-1} a_{k'}^2$, so the variation

$$\delta p_k = \frac{2}{\sum_{k''=1}^{d-1} a_{k''}^2}\left(a_k \delta a_k - p_k \sum_{k'=1}^{d-1} a_{k'} \delta a_{k'}\right) \tag{32}$$

implies that

$$\frac{\delta \mathcal{D}_M}{\delta a_k} \propto a_k \left(A_{M,k}(\boldsymbol{p}) - \sum_{k'=1}^{d-1} A_{M,k'}(\boldsymbol{p}) p_{k'}\right),$$

$$A_{M,k}(\boldsymbol{p}) = \left[(1-\epsilon_0)^M - \frac{(1-\epsilon_0)^{2M}}{\mathcal{N}_{M,\epsilon 0}} - \frac{\epsilon_0^{2M}}{\mathcal{N}_{M,\epsilon 0}} \sum_{k'=1}^{d-1} p_{k'}^{2M}\right] p_k^{M-1} + \epsilon_0^M p_k^{2M-1} \tag{33}$$

after some straightforward calculations. Upon setting the gradient to zero and solving the extremal equation $p_k \sum_{k'=1}^{d-1} A_{M,k'}(\boldsymbol{p}) p_{k'} = p_k A_{M,k}(\boldsymbol{p})$, it is easy to verify that $p_k = 1/(d-1)$ is the solution.[1] So the optimal strategy

[1]The interested reader may even try taking the gradient in (33) to set up a steepest-descent method and obtain the uniform probability distribution after many iterations.

should give a rank-$(d-1)$ ρ_{err} that possesses a uniform eigenspectrum.

The purpose of this section is to emphasize that, while the special situation in which the user can freely change noise-coping strategies, thereby varying the p_k, and simultaneously fix the error rate is an especially easy one to analyze, there are practical scenarios where the error rate will also be inevitably varied as a consequence of such a strategy change. An example is noise dilution on the independent and identically distributed loss channel, as cautioned at the end of Sec. IV. Increasing the number of diluted noise layers L_{err} not only changes the error component, but also alters the effective error rate as a result of the non-trace-preserving character of circuit operations on lossy states [see Eq. (19)]. Therefore, all previous arguments leading to (33) are no longer valid with this channel and so a better noise-coping strategy does not necessarily lead to error components of eigenspectra closer to a uniform distribution.

To systematically compare the difference between the eigenspectrum $\boldsymbol{\lambda}$ of a $\tilde{d}$-dimensional error component ρ_{err} and the uniform distribution $\boldsymbol{\lambda}_{\text{unif}} = (1\,1\cdots1)^{\top}/\tilde{d}$, we take the Hellinger distance [109]

$$H(\boldsymbol{\lambda}, \boldsymbol{\lambda}_{\text{unif}}) = \frac{1}{\sqrt{2}}\sqrt{\sum_{k=0}^{\tilde{d}-1}\left(\sqrt{\lambda_k} - 1/\sqrt{\tilde{d}}\right)^2} \tag{34}$$

as the figure of merit. For n-qubit independent and identically distributed Pauli channels, the dimension $\tilde{d} = d = 2^n$. On the other hand, for n-qubit independent and identically distributed lossy scenarios, the inclusion of the third vacuum sector as in (11) as a way to account for losses is equivalent to $\tilde{d} = 3^n$. If the noiseless n-qubit target $\rho = |\ \rangle\langle\ |$ is a product state ($|\ \rangle = |\psi_1\rangle|\psi_2\rangle\cdots|\psi_n\rangle$), then the *exact* Hellinger distance is given by

$$H_x^{\text{i.i.d. loss}} = \left\{1 + \frac{1}{2\times 3^n} - \frac{(1-x)^{n/2}}{3^{n/2}\sqrt{1-(1-x)^n}} \times \left[\left(1+\sqrt{\frac{x}{1-x}}\right)^n - 1\right]\right\}^{1/2}, \quad x = \frac{\epsilon}{L_{\text{err}}}. \tag{35}$$

One can refer to Appendix D for the derivation and verify indeed that $H_{\epsilon/L_{\text{err}}}^{\text{i.i.d. loss}}$ increases with increasing L_{err}, which is in direct contrast with the special case discussed previously.

Figure 8(a) shows very similar behaviors in the Hellinger distances for general random states. On the other hand, the independent and identically distributed depolarizing channel [see Fig. 8(b)] yields distances that agree in trend with that for the special case. An important distinction from the independent and identically distributed loss channel is that the overall error rate in (17) remains as $n\epsilon$ for any L_{err}, as opposed to that in (19).

C. Quantum clusters of hardware-efficient networks

Finally, we give a practical application to virtual distillation with noise dilution. Suppose one is interested in designing a quantum-computer cluster (see Fig. 9) that houses several independent quantum circuits that are accessible by the public

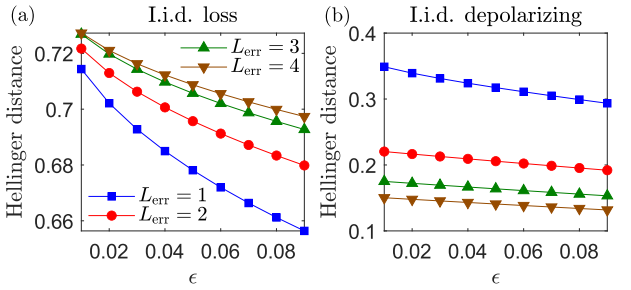

FIG. 8. Hellinger distance between the error-component eigenspectrum and uniform distribution for the independent and identically distributed (a) loss and (b) depolarizing channels, respectively. All graphs are averaged over the respective numbers of Haar circuit-unitary operators as stated in all previous simulation figures.

domain. At any given instance, a user may log in to the cluster and use one such n-qubit circuit. Additionally, we assume a limited number of detectors, so the output qubits need to be queued with delay lines for the final measurement. In a typical NISQ cluster, each unitary operator W_l describes a

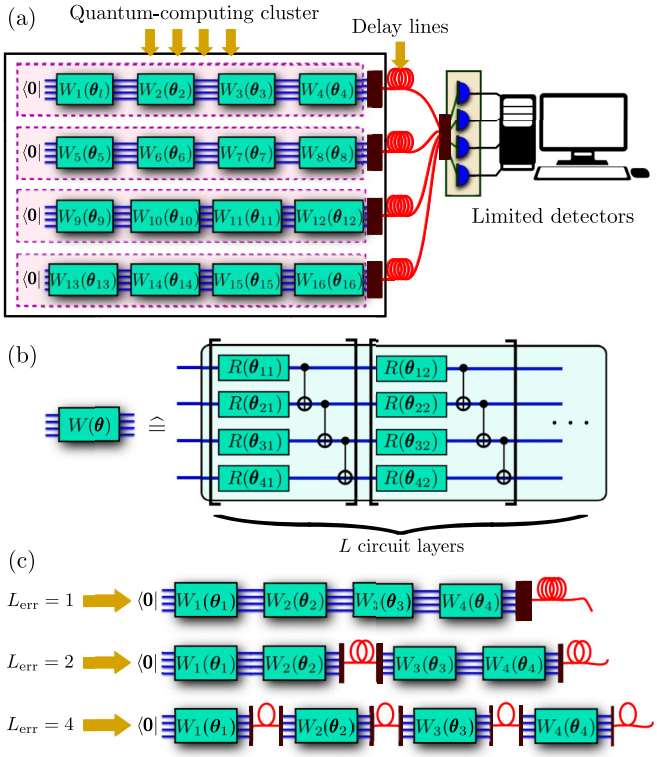

FIG. 9. (a) Quantum-computing cluster consisting of four separate NISQ circuits, where each circuit accepts a maximum of four qubits and may be employed independently by a user at any given instance. Delay lines are used to queue the output qubits for the final measurement with a (limited) set of four qubit detectors. (b) Each circuit unitary module $W_l(\theta_l)$, with θ_l the trainable parameters for the gate W_l, is made up of L layers of single-qubit rotations and CNOT gates. (c) Noise dilution is done by splitting the delay lines and distributing them equally among the circuit modules. For small error rates, dividing the delay length by L_{err} is equivalent to dividing the unsplit error rate by L_{err} in each diluted noise layer.

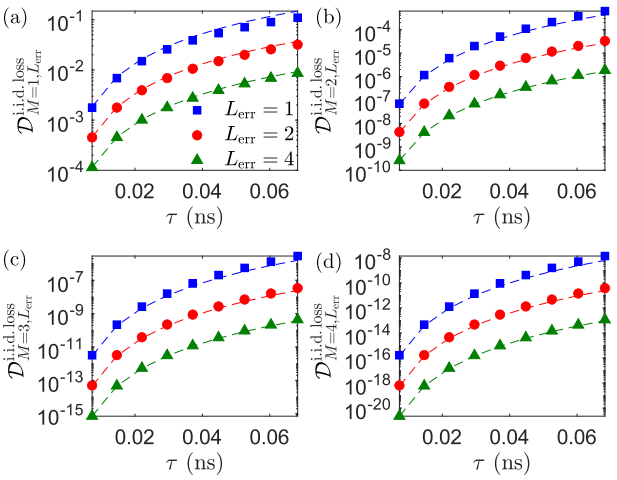

FIG. 10. Virtual-distillation MSE curves for four-qubit hardware-efficient circuits ($L = 2$) with respect to the total delay time τ under the independent and identically distributed loss channel. Performances for various M and L_{err} are illustrated.

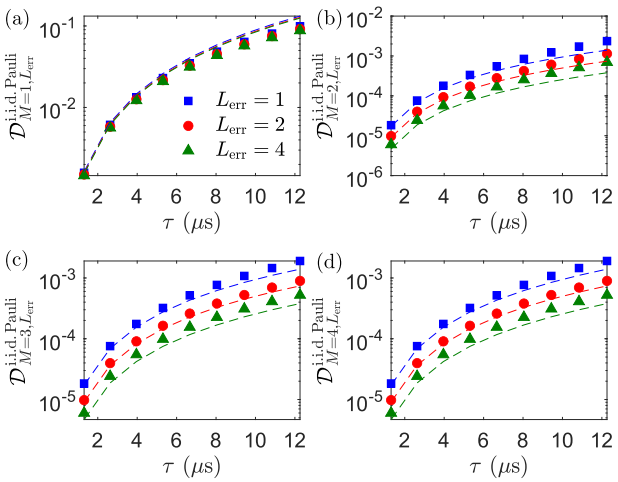

FIG. 11. Virtual-distillation MSE curves for four-qubit hardware-efficient circuits ($L = 2$) with respect to the total delay time τ under the independent and identically distributed depolarizing channel. Performances for various M and L_{err} are illustrated.

circuit comprising, say, L layers of passive components, which are single-qubit and controlled-NOT (CNOT) gates. From [85], randomized circuits of this kind are approximately 2-designs if $L = O(\text{poly}(n))$.

We will illustrate the results of noise dilution under such a practical situation by investigating two separate scenarios. The first scenario is where photon loss is a dominant source of noise in the delay lines, which could be the case when the cluster is integrated into a photonic chip. For more concrete simulations, we take the decay rate $\gamma_{\text{loss}} = 0.2$ dB cm^{-1} [116] or equivalently $\gamma_{\text{loss}} = 6$ dB ns^{-1}, where $\epsilon_{\text{loss}} = 1 - 10^{-\gamma_{\text{loss}}\tau/10}$ is related to the delay time τ (in nanoseconds). The second scenario is where polarization drifts occur more frequently in regular optical-fiber-based delay lines [81–83] such that the depolarizing channel serves as an appropriate noise model. As an example, we quote from [117] the depolarizing rate of about $\gamma_{\text{depol}} = 1.3\times10^{-4}$ s^{-1}, which is equivalent to a depolarization error rate of about 0.01 for a 400-m optical fiber. In other words, $\epsilon_{\text{depol}} = 1 - e^{-\gamma_{\text{depol}}\tau}$.

To present the simulation findings, we study virtual distillation performances on four-qubit hardware-efficient circuits, where the number of single-qubit-CNOT layers $L = 2$ such that a total of eight circuit layers is considered for each entire computation circuit. Hence, $L_{\text{err}} = 2$, for instance, implies that one of the diluted noise layers is sandwiched between two unitary subcircuits, each with $2L = 4$ circuit layers [see Fig. 9(c)]. Figures 10 and 11 plot the virtual-distillation performances on both independent and identically distributed channels based on the aforementioned specifications.

An interesting conclusion out of these numerical findings is that even though each circuit unitary operator W_l is shallow ($L = 2$), noise dilution can still effectively enhance error mitigation as L_{err} increases. Therefore, whenever the situation permits, the lesson here is that a uniform distribution of peripherals can in practice improve the performance of virtual distillation, at least for the independent and identically distributed loss and Pauli channels. For the case where the peripherals are delay lines, within the same total delay time, our results suggest that transmitting output qubits from uniformly delayed circuit operations to the detectors is a better error-mitigation strategy than delaying qubits from instantaneous circuit operations.

VII. DISCUSSION

Virtual distillation is a technically easy-to-use technique that mitigates errors in a noise-model-agnostic manner. Recent developments have further enhanced the feasibility of this error-mitigation technique. We applied this distillation procedure to practical situations in which general quantum-computing circuits are used in conjunction with additional peripherals that could introduce excess noise.

Under the general 2-design assumption about quantum circuits, and multiqubit loss and Pauli channels as models for the excess noise, we analytically showed (supported with numerical simulations) that for the same total error rate, distributing the peripherals homogeneously across the entire quantum circuit improves the quality of the error-mitigated output state as opposed to collecting all peripherals at one place: The more noise layers the peripheral excess-noise channel is diluted into, the better the mitigation results (Figs. 3 and 6). Furthermore, it turns out that virtual distillation exponentially reduces errors due to losses as the order increases, but gives the same magnitude of error reduction under the Pauli channel for all distillation orders greater than one up to the leading order of the error rate (Figs. 2 and 5). These two different behaviors stem from the types of error component generated by each noise model, where the one from losses is orthogonal to the multiqubit Hilbert-space sector and that from the Pauli channel generally does not commute with the target state. The latter error component is therefore persistent against virtual distillation.

Such findings come in handy when designing a quantum-computing cluster consisting of several noisy intermediate-scale quantum-circuit networks, where having a huge number of measurement detectors accommodating all these networks

is economically difficult. In this case, delay lines are generally used to queue output qubits for acquiring the final measurement data set. Our results suggest that splitting the delay lines homogeneously across the circuits can typically improve the error-mitigative power with virtual distillation applied to such situations, even when circuit depths are not deep.

ACKNOWLEDGMENTS

We thank Yosep Kim for fruitful discussions. This work was supported by Hyundai Motor Company, the Brain Korea 21 FOUR Project grant funded by the Korean Ministry of Education, the National Research Foundation of Korea grants funded by the Korea government (Grants No. NRF-2020R1A2C1008609, No. NRF-2020K2A9A1A06102946, No. NRF-2022M3K4A1097117, No. NRF-2019R1A6A1A10073437, and No. NRF-2022M3E4A1076099) via the Institute of Applied Physics at Seoul National University, and by the Institute of Information and Communications Technology Planning and Evaluation grants funded by the Korea government (MSIT) (Grants No. IITP-2021-0-01059 and No. IITP-2022-2020-0-01606).

APPENDIX A: SMALL-ERROR REGIME

The error component $\rho_{\rm err}$, being a quantum state itself, may be parametrized with the auxiliary complex operator $A_{\rm err}$ inasmuch as $\tilde{\rho}_{\rm err}(\mu, |\ \rangle\langle\ |) = A_\mu^\dagger A_\mu / {\rm tr}\{A_\mu^\dagger A_\mu\}$. When $\mu \ll 1$, we have the Taylor expansion

$$A_\mu \cong \Delta_0 + \mu\Delta_1 \tag{A1}$$

about $\mu = 0^+$, where $\Delta_0 = A_{0^+}$ and $\Delta_1 = \partial A_\mu/\partial\mu|_{\mu=0^+}$, so that

$$\tilde{\rho}_{\rm err}(\mu, |\ \rangle\langle\ |) \cong \frac{\Delta_0^\dagger\Delta_0}{{\rm tr}\{\Delta_0^\dagger\Delta_0\}} + \mu\left(\frac{\Delta_1^\dagger\Delta_0 + \Delta_0^\dagger\Delta_1}{{\rm tr}\{\Delta_0^\dagger\Delta_0\}} - \frac{\Delta_0^\dagger\Delta_0 {\rm tr}\{\Delta_1^\dagger\Delta_0 + \Delta_0^\dagger\Delta_1\}}{{\rm tr}\{\Delta_0^\dagger\Delta_0\}^2}\right). \tag{A2}$$

Similarly, by taking $\gamma_0^{(0,\mu)}$ in (3) to be real without loss of generality, $\gamma_0^{(0,\mu)} \cong \sqrt{d} + \mu\delta_1$, where we note that $\delta_1 = \partial\gamma_0^{(0,\mu)}/\partial\mu|_{\mu=0^+} < 0$. This gives us

$$\epsilon(\mu) \cong -\frac{2\mu\delta_1}{\sqrt{d}} \equiv \epsilon. \tag{A3}$$

Upon defining the constant $\rho_{\rm err} \equiv \Delta_0^\dagger\Delta_0/{\rm tr}\{\Delta_0^\dagger\Delta_0\}$, we obtain Eq. (5).

APPENDIX B: VIRTUAL DISTILLATION UNDER INDEPENDENT AND IDENTICALLY DISTRIBUTED LOSS CHANNEL

As a brief review demonstration, the solution to (8) is given by

$$\rho' = e^{{\rm t}\mathcal{L}}[\rho], \quad \mathcal{L}[\rho] = \gamma\sum_{j=0}^{1}\left(a_j\rho a_j^\dagger - \frac{1}{2}a_j^\dagger a_j\rho - \frac{1}{2}\rho a_j^\dagger a_j\right). \tag{B1}$$

Now, taking the number ket $|n_0 = 1, n_1 = 0\rangle \equiv |1,0\rangle \leftrightarrow |0\rangle$ for the zeroth (horizontal) polarization ket as the initial ket for ρ at $t = 0$, it is easy to see that

$$\begin{aligned}\mathcal{L}[|0\rangle\langle 0|] &\leftrightarrow \gamma(a_0|1,0\rangle\langle 1,0|a_0^\dagger - a_0^\dagger a_0|1,0\rangle\tfrac{1}{2}\langle 1,0| \\ &\quad - |1,0\rangle\tfrac{1}{2}\langle 1,0|a_0^\dagger a_0) \\ &= -\gamma(|0\rangle\langle 0| - |{\rm vac}\rangle\langle{\rm vac}|), \\ \mathcal{L}[|{\rm vac}\rangle\langle{\rm vac}|] &= 0.\end{aligned} \tag{B2}$$

The corresponding solution for the noisy state ρ' is therefore given by

$$\begin{aligned}\rho' &= |0\rangle\langle 0| + \sum_{l=1}^{\infty}\frac{t^l}{l!}\mathcal{L}^l[|0\rangle\langle 0|] \\ &= |0\rangle\langle 0| + \sum_{l=1}^{\infty}\frac{(-\gamma t)^l}{l!}(|0\rangle\langle 0| - |{\rm vac}\rangle\langle{\rm vac}|) \\ &= |{\rm vac}\rangle\langle{\rm vac}| + e^{-\gamma{\rm t}}(|0\rangle\langle 0| - |{\rm vac}\rangle\langle{\rm vac}|) \\ &= |0\rangle(1-\epsilon)\langle 0| + |{\rm vac}\rangle\epsilon\langle{\rm vac}|.\end{aligned} \tag{B3}$$

The reader may feel free to go through the same exercise and obtain all other actions stated in (9).

When $L_{\rm err} = 1$ and $\rho = |\ \rangle\langle\ |$ of dimension $d = 2^n$, Eq. (12) leads to

$$\begin{aligned}\rho'^M &\cong (1 - Mn\epsilon)\rho + \epsilon^M(|{\rm vac}\rangle\langle{\rm vac}| \otimes {\rm tr}_1\{\rho\}^M + \cdots \\ &\quad + {\rm tr}_n\{\rho\}^M \otimes |{\rm vac}\rangle\langle{\rm vac}|)\end{aligned} \tag{B4}$$

so that

$$\begin{aligned}\frac{\rho'^M}{{\rm tr}\{\rho'^M\}} &\cong \rho + \epsilon^M\Bigg(|{\rm vac}\rangle\langle{\rm vac}| \otimes {\rm tr}_1\{\rho\}^M + \cdots \\ &\quad + {\rm tr}_n\{\rho\}^M \otimes |{\rm vac}\rangle\langle{\rm vac}| - \rho\sum_{j=1}^{n}{\rm tr}\{{\rm tr}_j\{\rho\}^M\}\Bigg).\end{aligned} \tag{B5}$$

The MSE for $L_{\rm err} = 1$ therefore reads

$$\begin{aligned}\mathcal{D}^{\text{i.i.d. loss}}_{M,L_{\rm err}=1} &= \epsilon^{2M}\Bigg[n\langle{\rm tr}\{{\rm tr}_1\{\rho\}^{2M}\}\rangle \\ &\quad + \left\langle\left(\sum_{j=1}^{n}{\rm tr}\{{\rm tr}_j\{\rho\}^M\}\right)^2\right\rangle\Bigg].\end{aligned} \tag{B6}$$

A simplified expression is available for $M = 1$, namely, beginning with

$$\mathcal{D}^{\text{i.i.d. loss}}_{M=1,L_{\rm err}=1} = \epsilon^2\{n\langle{\rm tr}\{{\rm tr}_1\{\rho\}^2\}\rangle + n^2\}, \tag{B7}$$

we realize that

$${\rm tr}\{{\rm tr}_1\{\rho\}^2\} = \sum_{m_1,m_1'=0}^{1}\sum_{m_2=0}^{2^{n-1}-1}{\rm tr}\left\{\rho F_{m_1,m_1'}\rho G_{m_1,m_1',m_2}\right\}, \tag{B8}$$

with

$$\begin{aligned}F_{m_1,m_1'} &= |m_1\rangle\langle m_1'| \otimes 1_{n-1}, \\ G_{m_1,m_1',m_2} &= |m_1'\rangle|m_2\rangle\langle m_1|\langle m_2|.\end{aligned} \tag{B9}$$

Therefore, in order to average

$$\mathrm{tr}\{\rho F_{m_1,m_1'}\rho G_{m_1,m_1',m_2}\} = \mathrm{tr}\{U|0\rangle\langle 0|U^\dagger F_{m_1,m_1'}U|0\rangle\langle 0|U^\dagger G_{m_1,m_1',m_2}\}, \quad \text{(B10)}$$

we make use of Eq. (23) for any 2-design unitary U and find that

$$\langle \mathrm{tr}\{\rho F_{m_1,m_1'}\rho G_{m_1,m_1',m_2}\}\rangle = \frac{1}{d(d+1)}\left(\frac{d}{2}\delta_{m_1,m_1'}+1\right), \quad \text{(B11)}$$

so that

$$\langle \mathrm{tr}\{\mathrm{tr}_1\{\rho\}^2\}\rangle = \frac{d+4}{2(d+1)} \quad \text{(B12)}$$

and

$$\mathcal{D}^{\text{i.i.d. loss}}_{M=1,L_{\text{err}}=1} = \epsilon^2\left[\frac{n(d+4)}{2(d+1)}+n^2\right] \quad \text{(B13)}$$

for $M=1$.

For any $L_{\text{err}} \geqslant 1$ the case is governed by the action

$$\rho'_{L_{\text{err}}} = \Phi^{\text{multiqubit loss}}_{\epsilon/L_{\text{err}}}[W_{L_{\text{err}}}\cdots \times \Phi^{\text{multiqubit loss}}_{\epsilon/L_{\text{err}}}[W_2\Phi^{\text{multiqubit loss}}_{\epsilon/L_{\text{err}}}[W_1|\mathbf{0}\rangle\langle\mathbf{0}|W_1^\dagger]W_2^\dagger] \cdots W_{L_{\text{err}}}^\dagger]. \quad \text{(B14)}$$

By recalling that the unitary operators W_l act on the Hilbert-space sector that is orthogonal to $|\text{vac}\rangle\langle\text{vac}|$, it is straightforward to see that

$$\begin{aligned}\rho'_{L_{\text{err}}} &\cong \left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{L_{\text{err}}n}\rho + \frac{\epsilon}{L_{\text{err}}}\left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{L_{\text{err}}n-1} \times (|\text{vac}\rangle\langle\text{vac}|\otimes\mathrm{tr}_1\{\rho\}+\cdots+\mathrm{tr}_n\{\rho\}\otimes|\text{vac}\rangle\langle\text{vac}|) \\ &\cong (1-n\epsilon)\rho + \frac{\epsilon}{L_{\text{err}}}(|\text{vac}\rangle\langle\text{vac}|\otimes\mathrm{tr}_1\{\rho\}+\cdots \\ &\quad + \mathrm{tr}_n\{\rho\}\otimes|\text{vac}\rangle\langle\text{vac}|). \end{aligned} \quad \text{(B15)}$$

Then, as all vacuum-related terms and the noiseless state ρ are mutually orthogonal to each other, raising ρ' to the Mth power amounts simply to

$$\rho'^M_{L_{\text{err}}} \cong (1-Mn\epsilon)\rho + \left(\frac{\epsilon}{L_{\text{err}}}\right)^M(|\text{vac}\rangle\langle\text{vac}|\otimes\mathrm{tr}_1\{\rho\}^M+\cdots + \mathrm{tr}_n\{\rho\}^M\otimes|\text{vac}\rangle\langle\text{vac}|), \quad \text{(B16)}$$

along with its resulting normalized state

$$\frac{\rho'^M_{L_{\text{err}}}}{\mathrm{tr}\{\rho'^M_{L_{\text{err}}}\}} \cong \rho + \left(\frac{\epsilon}{L_{\text{err}}}\right)^M\left(|\text{vac}\rangle\langle\text{vac}|\otimes\mathrm{tr}_1\{\rho\}^M+\cdots + \mathrm{tr}_n\{\rho\}^M\otimes|\text{vac}\rangle\langle\text{vac}| - \rho\sum_{j=1}^n \mathrm{tr}\{\mathrm{tr}_j\{\rho\}^M\}\right). \quad \text{(B17)}$$

Hence, one obtains the more general MSE expression for small ϵ,

$$\mathcal{D}^{\text{i.i.d. loss}}_{M,L_{\text{err}}} = \left(\frac{\epsilon}{L_{\text{err}}}\right)^{2M}\left[n\langle\mathrm{tr}\{\mathrm{tr}_1\{\rho\}^{2M}\}\rangle + \left\langle\left(\sum_{j=1}^n\mathrm{tr}\{\mathrm{tr}_j\{\rho\}^M\}\right)^2\right\rangle\right] \quad \text{(B18)}$$

TABLE I. Values of $\langle\mathrm{tr}\{\mathrm{tr}_1\{\rho\}^{2M}\}\rangle$ for various n and M. For $M>1$, Monte Carlo simulations with 10 000 samples are conducted for each n-M pair. Random circuit unitary operators are assumed to follow the Haar distribution.

$n \setminus M$	1	2	3	4	5
2	$\frac{4}{5}$	0.6283	0.5191	0.4490	0.3885
3	$\frac{2}{3}$	0.3934	0.2605	0.1808	0.1311
4	$\frac{10}{17}$	0.2648	0.1356	0.0754	0.0431
5	$\frac{6}{11}$	0.1940	0.0794	0.0352	0.0167
6	$\frac{34}{65}$	0.1604	0.0544	0.0201	0.0075

or

$$\mathcal{D}^{\text{i.i.d. loss}}_{M=1,L_{\text{err}}} = \left(\frac{\epsilon}{L_{\text{err}}}\right)^2\left[\frac{n(d+4)}{2(d+1)}+n^2\right]. \quad \text{(B19)}$$

This gives us the ratio $\mathcal{D}^{\text{i.i.d. loss}}_{M,L_{\text{err}}}/\mathcal{D}^{\text{i.i.d. loss}}_{M,L_{\text{err}}=1} = O(L_{\text{err}}^{-2M})$ for any M. For completeness, we tabulate values of $\langle\mathrm{tr}\{\mathrm{tr}_1\{\rho\}^{2M}\}\rangle$ and $(\{\sum_{j=1}^n\mathrm{tr}\{\mathrm{tr}_j\{\rho\}^M\}\}^2)$ in Tables I and II, which are used to generate Figs. 2–4. To this end, all random unitary operators are assumed to follow the Haar distribution for $M>1$, which is arguably the most common 2-design distribution.

APPENDIX C: VIRTUAL DISTILLATION UNDER THE INDEPENDENT AND IDENTICALLY DISTRIBUTED PAULI CHANNEL

The general MSE expression for arbitrary $\epsilon = \sum_{l=1}^3\epsilon_l$ and L_{err} may be obtained from the map action in (14) by sketching a flowchart of how ρ' evolves as noise dilution proceeds for L_{err} layers, each with a diluted error rate of ϵ/L_{err} (see Fig. 12). We therefore find the following general noisy state $\rho' \equiv \rho'_{L_{\text{err}}}$, with respect to the target $\rho \equiv \rho_{L_{\text{err}}}$, expanded up to first order in all the error rates ϵ_1, ϵ_2, and ϵ_3:

$$\begin{aligned}\rho' &\cong (1-n\epsilon)\rho + \sum_{l=1}^3\frac{\epsilon_l}{L_{\text{err}}}\sum_{j=1}^n T_j^{(l)}, \\ T_j^{(l)} &= P_j^{(l)}\rho P_j^{(l)} + W_{L_{\text{err}}}P_j^{(l)}\rho_{L_{\text{err}}-1}P_j^{(l)}W_{L_{\text{err}}}^\dagger + \cdots \\ &\quad + W_{L_{\text{err}}}W_{L_{\text{err}}-1}\cdots W_2P_j^{(l)}\rho_1P_j^{(l)}W_2^\dagger\cdots W_{L_{\text{err}}-1}^\dagger W_{L_{\text{err}}}^\dagger. \end{aligned} \quad \text{(C1)}$$

TABLE II. Values of $\langle(\sum_{j=1}^n\mathrm{tr}\{\mathrm{tr}_j\{\rho\}^M\})^2\rangle$ for various n and M. For $M>1$, Monte Carlo simulations with 10 000 samples are conducted for each n-M pair. Random circuit unitary operators are assumed to follow the Haar distribution.

$n \setminus M$	1	2	3	4	5
2	4	2.6278	2.0963	1.7994	1.5547
3	9	4.0678	2.3808	1.5438	1.0734
4	16	5.5635	2.3935	1.1653	0.6196
5	25	7.4374	2.5432	0.9705	0.3969
6	36	9.8520	2.9229	0.9272	0.3104

$$\rho_0 = |\mathbf{0}\rangle\langle\mathbf{0}|$$

$$\downarrow W_1,\ \Phi^{\text{i.i.d. Pauli}}_{\epsilon/L_{\text{err}}}$$

$$\rho_1' \cong \left(1-\frac{\epsilon}{L_{\text{err}}}\right)^n \rho_1 + \sum_{l=1}^{3}\frac{\epsilon_l}{L_{\text{err}}}\sum_{j=1}^{n} P_j^{(l)}\rho_1 P_j^{(l)}$$

$$\downarrow W_2,\ \Phi^{\text{i.i.d. Pauli}}_{\epsilon/L_{\text{err}}}$$

$$\rho_2' \cong \left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{2n} \rho_2 + \sum_{l=1}^{3}\frac{\epsilon_l}{L_{\text{err}}}\sum_{j=1}^{n}\left[P_j^{(l)}\rho_2 P_j^{(l)} + W_2 P_j^{(l)}\rho_1 P_j^{(l)} W_2^\dagger\right]$$

$$\downarrow W_3,\ \Phi^{\text{i.i.d. Pauli}}_{\epsilon/L_{\text{err}}}$$

$$\rho_3' \cong \left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{3n} \rho_3 + \sum_{l=1}^{3}\frac{\epsilon_l}{L_{\text{err}}}\sum_{j=1}^{n}\left[P_j^{(l)}\rho_3 P_j^{(l)} + W_3 P_j^{(l)}\rho_2 P_j^{(l)} W_3^\dagger + W_3 W_2 P_j^{(l)}\rho_1 P_j^{(l)} W_2^\dagger W_3^\dagger\right]$$

$$\vdots$$

$$\downarrow W_{L_{\text{err}}},\ \Phi^{\text{i.i.d. Pauli}}_{\epsilon/L_{\text{err}}}$$

$$\rho_{L_{\text{err}}}' \cong \left(1-\frac{\epsilon}{L_{\text{err}}}\right)^{L_{\text{err}}n} \rho_{L_{\text{err}}} + \sum_{l=1}^{3}\frac{\epsilon_l}{L_{\text{err}}}\sum_{j=1}^{n}\Big[P_j^{(l)}\rho_{L_{\text{err}}} P_j^{(l)} + W_{L_{\text{err}}} P_j^{(l)}\rho_{L_{\text{err}}-1} P_j^{(l)} W_{L_{\text{err}}}^\dagger + \dots + W_{L_{\text{err}}} W_{L_{\text{err}}-1}\dots W_2 P_j^{(l)}\rho_1 P_j^{(l)} W_2^\dagger \dots W_{L_{\text{err}}-1}^\dagger W_{L_{\text{err}}}^\dagger\Big]$$

FIG. 12. Evolution flowchart of ρ' in an L_{err}-layered dilution setting. Here $\rho_l \equiv W_l\cdots W_2 W_1|0\rangle\langle 0|W_1^\dagger W_2^\dagger\cdots W_l^\dagger$ and $P_j^{(l)}$ is one of the single-qubit Pauli operators.

After raising $\rho'_{L_{\text{err}}}$ to the Mth power, upon a further trace normalization, we get

$$\frac{\rho'^M}{\text{tr}\{\rho'^M\}} \cong \rho + \sum_{l=1}^{3}\frac{\epsilon_l}{L_{\text{err}}}\sum_{j=1}^{n}\big[\rho T_j^{(l)} + (M-2)\rho T_j^{(l)}\rho + T_j^{(l)}\rho - M\rho\,\text{tr}\{\rho T_j^{(l)}\}\big]. \tag{C2}$$

Hence, unlike the loss channel, where the error term goes as ϵ^M, the Pauli channel gives rise to a persistent error term that *does not* go away by simply increasing M to infinity. This implies the MSE expressions

$$\begin{aligned}\mathcal{D}^{\text{i.i.d. Pauli}}_{M=1,L_{\text{err}}} &= (n\epsilon)^2 - 2n\epsilon\sum_{l=1}^{3}\frac{\epsilon_l}{L_{\text{err}}}\sum_{j=1}^{n}\langle\text{tr}\{\rho T_j^{(l)}\}\rangle \\ &\quad + \sum_{l,l'=1}^{3}\frac{\epsilon_l\epsilon_{l'}}{L_{\text{err}}^2}\sum_{j,j'=1}^{n}\langle\text{tr}\{T_j^{(l)}T_{j'}^{(l')}\}\rangle, \\ \mathcal{D}^{\text{i.i.d. Pauli}}_{M\geqslant 2,L_{\text{err}}} &= 2\sum_{l,l'=1}^{3}\frac{\epsilon_l\epsilon_{l'}}{L_{\text{err}}^2}\sum_{j,j'=1}^{n}\big[\langle\text{tr}\{\rho T_j^{(l)}T_{j'}^{(l')}\}\rangle \\ &\quad - \langle\text{tr}\{\rho T_j^{(l)}\}\text{tr}\{\rho T_{j'}^{(l')}\}\rangle\big].\end{aligned} \tag{C3}$$

In other words, there is a difference in MSE in raising the virtual distillation order from $M=1$ to $M=2$. However, up to first order in error rates, carrying out virtual distillation with orders beyond $M=2$ offers no further reduction in MSE. This is a manifestation of the noncommutativity between $T_j^{(l)}$ and ρ.

The remaining task is the evaluation of circuit averages. First, recall that $T_j^{(l)}$ is a sum of L_{err} pure states. The average

$$\langle\text{tr}\{\rho T_j^{(l)}\}\rangle = \langle\text{tr}\{\rho P_j^{(l)}\rho P_j^{(l)}\}\rangle + \cdots \tag{C4}$$

consists of the term

$$\langle\text{tr}\{\rho P_j^{(l)}\rho P_j^{(l)}\}\rangle = \langle\text{tr}\{W_{L_{\text{err}}}\rho_{L_{\text{err}}-1}W_{L_{\text{err}}}^\dagger P_j^{(l)} W_{L_{\text{err}}} \times \rho_{L_{\text{err}}-1}W_{L_{\text{err}}}^\dagger P_j^{(l)}\}\rangle = \frac{1}{d+1}, \tag{C5}$$

where we have put (23) to good use, and all other $L_{\text{err}}-1$ terms contribute precisely the same result, so

$$\langle\text{tr}\{\rho T_j^{(l)}\}\rangle = \frac{L_{\text{err}}}{d+1}. \tag{C6}$$

The next average $\langle\text{tr}\{T_j^{(l)}T_{j'}^{(l')}\}\rangle$ would depend on the indices j, j', l, and l'. Generally speaking, this term consists of two types of averages, the self-terms such as $\langle\text{tr}\{P_j^{(l)}\rho P_j^{(l)}P_{j'}^{(l')}\rho P_{j'}^{(l')}\}\rangle$, and cross terms like $\langle\text{tr}\{P_j^{(l)}\rho P_j^{(l)}W_{L_{\text{err}}}P_{j'}^{(l')}\rho_{L_{\text{err}}-1}P_{j'}^{(l')}W_{L_{\text{err}}}^\dagger\}\rangle$. Since every term of the same type gives the same result, we simply calculate one of each. Starting with the latter, for any indices, the combined use of (22) and (23) leads to

$$\begin{aligned}&\langle\text{tr}\{P_j^{(l)}\rho P_j^{(l)}W_{L_{\text{err}}}P_{j'}^{(l')}\rho_{L_{\text{err}}-1}P_{j'}^{(l')}W_{L_{\text{err}}}^\dagger\}\rangle \\ &\quad = \frac{d}{d^2-1} - \frac{1}{d^2-1}\langle\text{tr}\{\rho_{L_{\text{err}}-1}P_{j'}^{(l')}\rho_{L_{\text{err}}-1}P_{j'}^{(l')}\}\rangle \\ &\quad = \frac{d^2+d-1}{(d+1)(d^2-1)}.\end{aligned} \tag{C7}$$

TABLE III. Values of $\langle \mathrm{tr}\{\rho T_j^{(l)} T_{j'}^{(l')}\}\rangle$ for various n and L_{err}. Monte Carlo simulations with 10 000 samples are conducted for each n-L_{err} pair. Random circuit unitary operators are assumed to follow the Haar distribution.

$n \setminus L_{\mathrm{err}}$	1	2	3	4
		$l \neq l', j = j'$		
2	0.0000	0.0791	0.2422	0.4802
3	0.0000	0.0246	0.0739	0.1486
4	0.0000	0.0071	0.0204	0.0415
5	0.0000	0.0019	0.0054	0.0110
6	0.0000	0.0005	0.0014	0.0028
		Any l and l', $j \neq j'$		
2	0.0696	0.2090	0.4363	0.7526
3	0.0226	0.0703	0.1417	0.2336
4	0.0061	0.0201	0.0397	0.0679
5	0.0017	0.0053	0.0109	0.0175
6	0.0005	0.0014	0.0027	0.0044

For the former,

$$\left\langle \mathrm{tr}\left\{P_j^{(l)}\rho P_j^{(l)} P_{j'}^{(l')}\rho P_{j'}^{(l')}\right\}\right\rangle = \begin{cases} 1 & \text{for } j = j',\ l = l' \\ \dfrac{1}{d+1} & \text{otherwise.} \end{cases}$$

So

$$\left\langle \mathrm{tr}\left\{T_j^{(l)} T_{j'}^{(l')}\right\}\right\rangle = L_{\mathrm{err}} \frac{d\delta_{j,j'}\delta_{l,l'} + 1}{d+1} + L_{\mathrm{err}}(L_{\mathrm{err}} - 1)\frac{d^2 + d - 1}{(d+1)(d^2-1)}. \quad \text{(C8)}$$

Equations (C6) and (C8) therefore supply the exact answer

$$\mathcal{D}^{\text{i.i.d. Pauli}}_{M=1,L_{\mathrm{err}}} = (n\epsilon)^2\left[\frac{d^3}{(d+1)(d^2-1)} - \frac{d}{L_{\mathrm{err}}(d+1)(d^2-1)}\right] + \frac{nd}{L_{\mathrm{err}}(d+1)}\sum_{l=1}^{3}\epsilon_l^2 \quad \text{(C9)}$$

for any 2-design unitary operators W_l.

When $M \geqslant 2$, only certain averages are calculable solely from the 2-design properties. For instance, the term $\langle \mathrm{tr}\{\rho T_j^{(l)} T_{j'}^{(l')}\}\rangle$ may again be found by considering different index conditions. If $j = j'$ and $l = l'$, then the same stepsas before produce

$$\left\langle \mathrm{tr}\left\{\rho T_j^{(l)2}\right\}\right\rangle = \frac{L_{\mathrm{err}}}{d+1} + \frac{L_{\mathrm{err}}(L_{\mathrm{err}} - 1)}{(d+1)^2}. \quad \text{(C10)}$$

Otherwise, if $l \neq l'$, there exists an interesting analytical observation for $\langle \mathrm{tr}\{\rho T_j^{(l)} T_j^{(l')}\}\rangle$, that is, when $j = j'$, one finds

TABLE IV. Values of $\langle \mathrm{tr}\{\rho T_j^{(l)}\}\mathrm{tr}\{\rho T_{j'}^{(l')}\}\rangle$ for various n and L_{err}. Monte Carlo simulations with 10 000 samples are conducted for each n-L_{err} pair. Random circuit unitary operators are assumed to follow the Haar distribution.

$n \setminus L_{\mathrm{err}}$	1	2	3	4
		$l = l', j = j'$		
2	0.0867	0.2560	0.5036	0.8145
3	0.0298	0.0862	0.1649	0.2715
4	0.0092	0.0268	0.0489	0.0809
5	0.0026	0.0070	0.0134	0.0217
6	0.0007	0.0018	0.0034	0.0056
		$l \neq l', j = j'$		
2	0.0295	0.1371	0.3296	0.5993
3	0.0101	0.0445	0.1027	0.1868
4	0.0030	0.0131	0.0301	0.0541
5	0.0009	0.0036	0.0080	0.0144
6	0.0002	0.0009	0.0021	0.0037
		Any l and l', $j \neq j'$		
2	0.0480	0.1744	0.3848	0.6821
3	0.0142	0.0518	0.1161	0.2021
4	0.0038	0.0144	0.0316	0.0559
5	0.0010	0.0038	0.0084	0.0150
6	0.0002	0.0010	0.0021	0.0038

that each of the L_{err} self-terms, say,

$$\begin{aligned}&\left\langle \mathrm{tr}\left\{\rho P_j^{(l)}\rho P_j^{(l)} P_j^{(l')}\rho P_j^{(l')}\right\}\right\rangle \\ &\quad = \left\langle\left\langle |P_j^{(l)}|\right\rangle\left\langle |P_j^{(l')}|\right\rangle\left\langle |P_j^{(l)} P_j^{(l')}|\right\rangle\right\rangle \\ &\quad = -\left\langle\left\langle |P_j^{(l)}|\right\rangle\left\langle |P_j^{(l')}|\right\rangle\left\langle |P_j^{(l')} P_j^{(l)}|\right\rangle\right\rangle, \end{aligned} \quad \text{(C11)}$$

is purely imaginary owing to the anticommutativity of the single-qubit Pauli operators. Apart from this, both $\langle \mathrm{tr}\{\rho T_j^{(l)} T_{j'}^{(l')}\}\rangle$ and $\langle \mathrm{tr}\{\rho T_j^{(l)}\}\mathrm{tr}\{\rho T_{j'}^{(l')}\}\rangle$ appearing in $\mathcal{D}^{\text{i.i.d. Pauli}}_{M\geqslant 2,L_{\mathrm{err}}}$ involve third moments that depend on the 2-design distribution.

By assuming the Haar distribution for all circuit unitary operators, we present some lists of values for these two averages in order to compare the small-ϵ MSE analytical formulas with the simulation result for $M \geqslant 2$ in Tables III and IV. These are used to produce Figs. 5–7.

TABLE V. Eigenvalues of ρ_{err} for the independent and identically distributed loss channel acting on a product state and their multiplicities. Note that the complete dimension of ρ_{err} is 3^n when the vacuum sector is accounted for.

Eigenvalues of $\rho_{L_{\mathrm{err}}}$	Multiplicity
0	$3^n - 2^n + 1$
$\frac{\epsilon}{L_{\mathrm{err}}}(1 - \frac{\epsilon}{L_{\mathrm{err}}})^{n-1}$	$\binom{n}{1}$
$(\frac{\epsilon}{L_{\mathrm{err}}})^2(1 - \frac{\epsilon}{L_{\mathrm{err}}})^{n-2}$	$\binom{n}{2}$
$\vdots$	$\vdots$
$(\frac{\epsilon}{L_{\mathrm{err}}})^{n-1}(1 - \frac{\epsilon}{L_{\mathrm{err}}})$	$\binom{n}{n-1}$
$(\frac{\epsilon}{L_{\mathrm{err}}})^n$	$\binom{n}{n}$

APPENDIX D: HELLINGER DISTANCE FOR INDEPENDENT AND IDENTICALLY DISTRIBUTED LOSSES ON A PRODUCT STATE

If $|\ \rangle = |\psi_1\rangle|\psi_2\rangle\cdots|\psi_n\rangle$ is an n-qubit product ket, then recalling once again that $|\mathrm{vac}\rangle\langle\mathrm{vac}|$ is invisible to *all* circuit unitary operators, the exact lossy state

$$\begin{aligned}\rho'_{L_{\mathrm{err}}} =& |\ \rangle\left(1-\frac{\epsilon}{L_{\mathrm{err}}}\right)^{L_{\mathrm{err}}n}\langle\ | + \frac{\epsilon}{L_{\mathrm{err}}}\left(1-\frac{\epsilon}{L_{\mathrm{err}}}\right)^{L_{\mathrm{err}}n-1}\underbrace{(\Psi_1+\cdots+\Psi_n)}_{\binom{n}{1}\text{ terms}}\\ &+\left(\frac{\epsilon}{L_{\mathrm{err}}}\right)^2\left(1-\frac{\epsilon}{L_{\mathrm{err}}}\right)^{L_{\mathrm{err}}n-2}\underbrace{(\Psi_{1,2}+\Psi_{1,3}+\cdots+\Psi_{n-1,n-2}+\Psi_{n-1,n})}_{\binom{n}{2}\text{ terms}}+\cdots\\ &\left(\frac{\epsilon}{L_{\mathrm{err}}}\right)^{n-1}\left(1-\frac{\epsilon}{L_{\mathrm{err}}}\right)^{L_{\mathrm{err}}n-n+1}\underbrace{(\Psi_{1,2,\ldots,n-1}+\Psi_{1,2,\ldots,n-2,n}+\cdots+\Psi_{2,3,\ldots,n})}_{\binom{n}{n-1}\text{ terms}}\\ &+|\mathrm{vac}\rangle\left(\frac{\epsilon}{L_{\mathrm{err}}}\right)^{n}\left(1-\frac{\epsilon}{L_{\mathrm{err}}}\right)^{L_{\mathrm{err}}n-n}\langle\mathrm{vac}|\\ =& |\ \rangle\left(1-\frac{\epsilon}{L_{\mathrm{err}}}\right)^{L_{\mathrm{err}}n}\langle\ | + \left(1-\frac{\epsilon}{L_{\mathrm{err}}}\right)^{(L_{\mathrm{err}}-1)n}\left[1-\left(1-\frac{\epsilon}{L_{\mathrm{err}}}\right)^{n}\right]\rho_{\mathrm{err}}\end{aligned} \tag{D1}$$

is a linear combination of orthonormal vacuum-substituted projectors $\Psi_\cdot$, where $\Psi_{1,2} = |\mathrm{vac}\rangle\langle\mathrm{vac}|\otimes|\mathrm{vac}\rangle\langle\mathrm{vac}|\otimes|\psi_3\rangle\langle\psi_3|\otimes\cdots\otimes|\psi_n\rangle\langle\psi_n|$, for instance.

It is clear that $\mathrm{tr}\{\rho'_{L_{\mathrm{err}}}\} = (1-\epsilon/L_{\mathrm{err}})^{(L_{\mathrm{err}}-1)n} < 1$ whenever $L_{\mathrm{err}} > 1$, since discarding noncoincidental data means that every action by a subcircuit unitary operator W_l loses information about the error component at every dilution layer. From Eq. (D1), the trace-normalized error component has degenerate eigenvalues according to the multiplicities listed in Table V. From the definition of the Hellinger distance in Eq. (34), these eigenvalues and multiplicities lead to (35).

[1] D. E. Deutsch, A. Barenco, and A. Ekert, Universality in quantum computation, Proc. R. Soc. A **449**, 669 (1995).

[2] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, Elementary gates for quantum computation, Phys. Rev. A **52**, 3457 (1995).

[3] B.-G. Englert, C. Kurtsiefer, and H. Weinfurter, Universal unitary gate for single-photon two-qubit states, Phys. Rev. A **63**, 032303 (2001).

[4] S. D. Bartlett, B. C. Sanders, S. L. Braunstein, and K. Nemoto, Efficient Classical Simulation of Continuous Variable Quantum Information Processes, Phys. Rev. Lett. **88**, 097904 (2002).

[5] A. Sawicki, L. Mattioli, and Z. Zimborás, Universality verification for a set of quantum gates, Phys. Rev. A **105**, 052602 (2022).

[6] I. Chuang and M. Nielsen, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[7] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, Quantum computers, Nature (London) **464**, 45 (2010).

[8] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature (London) **549**, 172 (2017).

[9] B. Lekitsch, S. Weidt, A. G. Fowler, K. Mølmer, S. J. Devitt, C. Wunderlich, and W. K. Hensinger, Blueprint for a microwave trapped ion quantum computer, Sci. Adv. **3**, e1601540 (2017).

[10] L. K. Grover, *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* (ACM Press, New York, 1996), pp. 212–219.

[11] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput. **26**, 1484 (1997).

[12] R. Raussendorf and H. J. Briegel, A One-Way Quantum Computer, Phys. Rev. Lett. **86**, 5188 (2001).

[13] A. Kitaev, Fault-tolerant quantum computation by anyons, Ann. Phys. (NY) **303**, 2 (2003).

[14] R. Raussendorf, J. Harrington, and K. Goyal, Topological fault-tolerance in cluster state quantum computation, New J. Phys. **9**, 199 (2007).

[15] A. Sehrawat, L. H. Nguyen, and B.-G. Englert, Test-state approach to the quantum search problem, Phys. Rev. A **83**, 052311 (2011).

[16] A. Montanaro, Quantum algorithms: An overview, npj Quantum Inf. **2**, 15023 (2016).

[17] E. Knill, R. Laflamme, and W. H. Zurek, Resilient quantum computation, Science **279**, 342 (1998).

[18] D. Franklin and F. T. Chong, in *Nano, Quantum and Molecular Computing: Implications to High Level Design and Validation*, edited by S. K. Shukla and R. I. Bahar (Springer US, Boston, 2004), pp. 247–266.

[19] D. Aharonov and M. Ben-Or, Fault-tolerant quantum computation with constant error rate, SIAM J. Comput. **38**, 1207 (2008).

[20] E. Knill, Approximation by quantum circuits, arXiv:quant-ph/9508006.

[21] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[22] T. R. Bromley, J. M. Arrazola, S. Jahangiri, J. Izaac, N. Quesada, A. D. Gran, M. Schuld, J. Swinarton, Z. Zabaneh, and N. Killoran, Applications of near-term photonic quantum computers: Software and algorithms, Quantum Sci. Technol. **5**, 034010 (2020).

[23] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[24] A. Finnila, M. Gomez, C. Sebenik, C. Stenson, and J. Doll, Quantum annealing: A new method for minimizing multidimensional functions, Chem. Phys. Lett. **219**, 343 (1994).

[25] T. Kadowaki and H. Nishimori, Quantum annealing in the transverse ising model, Phys. Rev. E **58**, 5355 (1998).

[26] S. Aaronson and A. Arkhipov, *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing* (ACM Press, New York, 2011), pp. 333–342.

[27] S. Aaronson, A linear-optical proof that the permanent is #P-hard, Proc. R. Soc. A **467**, 3393 (2011).

[28] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Gaussian Boson Sampling, Phys. Rev. Lett. **119**, 170501 (2017).

[29] A. Trabesinger, Quantum simulation, Nat. Phys. **8**, 263 (2012).

[30] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, Rev. Mod. Phys. **86**, 153 (2014).

[31] J. Biamonte, Universal variational quantum computation, Phys. Rev. A **103**, L030401 (2021).

[32] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[33] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, N. P. D. Sawaya, S. Sim, L. Veis, and A. Aspuru-Guzik, Quantum chemistry in the age of quantum computing, Chem. Rev. **119**, 10856 (2019).

[34] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, J. Phys. Soc. Jpn. **90**, 032001 (2021).

[35] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[36] Y. S. Teo, Optimized gradient and Hessian estimators for scalable variational quantum algorithms, arXiv:2206.12643.

[37] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[38] D. Wecker, M. B. Hastings, and M. Troyer, Progress towards practical quantum variational algorithms, Phys. Rev. A **92**, 042303 (2015).

[39] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

[40] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[41] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices, Phys. Rev. X **10**, 021067 (2020).

[42] M. Schuld, I. Sinayskiy, and F. Petruccione, An introduction to quantum machine learning, Contemp. Phys. **56**, 172 (2015).

[43] M. Schuld and N. Killoran, Quantum Machine Learning in Feature Hilbert Spaces, Phys. Rev. Lett. **122**, 040504 (2019).

[44] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[45] P. Date and W. Smith, Quantum discriminator for binary classification, arXiv:2009.01235.

[46] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Data re-uploading for a universal quantum classifier, Quantum **4**, 226 (2020).

[47] T. Dutta, A. Pérez-Salinas, J. P. S. Cheng, J. I. Latorre, and M. Mukherjee, Single-qubit universal classifier implemented on an ion-trap quantum device, Phys. Rev. A **106**, 012411 (2022).

[48] T. Goto, Q. H. Tran, and K. Nakajima, Universal Approximation Property of Quantum Machine Learning Models in Quantum-Enhanced Feature Spaces, Phys. Rev. Lett. **127**, 090506 (2021).

[49] S. Shin, Y. S. Teo, and H. Jeong, Exponential data encoding for quantum supervised learning, Phys. Rev. A **107**, 012422 (2023).

[50] *Quantum State Estimation*, edited by M. G. A. Paris and J. Řeháček, Lecture Notes in Physics Vol. 649 (Springer, Berlin, 2004).

[51] Y. S. Teo, *Introduction to Quantum-State Estimation* (World Scientific, Singapore, 2015).

[52] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert, Quantum State Tomography via Compressed Sensing, Phys. Rev. Lett. **105**, 150401 (2010).

[53] D. Ahn, Y. S. Teo, H. Jeong, F. Bouchard, F. Hufnagel, E. Karimi, D. Koutný, J. Řeháček, Z. Hradil, G. Leuchs, and L. L. Sánchez-Soto, Adaptive Compressive Tomography with No *a priori* Information, Phys. Rev. Lett. **122**, 100404 (2019).

[54] Y. S. Teo and L. L. Sánchez-Soto, Modern compressive tomography for quantum information science, Int. J. Quantum Inf. **19**, 2140003 (2021).

[55] D. Qin, X. Xu, and Y. Li, An overview of quantum error mitigation formulas, Chin. Phys. B **31**, 090306 (2022).

[56] Y. Li and S. C. Benjamin, Efficient Variational Quantum Simulator Incorporating Active Error Minimization, Phys. Rev. X **7**, 021050 (2017).

[57] K. Temme, S. Bravyi, and J. M. Gambetta, Error Mitigation for Short-Depth Quantum Circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[58] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, *Proceedings of the 2020 IEEE International Conference on Quantum Computing and Engineering* (IEEE, Piscataway, 2020), pp. 306–316.

[59] D. Greenbaum, Introduction to quantum gate set tomography, arXiv:1509.02921.

[60] 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. **5**, eaaw5686 (2019).

[61] S. Zhang, Y. Lu, K. Zhang, W. Chen, Y. Li, J.-N. Zhang, and K. Kim, Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system, Nat. Commun. **11**, 587 (2020).

[62] H. Kwon and J. Bae, A hybrid quantum-classical approach to mitigating measurement errors in quantum algorithms, IEEE Trans. Comput. **70**, 1401 (2021).

[63] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states, Phys. Rev. A **95**, 042308 (2017).

[64] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, Decoding quantum errors with subspace expansions, Nat. Commun. **11**, 636 (2020).

[65] P. Suchsland, F. Tacchino, M. H. Fischer, T. Neupert, P. K. Barkoutsos, and I. Tavernelli, Algorithmic error mitigation scheme for current quantum processors, Quantum **5**, 492 (2021).

[66] K. Wang, Y.-A. Chen, and X. Wang, Mitigating quantum errors via truncated Neumann series, arXiv:2111.00691.

[67] B. Koczor, Exponential Error Suppression for Near-Term Quantum Devices, Phys. Rev. X **11**, 031057 (2021).

[68] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual Distillation for Quantum Error Mitigation, Phys. Rev. X **11**, 041036 (2021).

[69] K. Yamamoto, S. Endo, H. Hakoshima, Y. Matsuzaki, and Y. Tokunaga, Error-Mitigated Quantum Metrology via Virtual Purification, Phys. Rev. Lett. **129**, 250503 (2022).

[70] A. Seif, Z.-P. Cian, S. Zhou, S. Chen, and L. Jiang, Shadow distillation: Quantum error mitigation with classical shadows for near-term quantum processors, PRX Quantum **4**, 010303 (2023).

[71] D. Mogilevtsev and V. S. Shchesnovich, Single-photon generation by correlated loss in a three-core optical fiber, Opt. Lett. **35**, 3375 (2010).

[72] J. Carpenter, C. Xiong, M. J. Collins, J. Li, T. F. Krauss, B. J. Eggleton, A. S. Clark, and J. Schröder, Mode multiplexed single-photon and classical channels in a few-mode fiber, Opt. Express **21**, 28794 (2013).

[73] I. Shomroni, S. Rosenblum, Y. Lovsky, O. Bechler, G. Guendelman, and B. Dayan, All-optical routing of single photons by a one-atom switch controlled by a single photon, Science **345**, 903 (2014).

[74] D. Bonneau, G. J. Mendoza, J. L. O'Brien, and M. G. Thompson, Effect of loss on multiplexed single-photon sources, New J. Phys. **17**, 043057 (2015).

[75] G. J. Mendoza, R. Santagati, J. Munns, E. Hemsley, M. Piekarek, E. Martín-López, G. D. Marshall, D. Bonneau, M. G. Thompson, and J. L. O'Brien, Active temporal and spatial multiplexing of photons, Optica **3**, 127 (2016).

[76] D. E. Jones, B. T. Kirby, and M. Brodsky, *Proceedings of the 2018 Optical Fiber Communications Conference and Exposition (OFC), San Diego, 2018* (IEEE, Piscataway, 2018).

[77] S. Omkar, Y. S. Teo, and H. Jeong, Resource-Efficient Topological Fault-Tolerant Quantum Computation with Hybrid Entanglement of Light, Phys. Rev. Lett. **125**, 060501 (2020).

[78] S. Bartolucci, P. Birchall, D. Bonneau, H. Cable, M. Gimeno-Segovia, K. Kieling, N. Nickerson, T. Rudolph, and C. Sparrow, Switch networks for photonic fusion-based quantum computing, arXiv:2109.13760.

[79] S. Omkar, S.-H. Lee, Y. S. Teo, S.-W. Lee, and H. Jeong, All-photonic architecture for scalable quantum computing with Greenberger-Horne-Zeilinger states, PRX Quantum **3**, 030309 (2022).

[80] J.-H. Kim, J.-W. Chae, Y.-C. Jeong, and Y.-H. Kim, Quantum communication with time-bin entanglement over a wavelength-multiplexed fiber network, APL Photon. **7**, 016106 (2022).

[81] A. Dragan and K. Wódkiewicz, Depolarization channels with zero-bandwidth noises, Phys. Rev. A **71**, 012322 (2005).

[82] A. Bayat, V. Karimipour, and I. Marvian, Threshold distances for transmission of EPR pairs through Pauli channels, Phys. Lett. A **355**, 81 (2006).

[83] M. Karpiński, C. Radzewicz, and K. Banaszek, Fiber-optic realization of anisotropic depolarizing quantum channels, J. Opt. Soc. Am. B **25**, 668 (2008).

[84] G. C. Amaral and G. P. Temporão, Characterization of depolarizing channels using two-photon interference, Quantum Inf. Process. **18**, 342 (2019).

[85] A. W. Harrow and R. A. Low, Random quantum circuits are approximate 2-designs, Commun. Math. Phys. **291**, 257 (2009).

[86] R. V. Mises and H. Pollaczek-Geiringer, Praktische verfahren der gleichungsauflösung, Z. Angew. Math. Mech. **9**, 152 (1929).

[87] 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. **3**, 033098 (2021).

[88] J. Cotler, S. Choi, A. Lukin, H. Gharibyan, T. Grover, M. E. Tai, M. Rispoli, R. Schittko, P. M. Preiss, A. M. Kaufman, M. Greiner, H. Pichler, and P. Hayden, Quantum Virtual Cooling, Phys. Rev. X **9**, 031013 (2019).

[89] Z. Cai, Resource-efficient purification-based quantum error mitigation, arXiv:2107.07279.

[90] M. Huo and Y. Li, Dual-state purification for practical quantum error mitigation, Phys. Rev. A **105**, 022427 (2022).

[91] P. Czarnik, A. Arrasmith, L. Cincio, and P. J. Coles, Qubit-efficient exponential suppression of errors, arXiv:2102.06056.

[92] Y. Xiong, S. X. Ng, and L. Hanzo, Quantum error mitigation relying on permutation filtering, IEEE Trans. Commun. **70**, 1927 (2022).

[93] S. Aaronson, Shadow tomography of quantum states, SIAM J. Comput., **49**, STOC18-368-STOC18-394 (2020).

[94] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[95] M. Paini, A. Kalev, D. Padilha, and B. Ruck, Estimating expectation values using approximate quantum states, Quantum **5**, 413 (2021).

[96] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, Robust shadow estimation, PRX Quantum **2**, 030348 (2021).

[97] S. J. D. Phoenix, Wave-packet evolution in the damped oscillator, Phys. Rev. A **41**, 5132 (1990).

[98] A. Fujiwara and H. Imai, Quantum parameter estimation of a generalized Pauli channel, J. Phys. A: Math. Gen. **36**, 8093 (2003).

[99] A. Chiuri, V. Rosati, G. Vallone, S. Pádua, H. Imai, S. Giacomini, C. Macchiavello, and P. Mataloni, Experimental Realization of Optimal Noise Estimation for a General Pauli Channel, Phys. Rev. Lett. **107**, 253602 (2011).

[100] S. Omkar, R. Srikanth, and S. Banerjee, Dissipative and non-dissipative single-qubit channels: Dynamics and geometry, Quantum Inf. Process. **12**, 3725 (2013).

[101] S. T. Flammia and J. J. Wallman, Efficient estimation of Pauli channels, ACM Trans. Quantum Comput. **1**, 1 (2020).

[102] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[103] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).

[104] Y. R. Sanders, J. J. Wallman, and B. C. Sanders, Bounding quantum gate error rate based on reported average fidelity, New J. Phys. **18**, 012002 (2015).

[105] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia, Comparing Experiments to the Fault-Tolerance Threshold, Phys. Rev. Lett. **117**, 170502 (2016).

[106] E. Huang, A. C. Doherty, and S. Flammia, Performance of quantum error correction with coherent errors, Phys. Rev. A **99**, 022313 (2019).

[107] K. Siudzińska, Classical capacity of generalized pauli channels, J. Phys. A: Math. Theor. **53**, 445301 (2020).

[108] S. Chen, S. Zhou, A. Seif, and L. Jiang, Quantum advantages for Pauli channel estimation, Phys. Rev. A **105**, 032435 (2022).

[109] K.-J. Miescke and F. Liese, *Statistical Decision Theory: Estimation, Testing, and Selection* (Springer, New York, 2008).

[110] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Exact and approximate unitary 2-designs and their application to fidelity estimation, Phys. Rev. A **80**, 012304 (2009).

[111] B. Collins and P. Śniady, Integration with respect to the Haar measure on unitary, orthogonal and symplectic group, Commun. Math. Phys. **264**, 773 (2006).

[112] Z. Puchała and J. Miszczak, Symbolic integration with respect to the Haar measure on the unitary groups, Bull. Polish Acad. Sci. **65**, 21 (2017).

[113] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum **3**, 010313 (2022).

[114] F. Mezzadri, How to generate random matrices from the classical compact groups, Notices of the AMS **54**, 592 (2007).

[115] C. Cafaro and S. Mancini, Quantum stabilizer codes for correlated and asymmetric depolarizing errors, Phys. Rev. A **82**, 012306 (2010).

[116] J. M. Arrazola, V. Bergholm, K. Brádler, T. R. Bromley, M. J. Collins, I. Dhand, A. Fumagalli, T. Gerrits, A. Goussev, L. G. Helt, J. Hundal, T. Isacsson, R. B. Israel, J. Izaac, S. Jahangiri, R. Janik, N. Killoran, S. P. Kumar, J. Lavoie, A. E. Lita *et al.*, Quantum circuits with many photons on a programmable nanophotonic chip, Nature (London) **591**, 54 (2021).

[117] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, Violation of Bell's Inequality under Strict Einstein Locality Conditions, Phys. Rev. Lett. **81**, 5039 (1998).