

**Direct estimation of minimum gate fidelity**Yiping Lu,<sup>1,2</sup> Jun Yan Sim,<sup>2</sup> Jun Suzuki <sup>3</sup>, Berthold-Georg Englert <sup>2,4,5</sup>, and Hui Khoon Ng<sup>1,2,5,\*</sup><sup>1</sup>*Yale-NUS College, Singapore 138527, Singapore*<sup>2</sup>*Centre for Quantum Technologies, National University of Singapore, Singapore 117543, Singapore*<sup>3</sup>*Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan*<sup>4</sup>*Department of Physics, National University of Singapore, Singapore 117551, Singapore*<sup>5</sup>*MajuLab, CNRS-UCA-SU-NUS-NTU International Joint Research Unit, Singapore 117543, Singapore*

(Received 6 April 2020; accepted 22 July 2020; published 18 August 2020)

With the current interest in building quantum computers, there is a strong need for accurate and efficient characterization of the noise in quantum gate implementations. A key measure of the performance of a quantum gate is the minimum gate fidelity, i.e., the fidelity of the gate, minimized over all input states. Conventionally, the minimum fidelity is estimated by first accurately reconstructing the full gate process matrix using the experimental procedure of quantum process tomography (QPT). Then, a numerical minimization is carried out to find the minimum fidelity. QPT is, however, well known to be costly, and it might appear that we can do better if the goal is only to estimate one single number. In this work, we propose a hybrid numerical-experimental scheme that employs a numerical gradient-free minimization (GFM) and an experimental target-fidelity estimation procedure to directly estimate the minimum fidelity without reconstructing the process matrix. We compare this to an alternative scheme, referred to as QPT fidelity estimation, that does use QPT, but directly employs the minimum gate fidelity as the termination criterion. Both approaches can thus be considered as direct estimation schemes. General resource estimates suggest a significant resource savings for the GFM scheme over QPT fidelity estimation; numerical simulations for specific classes of noise, however, show that both schemes have similar performance, reminding us of the need for caution when using general bounds for specific examples. The GFM scheme, however, presents potential for future improvements in resource cost, with the development of even more efficient GFM algorithms.

DOI: [10.1103/PhysRevA.102.022410](https://doi.org/10.1103/PhysRevA.102.022410)**I. INTRODUCTION**

Quantum computing is much in the news these days, with the recent achievement of quantum supremacy by the Google device [1] and the renewed interest in building a quantum computer (see, for example, Refs. [2,3]). The main obstacle to realizing a quantum computer of a useful scale and accuracy is the noise that threatens to destroy the quantum features that give quantum computers their power. A key focus of any implementation of a quantum information processing device is hence the characterization, and subsequent control and mitigation, of the noise that unavoidably accompanies the operation of a quantum gate.

An often used measure of the quality of a quantum gate implementation is the fidelity of the gate, i.e., a quantification of how close the action of the actual gate is to the theoretical ideal. Efficient and easy-to-implement procedures such as randomized benchmarking [4–6] can offer information on the *average* fidelity of a gate, averaged over input states according to some distribution of interest (e.g., Haar-distributed pure states for randomized benchmarking). What is often more telling about the gate performance, however, is the *minimum* fidelity, i.e., the fidelity of the gate operation, minimized

over all possible input states. This is a state-distribution-independent quantity and gives a minimum guarantee for the quality of the gate, which is important to assure that the quantum computer functions correctly in all scenarios.

Conventionally, the minimum fidelity is estimated by first a full quantum process tomography (QPT) of the noisy gate operation, followed by a numerical minimization, using the estimated process matrix, to find the minimum fidelity over all possible input states. This is a potentially costly procedure: For a  $d$ -dimensional system, QPT requires an accurate estimation of  $d^4$  real parameters, all to yield one number, i.e., the minimum fidelity. Resource estimates of the number of uses of the channel required for an accurate full QPT can be deduced from bounds for state tomography [7,8], giving  $O(d^6/\epsilon'^2)$ , where  $\epsilon'$  is the accuracy of the reconstruction, measured by a distance between quantum processes.

One expects to be able to do better if the desire is only to estimate the minimum fidelity to some desired accuracy  $\epsilon$ , without the full reconstruction of the process matrix of the noisy gate. In this work, we propose a direct route to estimating the minimum fidelity, without the use of QPT. It combines a numerical gradient-free minimization (GFM) algorithm with the direct target fidelity estimation scheme of Refs. [9,10] to perform a hybrid numerical-experimental descent of the fidelity function to the minimum. We compare this GFM scheme with an alternative route, which we name

\*huikhooon.ng@yale-nus.edu.sg

*QPT fidelity estimation.* QPT fidelity estimation does go through QPT, but rather than employing a process distance as the figure of merit as done in standard QPT, we make direct use of the minimum fidelity as the stopping rule. In this sense, this second route can also be regarded as a direct estimation of the minimum fidelity and offers fair comparison with the GFM scheme.

Resource estimates suggest that QPT fidelity estimation needs  $O(d^8/\epsilon^2)$  uses of the gate to achieve an estimate of the minimum fidelity to accuracy  $\epsilon$ . This follows from an argument (see Sec. IV A) that relates the desired minimum fidelity accuracy of  $\epsilon$  to the Choi-state trace-distance accuracy of  $\epsilon'$  used in Refs. [7,8]. This  $O(d^8)$  uses of the gate, for fixed  $\epsilon$ , is a prohibitively high resource cost. The GFM approach, in contrast, is expected (see Sec. IV A) to require much milder  $O(d^4)$  gate uses. It is thus, from this perspective, a much more efficient procedure for estimating the minimum fidelity.

For specific classes of noisy gates, however, our numerical results suggest a somewhat different conclusion. We numerically simulated both GFM and QPT fidelity estimation procedures for two natural classes of noisy gates: noisy gates random in the Hilbert-Schmidt sense, and gates with random Pauli and amplitude-damping noise. For both classes, our GFM scheme performs close to (though better than) the resource estimates; the QPT fidelity estimation scheme, surprisingly, took only  $O(d^4)$  gate uses, much fewer than the  $O(d^8)$  prediction. The performance of the two schemes is hence comparable for these two classes of noisy gates, with only slightly better scaling for the GFM scheme. This better performance for QPT is especially unexpected as our simulations do not follow the optimal procedure behind the theoretical bounds obtained in Refs. [7,8]. This reminds us that for specific classes of noisy gates, the resource estimates, which account even for worst-case scenarios, may not provide a reliable gauge of typical performance.

Nevertheless, while there is little room for improvement in the resource scaling for QPT fidelity estimation, the scaling for the GFM scheme is limited largely by the efficiency of the numerical GFM algorithm, a subject of intense study in the field of numerical optimization. Our work also emphasizes the importance of making direct use of the quantity of interest, in this case the minimum fidelity, in the measurement procedure, rather than a secondary quantity such as a process distance. Both schemes explored here, the GFM scheme as well as the QPT fidelity estimation, can be considered when looking for methods of direct estimation of the minimum gate fidelity, with the GFM offering potential for further reduction in resource cost, while the QPT fidelity estimation offers the advantage of a measurement setup familiar from standard QPT.

Whether the observation of Ref. [11], namely, that very few measurements chosen at random may be sufficient for the estimation of particular properties, has a bearing on estimating the minimum gate fidelity is currently unknown, and this deserves to be explored. The matter is, however, not within the scope of this work.

Below, we begin in Sec. II with a description of our GFM algorithm, also providing a reminder of the direct target fidelity estimation scheme following the analysis of Ref. [9]. We then explain the QPT fidelity estimation procedure in

Sec. III. Section IV gives the resource estimates and numerical simulations that compare the performance of the two schemes. We conclude in Sec. V.

## II. DIRECT ESTIMATION WITH GRADIENT-FREE MINIMIZATION

For a known ideal gate  $\mathcal{G}$ , acting on an  $n$ -qubit system, and its noisy implementation  $\tilde{\mathcal{G}}$ , we are interested in how close  $\tilde{\mathcal{G}}$  is to  $\mathcal{G}$ . One way to quantify this is to compare the fidelity between the output state of  $\tilde{\mathcal{G}}$  with the ideal output state after  $\mathcal{G}$ . The minimum fidelity, minimized over all input states, is defined as

$$F_{\min} \equiv \min_{\psi} F(\mathcal{G}(\psi), \tilde{\mathcal{G}}(\psi)) = \min_{\psi} F(\psi, \mathcal{E}(\psi)). \quad (1)$$

Here,  $\psi \equiv |\psi\rangle\langle\psi|$  is a pure state on the  $d(=2^n)$ -dimensional Hilbert space of the quantum system,  $F(|\phi\rangle\langle\phi|, \rho) \equiv \langle\phi|\rho|\phi\rangle$  denotes the squared fidelity between the pure state  $|\phi\rangle$  and the (possibly mixed) state  $\rho$ , and  $\mathcal{E} \equiv \mathcal{G}^{-1} \circ \tilde{\mathcal{G}}$  is the noise process that describes the imperfections in the gate implementation  $\tilde{\mathcal{G}}$ .  $\mathcal{G}^{-1}$  is the inverse of the ideal unitary gate  $\mathcal{G}(\cdot) = U(\cdot)U^\dagger$ , for unitary  $U$ , i.e.,  $\mathcal{G}^{-1}(\cdot) = U^\dagger(\cdot)U$ . We assume that  $\tilde{\mathcal{G}}$  is a completely positive (CP) and trace-preserving (TP) map—also referred to as a quantum channel—and hence so is  $\mathcal{E}$ . Even though the minimization in Eq. (1) appears to only be over pure states  $\psi$ ,  $F_{\min}$  is, in fact, the minimum fidelity over all states, pure or mixed, as the concavity of  $F$  ensures that the minimum is attained on a pure state.

The goal here is to estimate  $F_{\min}$  without first estimating the full process matrix of the noisy gate  $\tilde{\mathcal{G}}$ , or, equivalently, of the noise channel  $\mathcal{E}$ . We assume the following experimental capabilities: (i) We can prepare any input state of our choice; (ii) we can send that input state through the noisy gate and access the output state; (iii) we can perform product-Pauli measurements on the output state. The noisy gate is regarded here as a black box  $\tilde{\mathcal{G}}$  in the laboratory that takes the input  $\psi$  and gives back the output  $\tilde{\mathcal{G}}(\psi)$ .

We need two additional ingredients. The first is the technique of direct estimation of target fidelity (DTFE) invented in Ref. [9], allowing the estimation of the fidelity of an  $n$ -qubit state  $\rho$  with some target pure state  $|\psi\rangle$  without full state tomography. One needs only make probabilistic product-Pauli measurements, according to a distribution determined by the target pure state (see Sec. II A for more details).

The second ingredient is a numerical method for gradient-free minimization (GFM), implementable on a classical computer. A GFM method finds a local minimum of a function in the case where function values, but not the gradient values, are easily available as inputs to the algorithm. A common situation is one where the function itself cannot be written down explicitly, but is accessible only through a numerical procedure. The gradient of the function is hence also not available as a function that can be written down explicitly, and methods of numerically approximating the gradient typically do not work well or are prohibitively expensive to evaluate. GFM methods incorporate the gradient estimation with the minimization by choosing trial points in the domain space, deducing some gradient information (often only a rough estimate) from the function values evaluated at those points,

taking a small step in a direction expected to lower the function value according to that gradient information, and then repeating the process with new trial points at the new location. With standard regularity criteria on the function, one eventually arrives at a local minimum and a repetition of the procedure sufficiently many times gives a good chance of finding the global minimum.

We now put the ingredients together for our GFM scheme of direct estimation of the minimum fidelity. Again, the goal is to estimate  $F_{\min}$  without full knowledge of  $\tilde{\mathcal{G}}$ , apart from access to it as an input-output black box in the laboratory. As is typical in such problems, the requirement is to estimate  $F_{\min}$  to within a target accuracy of  $\epsilon$  from the true value with high probability, i.e.,  $|\widehat{F}_{\min} - F_{\min}| \leq \epsilon$  for some small  $\epsilon > 0$ , where  $\widehat{F}_{\min}$  is the estimate while  $F_{\min}$  is the (unknown) true value. The function we are minimizing here is  $f(\psi) \equiv F(\psi, \mathcal{E}(\psi))$ , over the domain of pure states  $\psi$ . A full description of  $f$  and its gradient for minimization using gradient-descent methods requires knowledge of  $\mathcal{E}$ , which we do not possess. However, we can evaluate the function value in the laboratory by preparing  $\psi$ , feeding it into the black box  $\tilde{\mathcal{G}}$ , and then estimating  $f(\psi)$  by carrying out the DTFE scheme in the laboratory, treating  $\mathcal{G}(\psi)$  as the target (pure) state and  $\tilde{\mathcal{G}}(\psi)$  as the state to be compared with the target. In this way, we have access, from measurements in the laboratory on the black box  $\tilde{\mathcal{G}}$ , to the function values  $f(\psi)$ .

Our scheme then proceeds as follows: We run, on a classical computer, a GFM algorithm to minimize  $f(\psi)$ . At each iterative step, the GFM suggests a set of trial states  $\psi$ , for which it requires the values  $f(\psi)$ . These values are obtained from the experiment by preparing those  $\psi$  states and then performing the DTFE algorithm in the laboratory for each  $\psi$ . We iterate the GFM algorithm until a stopping rule is satisfied to confirm the attainment of a local minimum. The whole procedure is repeated sufficiently many times to find the global minimum with a specified (high) probability. The stopping rule is carefully tuned, as we will describe below, to attain the desired  $\epsilon$  accuracy. The scheme is summarized in Fig. 1.

In the following two sections, we describe, in greater detail, first the DTFE scheme of Ref. [9] and then the GFM algorithm employed in our numerical examples.

**A. Direct estimation of the target fidelity**

We describe here the basic ideas of the DTFE scheme of Ref. [9] needed to understand our work. The analysis and formulas here are taken from that reference, and we refer the reader to the original paper for further details. The target fidelity, i.e., the quantity of interest in Ref. [9], is the fidelity between a particular state  $\rho$ , imagined to be the actual state produced by a source, and a pure state  $|\psi\rangle$ , taken to be the target state the source is designed to produce in the ideal situation. The target fidelity hence quantifies how close the source is to that ideal. That the target state is pure is important technically for the linearity, in  $\rho$  and  $\psi$ , of the squared fidelity used as the logical basis of the scheme; a pure target state is anyway the often-encountered situation in many quantum information processing tasks. For our purpose here, it suffices as well, as our minimization is over pure input states only.

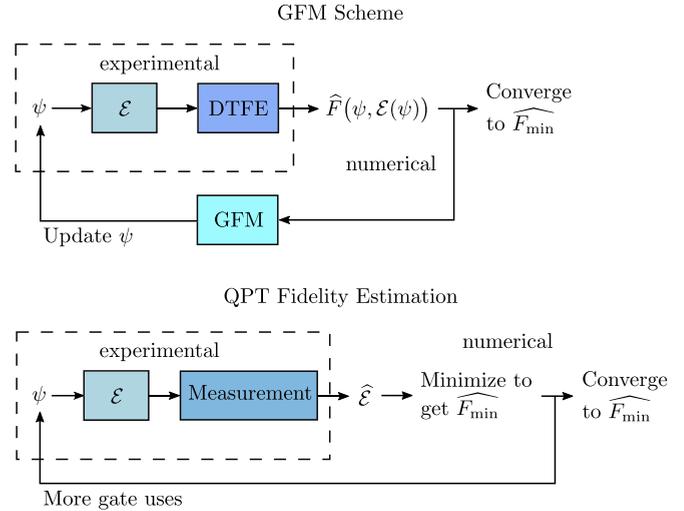


FIG. 1. Schematic depictions of the two approaches explored in our paper: GFM scheme and QPT fidelity estimation. Both approaches involve interwoven experimental and numerical steps. In the GFM approach, the experimental—via the direct target fidelity estimation procedure—and the numerical—via gradient-free descent—steps alternate seamlessly in a hybrid minimization algorithm, until convergence to the minimum fidelity  $\widehat{F}_{\min}$ . In QPT fidelity estimation, the experimental reconstruction of the process matrix occurs first, followed by a numerical minimization to obtain  $\widehat{F}_{\min}$ , and this is repeated, if needed, with more uses of the gate until the estimated accuracy of  $F_{\min}$  meets our set target.

We write the square of the target fidelity, for a  $d = 2^n$ -dimensional situation, as

$$F(\psi, \rho) = \text{tr}(\psi\rho) = \sum_k x_k(\psi)x_k(\rho), \tag{2}$$

where  $x_k(\cdot) \equiv \frac{1}{\sqrt{d}}\text{tr}\{W_k(\cdot)\}$ , with  $W_k$  an  $n$ -qubit product Pauli operator, so that the set  $\{\frac{1}{\sqrt{d}}W_k\}_{k=1}^{d^2}$  is an orthonormal operator basis.  $x_k(\tau)$  for any state  $\tau$  is then the (real) coefficient of  $\frac{1}{\sqrt{d}}W_k$  when writing  $\tau$  as a linear combination of elements of that Pauli operator basis. The sum over  $k$  in Eq. (2) above is understood to be only over those  $k$  values with  $x_k(\psi) \neq 0$ . In our present situation, the  $x_k(\psi)$ s are known, while the  $x_k(\rho)$ 's are not.

The fidelity can be rewritten as

$$F(\psi, \rho) = \sum_k p_k X_k, \tag{3}$$

with  $p_k \equiv [x_k(\psi)]^2$  and  $X \equiv \frac{x_k(\rho)}{x_k(\psi)}$ . Note that  $\{p_k\}$  is a probability distribution:  $\sum_k p_k = F(\psi, \psi) = 1$ , and  $p_k > 0$ . Let us define a random variable  $X$  which takes value  $X_k$  with probability  $p_k$ . Then, observe that  $F(\psi, \rho) = \langle X \rangle$ , the expectation value of  $X$ , which depends on the unknown  $\rho$ . Now, for each  $k$ ,  $x_k(\rho)$  can be estimated from the experiment by measuring  $W_k$ , a product-Pauli measurement, on  $\rho$ . Then, the target fidelity, now understood to be equal to  $\langle X \rangle$ , can be estimated by repeated trials where  $W_k$  is chosen as the measurement to be performed on  $\rho$  with probability  $p_k$ .

How good is this estimate? Suppose we do  $h \equiv 1/(\eta\delta^2)$  trials, with  $k_1, k_2, \dots, k_h$  chosen according

to the distribution  $\{p_k\}$ , and so obtain estimates for  $X_{k_1}, X_{k_2}, \dots, X_{k_h}$ . Then,  $Y \equiv \frac{1}{h} \sum_{\ell=1}^h X_{k_\ell}$  satisfies the inequality:  $\text{Probability}(|Y - \langle X \rangle| \geq \eta) \leq \delta$ . We cannot, however, know  $X_{k_\ell}$  precisely with only a finite number of copies of  $\rho$ . To estimate  $X_{k_\ell}$ , it suffices to measure  $W_{k_\ell}$  on  $t_{k_\ell}$  copies of  $\rho$ , where

$$t_{k_\ell} \equiv \frac{2 \ln(2/\delta)}{d p_{k_\ell} h \eta^2}. \quad (4)$$

Then,  $Y$ , now built from estimates of  $X_{k_\ell}$  using the above number of copies, satisfies instead the inequality

$$\text{Probability}(|Y - \langle X \rangle| \geq 2\eta) \leq 2\delta. \quad (5)$$

The expected total number of copies of  $\rho$  required to attain an estimate of the target fidelity  $2\eta$  away from the true value is then

$$h \sum_k p_k t_k = \sum_k \frac{2 \ln(2/\delta)}{d \eta^2} \leq 2 \ln(2/\delta) \frac{d}{\eta^2}, \quad (6)$$

giving a scaling of  $O(d/\eta^2)$  copies, for fixed  $\delta$ . A more accurate estimate of the expected number of copies is given in Ref. [9] by taking into account that  $t_{k_\ell}$  has to be an integer, but that does not change the  $O(d/\eta^2)$  scaling with  $d$ .

### B. Gradient-free minimization with CMA-ES

Our GFM direct estimation scheme accommodates the use of any GFM algorithm. The efficiency of the GFM algorithm is of crucial importance in the present context, but other considerations such as ease of coding and number of tuning parameters can also affect one's choice. There are many known GFM algorithms, including the downhill simplex method [12,13], the directional direct search [14–16], the stochastic method [17,18], and the covariance matrix adaptation evolution strategy [19–22] (CMA-ES), among others. The downhill simplex method is arguably the most well-studied one, but may not perform well in high-dimensional problems [23]. Trying out a few methods on our problem, we found CMA-ES to work well, converging more quickly than the downhill simplex method in our problem instances, with more stable performance than the directional direct search, and having fewer tuning parameters than the stochastic method. We thus focus on CMA-ES in our numerical examples below. We emphasize that a different user can choose a different GFM algorithm and observe a different efficiency performance; therein lies the potential for improvement beyond what we report here, with more efficient GFM methods.

Here, we provide some pertinent details of the CMA-ES algorithm, referring the reader to the original papers (Refs. [19–22]) for further explanation. We pay special attention to how we choose the stopping criteria that terminates the CMA-ES algorithm, as they are crucial for attaining the desired accuracy for  $F_{\min}$ .

To implement the CMA-ES algorithm, we parametrize the domain space, the  $d$ -dimensional Hilbert space of pure ( $n$ -qubit) states, as  $\psi = \ell \ell^\dagger / \text{tr}(\ell \ell^\dagger)$ , where  $\ell$  is the complex column vector,

$$\ell = (\ell_1 + i\ell_2, \ell_3 + i\ell_4, \dots, \ell_{2d-1} + i\ell_{2d})^T. \quad (7)$$

The domain space is  $2d$ -dimensional, but there is a single extra parameter—the trace of  $\psi$ , constrained to be 1—when regarded as a parametrization for  $\psi$ . This extra parameter does no harm to the minimization and, in fact, we observe this  $2d$ -parameter approach to work better for CMA-ES in our numerical examples than an alternative  $(2d - 1)$ -dimensional parametrization with spherical coordinates.

In the  $k$ th iterative step of the CMA-ES algorithm, a set of  $\lambda$  points in the domain space,  $L^{(k)} \equiv \{\ell_a^{(k)} \equiv (\ell_{a,1}^{(k)}, \ell_{a,2}^{(k)}, \dots, \ell_{a,2d}^{(k)})^T\}_{a=1}^\lambda$ , is drawn from the normal distribution,

$$\ell_a^{(k)} \sim m^{(k)} + \sigma^{(k)} \mathcal{N}(0, C^{(k)}) \quad \text{for } a = 1, \dots, \lambda, \quad (8)$$

where  $m^{(k)}$ ,  $\sigma^{(k)}$ , and  $C^{(k)}$  are the mean, step-size, and covariance matrix, respectively, for the  $k$ th step. At the initialization step,  $m^{(1)}$  is set as a Haar-random pure state,  $C^{(1)}$  is set equal to the identity matrix, and  $\sigma^{(1)}$  is set to be 0.3.  $m^{(1)}$  is set to be a Haar-random pure state to reflect our lack of knowledge of the state which attains the minimum value of fidelity. We observed empirically that the performance remains roughly the same even if we increase or decrease  $\sigma^{(1)}$  by an order of magnitude. As long as  $\sigma^{(1)}$  is not set to be too large or too small, the performance will not be affected too much. For the case with many local minima, choosing a larger value of  $\sigma^{(1)}$  might increase the chance of finding the global minimum. The fidelity value  $f_a^{(k)} \equiv f(\psi_a^{(k)})$ , for  $\psi_a^{(k)}$  built from  $\ell_a^{(k)}$ , is estimated using the DTFE scheme for each  $\ell_a^{(k)} \in L^{(k)}$ . The points are then ranked according to the  $f_a^{(k)}$  values, with  $\ell_{a:\lambda}^{(k)}$  having the  $a$ th smallest value of  $f$ , i.e.,  $f(\psi_{1:\lambda}^{(k)}) \leq f(\psi_{2:\lambda}^{(k)}) \leq \dots \leq f(\psi_{\lambda:\lambda}^{(k)})$ . The mean for the normal distribution is then updated by the weighted average,

$$m^{(k+1)} = \sum_{a=1}^{\mu} w_a \ell_{a:\lambda}^{(k)}, \quad (9)$$

where  $w_k > 0$  and  $\mu \leq \lambda$ . The step-size and covariance matrix are also updated according to rules based on the ranking of the  $f$  values. Further details of the algorithm, as well as the appropriate choice of parameters, can be found in Ref. [24].

In effect, the updates of  $m^{(k)}$ ,  $\sigma^{(k)}$ , and  $C^{(k)}$  move the “region of interest” within the domain space in the direction of smaller  $f$  values, in correspondence with our goal of minimizing  $f$ . In the next iterative step, we draw the domain points  $L^{(k+1)}$  from the updated region of interest and continue the move towards small  $f$  values. This continues until the stopping criteria (more on these below) are met. Our estimate for  $F_{\min}$ , for one run of CMA-ES, is then the fidelity for the best (smallest  $f$ ) point found in the final iterative step,  $\ell_{1:\lambda}^{(\dots)}$ . Each run of the CMA-ES algorithm returns an estimate of a local minimum value for  $f$ ; the entire procedure is repeated sufficiently many times, with different initial conditions, to have confidence that one of those local minima is the global minimum. We estimate the number of repeats needed, for a 95% chance of obtaining the global minimum, using random channels in our numerical simulation (see the numerical examples in Sec. IV B).

To determine when to stop the CMA-ES descent, we calculate two quantities in each iterative step. The first is the

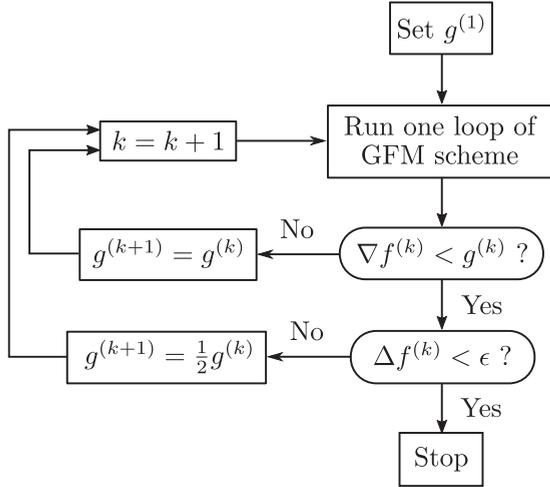


FIG. 2. Diagrammatic representation of the stopping criteria used in the GFM scheme. See main text for a description of the various quantities.

range of fidelities for the current set of sample points  $L^{(k)}$ ,

$$\Delta f^{(k)} \equiv \max_a \{f_a^{(k)}\} - \min_a \{f_a^{(k)}\}. \quad (10)$$

The second is

$$\nabla f^{(k)} \equiv \frac{1}{\lambda} \sum_{a=1}^{\lambda} \frac{|f_c^{(k)} - f_a^{(k)}|}{|\ell_c^{(k)} - \ell_a^{(k)}|}, \quad (11)$$

where  $\ell_c^{(k)}$  is the centroid of  $L^{(k)}$ , and  $f_c^{(k)}$  is the corresponding fidelity.  $\nabla f^{(k)}$  can be taken as an estimation of the magnitude of the gradient at  $\ell_c^{(k)}$ . The stopping criteria are then imposed adaptively to accommodate the variety of possible behaviors of the descent function  $f(\psi)$  for different noisy channels. We set an initial threshold  $g^{(1)}$  for the gradient estimate  $\nabla f^{(1)}$ . The target  $F_{\min}$  accuracy is  $\epsilon$ , as before. In subsequent iterative steps, if  $\nabla f^{(k)} > g^{(k)}$ , we retain the gradient threshold level, i.e., set  $g^{(k+1)} = g^{(k)}$ , and continue with the next iterative step. If, instead,  $\nabla f^{(k)} < g^{(k)}$  but  $\Delta f^{(k)} > \epsilon$ , we halve the gradient threshold, i.e., set  $g^{(k+1)} = g^{(k)}/2$ , before taking the next iterative step. If  $\nabla f^{(k)} < g^{(k)}$  and  $\Delta f^{(k)} < \epsilon$ , the algorithm is terminated. A schematic of the stopping criteria is given in Fig. 2.

The rationale for the adaptive stopping criteria is as follows. The algorithm should terminate when it gets to a stationary point, a minimum, where both the gradient and the range of fidelity values in the set of sample points are small. That the range is small is demanded by our set target accuracy; that the gradient is small enough is needed to ensure that we have arrived at the minimum value. If we know the  $f$  function, requiring the gradient alone to be small enough would also guarantee that the range is small, as the sample points are picked from a local region around the current position. However, since we do not have information about  $f(\psi)$ , we also do not know how small a gradient threshold is needed for sufficient convergence in the range. Instead, the gradient threshold has to be adjusted adaptively according to the noise channel at hand. We make use of the range of the fidelity values to judge whether the current gradient threshold

is sufficiently stringent to ensure that we have arrived at the minimum point. If the gradient is below the set threshold, but the range is still beyond  $\epsilon$ , this indicates that our gradient criterion is simply not stringent enough;  $g^{(k)}$  is thus halved and the iteration continues.

Note that the value of initial gradient threshold  $g^{(1)}$  has to be prechosen according to the expected channel distribution and the target  $\epsilon$  value, to prevent early termination of the algorithm. In our numerical examples, the value of  $g^{(1)}$  was chosen, for each distribution of  $n$ -qubit channels, so that the probability (over the channel distribution) that  $|\widehat{F}_{\min} - F_{\min}| \leq \epsilon$  is around 0.95, with  $\epsilon = 0.01$ , with the 0.95 judged from numerical simulations using a number of trial true channels.

There is one more important detail that has to be addressed: the accuracy of the DTFE procedure needed for the CMA-ES algorithm, as quantified by the  $\delta$  and  $\eta$  parameters of Sec. II A. The DTFE, with a given number of copies of the noisy state, estimates  $f(\psi)$  only to a certain accuracy. If the DTFE accuracy is set too high, the resource cost of our GFM direct estimation scheme will become very high; if it is set too low, the CMA-ES algorithm may not converge to the correct minimum value, if it converges at all. CMA-ES, like many evolutionary algorithms of a similar flavor, is relatively robust to noise in function evaluations [25], but proper handling of the accuracy of the function evaluation is still needed. In our numerical examples, we set  $\delta = 0.05$  in our DTFE subroutine and update the needed  $\eta$  value in the  $k$ th iteration of the CMA-ES according to the uncertainty handling algorithm [26,27] described next.

The CMA-ES update rule of Eq. (9) is based on the ranking of the  $f$  values. This means that noisy function evaluation will not affect it as long as the inaccuracies are not large enough to change the  $f$  ranking. One can judge whether the function evaluation is sufficiently accurate by checking for rank changes after reevaluations of the function. In our specific context, reevaluation of  $f(\psi)$  refers to repeating the DTFE procedure with the same parameters, to get a second estimate of the value of  $f(\psi)$  for a given target state  $\psi$ . The inherent randomness of the DTFE procedure will yield different values, with larger variations from fewer number of uses of the channels, and hence more inaccurate evaluation of  $f$ . We follow the uncertainty handling algorithm proposed in [26,27]. At each iterative step  $k$ , we reevaluate  $f(\psi_a)$  for each  $\psi_a$  built from  $\ell_a^{(k)} \in L^{(k)}$ , calculate the rank changes after the reevaluations, compute a measure of the uncertainty level, and adjust the accuracy of the DTFE procedure as needed. Specifically, we carry out the following steps:

*Uncertainty handling algorithm.*

(1) For each  $L^{(k)}$ , compute  $f$  for each  $\ell_a^{(k)} \in L^{(k)}$  twice. Denote the two values obtained as  $f_a^{(k)}$  and  $\tilde{f}_a^{(k)}$ .

(2) For each  $a = 1, \dots, \lambda$ , compute the rank change  $\Delta R_a^{(k)} = |R(f_a^{(k)}) - R(\tilde{f}_a^{(k)})| - 1$ , where  $R(f_a^{(k)})$  and  $R(\tilde{f}_a^{(k)})$  are the ranks of  $f_a^{(k)}$  and  $\tilde{f}_a^{(k)}$ , respectively, in the combined set  $\{f_a^{(k)}, \tilde{f}_a^{(k)}\}_{a=1}^{\lambda}$ .

(3) Compute the uncertainty level,

$$s^{(k)} \equiv \frac{1}{\lambda} \sum_{a=1}^{\lambda} \{2\Delta R_a^{(k)} - \pi_{\theta} [R(\tilde{f}_a^{(k)}) - H(\tilde{f}_a^{(k)} - f_a^{(k)})] - \pi_{\theta} [R(f_a^{(k)}) - H(f_a^{(k)} - \tilde{f}_a^{(k)})]\}, \quad (12)$$

where  $H(\cdot)$  is the step function and  $\pi_\theta(R)$  is the  $(50\theta)$ -th percentile of all possible rank changes (given by the set  $\{|1 - R|, |2 - R|, \dots, |2\lambda - 1 - R|\}$ ) for a given rank  $R$ .

(4) If  $s^{(k)} > 0$ , increase the accuracy in DTFE in the  $(k + 1)$ -th iteration by setting  $\eta^{(k+1)} = \alpha\eta^{(k)}$ , with  $0 < \alpha < 1$ . If  $s^{(k)} < 0$ , set  $\eta^{(k+1)} = \eta^{(k)}/\alpha$ .

(5) For  $a = 1, \dots, \lambda$ , set  $f_a^{(k)} = \frac{1}{2}(f_a^{(k)} + \tilde{f}_a^{(k)})$ .

In our numerical examples,  $\theta = 0.7$  and  $\alpha = \frac{1}{\sqrt{2}}$  were found to be good choices.

### III. QPT FIDELITY ESTIMATION

The performance of our GFM direct estimation scheme has to be compared with the standard alternative of quantum process tomography (QPT), whereby the full process matrix of the unknown channel  $\mathcal{E}$  is first estimated, and then the  $F_{\min}$  value is numerically computed from the obtained process matrix. For completeness, we remind the reader of a few aspects of QPT important for our work; of course, many textbooks and papers are available on the subject (see, for example, Ref. [28]).

QPT attempts to reconstruct the full description of an unknown quantum channel (or process)  $\mathcal{E}$  from a finite number of uses of the channel. A chosen set of states  $\{\rho_k\}$  is sent in as inputs to the channel, and state tomography is done on the outputs of the channel, using a measurement (a positive operator-valued measure, or POVM)  $\{\Pi_\ell\}$ . The probability of getting outcome  $\Pi_\ell$  if the input state  $\rho_k$  was sent is given by the Born rule,

$$p_{k\ell} = \text{tr}[\Pi_\ell \mathcal{E}(\rho_k)]. \quad (13)$$

One estimates the values of the  $p_{k\ell}$ 's from the experiment using the channel  $N_k$  times for each input state  $\rho_k$ , amounting to a total of  $N = \sum_k N_k$  uses of the channel. The sets  $\{\rho_k\}$  and  $\{\Pi_\ell\}$  are chosen to be informationally complete, i.e., there is a one-to-one mapping between the  $p_{k\ell}$ 's and the quantum channel  $\mathcal{E}$ . Once we have the estimate for  $\{p_{k\ell}\}$ , we apply the mapping to get an estimate for  $\mathcal{E}$ .

In our numerical examples below, we used a particular choice of  $\{\rho_k\}$  and  $\{\Pi_\ell\}$ , namely, the product tetrahedron states and the product tetrahedron measurement. The set of  $n$ -qubit product tetrahedron states is the pure states  $\psi_{k_1, k_2, \dots, k_n} \equiv \psi_{k_1} \otimes \psi_{k_2} \otimes \dots \otimes \psi_{k_n}$ , with each  $\psi_k$  a single-qubit state written in the Bloch-sphere representation as

$$\psi_k = \frac{1}{2}(\mathbb{1} + \mathbf{a}_k \cdot \boldsymbol{\sigma}), \quad k = 1, 2, 3, \text{ and } 4, \quad (14)$$

where the four  $\mathbf{a}_k$ s are three-dimensional unit vectors subtending a tetrahedron in the Bloch sphere [29], and  $\boldsymbol{\sigma}$  is the vector of Pauli operators. The product tetrahedron measurement is also defined in terms of these tetrahedron states:  $\Pi_{\ell_1, \ell_2, \dots, \ell_n} \equiv \Pi_{\ell_1} \otimes \Pi_{\ell_2} \otimes \dots \otimes \Pi_{\ell_n}$ , with each  $\Pi_\ell \equiv \frac{1}{2}\psi_\ell$ , for  $\ell = 1, 2, 3, \text{ and } 4$ . The tetrahedron states and measurements form an informationally complete QPT scheme. For  $n$  qubits, the scheme requires  $4^{2n}$  different settings of input state and output measurement. Owing to the tetrahedron geometry, this scheme uses the minimal number of different input states and measurement outcomes in a symmetric way.

The data from the QPT experiment consist of a sequence of detector clicks, which we summarize into  $\{n_{k\ell}\}$ , with  $n_{k\ell}$  being the number of clicks in the detector for  $\Pi_\ell$  for input

state  $\rho_k$ , where  $k, \ell = 1, 2, \dots, 4^n$ . The total number of uses of the channel is  $N = \sum_{k\ell} n_{k\ell}$ . To reconstruct the channel  $\mathcal{E}$  from the data, we first do linear inversion, i.e., solving for  $\mathcal{E}$  (using the Choi-state representation) by replacing, in the Born rule given by Eq. (13),  $p_{k\ell}$  with the relative frequency  $n_{k\ell}/N$ . This does not guarantee that the  $\mathcal{E}$  obtained from the linear inversion will be a valid, i.e., completely positive, channel. Complete positivity is then enforced by projecting the solution from linear inversion onto the nearest CPTP map using the algorithm of Ref. [30], thereby giving us an estimate  $\hat{\mathcal{E}}$  of the channel. The minimum fidelity  $F_{\min}$ , our quantity of interest, is then estimated by numerically minimizing  $F(\psi, \hat{\mathcal{E}}(\psi))$  over  $\psi$ , with standard conjugate-gradient methods.

We want to obtain an estimate of  $F_{\min}$  that is within  $\epsilon$  of the true value. The accuracy of  $F_{\min}$  is controlled by the accuracy of the estimate  $\hat{\mathcal{E}}$ —a better estimate of  $\mathcal{E}$ , obtained with a larger  $N$ , will give a more accurate estimate of  $F_{\min}$ . *A priori*, we do not know the  $N$  needed to attain the specified accuracy on  $F_{\min}$ . That depends on the unknown channel  $\mathcal{E}$ . For a fair comparison with our GFM scheme, we need  $N$  just large enough so that  $F_{\min}$  attains the desired accuracy. We thus increase  $N$  slowly and look for convergent behavior. See Fig. 1 for the summary of the scheme.

We begin with a small number of uses of the channel  $N_1$  to estimate  $\mathcal{E}$  and then  $F_{\min}$ . In the next iteration, we double the number of uses by measuring a further  $N_1$  uses, combining the obtained data with that from the previous iteration to obtain a second estimate of  $F_{\min}$ . In subsequent iterations, the amount of data used to estimate  $F_{\min}$  is doubled each time by doubling the number of uses of the channel. We continue the iteration, obtaining (hopefully) more and more accurate estimates of  $F_{\min}$  until both stopping criteria are met. The first requires that the error bar  $\hat{\sigma}^{(k)}$ —estimated through bootstrapping of the obtained data—of  $F_{\min}$  is smaller than  $2\epsilon$ . The second criterion demands that the change of  $F_{\min}$  from the previous iteration (with half the number of channel uses) is smaller than some threshold  $g_{\text{thres}}$ . A schematic of the stopping criteria is given in Fig. 3.

Note that despite the similarity to our GFM approach where the procedure was also terminated according to range and gradient criteria, the reason why a gradient criterion is needed here for QPT is rather different. The  $F_{\min}$  estimation from QPT was observed in our numerical investigations to be biased, usually underestimating the  $F_{\min}$  unless  $N$  is large enough. Similar to how we chose our  $g^{(1)}$  values for the GFM scheme, suitable values of  $g_{\text{thres}}$ , different for different number of qubits and  $\epsilon$  values, were chosen by running tests for 100 randomly chosen trial channels and setting  $g_{\text{thres}}$  such that the  $F_{\min}$  value is within the desired accuracy for 95% of the test channels.

### IV. PERFORMANCE OF OUR SCHEME

How well does our GFM scheme perform, in terms of the number of uses of the channel, compared with the alternative using QPT fidelity estimation? We are interested, in particular, with the scaling of the resource cost with the number of qubits, for fixed target precision  $\epsilon$ . The resource cost of QPT is well known to scale poorly with the dimension; does our scheme do better in that respect?

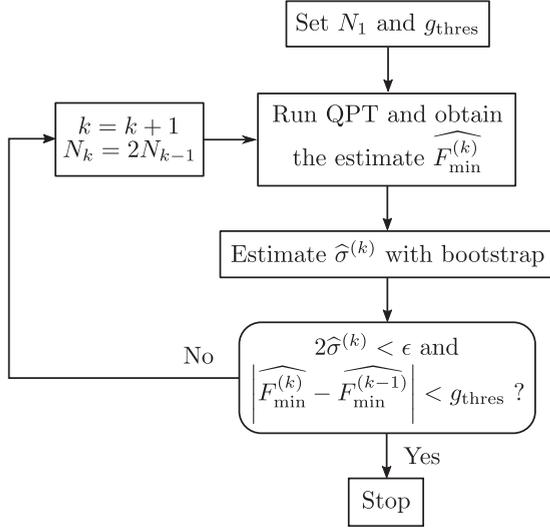


FIG. 3. Diagrammatic representation of the stopping criteria used in the QPT fidelity estimation. See main text for a description of the various quantities.

Intuitively, one would expect our GFM scheme to outperform QPT fidelity estimation. Our scheme directly estimates the one quantity of interest,  $F_{\min}$ , while QPT first estimates the full process matrix, which contains all information about the channel, and then estimates  $F_{\min}$ , discarding the rest of the gathered information. Our minimalistic approach should thus win over QPT. Indeed, resource estimates suggest this, as we explain below. However, our numerical examples, which take into full account the complexities of both schemes, as well as the particular class of channels considered, tell a somewhat different story. Below, we first discuss resource estimates to give some indication of potential performance and then describe the numerical comparison obtained from simulations of two specific classes of noise channels.

### A. Resource estimates

We can gain some insights into the possible performance by considering known theoretical bounds on the various components of our scheme as well as on QPT. We begin with QPT. A general scaling law for the resource cost of quantum state tomography was derived in Refs. [7,8]:  $O(dr^2/\epsilon'^2)$  number of copies are needed to estimate the density operator  $\rho$  of a  $d$ -dimensional system to accuracy  $\epsilon'$ , as measured by the trace-distance deviation from the true state.  $r$  is the rank of  $\rho$ . The channel-state duality via the Choi-Jamiołkowski isomorphism [31,32] allows us to apply this result directly to QPT. Specifically, a  $d$ -dimensional quantum channel corresponds to a  $d^2$ -dimensional Choi state. Further assuming that the Choi state is full rank (supposing we have no reason to assume otherwise), we see that the bound for the resource cost of QPT becomes  $O(d^6/\epsilon'^2)$  uses of the quantum channel, to estimate the Choi state within  $\epsilon'$  (trace) distance from that of the true channel.

We are, however, interested in the accuracy of  $F_{\min}$ , not in the Choi-state trace distance. Accurate reconstruction of the Choi state of the channel, of course, assures that the estimate

of  $F_{\min}$  will be accurate as well. Specifically, we can show a relationship between the deviation in minimum fidelity and the Choi-state trace distance (see Appendix A),

$$|F'_{\min} - F_{\min}| \leq 2d \|\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}\|_{\text{tr}}, \quad (15)$$

for the Choi states  $\rho_{\mathcal{E}}$  and  $\rho_{\mathcal{E}'}$  of two arbitrary channels  $\mathcal{E}$  and  $\mathcal{E}'$ , respectively, where  $F_{\min}$  and  $F'_{\min}$  are the minimum fidelity of the channels  $\mathcal{E}$  and  $\mathcal{E}'$ , respectively, and  $\|M\|_{\text{tr}} \equiv \frac{1}{2}\text{tr}(\sqrt{M^\dagger M})$  denotes the trace norm. The factor of  $d$  in the first term on the right-hand side of Eq. (15) is unavoidable; in fact, there exist (see Appendix A) one-parameter families of channels  $\mathcal{E}$  and  $\mathcal{E}'$  such that  $|F'_{\min} - F_{\min}| = d \|\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}\|_{\text{tr}}$ .

To attain accuracy  $\epsilon$  for  $F_{\min}$ , i.e.,  $|F_{\min} - F_{\min}| \leq \epsilon$ , Eq. (15) indicates that it is sufficient that the trace-distance accuracy of the QPT scheme satisfies  $\|\widehat{\rho}_{\mathcal{E}} - \rho_{\mathcal{E}}\|_{\text{tr}} \leq \epsilon'$ , where  $\epsilon' = \epsilon/(2d)$  and  $\widehat{\rho}_{\mathcal{E}}$  is the reconstructed Choi state from QPT. This yields a resource cost of  $O(d^6/(\epsilon/d)^2) = O(d^8/\epsilon^2)$ , i.e.,  $O(d^8)$  uses of the channel for fixed  $\epsilon$ , an altogether prohibitive scaling.

Next, we look at the resource estimates for our GFM scheme. Our two main ingredients, the DTFE scheme and the chosen GFM algorithm, directly determine the resource cost. The efficiency of the chosen GFM algorithm determines the number of function evaluations needed to arrive at the minimum point; the efficiency of the DTFE scheme determines how many uses of the channel are needed for each function evaluation to obtain an estimate of the required accuracy. For the latter, Ref. [9] gives  $O(d/\eta^2)$  for estimating the target fidelity to an accuracy  $\eta$ . We cannot, however, directly use this estimate. As explained earlier, in our GFM scheme,  $\eta$  is not a fixed value, but is a quantity that is adjusted as the GFM iteration proceeds. We thus also need to consider possible dimensional dependence of  $\eta$ . In all our numerical examples, we observe that  $\eta \sim 1/d$ , for the  $\eta$  values used in the course of the iterations, for  $F_{\min}$  accuracy  $\epsilon = 0.01$ . Using this empirical estimate, we thus have that each use of the DTFE subroutine requires  $O(d^3)$  uses of the channel.

For a GFM algorithm, there are generally two aspects to consider when estimating the resource cost: (i) the number of function evaluations needed in the course of the descent to a local minimum; (ii) the number of repeats of the descent to obtain the global minimum with high confidence. In our numerical examples fixed at  $\epsilon = 0.01$ , we observe that only a constant,  $d$ -independent number of repeats was needed to arrive at the global minimum with high confidence [33]. We thus disregard aspect (ii) in our considerations of the resource cost. For aspect (i), the CMA-ES algorithm has been observed empirically [21] to require  $O(d)$  to  $O(d^2)$  function evaluations in total to attain a fixed accuracy of the extremum function value. Combining this with the fact that each function evaluation invokes the DTFE scheme that requires  $O(d^3)$  uses of the channel, we see that our scheme requires  $O(d^4)$  to  $O(d^5)$  channel uses, a possible improvement over QPT fidelity estimation.

### B. Numerical examples

While resource estimates can provide initial clues to the performance of a scheme, a more accurate test comes from numerical simulations of the scheme. Numerical tests, in

particular, are able to give indications of variations in performance for different classes of channels, an aspect often not easily captured in a theoretical analysis. In this section, we present our numerical comparisons of the performance of our GFM scheme with that of QPT fidelity estimation to estimate the minimum fidelity for channels on one to five qubits.

The test channels used in our numerical simulations are drawn from two commonly encountered classes of random channels. It is possible to construct specific channels where one of the schemes outperforms the other. However, these examples are artificial and seldom relevant in practice. The first class, which we label as Class-HS, comprises channels chosen randomly according to the Hilbert-Schmidt measure on the space of quantum channels [34]. To sample from this class, Haar-random unitary matrices  $U$  are chosen on a  $d^2$ -dimensional Hilbert space, considered as the space of the system with an equal-dimensional ancilla  $A$ , and the channel is defined as  $\mathcal{E}(\cdot) \equiv \text{tr}_A[U(\cdot \otimes \psi_A)U^\dagger]$ , for  $\psi_A$  a pure state on  $A$ . This first class of channels is a well-studied class used in many discussions of generic properties of quantum channels.

The second class, which we label as Class-PA, imitates noise channels observed in many quantum experiments today:  $n$ -qubit weak noise channels composed of a random Pauli channel followed by a random amplitude-damping channel. Here, the Pauli channel is  $\mathcal{E}_{\text{Pauli}}(\cdot) \equiv \sum_{k=1}^{d^2} p_k W_k \rho W_k$ , for  $d \equiv 2^n$ , with  $W_k$  a tensor-product Pauli operator and  $\{p_k\}$  a probability distribution. The random Pauli channel is generated by drawing numbers  $q_2, q_3, \dots, q_{d^2}$  uniformly from a  $(d^2 - 2)$ -simplex, and then setting  $p_k = u q_k$ , for  $k = 2, 3, \dots, d^2$ , with  $u$  uniformly randomly chosen from the range 0 to 0.1, and  $p_1 \equiv 1 - \sum_{k=2}^{d^2} p_k$ . The upper limit of 0.1 on  $u$  ensures that we have a weak (i.e., close to the identity) Pauli channel. The amplitude-damping channel is  $\mathcal{E}_{\text{AD}}(\cdot) \equiv \sum_{j=0}^{d-1} E_j(\cdot)E_j^\dagger$ , with

$$E_0 \equiv \sum_{j=0}^{d-1} \sqrt{1 - \gamma_j} |\psi_j\rangle\langle\psi_j|,$$

$$\text{and } E_j \equiv \sqrt{\gamma_j} |\psi_0\rangle\langle\psi_j|, \quad j = 1, 2, \dots, d - 1, \quad (16)$$

where  $\gamma_0 = 0$ ,  $\gamma_1 = 0.1$ ,  $\gamma_2, \gamma_3, \dots, \gamma_{d-1}$  are uniformly randomly chosen from the range 0 and 0.1, and  $\{|\psi_j\rangle\}_{j=0}^{d-1}$  is a Haar-random basis for the system Hilbert space. The amplitude-damping channel models population decay of the  $n$ -qubit system towards some  $|\psi_0\rangle$  state. Physically,  $|\psi_0\rangle$  is typically the energetic ground state of the system, which may or may not be aligned with the chosen Pauli axes directions.

In each numerical experiment, we randomly (from the chosen class of channels) choose an  $n$ -qubit channel and estimate  $F_{\min}$  using our GFM scheme and, separately, using QPT fidelity estimation. The target accuracy for  $F_{\min}$  is set to  $\epsilon = 0.01$ . The experiment is repeated 100 times (i.e., 100 different channels) for a fixed channel class and fixed  $n$ . The tuning parameters, namely, the parameters in the CMA-ES algorithm and the DTFE procedure, and those in the stopping criteria for the GFM scheme and the QPT fidelity estimation scheme are prechosen for each channel class and  $n$  (see Sec. II) and fixed throughout the 100 experiments. The histograms for the experiments are given in Fig. 4.

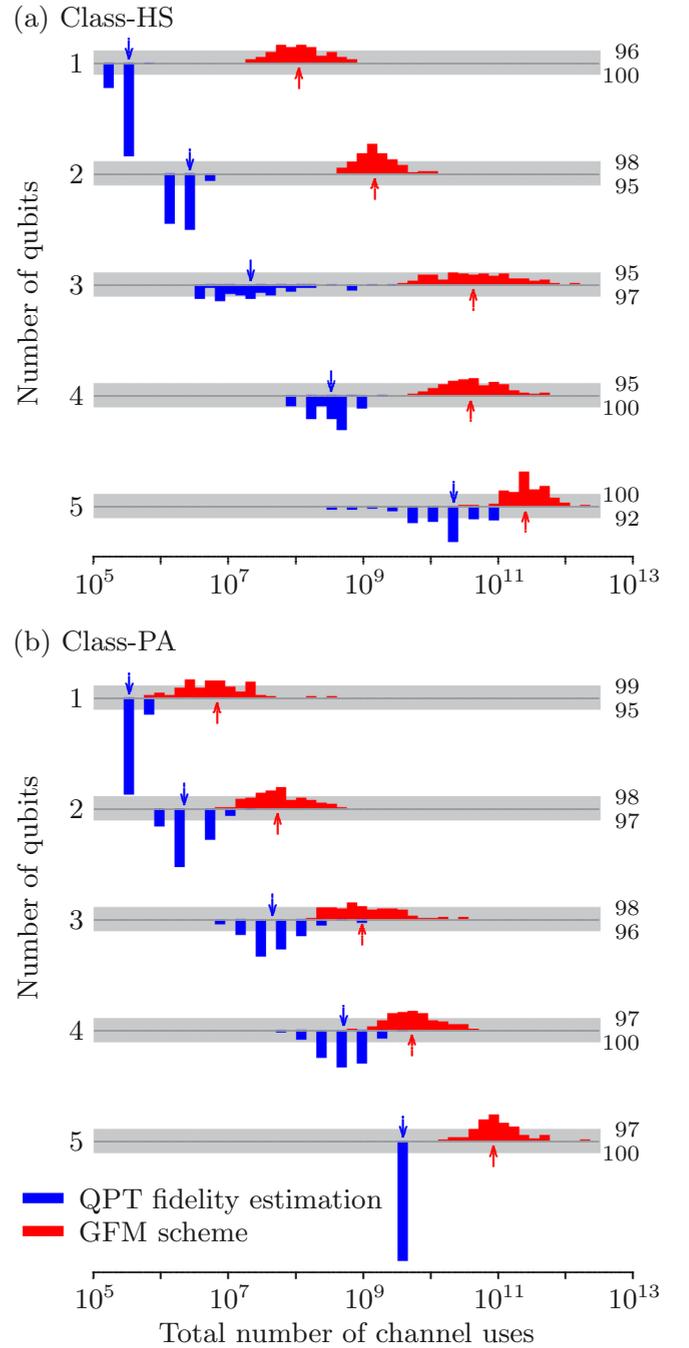


FIG. 4. The total number of channel uses for our GFM scheme and for QPT fidelity estimation for  $n$  qubits and fixed target  $F_{\min}$  accuracy of  $\epsilon = 0.01$ . Two classes of random channels were used: (a) Class-HS, and (b) Class-PA (see main text). Each histogram is generated from 100 randomly chosen channels, with the same 100 channels used for both schemes. The medians are indicated by the arrows; the gray bars mark the height of 10 counts. The numbers on the right (top number for GFM, bottom for QPT) of the histograms give the number of channels, out of 100, for which the estimated  $F_{\min}$  is within the desired accuracy. The stopping criteria are adjusted so that these numbers are around 95. The scaling of the number of channel uses with the dimension  $d = 2^n$  is observed, using the medians of the histograms, as (a) Class-HS:  $O(d^{4.2})$  for QPT and  $O(d^{2.7})$  for GFM; and (b) Class-PA:  $O(d^{3.8})$  for QPT and  $O(d^{3.4})$  for GFM.

The numerical experiments tell a rather different story than the resource estimates of the previous section. The most striking feature of Fig. 4 is that QPT fidelity estimation requires fewer uses of the channel than our GFM scheme, at least for up to the tested five-qubit situation. In terms of scaling with the dimension of the system, the number of uses of the channel for our GFM scheme is about  $O(d^3)$ , not far from the resource estimates; that for QPT is, however, closer to  $O(d^4)$ , rather than the  $O(d^8)$  behavior of the resource estimates. Overall, the numerics suggest that our GFM scheme shows some advantage over QPT in terms of scaling with the size of the system, with a larger advantage for Class-HS than for Class-PA (see Fig. 4). However, when the system size is small, i.e., the case of practical interest in the near future, QPT outperforms our scheme in actual number of uses of the channel.

One possible reason behind the significant difference between the observed numerical scaling and the resource estimates on the number of channel uses is that the inequality relation in Eq. (15) merely provides an upper bound on the  $F_{\min}$  deviation. Given a true channel  $\mathcal{E}$  and its estimate  $\hat{\mathcal{E}}$  from QPT, the deviation in minimum fidelity of  $\mathcal{E}$  and  $\hat{\mathcal{E}}$  is typically much smaller than the upper bound indicated in Eq. (15). This is particularly true for the two classes of channels in our numerical examples, as we observed empirically. Therefore, the use of Eq. (15) grossly overestimates how stringent we need to be in the trace-distance deviation to achieve a desired fidelity deviation. This highlights the importance of directly incorporating the desired figure of merit—the minimum fidelity in our case—into the QPT scheme, using it within the stopping criteria, rather than imposing a target accuracy on the deviation of the full process matrix.

Note that the spread in the number of channel uses visible in Fig. 4 is not solely due to our random choice of channels. The fluctuations in the data (and hence the final estimated  $F_{\min}$  values) for each run of both schemes also contribute to the observed spread. The histograms shown in Fig. 4 show only one run per scheme per random channel. Additional numerical studies indicate that if each procedure were repeated 100 times on the *same* channel, the spread in the number of channel uses would be roughly as large as what is seen in Fig. 4. Also, note that the gaps in the histograms for the QPT fidelity estimation in Fig. 4 (blue bars) are due to our choices of  $N_1$ , the base number of channel uses, subsequently doubled in each round of the iterative procedure.

A further remark concerns the resource scaling of our GFM scheme for Class-HS. Observe in Fig. 4(a) that we need  $O(d^{2.7})$  channel uses for our scheme, a more favorable scaling than the  $O(d^{4.2})$  of QPT fidelity estimation. In fact, we suspect a further slowdown in the increase in channel uses with  $d$  beyond five qubits for our scheme, gaining further advantage over QPT fidelity estimation. This is because, numerically, we observe that for a typical channel from Class-HS, the range of fidelity values over all pure states shrinks logarithmically as  $d$  increases. This means that for large  $d$  and fixed target accuracy for the estimation of  $F_{\min}$ , we only need to pick *any* pure state  $\psi$ , estimate  $f(\psi)$  using the DTFE scheme, and that is already close to the true  $F_{\min}$  value, even without further minimization using the GFM algorithm. We thus expect the number of uses of the channel for our GFM scheme to be

much reduced in that case. Note that this reduction in the fidelity range is not observed for Class-PA.

As the classical computational resource remains one of the limiting factors of tomography, it is also important to compare the two schemes in this aspect. The classical computational resource cost of the GFM scheme is estimated to be  $O(d^5)$ . For the QPT fidelity estimation scheme, the estimated classical computational resource cost is at least  $O(d^6)$ . See Appendix B for the discussion. In this aspect, the GFM scheme has better performance compared with the QPT fidelity estimation scheme.

## V. CONCLUSION

We have explored two different direct schemes for estimating the minimum gate fidelity. The GFM scheme presents an interesting hybrid application of numerical GFM algorithms together with the experimental procedure of direct target fidelity estimation. We compared this to the alternative approach of QPT fidelity estimation, a key difference of which from standard QPT procedures is the direct incorporation of the quantity of interest as the stopping criterion in the iteration. Resource estimates suggest an extremely high cost for QPT fidelity estimation, with a scaling for  $O(d^8)$  gate uses, compared with  $O(d^4)$  for the GFM approach. This large difference in performance, however, was not seen for the numerical tests carried out on specific classes of noise channels, with both schemes showing a gate-use scaling closer to  $O(d^4)$ . This reminds us of the need to examine the performance of every procedure for the particular context at hand, in addition to general results that apply in all situations.

In practice, one could consider using either scheme. QPT fidelity estimation has the advantage that the QPT experiment can be easier to perform, using familiar measurement setups. Its performance, however, appears limited to  $O(d^4)$  gate uses (for our specific numerical examples) to achieve an estimate of the minimum fidelity of a specified accuracy. On the other hand, while the current performance in our numerical examples is also  $O(d^4)$ , the GFM approach presents a potential for future improvement with an increase in the efficiency of the numerical GFM scheme used. The GFM approach is also more efficient in terms of classical computational resource cost. The downside, though, is that the GFM scheme, with the need for preparing arbitrary input states suggested by the GFM iteration, can be more difficult to implement experimentally. That may, nevertheless, be well worth the effort if the number of gate uses can be significantly lowered for large  $d$ .

## ACKNOWLEDGMENTS

This work is supported by the Ministry of Education, Singapore (through Grant No. MOE2016-T2-1-130). The Centre for Quantum Technologies is a Research Centre of Excellence funded by the Ministry of Education and the National Research Foundation of Singapore.

## APPENDIX A: DERIVATION OF EQ. (15)

Here, we explain the steps leading to Eq. (15) which relates  $|F'_{\min} - F_{\min}|$  to  $\|\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}\|_{\text{tr}}$ . We are concerned with two arbitrary CPTP channels,  $\mathcal{E}$  and  $\mathcal{E}'$ . The Choi state of  $\mathcal{E}$  is

denoted as  $\rho_{\mathcal{E}} \equiv (\mathbb{1} \otimes \mathcal{E})(\Phi)$ , where  $\Phi \equiv |\Phi\rangle\langle\Phi|$  with  $|\Phi\rangle \equiv \frac{1}{\sqrt{d}} \sum_{i=1}^d |i\rangle|i\rangle$ , a specific choice of a maximally entangled bipartite state. The fidelity of state  $|\psi\rangle$  under the action of  $\mathcal{E}$  is  $F_{\mathcal{E}}(\psi) \equiv F(|\psi\rangle, \mathcal{E}(\psi)) = \langle\psi|\mathcal{E}(\psi)|\psi\rangle$ , and the minimum fidelity  $F_{\min}$  is attained by state  $|\psi_{\mathcal{E}}\rangle$ . Analogous definitions apply for the channel  $\mathcal{E}'$ , with Choi state  $\rho_{\mathcal{E}'}$  and minimum fidelity  $F'_{\min}$ , attained by state  $|\psi_{\mathcal{E}'}\rangle$ .

We begin with the fidelity difference, and employ the triangle inequality,

$$\begin{aligned}
 & |F'_{\min} - F_{\min}| \\
 &= |F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}}(\psi_{\mathcal{E}})| \\
 &= \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}}(\psi_{\mathcal{E}'}) + F_{\mathcal{E}}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}}(\psi_{\mathcal{E}}) \\
 &\quad + F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}'}(\psi_{\mathcal{E}}) + F_{\mathcal{E}'}(\psi_{\mathcal{E}}) - F_{\mathcal{E}}(\psi_{\mathcal{E}})| \\
 &\leq \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}}(\psi_{\mathcal{E}'})| + \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}}) - F_{\mathcal{E}}(\psi_{\mathcal{E}})| \\
 &\quad + \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}'}(\psi_{\mathcal{E}})| + \frac{1}{2} |F_{\mathcal{E}}(\psi_{\mathcal{E}'} - F_{\mathcal{E}}(\psi_{\mathcal{E}})| \\
 &= \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}}(\psi_{\mathcal{E}'})| + \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}}) - F_{\mathcal{E}}(\psi_{\mathcal{E}})| \\
 &\quad + \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}}) - F_{\mathcal{E}'}(\psi_{\mathcal{E}'})| + \frac{1}{2} |F_{\mathcal{E}}(\psi_{\mathcal{E}'} - F_{\mathcal{E}}(\psi_{\mathcal{E}})| \\
 &\leq \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}}(\psi_{\mathcal{E}'})| + \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}}) - F_{\mathcal{E}}(\psi_{\mathcal{E}})| \\
 &\quad + \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}}) - F_{\mathcal{E}}(\psi_{\mathcal{E}})| + \frac{1}{2} |F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}}(\psi_{\mathcal{E}'})| \\
 &= |F_{\mathcal{E}'}(\psi_{\mathcal{E}'}) - F_{\mathcal{E}}(\psi_{\mathcal{E}'})| + |F_{\mathcal{E}'}(\psi_{\mathcal{E}}) - F_{\mathcal{E}}(\psi_{\mathcal{E}})|. \quad (\text{A1})
 \end{aligned}$$

Above, we noted that  $F_{\mathcal{E}}(\psi_{\mathcal{E}'}) \geq F_{\mathcal{E}}(\psi_{\mathcal{E}})$  and  $F_{\mathcal{E}'}(\psi_{\mathcal{E}}) \geq F_{\mathcal{E}'}(\psi_{\mathcal{E}'})$  since  $F_{\mathcal{E}}$  and  $F_{\mathcal{E}'}$  attain their minimum values on  $\psi_{\mathcal{E}}$  and  $\psi_{\mathcal{E}'}$ , respectively.

Now, straightforward calculation tells us that  $F_{\mathcal{F}}(\psi) = d \langle \bar{\psi} | \langle \psi | \rho_{\mathcal{F}} | \bar{\psi} \rangle | \psi \rangle$  for any state  $|\psi\rangle$ , a CPTP channel  $\mathcal{F}$ , and its associated Choi state  $\rho_{\mathcal{F}}$ . Here,  $|\bar{\psi}\rangle$  is a state associated with  $|\psi\rangle$  by the Choi-Jamiołkowski isomorphism (see, for example, Ref. [35], Sec. II for further explanation). Then, for any state  $|\psi\rangle$ ,

$$\begin{aligned}
 |F_{\mathcal{E}'}(\psi) - F_{\mathcal{E}}(\psi)| &= d \langle \bar{\psi} | \langle \psi | (\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}) | \bar{\psi} \rangle | \psi \rangle \\
 &\leq d \max_{|\bar{\psi}\rangle} \langle \bar{\psi} | (\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}) | \bar{\psi} \rangle. \quad (\text{A2})
 \end{aligned}$$

We can relate this to the trace distance  $\|\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}\|_{\text{tr}} = \frac{1}{2} \sum_i |\lambda_i|$ , where  $\lambda_i$ 's are the eigenvalues of  $\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}$ . Since  $\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}$  is traceless, i.e.,  $\sum_i \lambda_i = 0$ , we know that  $\sum_i |\lambda_i^{(+)}| = \sum_i |\lambda_i^{(-)}|$ , where  $\lambda_i^{(+)}$  and  $\lambda_i^{(-)}$  are, respectively, the positive and negative eigenvalues of  $\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}$ . Hence,  $\|\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}\|_{\text{tr}} = \sum_i |\lambda_i^{(+)}| = \sum_i |\lambda_i^{(-)}| \geq \max_{|\bar{\psi}\rangle} \langle \bar{\psi} | (\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}) | \bar{\psi} \rangle$ , so that we have

$$|F_{\mathcal{E}'}(\psi) - F_{\mathcal{E}}(\psi)| \leq d \|\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}\|_{\text{tr}}. \quad (\text{A3})$$

Applying this inequality to the two terms in the last line of Eq. (A1), we find

$$|F'_{\min} - F_{\min}| \leq 2d \|\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}\|_{\text{tr}}, \quad (\text{A4})$$

yielding the inequality of Eq. (15).

The factor of  $d$  on the right-hand side is unavoidable: Let  $\mathcal{D}(\cdot) \equiv \text{tr}(\cdot)\mathbb{1}/d$  denote the  $d$ -dimensional erasure channel, and let  $\mathcal{D}'(\cdot) \equiv \mathcal{D}(\cdot) - \eta Z|0\rangle\langle 0|(\cdot)|0\rangle$ , for  $Z \equiv |0\rangle\langle 0| - |1\rangle\langle 1|$ , and  $|0\rangle$  and  $|1\rangle$  are orthonormal states. Let  $\mathcal{I}$  denote the identity map. Then, for any  $p \in [0, 1]$ ,  $\mathcal{E} \equiv (1-p)\mathcal{I} + p\mathcal{D}$  and  $\mathcal{E}' \equiv (1-p)\mathcal{I} + p\mathcal{D}'$  saturate the inequality Eq. (A4), up

to a constant factor of 2:  $|F'_{\min} - F_{\min}| = d \|\rho_{\mathcal{E}'} - \rho_{\mathcal{E}}\|_{\text{tr}}$ , for these two maps. Small  $p$  values give channels that describe weak noise, the case of interest in this work.

## APPENDIX B: ESTIMATION OF CLASSICAL COMPUTATIONAL RESOURCE COST

In each iteration of the GFM scheme, the classical computational resource is spent on computing fidelity values with the DTFE scheme, followed by the proposal of the next set of states with the CMA-ES algorithm. To estimate a single value of fidelity using the DTFE scheme, one would compute the average of  $X$  over  $d^2$  bases [see Eq. (3)]. For each basis, to estimate the value of  $X$  from experimental data, one would take an average over the number of different outcomes, which is bounded by  $d$ . Therefore, the computational complexity of the DTFE scheme is  $O(d^3)$ . According to Ref. [22], the computational complexity of each iteration of the CMA-ES algorithm is  $O(d^2)$ . Since the required number of function evaluations is observed empirically to be  $O(d)$  to  $O(d^2)$  [21], the classical computational resource cost of the GFM scheme is estimated to be  $O(d^5)$ .

The QPT fidelity estimation scheme is composed of three computational tasks, i.e., linear inversion, projection of the solution from linear inversion to the nearest CPTP map, and numerical minimization of fidelity of the reconstructed channel. In linear inversion, one tries to invert a system of linear equations,  $Ax = f$ . Here,  $x$  is a  $d^4 \times 1$  column vector of the parameters of the channel.  $A$  is a  $m \times d^4$  matrix determined by the input states and measurements, where  $m$  denotes the total number of outcomes. For the product tetrahedron input states and measurements used in our scheme,  $m = d^4$ .  $f$  is a  $m \times 1$  column vector of the relative frequencies of the  $m$  outcomes. Due to the use of product input states and product measurements in our scheme, the matrix  $A$  can be written as a tensor product of matrices, i.e.,  $A = A_1 \otimes A_2 \otimes \cdots \otimes A_{2n}$ . As a result, the computational cost is greatly reduced. Inverting the  $2n$  matrices,  $\{A_i\}_{i=1}^{2n}$  costs  $O(n)$ . The computation of  $A^{-1}f = A_1^{-1} \otimes A_2^{-1} \otimes \cdots \otimes A_{2n}^{-1} f$  costs  $O(nd^4)$  (see Appendix E of Ref. [36]).

The projection algorithm in Ref. [30] works by projecting the channel onto CP space and TP space alternatively and iteratively. In each iteration, the most computationally expensive operation is the eigendecomposition of the Choi state, which has the computational complexity of  $O(d^6)$ . It is, however, not clear how the number of iterations in the projection algorithm scales with the dimension. The computational complexity of the projection algorithm is thus estimated to be at least  $O(d^6)$ .

The numerical minimization of the fidelity of the reconstructed channel is performed using the conjugate-gradient algorithm. The calculation of the gradient in each iteration of the conjugate-gradient algorithm involves a matrix-vector multiplication which cost  $O(d^4)$ . The observed scaling of the number of iterations required for the conjugate-gradient algorithm implemented here is, at most,  $O(d)$ . Thus, the computational complexity of the numerical minimization is  $O(d^5)$ . Since the projection algorithm is the most expensive task among the three computational tasks, the classical computational resource cost of the QPT fidelity estimation scheme is estimated to be at least  $O(d^6)$ .

- [1] F. Arute *et al.*, Quantum supremacy using a programmable superconducting processor, *Nature (London)* **574**, 505 (2019).
- [2] K. Wright *et al.*, Benchmarking an 11-qubit quantum computer, *Nat. Commun.* **10**, 5464 (2019).
- [3] P. J. Karalekas, N. A. Tezak, E. C. Peterson, C. A. Ryan, M. P. da Silva, and R. S. Smith, A quantum-classical cloud platform optimized for variational hybrid algorithms, *Quantum Sci. Technol.* **5**, 024003 (2020).
- [4] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Symmetrized characterization of noisy quantum processes, *Science* **317**, 1893 (2007).
- [5] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized benchmarking of quantum gates, *Phys. Rev. A* **77**, 012307 (2008).
- [6] E. Magesan, J. M. Gambetta, and J. Emerson, Scalable and Robust Randomized Benchmarking of Quantum Processes, *Phys. Rev. Lett.* **106**, 180504 (2011).
- [7] R. Kueng, H. Rauhut, and U. Terstiege, Low rank matrix recovery from rank one measurements, *Appl. Comput. Harmon. Anal.* **42**, 88 (2017).
- [8] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, Sample-optimal tomography of quantum states, *IEEE Trans. Inf. Theory* **63**, 5628 (2017).
- [9] S. T. Flammia and Y. K. Liu, Direct Fidelity Estimation from Few Pauli Measurements, *Phys. Rev. Lett.* **106**, 230501 (2011).
- [10] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Practical Characterization of Quantum Devices without Tomography, *Phys. Rev. Lett.* **107**, 210404 (2011).
- [11] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, *Nat. Phys.* (2020), doi: 10.1038/s41567-020-0932-7.
- [12] J. A. Nelder and R. Mead, A simplex method for function minimization, *Comput. J.* **7**, 308 (1965).
- [13] S. Singer and S. Singer, Complexity analysis of Nelder-Mead search iterations, in *Proceedings of the First Conference on Applied Mathematics and Computation*, edited by M. Rogina, V. Hari, N. Limić, and Z. Tutek (PMF Matematički odjel, Zagreb, 1999).
- [14] C. Audet and J. E. Dennis, Analysis of generalized pattern searches, *SIAM J. Optim.* **13**, 889 (2003).
- [15] A. L. Custódio and L. N. Vicente, Using sampling and simplex derivatives in pattern search methods, *SIAM J. Optim.* **18**, 537 (2007).
- [16] A. L. Custódio, H. Rocha, and L. N. Vicente, Incorporating minimum Frobenius norm models in direct search, *Comput. Optim. Appl.* **46**, 265 (2010).
- [17] Y. Nesterov and V. G. Spokoiny, Random gradient-free minimization of convex functions, *Found. Comput. Math.* **17**, 527 (2017).
- [18] S. Ghadimi and G. Lan, Stochastic first- and zeroth-order methods for nonconvex stochastic programming, *SIAM J. Optim.* **23**, 2341 (2013).
- [19] Z. Li and Q. Zhang, A simple yet efficient evolution strategy for large-scale black-box optimization, *IEEE Trans. Evol. Comput.* **22**, 637 (2018).
- [20] I. Loshchilov, A computationally efficient limited memory CMA-ES for large scale optimization, in *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*, edited by C. Igel (ACM, New York, 2014).
- [21] N. Hansen and A. Ostermeier, Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.* **9**, 159 (2001).
- [22] R. Ros and N. Hansen, A simple modification in CMA-ES achieving linear time and space complexity, in *Proceedings of the 10th International Conference on Parallel Problem Solving from Nature*, edited by G. Rudolph *et al.* (Springer, Heidelberg, 2008).
- [23] V. J. Torczon, Multi-directional search: A direct search algorithm for parallel machines, Ph.D. thesis, Rice University, Houston, 1989.
- [24] N. Hansen, The CMA evolution strategy: A tutorial, [arXiv:1604.00772](https://arxiv.org/abs/1604.00772).
- [25] D. V. Arnold, *Noisy Optimization with Evolution Strategies* (Kluwer Academic, Boston, 2002).
- [26] N. Hansen, A. Niederberger, L. Guzzella, and P. Koumoutsakos, A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion, *IEEE Trans. Evol. Comput.* **13**, 180 (2009).
- [27] V. H. Meisner and C. Igel, Uncertainty handling in evolutionary direct policy search, in *Proceedings of the NIPS-08 Workshop on Model Uncertainty and Risk in Reinforcement Learning, Whistler, 2008* (unpublished).
- [28] *Quantum State Estimation*, edited by M. Paris and J. Řeháček, Lecture Notes in Physics, Vol. 649 (Springer, Heidelberg, 2004).
- [29] J. Řeháček, B.-G. Englert, and D. Kaszlikowski, Minimal qubit tomography, *Phys. Rev. A* **70**, 052321 (2004).
- [30] G. C. Knee, E. Bolduc, J. Leach, and E. M. Gauger, Quantum process tomography via completely positive and trace-preserving projection, *Phys. Rev. A* **98**, 062336 (2018).
- [31] M.-D. Choi, Completely positive linear maps on complex matrices, *Lin. Algebra Appl.* **10**, 285 (1975).
- [32] A. Jamiolkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, *Rep. Math. Phys.* **3**, 275 (1972).
- [33] For the Class-HS example in Sec. IV B, we observe that the number of repeats we need for one- to five-qubit systems is 4, 7, 3, 1, and 1, respectively. For the Class-PA example in Sec. IV B, we observe that the number of repeats we need is 1 for one- to five-qubit systems.
- [34] W. Bruzda, V. Cappellini, H.-J. Sommers, and K. Życzkowski, Random quantum operations, *Phys. Lett. A* **373**, 320 (2009).
- [35] J. Y. Sim, J. Suzuki, B.-G. Englert, and H. K. Ng, User-specified random sampling of quantum channels and its applications, *Phys. Rev. A* **101**, 022307 (2020).
- [36] J. Shang, Z. Zhang, and H. K. Ng, Superfast maximum likelihood reconstruction for quantum tomography, *Phys. Rev. A* **95**, 062336 (2017).