Optimal strategies for estimating the average fidelity of quantum gates

Daniel M. Reich, Giulia Gualdi, and Christiane P. Koch

Theoretische Physik, Universität Kassel, Heinrich-Plett-Str. 40, D-34132 Kassel, Germany
(Dated: May 27, 2013)

We show that the minimum experimental effort to characterize the proper functioning of a quantum device scales as $2^n$ for $n$ qubits and requires classical computational resources $\sim n^2 2^{3n}$. This represents an exponential reduction compared to the best currently available protocol, Monte Carlo characterization. The reduction comes at the price of either having to prepare entangled input states or obtaining bounds rather than the average fidelity itself. It is achieved by applying Monte Carlo sampling to so-called two-designs or two classical fidelities. For the specific case of Clifford gates, the original version of Monte Carlo characterization based on the channel-state isomorphism remains an optimal choice. We provide a classification of the available efficient strategies for device characterization in terms of the number of required experimental settings, average number of actual measurements and classical computational resources.

PACS numbers: 03.65.Wj, 03.67.Ac

Introduction

The development of quantum technologies is currently facing a number of obstacles. One of them is the difficulty to assess efficiently how well a quantum device implements the desired operation. The corresponding performance measure is the average fidelity which can be determined via quantum process tomography [1]. Full process tomography scales, however, super-exponentially in the number of qubits and provides the full process matrix, i.e., much more information than just the average fidelity. For practical applications, a more targeted and less resource-intensive approach is required.

Recent attempts at reducing the resources for device characterization employ stochastic sampling [2, 3]. The process matrix can be estimated efficiently if it is sparse [4–6]. For general unitary operations, Monte Carlo sampling combined with the channel-state isomorphism currently appears to be the most efficient approach [2, 3]. It comes with the advantage of separable input states and local measurements. A second, potentially efficient approach is based on rewriting the integral defining the average fidelity, $F_{av} = \int \langle \Psi | O + D (| \Psi \rangle \langle \Psi |) O | \Psi \rangle d \Psi$, as a finite sum over state fidelities in Hilbert space.

We have recently shown that a minimal set of states in Hilbert space is sufficient for device characterization [14], such that the number of states that enter the sum is determined only by the desired bounds.

Monte Carlo characterization

We first review Monte Carlo estimation of the average fidelity as introduced in Refs. [2, 3] before applying it to two-designs [5, 7] and classical fidelities [12]. Monte Carlo sampling estimates an unknown quantity $F$ from a small random sample of measurements. Formally, $F$ is decomposed into independent contributions $X_\kappa$, $F = \sum_{\kappa=1}^{T} \Pr(\kappa) X_\kappa$, where enter with probability $\Pr(\kappa)=\sum_{\kappa=1}^{T} \Pr(\kappa)=1$. $T$ is the size of the event space, and $\Pr(\kappa)$ is also called relevance distribution. The state or process to be characterized is described by the probability $\rho$. We can associate to $\rho$ its characteristic function $\chi_{\rho}(\kappa) = \text{Tr}[\rho W_\kappa]$ which

$$
\chi_{\rho}(\kappa) = \text{Tr}[\rho W_\kappa] = \sum_{\kappa=1}^{T} \Pr(\kappa) \chi_{\rho}(\kappa),
$$
allows to reconstruct \( \rho \) from the measurements of the \( W_k \). Note that \( \chi_\nu(\kappa) \) is simply the expectation value of measurement operator \( W_k \). \( F \) can be expressed in terms of the product of the characteristic functions for the ideal and actual state/process, \( F = \sum_\kappa \chi_\nu^{\text{id}}(\kappa) \chi_\nu^{\text{act}}(\kappa) \). The random variable \( X_\kappa \) therefore becomes

\[
X_\kappa = \frac{\chi_\nu^{\text{act}}(\kappa)}{\chi_\nu^{\text{id}}(\kappa)} = \frac{\text{Tr}[\rho^{\text{act}} W_k]}{\text{Tr}[\rho^{\text{id}} W_k]} \quad \text{and} \quad P(\kappa) = \frac{\chi_\nu^{\text{act}}(\kappa)^2}{\mathcal{N}}
\]

with \( \mathcal{N} \) ensuring proper normalization of \( P(\kappa) \). Two levels of stochastic sampling are involved in the Monte Carlo characterization of a quantum state or process: Equation (1) implies that \( F \) is the expectation value of a random variable \( X \) taking values \( X_\kappa \) with known probability \( P(\kappa) \) (top level). However, the \( X_\kappa \) cannot be accessed directly, since they depend on another random variable, the expectation value of \( W_k \) for \( \rho^{\text{act}} \) (bottom level). Due to the statistical nature of quantum measurements as well as random errors in the experiment, it will be necessary to repeatedly measure \( W_k \) in order to determine \( X_\kappa \). Suppose that \( X_\kappa \) has been determined with sufficient accuracy. Monte Carlo sampling then estimates the expectation value \( F \) of the random variable \( X \) by a finite number of realizations,

\[
F = \lim_{L \to \infty} F_L \quad \text{with} \quad F_L = \frac{1}{L} \sum_{\kappa=1}^L X_{\kappa l}.
\]

The sample size \( L \) is chosen to guarantee that the probability for \( F_L \) to differ from \( F \) by more than \( \epsilon \) is less than \( \delta \). The key point of the Monte Carlo approach is that \( L \) depends only on the desired accuracy \( \epsilon \) and confidence level \( \delta \) and is independent of the system size. However, the \( X_{\kappa l} \) are known only approximately. Denoting their approximate values by \( \tilde{X}_{\kappa l} \), this gives rise to an approximation of \( F_L \) by \( \tilde{F}_L = \frac{1}{L} \sum_{\kappa=1}^L \tilde{X}_{\kappa l} \). In addition to ensuring that \( F_L \) approximates \( F \) with an error of at most \( \epsilon \), one therefore also has to guarantee that \( \tilde{F}_L \) approximates \( F_L \) with the desired accuracy. This implies repeated measurements for a given element \( l \) (\( l = 1, \ldots, L \)) of the Monte Carlo sample. Denoting the number of respective measurements by \( N_l \), the total number of experiments is given by \( N_{\text{exp}} = \sum_{l=1}^L N_l \). It can be shown that a proper choice of \( (N_l) \) and \( \langle N_{\text{exp}} \rangle \) guarantees the approximations of \( F_L \) by \( \tilde{F}_L \) and of \( F \) by \( F_L \) to hold with the desired confidence level. Generally, this choice depends on the system size.

The average fidelity of a quantum process can be obtained by Monte Carlo estimation when combining it with the channel-state isomorphism [2, 3]. \( F_{av} \) is then expressed in terms of the entanglement fidelity \( F_e \) via \( F_{av} = (d F_e + 1)/(d + 1) \) [15]. Since \( F_e \) is a state fidelity in Liouville space and Liouville space vectors correspond to Hilbert space operators, this implies evaluation of \( F_{av} \) with respect to an operator basis, comparing input to output operators. Since a complete operator basis consists of \( 2^n \) elements and the size of the event space is given by all possible combinations of input and output operators, \( T = 2^{2n} \). The fact that only states, not operators can be prepared as input is remedied by randomly selecting eigenstates of the input operators. There are \( 6 \) eigenstates for the \( 3 \) Pauli operators for each qubit. Therefore the number of experimental settings, i.e., pairs of input state/output measurement operator, which one needs to be able to implement experimentally is given by \( N_{\text{setting}} = N_{\text{input}} \times N_{\text{meas}} = 6^n \cdot 2^n \). The random selection of settings for the actual experiments requires classical computational resources \( C_{\text{class}} \) that scale as \( n^2 2^{4n} \) [2, 3]. To summarize, in Monte Carlo estimation one randomly draws \( L \) samples from the total number \( T \) of possible events on a classical computer (with computational cost \( C_{\text{class}} \)). Although only some of the possible settings will be selected, the ability to implement all settings in the experiment is implied. Due to the statistical nature of measurements, all in all \( \langle N_{\text{exp}} \rangle \) runs of the experiment have to be carried out. For the experimental implementation, \( N_{\text{setting}}, \langle N_{\text{exp}} \rangle \) and \( C_{\text{class}} \) thus characterize the procedure.

Separating input states and measurement operators

Since state fidelities in Hilbert space as opposed to a state fidelity in Liouville space are sufficient to estimate \( F_{av} \) [12, 14], we distinguish in an event \( \kappa l \) between input states and measurement operators, \( \kappa l = (i_l, k_l) \). This allows for applying Monte Carlo sampling to the classical fidelities of Ref. [12] and the two-design approach [7]. In the spirit of Monte Carlo device characterization [2, 3], we consider only local measurements, i.e., we choose our \( W_k \) to be Pauli operators.

The two classical fidelities which yield an upper and a lower bound to the average fidelity [12] can be written as [14]

\[
F_{j} = \frac{1}{d} \sum_{i=1}^{d} \sum_{j} \text{Tr}[\rho_{i}^{j, id} \rho_{j, act}^{j, id}]
= \frac{1}{d} \sum_{i=1}^{d} \sum_{j} \text{Tr}[U |\Psi_{i}^{j} \rangle \langle \Psi_{j}^{j} | U^+ D(|\Psi_{i}^{j} \rangle \langle \Psi_{j}^{j} |)]
\]

with \( \{|\Psi_{i}^{j} \rangle \}_{i=1}^{d} (j = 1, 2) \) two mutually unbiased bases in \( d \)-dimensional Hilbert space (\( d = 2^n \)), \( D \) the dynamical map describing the actual evolution, and \( U \) the desired unitary. Expanding the states \( \rho_{i}^{j, id} \), \( \rho_{j, act}^{j, id} \) in terms of Pauli operators, Eq. (3) becomes

\[
F_{j} = \sum_{i=1}^{d} \sum_{k=1}^{2^n} \mathcal{P}_{i, k}^{j} \langle i, k \rangle \chi_{U}^{j}(i, k)
\]

with characteristic function \( \chi_{U}^{j}(i, k) = \text{Tr}[W_k U |\Psi_{i}^{j} \rangle \langle \Psi_{j}^{j} | U^+] \) and relevance distribution \( \mathcal{P}_{i, k}^{j} = \frac{1}{\mathcal{D}} \chi_{U}^{j}(i, k) \). Note that \( \text{Tr}[W_k W_{k'}] = d \delta_{k,k'} \).
We show in the supplementary material that $\Pr^j(i,k)$ is properly normalized such that we can estimate $\hat{F}_j$ ($j = 1, 2$) by Monte Carlo sampling.

The expression for the average fidelity when using two-designs, given in terms of $d + 1$ mutually unbiased bases [7],

$$ F_{av}^{2 \text{des}} = \frac{1}{d(d + 1)} \sum_{i=1}^{d(d+1)} \text{Tr}[U|\Psi_i\rangle \langle \Psi_i|U^+D(|\Psi_i\rangle \langle \Psi_i|)], $$

is formally similar to Eq. (3), i.e., it can be interpreted as the sum over $d + 1$ classical fidelities. Equation (5) can thus be rewritten

$$ F_{av}^{2 \text{des}} = \frac{1}{d(d + 1)} \sum_{i=1}^{d(d+1)} \sum_{k=1}^{d^2} \Pr^{2 \text{des}}(i,k) \chi^2_{\text{av}}(i,k), $$

with characteristic function $\chi^2_{\text{av}}$ analogous to $\chi^2_U$, and the relevance distribution differing only in normalization, $\Pr^{2 \text{des}}(i,k) = \frac{1}{d(d + 1)} \left[ \chi^2_{\text{av}}(i,k) \right]^2$. We show in the supplementary material that also $\Pr^{2 \text{des}}(i,k)$ is properly normalized.

### Resources for Monte Carlo estimation of general unitaries

Equations (4) or (6) are amenable to Monte Carlo estimation. This involves randomly selecting $L$ times a pair $(i, k_l)$ of input state/measurement operator. Compared to Refs. [2, 3], the number of input states is significantly reduced for the two approaches based on state fidelities in Hilbert space. This yields a correspondingly smaller number of settings that an experimentalist needs to be able to implement, cf. Table I. Moreover, the smaller number of input states yields an exponential reduction in the classical computational resources required for the random selection for the classical fidelities. This is due to $C_{\text{class}} = N_{\text{input}} \times C_{\text{single}}$ with $C_{\text{single}}$ the classical computational cost for sampling a single state fidelity in Hilbert space ($C_{\text{single}} \sim n^{2+2n}$ [3]). It reflects the fact that the relevance distribution for the classical fidelities depends on $O(d^3)$ parameters instead of the $O(d^4)$ parameters associated to the relevance distribution of Refs. [2, 3]. These parameters, $O(d^3)$, are sufficient to determine whether the actual evolution matches the desired unitary [14].

Analogously to Refs. [2, 3], we determine the sample size $L$ by Chebychev’s inequality. It provides an upper bound for the probability of a random variable $Z$ with variance $\sigma_Z$ to deviate from its mean,

$$ \Pr \left[ |Z - \langle Z \rangle| \geq \sigma_Z / \sqrt{\delta} \right] \leq \delta $$

(7)

\forall \delta > 0. In our case, $Z = F_L = \sum_{i=1}^{L} X_i$ and $\langle Z \rangle = F$, with $F = F_{av}^{2 \text{des}}$ or $F^j$. We show in the supplementary material that the variance of $X_i$, $\sigma^2 = \chi^2(i,k_l)/\chi^2_U(i,k_l)$, cf. Eqs. (4), (6), is smaller than one, and thus $\text{var}(F_L) \leq 1/L$. Then the choice $L = 1/(\epsilon^2 \delta)$ guarantees that the probability for our estimate $\hat{F}_L$ to differ from $F$ by more than $\epsilon$ is smaller than $\delta$. Choosing the desired accuracy $\epsilon$ and confidence level $\delta$ thus determines the sample size.

For each setting $l = 1, \ldots, L$, the observable $W_{l,i}$ has to be measured $N_l$ times to account for the statistical nature of the measurement. The corresponding approximation to $X_i$ is then given by

$$ \hat{X}_i = \frac{\chi^2_U(i,k_l)}{N_i} \sum_{j=1}^{N_i} w_{lj} $$

(8)

with $w_{lj}$ the measurement result for the $j$th repetition of experimental setting $l$, equal to either +1 or -1 for Pauli operators. Since $\hat{X}_i$ is given as the sum of independent random variables $w_{lj}$, $N_l$ can be determined using Hoeffding’s inequality. It provides an upper bound for the probability of a sum $S = \sum_{i=1}^{N_l} X_i$ of independent variables $Y_i$ with $a_i \leq Y_i \leq b_i$ to deviate from its expected value by more than $\epsilon$,

$$ \Pr \left[ |S - \langle S \rangle| \geq \epsilon \right] \leq 2 \exp \left( -\frac{2 \epsilon^2}{\sum_{i=1}^{N_l} (b_i - a_i)^2} \right) $$

(9)

\forall \epsilon > 0. In our case, $S = \hat{F}_L = \frac{1}{L} \sum_{i=1}^{L} \hat{X}_i$ and, using Eq. (8), $\sum_{i=1}^{N_l} (b_i - a_i)^2 = \sum_{i=1}^{L} 4N_l \chi^2_U(i,k_l)$. Obviously, the choice

$$ N_l = \frac{2}{L \epsilon^2 \chi^2_U(i,k_l)} \log \left( \frac{2}{\delta} \right) = N_l(i,k_l) $$

(10)

ensures the right-hand side of Eq. (9) to be $\leq \delta$. The setting $l$ is chosen with probability $\Pr^j/2 \text{des}(i,k_l)$. The average number of times that this specific experiment (with input state $i_l$ and measurement operator $W_{k_l}$) is

| approach | $C_{\text{class}}$ | $N_{\text{input}}$ | $N_{\text{setting}}$ | $\langle N_{\text{exp}} \rangle$ |
|----------|-------------------|-----------------|-------------------|-----------------|
| A        | $O(n^{2+2n})$    | 6               | $O(6^n2^{2n})$    | $O(2^{2n})$    |
| B        | $O(n^{2+2n})$    | $2n(2n+1)$     | $O(2^{2n})$      | $O(2n)$       |
| C        | $O(n^{2+2n})$    | 2 - 2n          | $O(2^n)$         | $O(2^n)$      |
carried out is therefore given by

\[
\langle N_i \rangle = \sum_{i=1}^{d} \sum_{k=1}^{d^2} \Pr(i, k) N_i(i, k) = \sum_{i=1}^{d} \sum_{k=1}^{d^2} \frac{4}{|\chi_U(i, k)|^2 L^2 \log \left( \frac{2}{\delta} \right)} \leq 1 + \frac{2d}{L^2 \log \left( \frac{2}{\delta} \right)}
\]

for the two classical fidelities \((j = 1, 2)\). The same \(\langle N_i \rangle\) is obtained for the two-designs due to normalization of \(\Pr^{2des}(i, k)\). The total number of experiments that need to be carried out to estimate the fidelities \(F_j\) and \(F_{av}^{2des}\) with accuracy \(\epsilon\) and confidence \(\delta\) is then estimated by

\[
\langle N_{exp} \rangle = \sum_{i=1}^{L} \langle N_i \rangle \leq L \left[ 1 + \frac{2d}{L^2 \log \left( \frac{2}{\delta} \right)} \right] \leq 1 + \frac{1}{c^2 \delta} + \frac{2d}{c^2} \log \left( \frac{2}{\delta} \right)
\]

This number is sufficient to account for both the sampling error due to finite \(L\) and statistical experimental errors in the measurement results. Notably, \(\langle N_{exp} \rangle \sim 2^n\) only, i.e., the average number of experiments to estimate \(F_{av}\) scales like that required for characterizing a general pure quantum state [4]. This represents an exponential reduction compared to Refs. [2, 3], cf. Table I. These savings come at the expense of (i) obtaining only bounds on the average fidelity when using the two classical fidelities \(F_j\) or (ii) the necessity to prepare entangled input states when using two-designs. The latter scales quadratically in \(n[7]\). Even factoring this additional cost in, Monte Carlo estimation of the average fidelity for a general unitary operation using two-designs is still exponentially more efficient than that based on the channel-state isomorphism [2, 3].

**Resources for Monte Carlo estimation of Clifford gates**

The scaling of \(\langle N_{exp} \rangle\) with the number of qubits changes dramatically for Clifford gates [2, 3]. This is due to the property of Clifford gates to map eigenstates of a \(d\)-dimensional set of commuting Pauli operators into eigenstates from the same set. The mutually unbiased bases in Eqs. [3, 5] can be chosen to be such eigenstates [10]. Given a generic eigenstate, \(|\Psi_i\rangle\), of a commuting set \(\mathcal{W}\) of Pauli operators, the characteristic function of a Clifford gate, \(U_{CI}\), becomes

\[
\chi_{U_{CI}}(i, k) = \text{Tr}[W_k U_{CI} |\Psi_i\rangle \langle \Psi_i| U_{CI}^\dagger] = \text{Tr}[W_k |\Psi_j\rangle \langle \Psi_j|] = \begin{cases} 1 & \text{if } W_k \in \mathcal{W} \\ 0 & \text{otherwise} \end{cases}
\]

Equation (13) implies the relevance distribution for Clifford gates, \(\Pr(i, k) \sim |\chi_{U_{CI}}(i, k)|^2\), to be zero for many settings \((i, k)\) and uniform otherwise. Since settings with \(\Pr(i, k) = 0\) will never be selected, the sampling complexity becomes independent of system size. Calculating \(\langle N_i \rangle\) according to Eq. (11) for a uniform relevance distribution, and accounting for the correct normalizations of \(\Pr(i, k)\), \(\langle N_i \rangle\) is found to be independent of \(d\), \(\langle N_i \rangle \leq 1 + \log(2)/L^2\), for all three approaches. Consequently, also \(\langle N_{exp} \rangle\) does not scale with system size, \(\langle N_{exp} \rangle \leq 1 + 1/(c^2 \delta) + 2 \log(2)/\delta^2\), cf. Table II. For Clifford gates, the three approaches require therefore a similar, size-independent number of measurements. A difference is found, however, for the number of possible experimental settings. For each input state \(i\), there are only \(d\) (instead of \(d^2\)) measurement operators \(W_k\) with non-zero expectation value. This leads to \(N_{setting} = N_{input} \times 2^n\) for Clifford gates, cf. Table II.

**Conclusions**

If one seeks an exact estimate of \(F_{av}\) for a general unitary operation with separable input states and local measurements, Refs. [2, 3, 10] provide the optimal strategy, requiring of the order \(2^n\) measurements for \(n\) qubits. This number of measurements can be reduced exponentially by determining bounds to \(F_{av}\) instead of the average fidelity itself using two classical fidelities or by allowing for entangled input states. Their preparation requires a number of single-qubit and two-qubit gates that scales as \(n^2[7]\). For the classical fidelities, the number of experimental settings that one needs to be able to prepare and the classical computational resources required for the sampling are also reduced exponentially. We have shown earlier [11] that the minimum number of input states for device characterization is of the order \(2^n\), corresponding to the number of states required by the classical fidelities. Therefore Monte Carlo sampling applied to classical fidelities represents the characterization strategy that requires the minimum amount of resources. For the special case of Clifford gates, the same scaling with the number of experiments is attained by the approach of Refs. [2, 3] which potentially provides a higher accuracy but requires a larger number of possible experimental settings. Our comprehensive classification should allow an experimentalist to choose the most suitable procedure to determine the average fidelity, defined in terms of the number of experimental settings, from which a Monte Carlo procedure randomly draws realizations, and the actual number of experiments to be carried out.

We would like to thank David Licht for helpful comments. Financial support from the EC through the FP7-
People IEF Marie Curie action Grant No. PIEF-GA-2009-254174 is gratefully acknowledged.

∗ present address: QSTAR - Quantum Science and Technology in Arcetri, Largo Enrico Fermi 2, I-50125 Florence, Italy

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

[2] S. T. Flammia and Y.-K. Liu, Phys. Rev. Lett. **106**, 230501 (2011).

[3] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, Phys. Rev. Lett. **107**, 210404 (2011).

[4] A. Shabani, R. L. Kosut, M. Mohseni, H. Rabitz, M. A. Broome, M. P. Almeida, A. Fedrizzi, and A. G. White, Phys. Rev. Lett. **106**, 100401 (2011).

[5] C. T. Schmiegelow, A. Bendersky, M. A. Larotonda, and J. P. Paz, Phys. Rev. Lett. **107**, 100502 (2011).

[6] M. Mohseni and A. T. Rezakhani, Phys. Rev. A **80**, 010101 (2009).

[7] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. Lett. **100**, 190403 (2008).

[8] W. K. Wootters and B. D. Fields, Ann. Phys. **191**, 363 (1989).

[9] J. Lawrence, Phys. Rev. A **84**, 022338 (2011).

[10] L. Steffen, M. P. da Silva, A. Fedorov, M. Baur, and A. Wallraff, Phys. Rev. Lett. **108**, 260506 (2012).

[11] C. T. Schmiegelow, M. A. Larotonda, and J. P. Paz, Phys. Rev. Lett. **104**, 123601 (2010).

[12] H. F. Hofmann, Phys. Rev. Lett. **94**, 160504 (2005).

[13] B. P. Lanyon, C. Hempel, D. Nigg, M. Müller, R. Gerritsma, F. Zähringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, et al., Science **334**, 57 (2011).

[14] D. M. Reich, G. Gualdi, and C. P. Koch, arXiv:1305.3222 (2013).

[15] M. Horodecki, P. Horodecki, and R. Horodecki, Phys. Rev. A **60**, 1888 (1999).

[16] J. Lawrence, C. Brukner, and A. Zeilinger, Phys. Rev. A **65**, 032320 (2002).
We provide here detailed proofs of the claims made in the paper.

The relevance distribution for a classical fidelity, Pr\(^l(i, k)\), Eq. (5) of the main paper, is normalized.

In order to prove normalization of the relevance distribution Pr\(^l(i, k)\), Eq. (5) of the main paper, we first show that

\[
\sum_{k=1}^{d^2} \langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle = d \delta_{im} \delta_{jn} \quad (1)
\]

with \(\{|\varphi_j\rangle\}\) the canonical basis. Note that for each pair \(|\varphi_i\rangle, |\varphi_j\rangle\), there are exactly \(d\) operators \(W_k\) with \(\langle \varphi_i | W_k | \varphi_j \rangle \neq 0\). This is seen easily in the bit representation \((W_k = \omega_k^1 \otimes \ldots \otimes \omega_k^n)\). For the \(m\)th qubit, the scalar product vanishes if there is a bit flip between the two states at the \(m\)th qubit and \(\omega_k^m = 1\) or \(\sigma_z\). Analogously, the scalar product vanishes if the \(m\)th qubit has the same value between the two states and \(\omega_k^m = \sigma_x\) or \(\sigma_y\). There are thus only two choices of \(W_k\) for each qubit that lead to a non-zero scalar product. Repeating the argument over all \(n\) qubits gives exactly \(2^n = d\) possible operators \(W_k\) for which \(\langle \varphi_i | W_k | \varphi_j \rangle \neq 0\).

Consider now \(\langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle\) for a certain \(W_k\) with \(|\varphi_j\rangle \neq |\varphi_m\rangle\) and \(|\varphi_j\rangle, |\varphi_n\rangle\) fixed. Since \(|\varphi_j\rangle \neq |\varphi_m\rangle\) there exists a qubit, \(l\), where the two states differ. We differentiate two cases for this qubit:

1. \(|\varphi_j\rangle\) and \(|\varphi_i\rangle\) take the same value on the \(l\)th qubit. Then \(\omega_k^l\) must be \(1\) or \(\sigma_z\) for \(\langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle\) not to vanish. However, there exists an operator \(W_{k'}\) such that the contribution of the two operators to the sum, \(\langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle + \langle \varphi_i | W_{k'} | \varphi_j \rangle \langle \varphi_n | W_{k'} | \varphi_m \rangle\), vanishes. This operator is identical to \(W_k\) except that \(\omega_k^l = \sigma_z\) and vice versa. Then

\[
\langle \varphi_i | W_{k'} = - \langle \varphi_i | W_k \quad \text{and} \quad W_{k'} | \varphi_m \rangle = W_k | \varphi_m \rangle
\]

with the minus sign due to \(\sigma_z\) on the \(l\)th qubit for either \(W_k\) or \(W_{k'}\).

2. Alternatively, \(|\varphi_j\rangle\) and \(|\varphi_i\rangle\) take different values on the \(l\)th qubit. Then \(\omega_k^l\) must be \(\sigma_x\) or \(\sigma_y\) for \(\langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle\) not to vanish. Again, there exists an operator \(W_{k'}\) such that the contribution of the two operators to the sum, \(\langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle\), vanishes. This operator is identical to \(W_k\) except that \(\omega_k^l = \sigma_x\) if \(\omega_k^l = \sigma_y\) and vice versa. If \(|\varphi_j^l\rangle = |1\rangle\) (and thus \(|\varphi_m^l\rangle = |0\rangle\)),

\[
\langle \varphi_i | W_{k'} = -i \langle \varphi_i | W_k \quad \text{and} \quad W_{k'} | \varphi_m \rangle = -i W_k | \varphi_m \rangle
\]

Otherwise, if \(|\varphi_j^l\rangle = |0\rangle\) (and thus \(|\varphi_m^l\rangle = |1\rangle\)),

\[
\langle \varphi_i | W_{k'} = i \langle \varphi_i | W_k \quad \text{and} \quad W_{k'} | \varphi_m \rangle = i W_k | \varphi_m \rangle
\]

In both cases, the terms in the sum cancel.

Consequently, for each \(W_k\) and \(|\varphi_i\rangle, |\varphi_j\rangle, |\varphi_m\rangle, |\varphi_n\rangle\) with \(|\varphi_i\rangle \neq |\varphi_j\rangle\) there exists a “pair operator” \(W_{k'}\) which cancels the contribution of \(W_k\) to Eq. (1) such that

\[
\sum_{k=1}^{d^2} \langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle = 0 \quad \text{if} \quad |\varphi_i\rangle \neq |\varphi_j\rangle
\]

Repeating the argument for \(|\varphi_j\rangle \neq |\varphi_n\rangle\) leads to

\[
\sum_{k=1}^{d^2} \langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle = d \delta_{im} \delta_{jn} \sum_{k=1}^{d^2} \langle \varphi_i | W_k | \varphi_j \rangle \langle \varphi_n | W_k | \varphi_m \rangle
\]

\[
= \delta_{im} \delta_{jn} \sum_{k=1}^{d^2} |\langle \varphi_i | W_k | \varphi_j \rangle|^2
\]

using Hermiticity of \(W_k\) in the last step. Now validity of Eq. (1) follows simply from the fact that there exist, for each pair \(|\varphi_i\rangle, |\varphi_j\rangle\), exactly \(d\) operators \(W_k\) with non-vanishing \(\langle \varphi_i | W_k | \varphi_j \rangle\), and, in the canonical basis, these matrix elements are equal to one.

Expanding a general vector \(|\Phi\rangle\) in the canonical basis and using Eq. (1), we find our second intermediate result,
It is now straightforward to prove normalization of $\Pr^i(i, k)$, starting from the definition

$$\Pr^i(i, k) = \frac{1}{d^2} \langle \sum_i^d \langle \Phi | W_k | \Phi \rangle \rangle^2 = \frac{1}{d^2} \left| \text{Tr} [W_k U^+ | \Psi_i^i] \langle \Phi | W_k | \Phi \rangle \right|^2$$

where we have used Eq. (2) in the last line with $|\Phi\rangle = U|\Psi_i^i\rangle$.

The variance of $X_i$ is smaller than one.

$X_i$, the random variable of the top level of sampling in the Monte Carlo estimation of the average fidelity,
has been defined in the main paper as the ratio of the measurement outcomes for the actual state and the ideal state, $$X_l = \frac{\chi_D(i_l, k_l)}{\chi_U(i_l, k_l)}.$$ We first show that the variance of each $X_l$ in the estimation of the classical fidelities is not too large,

$$\text{Var} \left( X_l \right) = \mathbb{E} \left( X_l^2 \right) - \mathbb{E} \left( X_l \right)^2 = \sum_{i=1}^{d} \sum_{k=1}^{d^2} \text{Pr}^j (i, k) \left( \frac{\chi_D^j(i_l, k_l)}{\chi_U^j(i_l, k_l)} \right)^2 - \left( \sum_{i=1}^{d} \sum_{k=1}^{d^2} \text{Pr}^j (i, k) \frac{\chi_D^j(i_l, k_l)}{\chi_U^j(i_l, k_l)} \right)^2$$

$$= \frac{1}{d^2} \sum_{i=1}^{d} \sum_{k=1}^{d^2} \left[ \sum_{m=1}^{d} \text{Tr} \left[ U |\Psi_i^j\rangle \langle \Psi_i^j| U^+ D \left( |\Psi_i^j\rangle \langle \Psi_i^j| \right) \right] \right]^2$$

$$= \frac{1}{d^2} \sum_{i=1}^{d} \sum_{k=1}^{d^2} \text{Tr} \left[ W_k D \left( |\Psi_i^j\rangle \langle \Psi_i^j| \right) \right]^2 - F_j^2$$ (5)

with $F_j$ the classical fidelities ($j = 1, 2$) defined in Eq. (5) of the main paper. Since $0 \leq F_j \leq 1$, $0 \leq F_j^2 \leq 1$. The same is true for the first term. This can be seen as follows. Each term $D \left( |\Psi_i^j\rangle \langle \Psi_i^j| \right)$ can be written as a density matrix $\rho_i$ with eigenvectors $|\phi_n^{(i)}\rangle$ and eigenvalues $\lambda_n^{(i)}$. Evaluating the trace for each $i$ in the corresponding eigenbasis yields

$$D \left( |\Psi_i^j\rangle \langle \Psi_i^j| \right) = \rho_i = \sum_{n=1}^{d} \lambda_n^{(i)} |\phi_n^{(i)}\rangle \langle \phi_n^{(i)}|.$$