Progress toward scalable tomography of quantum maps using twirling-based methods and information hierarchies

Citation: Lopez, Cecilia C. et al. "Progress toward scalable tomography of quantum maps using twirling-based methods and information hierarchies." Physical Review A 81.6 (2010): 062113. © 2010 The American Physical Society

As Published: http://dx.doi.org/10.1103/PhysRevA.81.062113

Publisher: American Physical Society

Persistent URL: http://hdl.handle.net/1721.1/60657

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Progress toward scalable tomography of quantum maps using twirling-based methods and information hierarchies

Cecilia C. López,[1,2,3] Ariel Bendersky,[4] Juan Pablo Paz,[4] and David G. Cory[1,5]
[1]*Department of Nuclear Science and Engineering, MIT, Cambridge, Massachusetts 02139, USA*
[2]*Theoretische Physik, Universität des Saarlandes, D-66041 Saarbrücken, Germany*
[3]*Departament de Física, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain*
[4]*Departamento de Física, FCEyN, UBA, Ciudad Universitaria, 1428 Buenos Aires, Argentina*
[5]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2J 2W9, Canada*
(Received 9 March 2010; published 11 June 2010)

We present in a unified manner the existing methods for scalable partial quantum process tomography. We focus on two main approaches: the one presented in Bendersky *et al.* [Phys. Rev. Lett. **100**, 190403 (2008)] and the ones described, respectively, in Emerson *et al.* [Science **317**, 1893 (2007)] and López *et al.* [Phys. Rev. A **79**, 042328 (2009)], which can be combined together. The methods share an essential feature: They are based on the idea that the tomography of a quantum map can be efficiently performed by studying certain properties of a twirling of such a map. From this perspective, in this paper we present extensions, improvements, and comparative analyses of the scalable methods for partial quantum process tomography. We also clarify the significance of the extracted information, and we introduce interesting and useful properties of the χ-matrix representation of quantum maps that can be used to establish a clearer path toward achieving full tomography of quantum processes in a scalable way.

DOI: 10.1103/PhysRevA.81.062113 PACS number(s): 03.65.Wj, 03.65.Yz, 03.67.Pp

I. INTRODUCTION

The number of parameters describing a quantum map scale exponentially with $\ln(D)$, with D the dimension of the Hilbert space $\mathcal{H}_D$ of the system. One can then argue that the resources required to obtain this exponentially large number of parameters will also necessarily increase exponentially. This is why the complete characterization of a quantum map is considered to be a *nonscalable* task. The task of characterizing a quantum map is known as quantum process tomography (QPT) [1] and the above is the main reason why full QPT is exponentially expensive. Moreover, many existing methods have another major defect as they are inefficient also in extracting partial information about the quantum process (for a review, see [2]). Recently, however, several works [3–10] have demonstrated that it is possible to extract partial but nevertheless relevant information in an efficient way [where by efficient we mean that it is done at a cost that scales at most polynomially with $\ln(D)$]. This has opened a new chapter in quantum information processing toward the scalable characterization of quantum processes. These new methods share a common feature. They are based on the idea that the relevant properties of the quantum map can be obtained by averaging properties of a family of maps which are obtained from the original one. The averaging is done by an operation denoted as twirling [11] (which will be defined in detail later) and involves the application of certain operations before and after the application of the map.

In this work we present a review of the recent methods for partial QPT, establishing connections between them and adding results. We not only present a unifying perspective of these methods but also develop a better understanding of the problem at hand—the tomographic characterization itself.

The paper is organized as follows: In Sec. II we introduce the χ-matrix description of a quantum process distinguishing completely positive (CP) maps and others that are not CP. In Sec. III we present the basic ideas behind the notion of a twirling operation. We show that the elements of the χ matrix can be obtained from this type of operation. Moreover, some important properties of such a matrix (in particular, some useful relations between diagonal and off-diagonal elements) are discussed in Secs. II A and III A.

In Sec. IV we review the method of "selective efficient quantum process tomography," originally presented in [8,10]. We reformulate this approach by using more general types of twirlings. Not only do we highlight the power of this method but also we establish the convenience of one type of twirl over another. Furthermore, we provide a clear prescription for its implementation when targeting the scalable measurement of several χ-matrix elements at a time.

In Sec. V we move to protocols utilizing simpler forms of twirling, which are substantially less demanding regarding their experimental implementation. We take the results from [6] and [9] and present them in a new compact form as a single protocol enabling us to obtain the diagonal elements of the χ matrix grouped by "how many" and "which" qubits are affected by the quantum map. By fully proving the method by construction, we aim to further clarify its simple implementation as well as its limitations.

Finally, in Sec. VI, we discuss the potential of these strategies toward achieving scalable complete tomography of a quantum process. We believe that the key to this lies in the hierarchization of the exponentially large number of parameters (in which the results of Secs. II A and III A play an important role). The methods described in Secs. IV and V retrieve the diagonal elements of the χ matrix, and in Sec. VI we show how the diagonal elements provide information not only about themselves but also about the off-diagonal ones. This is what we identify as an information hierarchy.

Furthermore, we hope that this article sets a practical path for experimentalists looking to implement quantum process

characterization in a quantum information setting, that is, when the scalability of the tomographic method matters.

II. THE χ-MATRIX DESCRIPTION OF A QUANTUM PROCESS

A general quantum process can be described by the action of an arbitrary map Λ on the state ρ in $\mathcal{H}_D$. Any linear map Λ can be expressed as

$$\mathbf{\Lambda}(\rho) = \sum_{l,l'=0}^{D^2-1} \chi_{l,l'} E_l \rho E_{l'}^{\dagger}, \tag{1}$$

where the operators $\{E_l, l = 0, \ldots, D^2 - 1\}$ form a basis for the space of operators $\mathcal{H}_D$. The complex numbers $\chi_{l,l}$ form the so-called χ matrix of the map. The χ matrix is obviously dependent on the operator basis. Without loss of generality, we can take this basis to be orthogonal, that is, to be such that $\mathrm{Tr}[E_l^{\dagger} E_{l'}] = D\delta_{l,l'}$ [12]. It is simple to show that the map preserves the hermiticity of ρ if and only if the χ matrix is Hermitian itself (i.e., if $\chi_{l,l'} = \chi_{l',l}^{*}$). Moreover, the map Λ is trace-preserving if and only if the condition $\sum_{l,l'} \chi_{l,l'} E_{l'}^{\dagger} E_l = I$ is satisfied. In such a case it is simple to count the number of independent real parameters defining the quantum map, which turns out to be $D^4 - D^2$ (with the trace preserving condition implying a reduction of the number of parameters in D^2 and also implying that the condition $\sum_l \chi_{l,l} = 1$ must be satisfied).

We remark that this description is valid for any linear map. For the case of Hermitian maps it is possible to uncover further structure. In such a case the χ matrix can be diagonalized by a unitary transformation. In matrix notation we can write $\chi = B^{\dagger} S B$, where S is a diagonal matrix with real eigenvalues. The columns of the unitary matrix B define the eigenvectors: The mth component of the lth eigenvector $\bar{b}_l$ is $(\bar{b}_l)_m = B_{m,l}$. By using this notation it is evident that the elements of the χ matrix can be obtained as $\chi_{l,l'} = \bar{b}_l^{\dagger} S \bar{b}_{l'} = \sum_m B_{m,l}^{*} S_{m,m} B_{m,l'}$.

Some simple but useful results follow from this expression. Replacing it in the original formula for the map given in Eq. (1) we obtain the following alternative expression for an arbitrary linear Hermitian map:

$$\mathbf{\Lambda}(\rho) = \sum_k S_{k,k} A_k \rho A_k^{\dagger}, \tag{2}$$

where the operators A_k form an orthonormal basis defined as $A_k = \sum_l B_{k,l}^{*} E_l$. (The orthonormality of A_k follows from the fact that these operators are a linear combination of the original E_l with coefficients that are elements of a unitary matrix.) It is worth noticing that the coefficients S_{kk}, which are the eigenvalues of the χ matrix, are necessarily real but can be either positive or negative. This representation for the quantum map is closely related to the so-called Kraus representation (which is obtained only if the eigenvalues S_{mm} are all positive, which is in turn valid for the case of CP maps only, as discussed below). In fact, Eq. (2) is a generalization of the Kraus representation valid for any linear Hermitian map.

More generally, these expressions make evident that an arbitrary linear Hermitian map can always be written as the difference between two CP maps [13] (since any matrix S can be expressed as the difference between two positive matrices).

A. The χ matrix of completely positive maps

Using the above results, we can derive some properties for the χ matrix of completely positive maps (i.e., when the map Λ and any trivial extension of it to a bigger Hilbert space preserve positivity). Since the matrix S is positive, it is clear that matrix elements χ are obtained as the inner product between two eigenvectors $\bar{b}_l$ defined through the positive matrix S,

$$\langle \bar{b}_l, \bar{b}_{l'} \rangle \equiv \bar{b}_l^{\dagger} S \bar{b}_{l'}.$$

From this observation we can conclude the following: First, it is evident that diagonal elements must be positive (i.e., $\chi_{l,l} \geqslant 0 \ \forall \ l$). Moreover, as the inner product satisfies the Cauchy-Schwarz inequality, we can obtain the following relation between diagonal and off-diagonal elements of the χ matrix:

$$|\chi_{l,l'}|^2 \leqslant \chi_{l,l} \, \chi_{l',l'}. \tag{3}$$

This means that, for any CP map, the diagonal elements of the χ matrix are always nonnegative and that they bound the corresponding off-diagonal elements. These two simple results are quite significant and they will prove very useful later on. Below, we will derive another relation between diagonal and off-diagonal coefficients for the χ matrix valid for positive (not necessarily CP) maps. In this way we will be also able to establish some conditions to distinguish these two important classes of maps.

III. TWIRLING OF A MAP, AND SAMPLING OF A TWIRL

The action of twirling a map is depicted in Fig. 1. We have a quantum process characterized by a map Λ that acts on a system (for example, a quantum information processor) originally prepared in an arbitrary state $|\phi_0\rangle$, as depicted in Fig. 1(a). We twirl the map by applying an operator U before the map, and an operator $U^{\dagger}$ after, as in Fig. 1(b). Typically the twirling is considered as the average of this over different elements U, resulting in a net map Λ^{T}, the twirled map. Different families of U's will return different types of twirl.

In particular, we are initially interested in the Haar twirl,

$$\mathbf{\Lambda}^{\mathrm{HT}}(\rho) = \int dU U^{\dagger} \mathbf{\Lambda}(U \rho U^{\dagger}) U, \tag{4}$$

where dU denotes the unitarily invariant Haar measure on U(D).

There is a version of this twirl where the average is over the Haar measure in state space,

$$\langle \phi_0 | \mathbf{\Lambda}^{\mathrm{HT}}(|\phi_0\rangle\langle\phi_0|) | \phi_0 \rangle = \int d\psi \langle \psi | \mathbf{\Lambda}(|\psi\rangle\langle\psi|) | \psi \rangle. \tag{5}$$

The relation between the two is straightforward if we notice that if U is randomly drawn according to the Haar measure

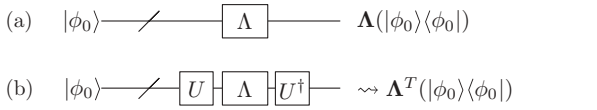

FIG. 1. Circuit representation of (a) the action of a map Λ and (b) the action of the map, now twirled by U.

on operator space, then $|\psi\rangle = U|\phi_0\rangle$ corresponds to the Haar measure on vector space—for any arbitrary fixed state $|\phi_0\rangle$.

There are several previous results concerning the Haar twirl, in its forms in both operator space [3,14] and in state space [15]. Summarizing this literature, we limit ourselves to state the following general mathematical formula:

$$\int dU \mathrm{Tr}[A_1 U^\dagger B_1 U A_2 U^\dagger B_2 U] = \frac{\mathrm{Tr}[A_1 A_2]}{D^2-1}\left(\mathrm{Tr}[B_1]\mathrm{Tr}[B_2] - \frac{\mathrm{Tr}[B_1 B_2]}{D}\right) + \frac{\mathrm{Tr}[A_1]\mathrm{Tr}[A_2]}{D^2-1}\left(\mathrm{Tr}[B_1 B_2] - \frac{\mathrm{Tr}[B_1]\mathrm{Tr}[B_2]}{D}\right) \quad (6)$$

for any operators A_1, A_2, B_1, B_2 in $\mathcal{H}_D$.

Given this and explicitly using the trace-preserving condition, the χ-matrix elements can be expressed as the outcome of a twirl,

$$\frac{D\chi_{l,l'} + \delta_{l,l'}}{D+1} = \int d\psi \langle\psi|\mathbf{\Lambda}(E_l^\dagger|\psi\rangle\langle\psi|E_{l'})|\psi\rangle, \quad (7)$$

as already stated in [8].

A. The χ matrix of positive (but not necessarily CP) maps

Equation (7) is valid for any map Λ under study. In particular, for processes that take positive operators into positive operators, Eq. (7) defines a valid inner product

$$\langle E_l, E_{l'}\rangle \equiv \int d\psi \langle\psi|\mathbf{\Lambda}(E_l^\dagger|\psi\rangle\langle\psi|E_{l'})|\psi\rangle.$$

Notice that we have $\langle E_l, E_l\rangle = (D\chi_{l,l}+1)/(D+1) \geqslant 0$. This implies that, for a positive (but not necessarily CP) map, we have that the diagonal elements of the χ matrix can be negative but only up to an exponentially vanishing value: $\chi_{l,l} \geqslant -1/D$. Also, notice that $\langle E_l, E_l\rangle$ is a survival probability: the probability of the system remaining in its initial state after applying the twirled map Λ^{HT} to it. Therefore, $(D\chi_{l,l}+1)/(D+1) \leqslant 1$, which implies $\chi_{l,l} \leqslant 1$.

Moreover, again using the Cauchy-Schwarz inequality on this inner product, we obtain that for $l \neq l'$

$$|\chi_{l,l'}|^2 \leqslant \chi_{l,l}\,\chi_{l',l'} + \frac{\chi_{l,l} + \chi_{l',l'}}{D} + \frac{1}{D^2}. \quad (8)$$

So for large systems where we can consider $D \gg 1$, the off-diagonal matrix elements are effectively bound by the diagonal ones. These bounds also suggest that non-CP but positive processes are "exponentially close" to CP ones. This is an interesting result in the framework of open quantum systems, where there are still important discussions about what mathematical conditions a physical map should fulfill [13,16,17].

B. (Approximate) sampling of a twirl

Equation (7) already demonstrates the usefulness of twirls in extracting the elements of the χ matrix. This is indeed what lies at the heart of the methods developed in [3–6, 8–10]. It is evident then that we will need to implement the twirl experimentally (in either operator or state space). Unfortunately, as we will see, the number of elements in the twirls that are of our interest is infinite or grows exponentially with $\ln(D)$ [18]. Thus implementing the twirl perfectly is a nonscalable task. However, it was initially suggested in [3,4] that we can approximate the twirl by sampling randomly over the family of U's, say M times, as depicted in Fig. 2. In the case of a state twirl, the approach would be the same, since in practice a state twirl results from implementing a series of operations (which would take the place of the U's) on a convenient initial state $|\phi_0\rangle$.

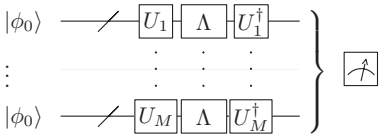

FIG. 2. Circuit representation of the twirling of Λ, approximated by sampling M times over the elements that constitute the family of twirl operators (the U's). Each time the system is prepared in the same initial state $|\phi_0\rangle$. The average of these M measurements will retrieve the desired probabilities.

If we are interested in measuring the probability of finding the system in any given state, this outcome will be a boolean variable retrieved with a standard deviation $\sigma \leqslant 1/\sqrt{M}$ (following the central limit theorem with $M \to \infty$), so for a desired precision ϵ we must have $M \geqslant \epsilon^{-2}$. On the other hand, the Chernoff bound tells us that for a desired precision ϵ and an error probability $\delta \ll 1$, we must have $M \geqslant \ln(2/\delta)/(2\epsilon^2)$, which is a stronger requirement when $\delta < 2e^{-2}$. This is a bound to the error probability and not to the error itself; however, it is rigorous for arbitrary M. In any case M should satisfy both conditions [19]. Since M is independent of the size of the system, this ensures the scalability of the experimental implementation if each of the M realizations themselves can be implemented efficiently. This holds of course unless the targeted probabilities are expected to be of the $O(1/\sqrt{D})$, in which case the estimation of each probability would require an exponentially large number of realizations. However, this would be the case of a process close to a random channel, and usually they are of no interest in quantum information and/or in relatively controllable quantum systems.

Finally, we note that we could separate the average of binary outcomes (the result of projective measurements) required to determine the probability for an experiment with a fixed twirl operator U from the average of experiments with different U's. This is useful in cases where repeating an experiment with a fixed U is trivial compared to running a new one with a different twirl operator. In this case, other interesting bounds to the error can be applied, as for example in the experimental work in [20].

In what follows we restrict ourselves to a Hilbert space that is an n-fold tensor product of a two-level system space, so $D = 2^n$. Moreover, we work with a specific set of operators $\{E_l\}$: the generalized Pauli operators (also called the product operator basis). We will specifically denote them as $\{P_l\}$, $P_l = \bigotimes_{j=1}^n P_l^{(j)}$. Each $P_l^{(j)}$ is an element of the Pauli group $\{I, \sigma_x, \sigma_y, \sigma_z\}$ for the jth qubit. $P_0 = \mathbb{I}$ is the identity operator in $\mathcal{H}_D$, and for $l > 0$ at least one factor in each P_l is a Pauli matrix. Notice that $P_l^\dagger = P_l$ and that $\mathrm{Tr}[P_l P_{l'}] = D\delta_{l,l'}$ indeed. From now on, the χ-matrix elements will be always associated to this basis.

IV. METHODS USING A FULL SPACE TWIRLING OF THE MAP UNDER STUDY

In this section we start by studying the methods utilizing a full twirl over U(D). If the twirl depicted in Fig. 2 is over U(D), the survival probability is the average fidelity of the original map Λ [5,8,10]. This is in fact Eq. (5), which is the definition of average fidelity of a quantum channel Λ [21], $F(\Lambda)$.

The tomographic methods in [8,10] are actually presented not in terms of twirl operators but rather in terms of the states of mutually unbiased bases (MUBs): $\{|\psi_{J,m}\rangle, J = 0,\ldots,D;\ m = 1,\ldots,D\}$. Here we introduce their equivalents using twirls in operator space. Nevertheless, further analysis will in turn lead to a slight preference toward the former one.

We rely on [5] to establish the equivalence between the Haar twirl in operator space and the Clifford twirl in U(D) for $D > 2$ (and we will explicitly prove it for $D = 2$ in Sec. V). On the other hand, the equivalence between a Haar twirl in state space and a twirl using MUB states, for dimensions that are powers of prime numbers, is presented in [22]. Altogether, we can write

$$\begin{aligned}&\langle\phi_0|\mathbf{\Lambda}^{\mathrm{HT}}(|\phi_0\rangle\langle\phi_0|)|\phi_0\rangle\\&\quad=\frac{1}{|\mathcal{C}|}\sum_{l=1}^{|\mathcal{C}|}\langle\phi_0|\mathcal{C}_l^\dagger\mathbf{\Lambda}(\mathcal{C}_l|\phi_0\rangle\langle\phi_0|\mathcal{C}_l^\dagger)\mathcal{C}_l|\phi_0\rangle\\&\quad=\frac{1}{D(D+1)}\sum_{J,m}\langle\psi_{J,m}|\mathbf{\Lambda}(|\psi_{J,m}\rangle\langle\psi_{J,m}|)|\psi_{J,m}\rangle,\end{aligned}\tag{9}$$

where the $\mathcal{C}_l$ are the Clifford operators in U(D) and $|\phi_0\rangle$ is an arbitrary fixed state. Both these twirls imply the same cost, as preparing MUB states starting from the computational basis and implementing the $\mathcal{C}_l$ require the same resources: $O(n^2)$ one-qubit and two-qubit gates [10,23,24]. And again, the number of Clifford operators $|\mathcal{C}|$ scales exponentially with $\ln(D)$, as does the number of MUB states (so in both cases we will resort to sampling the twirl).

In [8,10] it was shown how to selectively measure any diagonal χ-matrix element using an MUB twirl. There is an equivalent to this using a Clifford twirl in U(D). As presented in [8], if we implement an intermediary extra gate P_l before completing the twirl (see Fig. 3), the survival probability is

$$\mathrm{Tr}\big[|\phi_0\rangle\langle\phi_0|\mathbf{\Lambda}_l^{\mathrm{HT}}(|\phi_0\rangle\langle\phi_0|)\big]=\frac{D\chi_{l,l}+1}{D+1}.\tag{10}$$

This can be proven straightforwardly from Eq. (7).

We are thus able to measure efficiently one $\chi_{l,l}$ at a time (*selective* efficient quantum process tomography, SEQPT [8]). However, we can modify the protocol to automatically select and retrieve the largest $\chi_{l,l}$: the coefficients such that $\chi_{l,l} \geqslant 2/M$. The strategy goes as follows.

We first revisit the method as presented in [8] for an MUB twirl. As depicted in Fig. 4(a), we consider a single experiment where the system is prepared in a randomly chosen MUB state

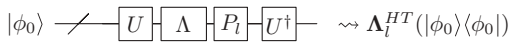

FIG. 3. Circuit representation of the action of a map Λ_l with $\Lambda_l(\rho) = P_l\mathbf{\Lambda}(\rho)P_l$, twirled by U.

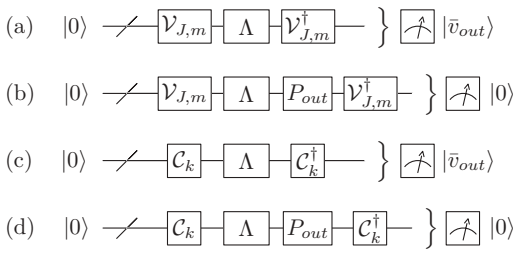

FIG. 4. Circuit representation of equivalent schemes to determine the largest $\chi_{l,l}$, by considering pairs of experiments.

$|J,m\rangle = \mathcal{V}_{J,m}|0\rangle$. $\mathcal{V}_{J,m}$ represents the change of basis operation between the computational state $|0\rangle$ and the targeted MUB state. We measure at the end the state in the computational (Zeeman) basis, obtaining then an n-bit string $|\bar{v}_{\mathrm{out}}\rangle$ (where $\bar{v}_{\mathrm{out}}$ is a boolean vector of length n that labels the states as binary numbers). Considering that $\mathcal{V}_{J,m}|\bar{v}_{\mathrm{out}}\rangle = |J,m'\rangle$ is just another state of the MUB, and that there are D possible Pauli operators that take $|J,m\rangle$ to $|J,m'\rangle$ (up to a global phase) [10], we can regard this experiment as equivalent to the one in Fig. 3, but where now we have D possible Pauli operators playing the role of the intermediary P_l.

To gain further insight into the mechanism of this result, we recall these dynamics using the stabilizer formalism [23]. We describe the state $|\bar{v}_{\mathrm{out}}\rangle$ with the subset $\mathcal{B}_Z$ formed by the n commuting Pauli operators $\{\sigma_z^{(1)},\ldots,\sigma_z^{(n)}\}$ and a string $\bar{s}_{\mathrm{out}}$ of n signs, ±1, corresponding to the eigenvalues of $|\bar{v}_{\mathrm{out}}\rangle$ for that subset. These n operators generate the maximally commuting (Abelian) group of D Pauli operators that stabilize the computational basis. On the other hand, the state $|0\rangle$ is described by $\mathcal{B}_Z$ with a string $\bar{s}_0$ of all $+1$ signs. The action of $\mathcal{V}_{J,m}$ on $|0\rangle$ is equivalent to changing $(\mathcal{B}_Z,\bar{s}_0)$ to $(\mathcal{B}_J,\bar{s}_0)$, where now $\mathcal{B}_J$ is another subset of n commuting Pauli operators—the generators of the group that stabilizes the D states corresponding to the MUB labeled by J. Also, the action of $\mathcal{V}_{J,m}$ on $|\bar{v}_{\mathrm{out}}\rangle$ is equivalent to changing $(\mathcal{B}_Z,\bar{s}_{\mathrm{out}})$ to $(\mathcal{B}_J,\bar{s}_{\mathrm{out}})$. We now use that the state $(\mathcal{B}_J,\bar{s}_{\mathrm{out}})$ can be thought as the result of a Pauli operator P_{out} acting on $(\mathcal{B}_J,\bar{s}_0)$, which leaves us with the scheme depicted in Fig. 4(b). P_{out} must fulfill the requisite of commuting (anticommuting) with the Pauli operators in $\mathcal{B}_J$ that have a corresponding $+1$ (-1) in $\bar{s}_{\mathrm{out}}$. We express this condition as the commutation relations

$$[P_{\mathrm{out}},\mathcal{V}_{J,m}^\dagger\sigma_z^{(j)}\mathcal{V}_{J,m}]_\pm = 0,\quad j = 1,\ldots,n,\tag{11}$$

where the $[,]_\pm$ stands for commutator or anticommutator, depending on the signs of $\bar{s}_{\mathrm{out}}$. But, as already stated before, there will be D possible candidates for the intermediary P_{out}. This can be seen as follows. First, we notice that Eq. (11) can be rewritten as $[P'_{\mathrm{out}},\sigma_z^{(j)}]_\pm = 0$, where we have defined

$$P'_{\mathrm{out}} \equiv \mathcal{V}_{J,m}P_{\mathrm{out}}\mathcal{V}_{J,m}^\dagger.\tag{12}$$

It is easy to see that the possible P'_{out} will be the tensor products that have I or σ_z for the qubits that have $+1$ in $\bar{s}_{\mathrm{out}}$ and σ_x or σ_y for the other qubits. There are $D = 2^n$ of these products, and then the actual P_{out}'s could be obtained by inverting Eq. (12). Therefore, we are indeed left with an experiment equivalent

to the one in Fig. 3 but with D possible intermediary Pauli operators.

The key here is that for two different sets $\mathcal{B}_{J_1}$ and $\mathcal{B}_{J_2}$ corresponding to two different MUBs, there can be only one P_{out} in common for both. This is because the $D+1$ subsets $\mathcal{B}_J$ are obtained by partitioning the D^2-1 nonidentity Pauli operators into $D+1$ different subsets of $D-1$ commuting operators. The $\mathcal{B}_J$ are then the generators of these subsets (plus the identity). Given their properties, any two $\mathcal{B}_{J_1}$ and $\mathcal{B}_{J_2}$, plus commutation relations with them [Eq. (11)], define a unique operator P_{out} [10]. Moreover, given the nature of the operators involved (Pauli gates and the operations involved in the change of basis for MUBs), and the number of equations (n), P_{out} can be established efficiently [10].

Therefore, if we consider together two experiments $(J_1,m_1,\bar{s}_{\text{out}}^{(1)})$ and $(J_2,m_2,\bar{s}_{\text{out}}^{(2)})$ [each like in Fig. 4(b), with $J_1 \neq J_2$], there will be only one possible intermediary Pauli gate compatible with both experiments, and it can be computed in a scalable way. In practice, we will perform M experiments, and analyzing all the possible $M(M-1)/2$ pairs we will establish the intermediary Pauli gates that have occurred at least twice, which will be at most $O(M^2)$. Then, we just count the number of experiments M_+ where those operators have potentially occurred among the D possible choices. The corresponding $\chi_{l,l}$ can be estimated as $(D\chi_{l,l}+1)/(D+1) = M_+/M$. Notice that $\chi_{l,l}$ is then estimated with a standard deviation $\leqslant 1/\sqrt{M}$. We also recall that $\sum_l \chi_{l,l} = 1$, so we can use this to estimate altogether the magnitude of the smaller coefficients.

This strategy can be also applied using Clifford gates acting on the initial state instead of using MUB states, as depicted in Fig. 4(c). Again, we use the stabilizer formalism as described before. Since the Pauli group is the normalizer of the Clifford group, indeed $\mathcal{C}P_k\mathcal{C}^\dagger \cong P_{k'}$ (where $\cong$ means equal up to a global phase). So this means that the action of $\mathcal{C}_j$ on a state is equivalent to changing $(\mathcal{B}_Z,\bar{s})$ to $(\mathcal{B}_P,\bar{s})$, where now $\mathcal{B}_P$ is another subset of n commuting Pauli operators. Again we use that the state $(\mathcal{B}_P,\bar{s}_{\text{out}})$ can be thought as the result of a Pauli operator P_{out} acting on $(\mathcal{B}_P,\bar{s}_0)$, which now leaves us in the scheme depicted in Fig. 4(d). Again, there are D possible operators that fulfill the requisite of commuting (anticommuting) with the Pauli operators in $\mathcal{B}_P$ that have a corresponding $+1$ (-1) in $\bar{s}_{\text{out}}$. The argument is completely analogous to the one for the MUB twirl.

We thus resort again on combining two experiments. However, the case of two Clifford twirl experiments is not as simple as the MUB twirl one. It no longer holds that given two experiments there is one single possible intermediary Pauli gate, because two different Clifford gates may map $\mathcal{B}_Z$ to two subsets $\mathcal{B}_{P_1}$ and $\mathcal{B}_{P_2}$ that generate two Pauli subgroups that have some operators in common. So not every pair of experiments, even if $\mathcal{C}_1 \neq \mathcal{C}_2$, will be useful toward establishing the $\chi_{l,l}$ above the threshold of $2/M$. In practice, we should determine the $2 \times n$ operators $\mathcal{C}_k\sigma_z^{(j)}\mathcal{C}_k^\dagger$ (where $k = 1,2$ are two randomly chosen Clifford gates) and check whether they constitute two independent sets of generators. If that is the case, then there is indeed a unique intermediary Pauli gate, as it is always the case with the MUB twirl. And thanks to the Gottesman-Knill theorem, this can be done efficiently with a classical computer.

To compare both methods, we consider the probability of successfully determining a unique intermediary Pauli gate given two different experiments drawn from a pool of M experiments. In the case of the MUB twirl, the probability of success is $\mathcal{P}_{\text{MUB}} = D/(D+1)$, since there are $D+1$ possible values of J_1, and having randomly withdrawn one, there are D different ones we could withdraw for J_2.

For the Clifford twirl case, this probability can be calculated as the probability that, given two randomly chosen maximal Abelian Pauli subgroups, the only common element in both groups is the identity, up to a phase. To compute this probability we proceed as follows. We fix the first maximal Abelian group and then compute the probability of adding one by one the operators belonging to the second group. The fixed group has, up to a phase, $D-1$ nonidentity Pauli operators. If we randomly choose a nonidentity Pauli operator, namely P_1, what is the probability of it not belonging to the fixed group? It is straightforward to see that this probability is $\frac{D^2-D}{D^2-1}$. Now, from the Pauli operators that commute with P_1, what is the probability of picking one Pauli operator P_2 that does not belong to the first group? Again, there are a total of $D^2/2-2$ Pauli operators which commute with P_1 and are neither P_1 nor the identity, but $D/2-1$ of those belong to the first group. So the probability of this happening is $\frac{D^2/2-1-D/2}{D^2/2-2}$. We proceed in the same way, computing the probability of picking a Pauli operator that does not belong either to the first group nor to the group generated by the previously chosen operators. The product of all those probabilities is the probability $\mathcal{P}_C$ of having only one intermediary Pauli operator given two Clifford twirl experiments:

$$\mathcal{P}_C = \prod_{j=0}^{n-1} \frac{D^2/2^j - 2^j - D/2^j}{D^2/2^j - 2^j}. \tag{13}$$

As shown in Fig. 5, this probability is smaller but asymptotically equivalent to $\mathcal{P}_{\text{MUB}}$. For the experiments that are being done nowadays, with only a few qubits, the MUB twirl still requires much fewer experimental runs to obtain the larger coefficients.

At this point we can conclude that the method introduced in [8] is indeed the most practical one and that it can retrieve the largest diagonal elements of the χ matrix in the sense that

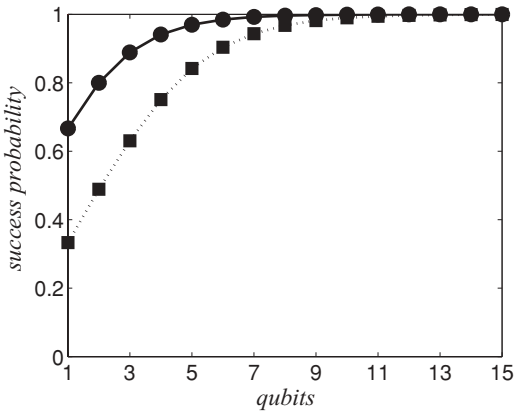

FIG. 5. Success probability of the two methods: ● $\mathcal{P}_{\text{MUB}}$, using an MUB twirl; ■ $\mathcal{P}_C$, using a Clifford twirl.

they are above the threshold of $2/M$, for M realizations of the MUB twirl. This is done in a scalable way, and with a standard deviation $\leqslant 1/\sqrt{M}$.

This protocol has been experimentally implemented recently, with photons, to characterize maps on a one-qubit space [20].

Nevertheless, although efficient, the method demands the errorless implementation of the $\mathcal{V}_{J,m}$ gates (or of the equally demanding Clifford gates)—at least relatively errorless compared with any errors in the implementation of Λ. If we have a functional quantum device that implements the Hadamard, phase, and controlled-NOT (CNOT) gates (which are the gates required to implement the Clifford gates [24] or the MUB states [10]) and the Pauli operators with enough accuracy, we will be in position to study more complex maps with twirls in U(D). If however we are still aiming to study gates and sequences whose complexity is comparable to the one of a Clifford gate in U(D), this method is unsuitable. In this case, a more practical alternative arises from the combination of the methods presented in [6] and [9] (which will be the object of the next section). This proposal allows us to establish the diagonal elements of the χ matrix coarse-grained in direction. Indeed, this information is particularly useful when seeking information for quantum error correction codes, where the particular type of error (σ_x, σ_y, or σ_z) is irrelevant.

The following method is experimentally quite less demanding, since it requires a twirl in $\mathrm{U}(2)^{\otimes n}$ rather than one in $\mathrm{U}(2^n)$. On the other hand, as we will see, it assumes a certain structure in the map under study. An example of such a scenario was explicitly shown in the experimental work in [9], using a liquid-state nuclear magnetic resonance (NMR) processor with four qubits. A relatively large number of qubits easily shows the significative difference between required resources; to our knowledge, this is the largest number of qubits on which a complete [20,25] or partial [6] quantum process characterization has been attempted.

V. METHODS USING A ONE-QUBIT TWIRLING OF THE MAP UNDER STUDY

In this section we concentrate on methods based on a one-qubit twirling of a map [4,6,9]. That is, the twirl is a tensor product of twirling operators U acting on each qubit. The two protocols presented in [6] and [9] can be actually merged into one. We review the compact method by proving it all together, which also shows clearly the simplicity and economy of its implementation (since both [6] and [9] include their experimental implementation).

In Sec. III we highlighted the promising role of the Haar twirl. This for example motivated the first works [3,4]. However, as mentioned before, the work by Dankert *et al.* [5] pointed out an equivalence between a Haar twirl and a Clifford twirl.

Rather than starting from the Haar twirl and crossing over to the Clifford twirl, we will work directly with the Clifford gates and prove everything from scratch. For this we will use that the Clifford operators can in turn be decomposed into Pauli operators (the normalizer of the Clifford group) and the so-called symplectic operators (the resulting quotient group).

We will follow the notation of [6]. Each index l carries the following information: w, ν_w, $\mathbf{i}_w$. w is the Pauli weight of P_l, that is, how many of the factors in P_l are nonidentity. The index ν_w in $\{1,\ldots,\binom{n}{w}\}$ counts the number of distinct ways that w nonidentity Pauli operators can be distributed over the n factor spaces. The index $\mathbf{i}_w$ is a vector of length w of the form $\mathbf{i}_w=(i_1,i_2,\ldots,i_w)$ with each component being $1=x$, $2=y$, or $3=z$ to denote which Pauli matrix occupies that respective factor position in the tensor product forming P_l. There are 3^w of these $\mathbf{i}_w$ for given w and ν_w.

We start first with a Pauli twirl (PT) of the map. Thus Λ becomes

$$\mathbf{\Lambda}^{\mathrm{PT}}(\rho)=\frac{1}{D^2}\sum_{m=0}^{D^2-1}P_m\mathbf{\Lambda}(P_m\rho P_m)P_m \tag{14}$$

$$=\frac{1}{D^2}\sum_{m=0}^{D^2-1}\sum_{l,l'}^{D^2-1}\chi_{l,l'}P_mP_lP_m\rho P_mP_{l'}P_m \tag{15}$$

$$=\sum_{l=0}^{D^2-1}\chi_{l,l}P_l\rho P_l. \tag{16}$$

This result was proven in [24]. It can be also seen as follows: For $l=l'$, $P_mP_lP_m\rho P_mP_lP_m=P_l\rho P_l$ since each P_l either commutes or anticommutes with each P_m. And if $l=l'$, for each jth factor in which they differ, we have $P_m^{(j)}P_l^{(j)}P_m^{(j)}\rho P_m^{(j)}P_{l'}^{(j)}P_m^{(j)}=\pm P_l^{(j)}\rho P_{l'}^{(j)}$, with each sign happening for half of the four possible $P_m^{(j)}$. Thus they cancel out in the sum.

We consider now a Symplectic one-qubit twirl (S1T) of the form

$$\mathbf{\Lambda}^{\mathrm{S1T}}(\rho)=\frac{1}{3^n}\sum_{m=1}^{3^n}S_m^\dagger\mathbf{\Lambda}(S_m\rho S_m^\dagger)S_m, \tag{17}$$

$$S_m=\bigotimes_{j=1}^{n}S_m^{(j)}, \tag{18}$$

where each $S_m^{(j)}$ is an element of the set given by $\{\exp[-i(\pi/4)\sigma_p],\ p=x,y,z\}$. It is straightforward to show that

$$\frac{1}{3}\sum_{m=1}^{3}S_m^{(j)\dagger}\sigma_jS_m^{(j)}\rho S_m^{(j)\dagger}\sigma_jS_m^{(j)}=\frac{\sigma_x\rho\sigma_x+\sigma_y\rho\sigma_y+\sigma_z\rho\sigma_z}{3},$$

so after a Clifford (Pauli+Symplectic) one-qubit twirl (C1T) [26] we get

$$\mathbf{\Lambda}^{\mathrm{C1T}}(\rho)=\frac{1}{3^n}\sum_{m=0}^{3^n}S_m^\dagger\mathbf{\Lambda}^{\mathrm{PT}}(S_m\rho S_m^\dagger)S_m \tag{19}$$

$$=\sum_{w=0}^{n}\sum_{\nu_w}^{\binom{n}{w}}\frac{\chi^{\mathrm{col}}_{w,\nu_w}}{3^w}\left(\sum_{\mathbf{i}_w}P_{w,\nu_w,\mathbf{i}_w}\rho P_{w,\nu_w,\mathbf{i}_w}\right), \tag{20}$$

where the collective coefficients $\chi^{\mathrm{col}}_{w,\nu_w}$ are just the diagonal χ-matrix coefficients $\chi_{l,l}$, re-labeled $\chi_{w,\nu_w,\mathbf{i}_w}$, after disregarding (averaging over) the information given by $\mathbf{i}_w$:

$$\chi^{\mathrm{col}}_{w,\nu_w}\equiv\sum_{\mathbf{i}_w}\chi_{w,\nu_w,\mathbf{i}_w}. \tag{21}$$

This is so far what was presented in [6], which can also be proven as in [4,9] using a different set of tools to handle the Clifford twirl as a Haar twirl [4,5,14].

Consider the computational state basis $|\bar{v}_h\rangle$, where $\bar{v}_h$ is a boolean vector of length n and Hamming weight h. (The Hamming weight h of a computational state is just the number of ones appearing in its binary representation.) The first result we can obtain is that the fidelity of a state $|\bar{v}_h\rangle$ undergoing this transformation is independent of the actual state,

$$f(\Lambda^{\mathrm{C1T}},|\bar{v}_h\rangle) = \mathrm{Tr}[|\bar{v}_h\rangle\langle\bar{v}_h|\mathbf{\Lambda}^{\mathrm{C1T}}(|\bar{v}_h\rangle\langle\bar{v}_h|)] \tag{22}$$

$$= \sum_{w=0}^{n}\sum_{\nu_w}^{\binom{n}{w}} \frac{\chi^{\mathrm{col}}_{w,\nu_w}}{3^w}\left(\sum_{\mathbf{i}_w}^{3^w} |\langle\bar{v}_h|P_{w,\nu_w,\mathbf{i}_w}|\bar{v}_h\rangle|^2\right) \tag{23}$$

$$= \sum_{w=0}^{n}\sum_{\nu_w}^{\binom{n}{w}} \frac{\chi^{\mathrm{col}}_{w,\nu_w}}{3^w}\left(\sum_{\mathbf{i}_w}^{3^w} \langle 0|P_{w,\nu_w,\mathbf{i}_w}|0\rangle|^2\right) \tag{24}$$

$$= \sum_{w=0}^{n}\sum_{\nu_w}^{\binom{n}{w}} \frac{\chi^{\mathrm{col}}_{w,\nu_w}}{3^w}. \tag{25}$$

To go from (23) to (24), we only need to realize that any computational state $|\bar{v}_h\rangle$ is a result of applying a Pauli operator $P_X^{\bar{v}_h}$ (that has σ_x where $\bar{v}_h$ has ones and nonidentity factors otherwise) to $|0\rangle$. This $P_X^{\bar{v}_h}$ will either commute or anticommute with $P_{w,\nu_w,\mathbf{i}_w}$ (and the $\pm$ will be absorbed by the modulus squared). The last equality (25) is obtained by realizing that the only nonidentity $P_{w,\nu_w,\mathbf{i}_w}$ that takes $|0\rangle$ back to it (up to a global phase) is the Pauli operator that has σ_z in all the positions indicated by ν_w (and thus only one of all the possible $\mathbf{i}_w$ given ν_w and w).

We must notice that although $f(\Lambda^{\mathrm{C1T}},|\bar{v}_h\rangle)$ is then equivalent to the average fidelity $F(\Lambda^{\mathrm{C1T}})$ of the process Λ^{C1T}, this is not the average fidelity of the process under study, namely $F(\Lambda) = (D\chi_{0,0}+1)/(D+1)$ (c.f. Sec. IV). However, this weaker twirl gives a different insight into the map structure. The first result we point out, presented in [6], is that we can obtain the diagonal elements of the χ matrix grouped by Pauli weight

$$p_w \equiv \sum_{\nu_w}^{\binom{n}{w}}\sum_{\mathbf{i}_w} \chi_{w,\nu_w,\mathbf{i}_w} = \sum_{\nu_w}^{\binom{n}{w}} \chi^{\mathrm{col}}_{w,\nu_w}. \tag{26}$$

The parameters p_w and $\chi^{\mathrm{col}}_{w,\nu_w}$ are just a coarse-graining of the diagonal elements of the χ matrix. The p_w relate to the probability $\mathrm{Prob}(\bar{v}_h,h)$ of obtaining any state $|\bar{v}_h\rangle$ with Hamming weight h when measuring the final state $\Lambda^{\mathrm{C1T}}(|0\rangle\langle 0|)$. We have

$$\mathrm{Prob}(\bar{v}_h,h) = \mathrm{Tr}[|\bar{v}_h\rangle\langle\bar{v}_h|\Lambda^{\mathrm{C1T}}(|0\rangle\langle 0|)]$$

$$= \sum_{w=0}^{n}\sum_{\nu_w}^{\binom{n}{w}} \frac{\chi^{\mathrm{col}}_{w,\nu_w}}{3^w}\left(\sum_{\mathbf{i}_w}^{3^w} \langle 0|P_{w,\nu_w,\mathbf{i}_w}|\bar{v}_h\rangle|^2\right).$$

For $\langle 0|P_{w,\nu_w,\mathbf{i}_w}|\bar{v}_h\rangle$ to be nonzero (i.e., ± 1), ν_w must indicate nonidentity factors at least where there are ones in $\bar{v}_h$ (so it must be $w \geqslant h$). Also the i_j in $\mathbf{i}_w$ must be $1 = x$ or $2 = y$ for the qubits with ones in $\bar{v}_h$, and $3 = z$ for the $w-h$ qubits that have zeros in $\bar{v}_h$ but have a nonidentity factor $P_{w,\nu_w,\mathbf{i}_w}$. There will be exactly 2^h of these operators for given w and ν_w, so

$$\mathrm{Prob}(\bar{v}_h,h) = \sum_{w=h}^{n}\sum_{\nu_w^*=1}^{\binom{n-h}{w-h}} \frac{2^h}{3^w}\chi^{\mathrm{col}}_{w,\nu_h+\nu_w^*}, \tag{27}$$

where ν_h indicates a $\chi^{\mathrm{col}}_{w,\nu_w}$ for Pauli operators that have a nonidentity factor for at least all the qubits whose corresponding component in $\bar{v}_h$ is a one. ν_w^* labels the $\binom{n-h}{w-h}$ coefficients with $w \geqslant h$ that fulfill this condition. If we now discard the "which qubit" information given by $\bar{v}_h$, summing over all the $\binom{n}{h}$ possibilities, then

$$\mathrm{Prob}(h) = \sum_{\bar{v}_h} \mathrm{Prob}\,(\bar{v}_h,h) \tag{28}$$

$$= \sum_{w=h}^{n} \frac{2^h}{3^w}\sum_{\nu_w^*=1}^{\binom{n-h}{w-h}}\sum_{\nu_h=1}^{\binom{n}{h}} \chi^{\mathrm{col}}_{w,\nu_h+\nu_w^*} \tag{29}$$

$$= \sum_{w=h}^{n} \frac{2^h}{3^w}\binom{w}{h}\left(\sum_{\nu_w=1}^{\binom{n}{w}} \chi^{\mathrm{col}}_{w,\nu_w}\right) \tag{30}$$

$$= \sum_{w=h}^{n} \frac{2^h}{3^w}\binom{w}{h} p_w. \tag{31}$$

In this way, all the p_w are related to the probabilities of measuring an outcome with Hamming weight h by an $n \times n$ matrix $R_{h,w} = \frac{2^h}{3^w}\binom{w}{h}$, as stated in [6].

We can also keep the "which qubit" information and use the probabilities $\mathrm{Prob}(\bar{v}_h,h)$ constructively to gain even more detail. This strategy was already suggested in [9] but more oriented to ensemble quantum information processors. We present it now in a different manner so it can be combined with the previous strategy.

Let us replace the descriptors w and ν_w by $\bar{v}_w$, a boolean vector of length n and Hamming weight w characterizing a Pauli operator P_l. $\bar{v}_w$ has a zero in the jth position if and only if $P_l^{(j)} = I$; otherwise it has a one. For example, the operator $\sigma_z^{(1)}\sigma_x^{(3)}$ for $n = 4$ qubits has $\bar{v}_2 = (1,0,1,0)$. There are of course $\sum_{w=0}^{n}\binom{n}{w} = 2^n = D$ of these vectors describing the P_l.

If we use Eq. (27) and start with the probability of having all the qubits flipped in the outcome, and go backward toward the survival probability (i.e., none of the qubits flipped), we find

$$\mathrm{Prob}(n) = \frac{2^n}{3^n}\chi^{\mathrm{col}}_{\bar{v}_n}, \tag{32a}$$

$$\mathrm{Prob}(\bar{v}_{n-1},n-1) = \frac{2^{n-1}}{3^{n-1}}\chi^{\mathrm{col}}_{\bar{v}_{n-1}} + \frac{2^{n-1}}{3^n}\chi^{\mathrm{col}}_{\bar{v}_n}, \tag{32b}$$

$$\mathrm{Prob}(\bar{v}_{n-2},n-2) = \frac{2^{n-2}}{3^{n-2}}\chi^{\mathrm{col}}_{\bar{v}_{n-2}} + \sum_{\bar{v}_{n-1}}\frac{2^{n-2}}{3^{n-1}}\chi^{\mathrm{col}}_{\bar{v}_{n-1}} + \frac{2^{n-2}}{3^n}\chi^{\mathrm{col}}_{\bar{v}_n}, \tag{32c}$$

$$\ldots.$$

So essentially we could determine $\chi^{\mathrm{col}}_{\bar{v}_n}$ using (32a), then insert it in (32b) and obtain the n possible $\chi^{\mathrm{col}}_{\bar{v}_{n-1}}$ from the different $\mathrm{Prob}(\bar{v}_{n-1},n-1)$, and then insert that in (32c), and so on and so forth. These equations define a triangular matrix that relates the probabilities $\mathrm{Prob}(\bar{v}_h,h)$ to the collective

coefficients $\chi^{\mathrm{col}}_{\bar{v}_w}$. Notice there is no need to perform different experiments to obtain the different probabilities: We only need to implement M realizations of the twirl and keep the outcome of the measurement for each of the realizations. This outcome should be a n-bit string indicating whether each jth qubit was found in $|0\rangle_j$ or $|1\rangle_j$.

The problem arises not in obtaining the experimental information, but in its posterior processing. The matrix given by Eqs. (32) is of size $D \times D$, therefore the cost of the processing would scale exponentially in n. For this strategy to work, it is key to relate it hierarchically to the determination of the p_w: The experimental information required is the same and can be obtained efficiently by sampling. The idea goes as follows. If we are analyzing a map Λ that is close to the identity (a noise channel) or a quantum gate involving a few qubits (typically one or two), then we would expect that above a certain cutoff Pauli weight w_{co}, the p_w will be null. This is a reasonable expectation: Since $\sum_{w=0}^{n} p_w = 1$ (the trace-preserving condition), the p_w cannot all be arbitrarily large, and thus it will be possible to bound the coefficients above the cutoff by a negligible amount. In this scenario, the matrix relating the $\mathrm{Prob}(\bar{v}_h,h)$ with the $\chi^{\mathrm{col}}_{\bar{v}_w}$ will have a size $M_{\mathrm{co}} \times M_{\mathrm{co}}$, $M_{\mathrm{co}} = \sum_{m=0}^{w_{\mathrm{co}}} \binom{n}{m}$, which scales polynomially in n [27]. There is a second caveat though. As explained in [6,9], respectively, the errors in determining the p_w or the $\chi^{\mathrm{col}}_{w,\nu_w}$ scale inefficiently with w, a consequence of the matrices relating them with the corresponding probabilities [Eqs. (31) and (32), respectively]. Although the measured probabilities will have a standard deviation $\leqslant 1/\sqrt{M}$, this error will propagate into the p_w or the $\chi^{\mathrm{col}}_{w,\nu_w}$ with a factor that grows polynomially with n but exponentially with w. Again, we must resort to neglecting the p_w after a certain cutoff. The system can be arbitrary large (arbitrary n), and as long as the p_w are negligible above a certain w_{co} (with w_{co} independent of or scaling efficiently with n) we will be able to obtain all the non-negligible $\chi^{\mathrm{col}}_{w,\nu_w}$ efficiently.

Notice that, in the previous section, the requirement that only a few ($\ll D$) coefficients $\chi_{w,\nu_w,\mathbf{i}_w}$ are non-negligible is not *a priori*. We can indeed run the protocol, efficiently, and arrive at this conclusion. However, the one-qubit twirling method poses a stronger condition, since the values of the $\chi^{\mathrm{col}}_{w,\nu_w}$ must respond to a hierarchy associated to their Pauli weights. Only then we can establish the Pauli weight cutoff and run the protocol [in particular, solve the system of equations (32)].

With the twirl in $\mathrm{U}(D)$ we obtain the coefficients directly with a standard deviation $\leqslant 1/\sqrt{M}$, while with the twirl in $\mathrm{U}(2)^{\otimes n}$ we only obtain probabilities $\mathrm{Prob}(\bar{v}_h,h)$ with standard deviations $\leqslant 1/\sqrt{M}$, which still need to be propagated in order to obtain the estimated error for the $\chi^{\mathrm{col}}_{w,\nu_w}$.

With the protocol of Sec. IV [8,10], the measurement of the largest $\chi_{l,l}$ can be done then more precisely, with no coarse-graining and with no restrictions on the map under study. Clearly, the protocol of this section [6,9] is quite less demanding, requiring the implementation of only $12n$ one-qubit gates instead of $O(n^2)$ one-qubit and CNOT gates. However, this advantage is counterbalanced: We have a more restricted and less precise tomographic method. In practice, nevertheless, the choice between the two will be given by the extent to which we can control our system experimentally.

Finally, we must notice that, in both approaches, the methods are universal in the sense that they do not require any prior knowledge on the specific dynamics of Λ. The protocol twirling in full space is valid for any linear Hermitian map, while the one with one-qubit twirling only has the extra requirement of having a structure with a cutoff Pauli weight. For an example of a characterization incorporating substantial prior knowledge of the dynamics or specific models for Λ, see [28].

VI. THE RELEVANCE OF THE DIAGONAL OF THE χ MATRIX

If we diagonalize the χ matrix, we will obtain the weights of an operator-sum representation, where the operators in the sum are the corresponding basis where the χ matrix is diagonal. Of course, this basis will not necessarily be the Pauli operator basis, but in principle a combination of them. Using the notation of Sec. II, take $\chi = R^\dagger S R$ to be the diagonalization of the χ matrix written in the Pauli operator basis. Let R be the change of basis, so

$$\mathbf{\Lambda}(\rho) = \sum_{m=0}^{D^2-1} S_{m,m} A_m \rho A_m^\dagger \quad A_m = \sum_{l=0}^{D^2-1} R^*_{m,l} P_l,$$

where the A_m form an orthonormal basis, but just as in an operator-sum representation, they are not necessarily unitary (otherwise any process would be unital) nor Hermitian. And, as we already mentioned, the $S_{m,m}$ are real but could be negative in principle. Thus in general neither the $\chi_{l,l}$ nor even the $S_{m,m}$ have a simple interpretation.

Nevertheless, despite the different ways of describing the process under study Λ in [5,6,8–10], in all the cases they determine specifically the diagonal elements of the χ matrix of the map in the generalized Pauli operator basis. Notice that either the one-qubit twirl or the full-space twirl implies a Pauli twirl (since the Pauli operators are a subgroup of the Clifford group in both cases) and that the Pauli twirl erases the information of the off-diagonal elements of the χ matrix. We ask then, what is the meaning of the diagonal? It was assumed in [6] that the p_w represented the probability of an operator of Pauli weight w happening in the process described by Λ. In [9], the $\chi^{\mathrm{col}}_{w,\nu_w}$ were regarded as indicators of the locality or range of the process, that is, the probability of an operator involving the qubits in ν_w happening. These are both quantities that are relevant to quantum error correction and fault-tolerant quantum computing.

Both these interpretations are fair when the χ matrix in the Pauli operator basis is approximately diagonal, at least block-diagonal in blocks characterized by w, ν_w. But that is not generally the case, in particular for maps that will be of our interest—such as quantum computing gates. For example, the CNOT gate for qubits a and b has a χ matrix with only a 4×4 nonzero block,

$$\chi_{\mathrm{CNOT}} = 0.25 \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{pmatrix},$$

corresponding to $P_l = I, \sigma_z^{(a)}, \sigma_x^{(b)}, \sigma_z^{(a)} \otimes \sigma_x^{(b)}$. Clearly, the off-diagonal coefficients carry critical information with equal weight, which for example differentiates the CNOT from a depolarizing channel with the same P_l.

Thus previous interpretations of p_w and $\chi^{\mathrm{col}}_{w,\nu_w}$ are arguable: We could even have in principle a process involving a set of qubits given by w, ν_w that has $\chi_{l,l'} \neq 0$ in that block but $\chi^{\mathrm{col}}_{w,\nu_w} = 0$ in the diagonal. However, as demonstrated in Secs. II A and III A, it is possible to draw a relation between the diagonal and off-diagonal elements of the χ matrix.

For CP maps, Eq. (3) guarantees that if either $\chi_{l,l} = 0$ or $\chi_{l',l'} = 0$, the off-diagonal $\chi_{l,l'}$ is null. And for positive maps in general, Eq. (8) gives us a bound that is exponentially close to this result. This a very powerful result, since once we have established the nonzero diagonal elements, in order to perform a full characterization we only need to worry about the off-diagonal elements that correspond to that resulting block. This hierarchization of the information could potentially allow for a complete quantum tomography of the process at a scalable cost—provided that the number of non-null matrix elements turns out to be $O(\mathrm{poly}(n))$.

It is in order here to point out though that the work in [8,10] also presents a strategy to measure the off-diagonal elements of the χ matrix. However, an ancillary qubit which is not twirled is required for this task. The ancilla is assumed to be error-free and outside the system we are looking to characterize. This does not imply an issue when it comes to scalability, since only one qubit ancilla is required for arbitrary D. Nonetheless, it puts this method in a different category regarding resources and assumptions when it comes to its implementation.

VII. CONCLUSIONS

By revisiting previous work [6,8,9] we have stated two scalable approaches for characterizing the diagonal elements of the χ matrix in the Pauli operator basis, for any arbitrary quantum process. We emphasize once more that the work in Secs. IV and V arises from the revision of these previous results and goes beyond, which we would like to summarize here: The general approach discussed in Sec. IV restates the method originally presented in [8], further clarifying its ability to measure the largest diagonal elements of the χ matrix together. We study this protocol by recognizing its familiarity with other twirling methods and present a natural alternative approach which, we conclude, is slightly less convenient if we work with only a few qubits. On the other hand, the approach discussed in Sec. V combines the two protocols originally presented in [6] and [9], by building and again proving both protocols, but simultaneously.

Furthermore, we have analyzed the two general approaches comparatively, establishing their advantages and disadvantages: While one is more powerful, the other is more realistic from the implementation point of view.

We have made the point that there are different ways of twirling that reproduce Eqs. (4) and (5). Moreover, we have shown that a deeper analysis may lead to advantages of one form of twirl over another, in particular for working with a small number of qubits. This is the case in Sec. IV when comparing the Clifford twirl and the MUB twirl in $\mathrm{U}(D)$. Another example of this, but twirling in $\mathrm{U}(2)^{\otimes n}$, can be found in [9,19], where it is shown that by carefully choosing the initial state of a twirl experiment, it is possible to reduce the total number of twirl operators from 12^n to 6^n.

On the other hand, in the light of Eqs. (3) and (8), our work establishes the relevance of the diagonal coefficients. We believe that this type of hierarchization of the information is key to achieve complete tomography in a scalable way. Since the number of parameters is indeed exponentially large, it is necessary to gather them or find relations among them, and then design protocols that will retrieve information about a whole group in one parameter.

The coarse-grained coefficients of Sec. V [Eqs. (21) and (26)] represent one example of grouping. When a sum of non-negative elements is null, we can conclude that all the elements in the sum are null. On the other hand, the bounding of the off-diagonal elements by the diagonals also gives us a form of grouping. When a diagonal element is null, we can conclude that all the elements corresponding to that row and column are also null.

If many of the parameters turn out to be null indeed in one shot, eventually leaving only $\mathrm{poly}(n)$ non-negligible ones, these strategies become an efficient way to measure all the coefficients. Nevertheless, notice that designing methods that retrieve specific partial information is not a trivial task, even when we assume that we can neglect all the other parameters. We should continue searching for bounds and relations between the characterization parameters of different types of maps. Also, we should further pursue the design of scalable methods to measure subgroups of information, while requiring the protocols to rely experimentally on minimum possible resources.

ACKNOWLEDGMENTS

CCL would like to thank the members of the Physics Department at the University of Buenos Aires for their warm hospitality. This work was supported in part by the National Security Agency NSA under Army Research Office ARO Contract No. W911NF-05-1-0469.

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

[2] M. Mohseni, A. T. Rezakhani, and D. A. Lidar, Phys. Rev. A **77**, 032322 (2008).

[3] J. Emerson, R. Alicki, and K. Życzkowski, J. Opt. B **7**, S347 (2005).

[4] B. Lévi, C. C. López, J. Emerson, and D. G. Cory, Phys. Rev. A **75**, 022314 (2007).

[5] C. Dankert, R. Cleve, J. Emerson, and E. Livine, Phys. Rev. A **80**, 012304 (2009); e-print arXiv:quant-ph/0606161v1 (2006).

[6] J. Emerson, M. Silva, O. Moussa, C. Ryan, M. Laforest, J. Baugh, D. G. Cory, and R. Laflamme, Science **317**, 1893 (2007).

[7] M. Silva, E. Magesan, D. W. Kribs, and J. Emerson, Phys. Rev. A **78**, 012347 (2008).
[8] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. Lett. **100**, 190403 (2008).
[9] C. C. López, B. Lévi, and D. G. Cory, Phys. Rev. A **79**, 042328 (2009).
[10] A. Bendersky, F. Pastawski, and J. P. Paz, Phys. Rev. A **80**, 032116 (2009).
[11] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).
[12] An example of such kind of basis is the one formed by the generalized Pauli operators. However, until Sec. IV we refrain from using a particular basis so our results remain general.
[13] A. Shabani and D. A. Lidar, Phys. Rev. Lett. **102**, 100402 (2009).
[14] S. Samuel, J. Math. Phys. **21**, 2695 (1980); P. A. Mello, J. Phys. A **23**, 4061 (1990); P. W. Brouwer and C. W. J. Beenakker, J. Math. Phys. **37**, 4904 (1996).
[15] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, J. Math. Phys. **45**, 2171 (2004).
[16] P. Pechukas, Phys. Rev. Lett. **73**, 1060 (1994); R. Alicki, *ibid.* **75**, 3020 (1995); P. Pechukas, *ibid.* **75**, 3021 (1995).
[17] H. A. Carteret, D. R. Terno, and K. Życzkowski, Phys. Rev. A **77**, 042113 (2008).
[18] D. Gross, K. Audenaert, and J. Eisert, J. Math. Phys. **48**, 052104 (2007).
[19] C. C. López, Ph.D. thesis, Massachusetts Institute of Technology, 2009 [http://hdl.handle.net/1721.1/51777].
[20] C. T. Schmiegelow, M. A. Larotonda, and J. P. Paz, Phys. Rev. Lett. **104**, 123601 (2010).
[21] M. A. Nielsen, Phys. Lett. A **303**, 249 (2002).
[22] A. Klappenecker and M. Roetteler, in *Proceedings of the IEEE International Symposium on Information Theory* (IEEE, Adelaide, SA, 2005), pp. 1740–1744.
[23] D. Gottesman, Ph.D. thesis, California Institute of Techonology, 1997, e-print arXiv:quant-ph/9705052v1; D. Gottesman, Phys. Rev. A **54**, 1862 (1996).
[24] C. Dankert, Master's thesis, University of Waterloo, 2005, e-print arXiv:quant-ph/0512217v2.
[25] Examples of process tomography using the largest systems in different setups are as follows: (a) with photons, M. W. Mitchell, C. W. Ellenor, S. Schneider, and A. M. Steinberg, Phys. Rev. Lett. **91**, 120402 (2003); J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, *ibid.* **93**, 080502 (2004); (b) in liquid-state NMR, Y. S. Weinstein, T. F. Havel, J. Emerson, N. Boulant, M. Saraceno, and S. Lloyd, J. Chem. Phys. **121**, 6117 (2004); (c) for the motion of optically trapped atoms, S. H. Myrskog, J. K. Fox, M. W. Mitchell, and A. M. Steinberg, Phys. Rev. A **72**, 013615 (2005); (d) with nitrogen vacancy centers, M. Howard, J. Twamley, C. Wittmann, T. Gaebel, F. Jelezko, and J. Wrachtrup, New J. Phys. **8**, 33 (2006); (e) with superconducting qubits, M. Neeley, M. Ansmann, R. C. Bialczak, M. Hofheinz, N. Katz1, E. Lucero, A. O'Connell, H. Wang, A. N. Cleland, and John M. Martinis, Nature Phys. **4**, 523 (2008); J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, Phys. Rev. Lett. **102**, 090502 (2009); (f) with ion traps, T. Monz, K. Kim, W. Hänsel, M. Riebe, A. S. Villar, P. Schindler, M. Chwalla, M. Hennrich, and R. Blatt, *ibid.* **102**, 040501 (2009); D. Hanneke, J. P. Home, J. D. Jost, J. M. Amini, D. Leibfried, and D. J. Wineland, Nature Phys. **6**, 13 (2009).
[26] The set $\{e^{-i\pi\sigma_q/4}, e^{-i\pi\sigma_q/4}\sigma_p\}$ with $p,q = x,y,z$ generates only half of the Clifford group for one qubit (up to a global phase). Nevertheless, this set of 12 operators is enough to implement the Clifford twirl we need.
[27] For example, for $w_{\mathrm{co}} < n/2$, it is trivial to prove that $M_{\mathrm{co}} \leqslant (w_{\mathrm{co}}+1)(ne/w_{\mathrm{co}})^{w_{\mathrm{co}}}$. In T. Worsch, Universität Karlsruhe, Fakultät für Informatik, Technical Report No. 31/94, 1994 [http://digbib.ubka.uni-karlsruhe.de/volltexte/181894], it is shown that $M_{\mathrm{co}} \leqslant (1+\epsilon)(\sqrt{n}/4\pi)(ne/w_{\mathrm{co}})^{w_{\mathrm{co}}}$, with $\epsilon \in O(1/n)$. For a simple case where we find only up to two-body coefficients, then trivially $M_{\mathrm{co}} = 1 + n + n(n-1)/2$.
[28] M. P. A. Branderhorst, J. Nunn, I. A. Walmsley, and R. L. Kosut, New J. Phys. **11**, 115010 (2009).