Optimal quantum learning of a unitary transformation

Alessandro Bisio
QUIT Group, Dipartimento di Fisica "A. Volta" and INFN Sezione di Pavia, via Bassi 6, I-27100 Pavia, Italy

Giulio Chiribella
Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario N2L 2Y5, Canada

Giacomo Mauro D'Ariano, Stefano Facchini, and Paolo Perinotti
QUIT Group, Dipartimento di Fisica "A. Volta" and INFN Sezione di Pavia, via Bassi 6, I-27100 Pavia, Italy
(Received 25 March 2009; published 25 March 2010)

We address the problem of learning an unknown unitary transformation from a finite number of examples. The problem consists in finding the learning machine that optimally emulates the examples, thus reproducing the unknown unitary with maximum fidelity. Learning a unitary is equivalent to storing it in the state of a quantum memory (the memory of the learning machine) and subsequently retrieving it. We prove that, whenever the unknown unitary is drawn from a group, the optimal strategy consists in a parallel call of the available uses followed by a "measure-and-rotate" retrieving. Differing from the case of quantum cloning, where the incoherent "measure-and-prepare" strategies are typically suboptimal, in the case of learning the "measure-and-rotate" strategy is optimal even when the learning machine is asked to reproduce a single copy of the unknown unitary. We finally address the problem of the optimal inversion of an unknown unitary evolution, showing also in this case the optimality of the "measure-and-rotate" strategies and applying our result to the optimal approximate realignment of reference frames for quantum communication.

DOI: 10.1103/PhysRevA.81.032324 PACS number(s): 03.67.Lx, 03.65.−w, 03.67.Hk

I. INTRODUCTION

A quantum memory would be an invaluable resource for quantum technology, and extensive experimental work is in progress for its realization [1–3]. On a quantum memory one can store unknown quantum states. Can we exploit it to store an unknown quantum transformation? In this way we could transmit the transformation to a distant party by just transmitting a state, without the need of transferring the device. More generally, we could process the transformation with the usual state manipulation techniques, as noticed by Vidal, Masanes, and Cirac, who addressed the problem in Ref. [4].

Storing-retrieving of transformations can also be seen as an instance of *quantum learning*, a topic which received increasing attention in the past few years (see, e.g., Refs. [5–7] for different approaches): Suppose that a user can dispose of N uses of a black box implementing an unknown unitary transformation U. Today the user is allowed to exploit the black box at his or her convenience, running an arbitrary quantum circuit that makes N calls to it. Tomorrow, however, the black box will no longer be available, and the user will be asked to reproduce U on a new input state $|\psi\rangle$ unknown to him or her. We refer to this scenario as to quantum learning of the unitary U from a finite set of N examples. Generally, the user may be required to reproduce U more than once, i.e., to produce $M \geqslant 1$ copies of U. In this case it is important to assess how the performance of learning decays with the number of copies required, as it was done in the case of quantum cloning [8].

Let us consider first the $M = N = 1$ case. Clearly, the only thing we can do today is to apply the black box to a known (generally entangled) state $|\varphi\rangle$. After that, what remains is the state $|\varphi_U\rangle = (U \otimes I)|\varphi\rangle$, that can be stored in a quantum memory. Then, when the new input state $|\psi\rangle$ becomes available, we send $|\psi\rangle$ and $|\varphi_U\rangle$ to an optimal *retrieving channel*, which emulates U applied to $|\psi\rangle$. If $N > 1$ input copies are available, we must also find the best storing strategy: we can, e.g., opt for a *parallel strategy* where U is applied on N different systems, yielding $(U^{\otimes N} \otimes I)|\varphi\rangle$, or for a *sequential strategy* where U is applied N times on the same system, alternated with other known unitaries, yielding $(U V_{N-1} \cdots V_2 U V_1 U \otimes I)|\varphi\rangle$. The most general storing strategy is described by a *quantum circuit board*, i.e., a quantum network with open slots where the input copies can be inserted [9,10]. In summary, solving the problem of the optimal quantum learning means finding the optimal storing board and the optimal retrieving channel.

An alternative to coherent retrieval is to estimate U, to store the outcome in a classical memory, and to perform the estimated unitary on the new input state. This incoherent estimation-based strategy has the double advantage of avoiding the expensive use of a quantum memory (which nowadays cannot store information for more than few milliseconds) and of allowing one to reproduce U an unlimited number of times with constant quality. However, estimation-based strategies are typically suboptimal for the similar task of quantum cloning [8], and, by analogy, one would expect a coherent retrieval to achieve better performances. Surprisingly, we find that whenever the unknown unitary is randomly drawn from a group the incoherent strategies already achieve the ultimate performances for quantum learning. In particular, we show that the performance of the optimal retrieving channel is equal to that of optimal estimation. For example, for a completely unknown qubit unitary the optimal fidelity behaves as $F = 1 - O(N^{-2})$ asymptotically for large N. Our result can be also extended to solve the problem of *optimal inversion* of the unknown U, in which the user is asked to perform $U^\dagger$.

In this case, we provide the optimal approximate realignment of reference frames for the quantum communication scenario considered by Ref. [11], reaching the above asymptotic fidelity without ancilla. The article is structured as follows: in Sec. II we introduce the notation and the theoretical framework used to solve the problem of optimal learning. The optimization is then presented in Sec. III by first addressing the case of a single output copy (Sec. III A), and subsequently showing how to generalize the argument to the case of $M > 1$ output copies (Sec. III B). In Sec. IV we discuss the problem of the optimal inversion of an unknown quantum dynamics, which can be regarded as a small variation of our learning problem. Section V concludes the article with a summary of the main results.

II. NOTATION AND THEORETICAL FRAMEWORK

To derive the optimal learning we use the method of *quantum combs* [9], briefly summarized here. For more details and for an extensive presentation of the method we refer to Ref. [10].

Let $\mathrm{Lin}(\mathcal{H})$ denote the space of linear operators acting on the Hilbert space $\mathcal{H}$, and $\mathrm{Lin}(\mathcal{H},\mathcal{K})$ be the space of linear operators from $\mathcal{H}$ to $\mathcal{K}$. In the following we will use the one-to-one correspondence between bipartite vectors $|A\rangle\!\rangle \in \mathcal{K}\otimes\mathcal{H}$ and linear operators $A \in \mathrm{Lin}(\mathcal{H},\mathcal{K})$ given by

$$|A\rangle\!\rangle = \sum_{m=1}^{\dim(\mathcal{K})}\sum_{n=1}^{\dim(\mathcal{H})} \langle m|A|n\rangle |m\rangle|n\rangle, \tag{1}$$

where $\{|m\rangle\}_{m=1}^{\dim(\mathcal{K})}$ and $\{|n\rangle\}_{n=1}^{\dim(\mathcal{H})}$ are two fixed orthonormal bases for $\mathcal{K}$ and $\mathcal{H}$, respectively.

If A and B are two commuting operators in $\mathrm{Lin}(\mathcal{H})$ it is simple to derive from Eq. (1) the equality

$$(A\otimes I_{\mathcal{H}})|B\rangle\!\rangle = (I_{\mathcal{H}}\otimes A^T)|B\rangle\!\rangle, \tag{2}$$

where $I_{\mathcal{H}}$ is the identity operator on $\mathcal{H}$ and A^T denotes the transpose of A with respect to the orthonormal basis $\{|n\rangle\}$.

A quantum channel $\mathcal{C}$ from $\mathrm{Lin}(\mathcal{H})$ to $\mathrm{Lin}(\mathcal{K})$ is a completely positive trace-preserving map, and is conveniently described by its Choi-Jamiołkowski operator, namely by the positive operator $C \in \mathrm{Lin}(\mathcal{K}\otimes\mathcal{H})$ defined by

$$C = (\mathcal{C}\otimes\mathcal{I}_{\mathcal{H}})(|I_{\mathcal{H}}\rangle\!\rangle\langle\!\langle I_{\mathcal{H}}|), \tag{3}$$

where $\mathcal{I}_{\mathcal{H}}$ is the identity map on $\mathrm{Lin}(\mathcal{H})$, and, according to Eq. (1), $|I_{\mathcal{H}}\rangle\!\rangle$ is the maximally entangled vector $|I_{\mathcal{H}}\rangle\!\rangle = \sum_{n=1}^{\dim(\mathcal{H})}|n\rangle|n\rangle \in \mathcal{H}^{\otimes 2}$.

The composition of two channels is represented in terms of their Choi-Jamiołkowski operators by the *link product* [9,10]. Precisely, if $\mathcal{D}$ is a channel from $\mathcal{K}$ to $\mathcal{L}$, the Choi operator of the channel $\mathcal{D}\circ\mathcal{C}$ resulting from the composition of $\mathcal{C}$ and $\mathcal{D}$ is given by the product

$$D * C = \mathrm{Tr}_{\mathcal{K}}[(D\otimes I_{\mathcal{H}})(I_{\mathcal{L}}\otimes C^{T_{\mathcal{K}}})], \tag{4}$$

with $\mathrm{Tr}_{\mathcal{K}}$ denoting partial transpose on $\mathcal{K}$. Viewing states as a special kind of channels with one-dimensional input space, Eq. (4) yields $\mathcal{C}(\rho) = C * \rho = \mathrm{Tr}_{\mathcal{H}}[C(I_{\mathcal{K}}\otimes\rho^T)]$. A channel $\mathcal{C}$ from $\mathcal{H}$ to $\mathcal{K}$ is trace preserving if and only if it satisfies the normalization condition

$$I_{\mathcal{K}} * C \equiv \mathrm{Tr}_{\mathcal{K}}[C] = I_{\mathcal{H}}. \tag{5}$$

For two channels with multipartite input and output, one can decide to connect only some particular output of the first channel to some input of the second one: for example, if $\mathcal{C}$ is a channel from $\mathrm{Lin}(\mathcal{H}\otimes\mathcal{A})$ to $\mathrm{Lin}(\mathcal{K}\otimes\mathcal{B})$ and $\mathcal{D}$ is a channel from $\mathrm{Lin}(\mathcal{A}'\otimes\mathcal{K})$ to $\mathrm{Lin}(\mathcal{B}'\otimes\mathcal{L})$ we can connect the wires with the same label $\mathcal{K}$, thus obtaining the new channel $(\mathcal{D}\otimes\mathcal{I}_{\mathcal{B}})(\mathcal{I}_{\mathcal{A}'}\otimes\mathcal{C})$, which is a channel from $\mathrm{Lin}(\mathcal{A}'\otimes\mathcal{H}\otimes\mathcal{A})$ to $\mathrm{Lin}(\mathcal{B}'\otimes\mathcal{L}\otimes\mathcal{B})$. Accordingly, the connections of quantum channels in a network will be encoded in the labels assigned to the Hilbert spaces: whenever two spaces have the same label, two channels acting on these spaces will be connected, and their Choi-Jamiołkowski operators will be contracted with the link product as in Eq. (4).

Remark (reordering of Hilbert spaces and commutativity of the link product). Encoding the connections in the labeling of the Hilbert spaces turns out to be very convenient in the treatment of multipartite quantum networks, because some formulas take a much simpler form if we suitably rearrange the ordering of the Hilbert spaces in the tensor product. For example, it may be convenient to rewrite the tensor product $\bigotimes_{i=1}^{2N+1}\mathcal{H}_i$ putting all spaces with even labels on the left and all spaces with odd labels on the right. This reordering can be done safely as long as different Hilbert spaces have different labels. Note that the link product of two Choi-Jamiołkowski operators is commutative up to this reordering of Hilbert spaces: for example, given two operators $C \in \mathrm{Lin}(\mathcal{K}\otimes\mathcal{H})$ and $D \in \mathrm{Lin}(\mathcal{L}\otimes\mathcal{K})$ with $\mathcal{H}\simeq\mathcal{K}\simeq\mathcal{L}$, we have $D * C = \mathrm{SWAP}\,(C * D)\,\mathrm{SWAP}$, where SWAP is the operator that exchanges the Hilbert spaces $\mathcal{L}$ and $\mathcal{H}$ in the tensor product $\mathcal{L}\otimes\mathcal{H}$. The reader should not be confused by fact that the link product is commutative (up to reordering of the Hilbert spaces), whereas the composition of channels is not ($\mathcal{C}\circ\mathcal{D}$ is in general different from $\mathcal{D}\circ\mathcal{C}$). The fact that the output of $\mathcal{C}$ is connected with the input of $\mathcal{D}$ (and not the other way round) is encoded in the fact that the output space of $\mathcal{C}$ has the same label of the input space of $\mathcal{D}$ (here they are both labeled as $\mathcal{K}$). In order to express the different composition of channels corresponding to $\mathcal{C}\circ\mathcal{D}$ we would have had to choose a different labeling, in which the output of $\mathcal{D}$ is identified with the input of $\mathcal{C}$.

A quantum circuit board is the quantum network resulting from a sequence of multipartite channels where some input of a channel is connected to some output of the previous one, as we just illustrated. A *quantum comb* is the Choi-Jamiołkowski operator associated to a quantum circuit board and is obtained as the link product of all component channels. The fact that the the circuit board represents a sequence of (trace-preserving) channels is expressed by a set of linear equations [9,10], and, therefore, optimizing a quantum circuit board is equivalent to optimizing a positive operator subject to these linear constraints. The constraints will be given explicitly for the case of learning in the next section.

III. OPTIMIZATION OF LEARNING

In this section we show that the optimal quantum learning of an unknown unitary randomly drawn from a group has a very simple and general structure: (i) in order to store the unitary it is

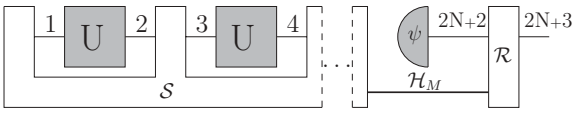

FIG. 1. The learning process is described by a quantum comb (in white) representing the storing board, in which the N uses of a unitary U are plugged, along with the state $|\psi\rangle$ (in gray). The wires represent the input-output Hilbert spaces. The output of the first comb is stored in a quantum memory, later used by the retrieving channel $\mathcal{R}$.

enough to apply the available examples in parallel on a suitable entangled state, (ii) the optimal state for storage has the same form of an optimal state for estimation of the unknown unitary, and (iii) the optimal retrieval can be achieved via estimation of the unknown unitary, namely by measuring the quantum memory, producing an estimate for the unknown unitary and, finally, applying the estimate M times.

A. The $M = 1$ case

We tackle the optimization of learning starting from the case where a single output copy is required. Referring to Fig. 1, we label the Hilbert spaces of quantum systems according to the following sequence: $(\mathcal{H}_{2n+1})_{n=0}^{N-1}$ are the inputs for the N examples of U, and $(\mathcal{H}_{2n+2})_{n=0}^{N-1}$ are the corresponding outputs. We denote by $\mathcal{H}_i = \bigotimes_{n=0}^{N-1} \mathcal{H}_{2n+1}$ ($\mathcal{H}_o = \bigotimes_{n=0}^{N-1} \mathcal{H}_{2n+2}$) the Hilbert spaces of all inputs (outputs) of the N examples. The input state $|\psi\rangle$ belongs to $\mathcal{H}_{2N+2}$, and the output state finally produced belongs to $\mathcal{H}_{2N+3}$. All spaces $\mathcal{H}_n$ considered here are d-dimensional, except the spaces $\mathcal{H}_0$ and $\mathcal{H}_{2N+1}$ which are one-dimensional and are introduced just for notational convenience. The comb of the whole learning process is an operator $L \geqslant 0$ on the tensor of all Hilbert spaces and satisfies the normalization condition [9,10]:

$$\mathrm{Tr}_{2k+1}[L^{(k)}] = I_{2k} \otimes L^{(k-1)} \qquad k = 0, 1, \ldots, N+1, \quad (6)$$

where $L^{(N+1)} = L$, $L^{(-1)} = 1$, and $L^{(k)}$ is a positive operator on the spaces $(\mathcal{H}_n)_{n=0}^{2k+1}$. When the N examples are connected with the learning board, the user obtains a channel $\mathcal{C}_U$ with Choi operator given by

$$\begin{aligned} C_U &= L * |U\rangle\!\rangle\langle\!\langle U|^{\otimes N} \\ &= \mathrm{Tr}_{i,o}[L(I_{2N+3} \otimes I_{2N+2} \otimes (|U\rangle\!\rangle\langle\!\langle U|^{\otimes N})^T)], \end{aligned} \quad (7)$$

as it follows from the definition of link product in Eq. (4).

As the figure of merit we maximize the fidelity of the output state $\mathcal{C}_U(|\psi\rangle\langle\psi|)$ with the target state $U|\psi\rangle\langle\psi|U^\dagger$, uniformly averaged over all input pure states $|\psi\rangle$ and all unknown unitaries U in the group G. Apart from irrelevant constants, such optimization coincides with the maximization of the channel fidelity between $\mathcal{C}_U$ and the target unitary (i.e., the fidelity between the Choi-Jamiołkowski states C_U/d and $|U\rangle\!\rangle\langle\!\langle U|/d$) averaged over U:

$$\begin{aligned} F &= \frac{1}{d^2} \int_G \mathrm{Tr}\{L[|U\rangle\!\rangle\langle\!\langle U| \otimes (|U\rangle\!\rangle\langle\!\langle U|^{\otimes N})^T]\}\, dU \\ &= \frac{1}{d^2} \int_G \langle\!\langle U|\langle\!\langle U^*|^{\otimes N} L |U^*\rangle\!\rangle^{\otimes N} |U\rangle\!\rangle\, dU, \end{aligned} \quad (8)$$

U^* being the complex conjugate of U in the computational basis, and dU denoting the normalized Haar measure. From the expression of F it is easy to prove that there is no loss of generality in requiring the commutation

$$[L, U_{2N+3} \otimes V_{2N+2}^* \otimes (U^* \otimes V)^{\otimes N}] = 0 \quad \forall U, \quad V \in G. \quad (9)$$

Moreover, using Eq. (6) for $k = N+1$ we obtain $\mathrm{Tr}_{\mathcal{H}_{2N+3}}[L] = I_{2N+2} \otimes L^{(N)}$, where $L^{(N)}$ is a positive operator acting on $\bigotimes_{n=0}^{2N+1} \mathcal{H}_n$ (recall that, however, $\mathcal{H}_0$ and $\mathcal{H}_{2N+1}$ are one-dimensional). Reordering the Hilbert spaces in the tensor product by putting all input spaces of the examples on the right and all output spaces on the left and using Eq. (9) we then get

$$\left[L^{(N)}, U_o^{*\otimes N} \otimes V_i^{\otimes N}\right] = 0 \quad \forall U, \quad V \in G. \quad (10)$$

Here the subscripts i, o recall that $U^{\otimes N}$ acts on the tensor product of all output spaces $\mathcal{H}_o = \bigotimes_{n=0}^{N-1} \mathcal{H}_{2n+1}$, while $V^{\otimes N}$ acts on the tensor product of all input spaces $\mathcal{H}_i = \bigotimes_{n=0}^{N-1} \mathcal{H}_{2n+1}$. This leads to the following.

Lemma 1 (optimality of parallel storage). The optimal storage of U can be achieved by applying $U_o^{\otimes N} \otimes I_i^{\otimes N}$ on a suitable input state $|\varphi\rangle \in \mathcal{H}_o \otimes \mathcal{H}_i$.

Proof. According to Fig. 1, the learning board $\mathcal{L}$ results from the connection of the storing board $\mathcal{S}$ with the retrieving channel $\mathcal{R}$. In terms of the corresponding Choi-Jamiołkowski operators L, S, R, respectively, one has $L = R * S$. Denoting by $\mathcal{H}_M$ the Hilbert space of the quantum memory in Fig. 1, we have that $\mathcal{R}$ is a channel from $(\mathcal{H}_{2N+2} \otimes \mathcal{H}_M)$ to $\mathcal{H}_{2N+3}$ and satisfies the normalization condition $I_{2N+3} * R = I_{2N+2} \otimes I_M$. Using this fact, one gets $\mathrm{Tr}_{2N+3}[L] \equiv I_{2N+3} * L = (I_{2N+3} * R) * S = (I_{2N+2} \otimes I_M) * S = I_{2N+2} \otimes \mathrm{Tr}_M[S]$, which compared with Eq. (6) for $k = N+1$ implies $\mathrm{Tr}_M[S] = L^{(N)}$. Now, without loss of generality we take the storing board $\mathcal{S}$ to be a sequence of isometries [9,10], which implies that S is rank one: $S = |\Phi\rangle\!\rangle\langle\!\langle\Phi|$. With this choice, the state S/d^N is a purification of $L^{(N)}/d^N$. Again, one can choose w.l.o.g. S/d^N to be a state on $(\mathcal{H}_o \otimes \mathcal{H}_i) \otimes (\mathcal{H}'_o \otimes \mathcal{H}'_i)$, with $\mathcal{H}'_o \simeq \mathcal{H}_o$ and $\mathcal{H}'_i \simeq \mathcal{H}_i$ and assume $|\Phi\rangle\!\rangle = |L^{(N)\frac{1}{2}}\rangle\!\rangle$. Taking $V = I$ in Eq. (10) and using Eq. (2) we get $(U_o^{\otimes N} \otimes I_{i,o',i'})|\Phi\rangle\!\rangle = (I_{o,i} \otimes U_{o'}^{T\otimes N} \otimes I_{i'})|\Phi\rangle\!\rangle$. When the examples of U are connected to the storing board, the output is the state $\rho_U = S * |U\rangle\!\rangle\langle\!\langle U|_{o,i}^{\otimes N}$. Using the above relation we find that ρ_U is the projector on the state $|\varphi_U\rangle = (U_{o'}^{\otimes N} \otimes I_{i'})|\varphi\rangle$, where $|\varphi\rangle = \langle\!\langle I^{\otimes N}|_{o,i}|\Phi\rangle\!\rangle \in \mathcal{H}_{o'} \otimes \mathcal{H}_{i'} \simeq \mathcal{H}_o \otimes \mathcal{H}_i$. This proves that every storing board gives the same output that would be obtained with a parallel scheme. In other words, every storing board can be simulated applying $(U_o^{\otimes N} \otimes I_i^{\otimes N})$ to a suitable input state $|\varphi\rangle \in \mathcal{H}_o \otimes \mathcal{H}_i$. ∎

Optimizing learning is then reduced to finding the optimal input state $|\varphi\rangle$ and the optimal retrieving channel $\mathcal{R}$. The fidelity can be computed substituting $L = R * S$ in Eq. (8) and using the relation $\langle\!\langle U|\langle\!\langle U^*|^{\otimes N}(R * S)|U\rangle\!\rangle|U^*\rangle\!\rangle^{\otimes N} = \langle\!\langle U|R|U\rangle\!\rangle * \langle\!\langle U^*|^{\otimes N} S |U^*\rangle\!\rangle^{\otimes N} = \langle\!\langle U|R|U\rangle\!\rangle * |\varphi_U\rangle\langle\varphi_U|$, which gives

$$F = \frac{1}{d^2} \int_G \langle\!\langle U|\langle\varphi_U^*|R|U\rangle\!\rangle|\varphi_U^*\rangle dU. \quad (11)$$

Lemma 2 (optimal states for storage). The optimal input state for storage can be taken of the form

$$|\varphi\rangle = \bigoplus_{j\in \mathrm{Irr}(U^{\otimes N})} \sqrt{\frac{p_j}{d_j}}|I_j\rangle\!\rangle \in \widetilde{\mathcal{H}}, \tag{12}$$

where p_j are probabilities, the index j runs over the set $\mathrm{Irr}(U^{\otimes N})$ of all irreducible representations $\{U_j\}$ contained in the decomposition of $\{U^{\otimes N}\}$, and $\widetilde{\mathcal{H}} = \bigoplus_{j\in \mathrm{Irr}(U^{\otimes N})}(\mathcal{H}_j \otimes \mathcal{H}_j)$ is a subspace of $\mathcal{H}_o \otimes \mathcal{H}_i$ carrying the representation $\widetilde{U} = \bigoplus_{j\in \mathrm{Irr}(U^{\otimes N})}(U_j \otimes I_j)$, I_j being the identity in $\mathcal{H}_j$.

Proof. Using Eqs. (2) and (10) it is possible to show that the marginal state $\rho = \mathrm{Tr}_i[|\varphi\rangle\langle\varphi|]$ is invariant under $U^{\otimes N}$. Decomposing $U^{\otimes N}$ into irreducible representations (irreps) we have $U^{\otimes N} = \bigoplus_j (U_j \otimes I_{m_j})$, where I_{m_j} is the identity on an m_j-dimensional multiplicity space $\mathbb{C}^{m_j}$. Therefore, ρ must have the form $\rho = \bigoplus_j p_j(I_j/d_j \otimes \rho_j)$, where ρ_j is an arbitrary state on the multiplicity space $\mathbb{C}^{m_j}$. Since $|\varphi\rangle$ is a purification of ρ, with a suitable choice of basis we have $|\varphi\rangle = |\rho^{\frac{1}{2}}\rangle\!\rangle = \bigoplus_j \sqrt{p_j/d_j}|I_j\rangle\!\rangle|\rho_j^{\frac{1}{2}}\rangle\!\rangle$, which after storage becomes $|\varphi_U\rangle = \bigoplus_j \sqrt{p_j/d_j}|U_j\rangle\!\rangle|\rho_j^{\frac{1}{2}}\rangle\!\rangle$. Hence, for every U the state $|\varphi_U\rangle$ belongs to the subspace $\widetilde{\mathcal{H}} = \bigoplus_j (\mathcal{H}_j^{\otimes 2} \otimes |\rho_j^{\frac{1}{2}}\rangle\!\rangle) \simeq \bigoplus_j \mathcal{H}_j^{\otimes 2}$. ∎

We can then restrict our attention to the subspace $\widetilde{\mathcal{H}}$ and consider retrieving channels $\mathcal{R}$ from $(\mathcal{H}_{2N+2} \otimes \widetilde{\mathcal{H}})$ to $\mathcal{H}_{2N+3}$. The normalization of the Choi operator is then

$$\mathrm{Tr}_{2N+3}[R] = I_{2N+2} \otimes I_{\widetilde{\mathcal{H}}}. \tag{13}$$

Combining the expression of the fidelity (8) with that of the input state (12), it is easy to see that one can always use a covariant retrieving channel, satisfying

$$[R, U_{2N+3} \otimes V_{2N+2}^* \otimes \widetilde{U}^*\widetilde{V}'] = 0 \quad \forall U, \quad V \in G, \tag{14}$$

where $\widetilde{V}' = \bigoplus_j (I_j \otimes V_j)$ acts on $\widetilde{\mathcal{H}}$. We now exploit the decompositions $U \otimes U_j^* = \bigoplus_{K\in \mathrm{Irr}(U\otimes U_j^*)}(U_K \otimes I_{m_K^{(j)}})$ and $V^* \otimes V_j = \bigoplus_{L\in \mathrm{Irr}(V^*\otimes V_j)}(V_L^* \otimes I_{m_L^{(j)}})$, which yield

$$U_{2N+3} \otimes V_{2N+2}^* \otimes \widetilde{U}^*\widetilde{V} = \bigoplus_{K,L} \left(U_K \otimes V_L^* \otimes I_{m_{KL}}\right). \tag{15}$$

Here $I_{m_{KL}}$ is given by $I_{m_{KL}} = \bigoplus_{j\in \mathsf{P}_{KL}}(I_{m_K^{(j)}} \otimes I_{m_L^{(j)}})$, where P_{KL} is the set of values of j such that the irrep $U_K \otimes V_L^*$ is contained in the decomposition of $U \otimes V^* \otimes U_j^* \otimes V_j$. Relations (14) and (15) then imply

$$R = \bigoplus_{K,L} (I_K \otimes I_L \otimes R_{KL}), \tag{16}$$

where R_{KL} is a positive operator on the multiplicity space $\mathbb{C}^{m_{JK}} = \bigoplus_{j\in \mathsf{P}_{KL}}(\mathbb{C}^{m_K^{(j)}} \otimes \mathbb{C}^{m_L^{(j)}})$. Moreover, using the equality $I \otimes I_j = \bigoplus_K (I_K \otimes I_{m_K^{(j)}})$ we obtain

$$\begin{aligned}
|I\rangle\!\rangle|\varphi^*\rangle &= \bigoplus_j \sqrt{\frac{p_j}{d_j}}|I\rangle\!\rangle|I_j\rangle\!\rangle \\
&= \bigoplus_j \bigoplus_{K\in \mathrm{Irr}(U\otimes U_j^*)} \sqrt{\frac{p_j}{d_j}}|I_K\rangle\!\rangle\left|I_{m_K^{(j)}}\right\rangle\!\rangle \\
&= \bigoplus_K \bigoplus_{j\in \mathsf{P}_{KK}} \sqrt{\frac{p_j}{d_j}}|I_K\rangle\!\rangle\left|I_{m_K^{(j)}}\right\rangle\!\rangle \\
&= \bigoplus_K |I_K\rangle\!\rangle|\alpha_K\rangle,
\end{aligned} \tag{17}$$

where $|I_K\rangle\!\rangle \in \mathcal{H}_K^{\otimes 2}$ and $|\alpha_K\rangle \in \mathbb{C}^{m_{KK}}$ is given by

$$|\alpha_K\rangle = \bigoplus_{j\in \mathsf{P}_{KK}} \sqrt{p_j/d_j}\,\left|I_{m_K^{(j)}}\right\rangle\!\rangle. \tag{18}$$

Exploiting Eqs. (16) and (17), the fidelity (11) can be rewritten as

$$F = \sum_K \frac{d_K}{d^2}\langle\alpha_K|R_{KK}|\alpha_K\rangle. \tag{19}$$

Theorem 1 (Optimal retrieving strategy). The optimal retrieving of U from the memory state $|\varphi_U\rangle$ is achieved by measuring the ancilla with the optimal POVM $P_{\hat{U}} = |\eta_{\hat{U}}\rangle\langle\eta_{\hat{U}}|$ given by $|\eta_{\hat{U}}\rangle = \bigoplus_j \sqrt{d_j}|\hat{U}_j\rangle\!\rangle$, and, conditionally on outcome $\hat{U}$, by performing the unitary $\hat{U}$ on the new input system.

Proof. Let us denote by $P_{KL}^{(j)}$ the projector on the tensor product $\mathbb{C}^{m_K^{(j)}} \otimes \mathbb{C}^{m_L^{(j)}}$ and by $R_{KL}^{(j)} = P_{KL}^{(j)}R_{KL}P_{KL}^{(j)}$ the corresponding diagonal block of R_{KL}. Using Schur's lemmas and Eq. (16) we obtain

$$\mathrm{Tr}_{2N+3}[R] = \sum_{K,L}\sum_{j\in \mathsf{P}_{KL}} \left(\frac{d_K}{d_j} I_j \otimes I_L \otimes \mathrm{Tr}_{m_K^{(j)}}\left[R_{KL}^{(j)}\right]\right). \tag{20}$$

Equation (13) then becomes $I_{m_L^{(j)}} = \sum_{K|\mathsf{P}_{KL}\ni j} \frac{d_K}{d_j}\mathrm{Tr}_{m_K^{(j)}}[R_{KL}^{(j)}]$ for all L, j, which for $K = L$ implies the bound

$$\mathrm{Tr}\left[R_{KK}^{(j)}\right] \leqslant \frac{d_j m_K^{(j)}}{d_K}. \tag{21}$$

For the fidelity (19) we then have the bound

$$F = \sum_K \frac{d_K}{d^2} \sum_{j,j'\in \mathsf{P}_{KK}} \sqrt{\frac{p_j p_{j'}}{d_j d_{j'}}}\left\langle\!\left\langle I_{m_K^{(j)}}\middle|R_{KK}\middle|I_{m_K^{(j')}}\right\rangle\!\right\rangle \tag{22}$$

$$\leqslant \sum_K \frac{d_K}{d^2}\left(\sum_{j\in \mathsf{P}_{KK}} \sqrt{\frac{p_j\left\langle\!\left\langle I_{m_K^{(j)}}\middle|R_{KK}^{(j)}\middle|I_{m_K^{(j)}}\right\rangle\!\right\rangle}{d_j}}\right)^2 \tag{23}$$

$$\leqslant \sum_K \frac{\left(\sum_{j\in \mathsf{P}_{KK}} m_K^{(j)}\sqrt{p_j}\right)^2}{d^2} = F_{\mathrm{est}}, \tag{24}$$

having used the positivity of R_{KK} for the first bound and Eq. (21) for the second. Regarding the last equality, it can be proved as follows. First, the Choi operator of the estimation-based strategy is $R_{\mathrm{est}} = \int_G |\hat{U}\rangle\!\rangle\langle\!\langle\hat{U}| \otimes |\eta_{\hat{U}}^*\rangle\langle\eta_{\hat{U}}^*|d\hat{U}$. Using Eq. (17) with $|\varphi^*\rangle$ replaced by $|\eta_I^*\rangle$ and performing the integral we obtain $R_{\mathrm{est}} = \bigoplus_K (I_K^{\otimes 2} \otimes \widetilde{R}_{KK})/d_K$, where

$\widetilde{R}_{KK} = |\beta_K\rangle\langle\beta_K|$, $|\beta_K\rangle = \bigoplus_{j\in \mathsf{P}_{KK}} \sqrt{d_j}|I_{m_K^{(j)}}\rangle\!\rangle$. Eq. (19) then gives

$$F_{\rm est} = \sum_K \frac{|\langle\alpha_K|\beta_K\rangle|^2}{d^2} = \sum_K \frac{\left(\sum_{j\in \mathsf{P}_{KK}} m_K^{(j)}\sqrt{p_j}\right)^2}{d^2}. \quad (25)$$

■

The above theorem shows that the optimal state for storing U is identical to the optimal state for estimating it [12] and, moreover, that the fidelity of unitary learning with $M = 1$ is precisely the fidelity of unitary estimation. Having reduced learning to estimation, we can then exploit the expressions for the optimal states and fidelities that are known in most relevant cases. For example, when U is an unknown qubit unitary in SU(2), learning becomes equivalent to optimal estimation of an unknown rotation in the Bloch sphere [13]. For large number of copies, the optimal input state is given by $|\varphi\rangle \approx \sqrt{4/N}\sum_{j=j_{\rm min}}^{N/2} \frac{\sin(2\pi j/N)}{\sqrt{2j+1}}|I_j\rangle\!\rangle$, with $j_{\rm min} = 0(1/2)$ for N even (odd), and the fidelity is $F \approx 1 - \pi^2/4N^2$. Remarkably, this asymptotic scaling can be achieved without using entanglement between the set of N qubits that are rotated and an auxiliary set of N rotationally invariant qubits: the optimal storing is achieved just by applying $U^{\otimes N}$ on the optimal N-qubit state [13]. Another example is that of an unknown phase-shift $U = \exp[i\theta\sigma_z]$. In this case, for large number of copies the optimal input state is $|\varphi\rangle = \sqrt{2/(N+1)}\sum_{m=-N/2}^{N/2}\sin[\pi(m+1/2)/(N+1)]|m\rangle$ and the fidelity is $F \approx 1 - 2\pi^2/(N+1)^2$ [14]. Again, the optimal state can be prepared using only N qubits.

B. Generalization to the $M > 1$ case

Our result can be extended to the case where the user must reproduce $M > 1$ copies of the unknown unitary U. In this case, there are two different notions of optimality induced by two different figures of merit, namely the single-copy and the global fidelity. In the following we will examine both cases.

1. Optimal learning according to the single-copy fidelity

Let $\mathcal{C}_U$ be the M-partite channel obtained by the user, and $\mathcal{C}^{(1)}_{U,\Sigma}$ be the local channel $\mathcal{C}^{(1)}_{U,\Sigma}(\rho) = {\rm Tr}_{\bar{1}}[\mathcal{C}_U(\rho\otimes\Sigma)]$, where ρ is the state of the first system, Σ is the state of the remaining $M-1$ systems, and ${\rm Tr}_{\bar{1}}$ denotes the trace over all systems except the first. The local channel $\mathcal{C}^{(1)}_{U,\Sigma}$ describes the evolution of the first input of $\mathcal{C}_U$ when the remaining $(M-1)$ inputs are prepared in the state Σ. Of course, the fidelity between $\mathcal{C}^{(1)}_{U,\Sigma}$ and the unitary U cannot be larger than the optimal fidelity $F_{\rm est}$ of Eq. (24), and the same holds for any local channel $\mathcal{C}^{(i)}_{U,\Sigma}$, in which all but the ith input system are discarded. Therefore, the measure-and-prepare strategy presented in Theorem 1 is optimal also for the maximization of the single-copy fidelity of all local channels, and such fidelity does not decrease with increasing M.

2. Optimal learning according to the global fidelity

The results of Sec. III A can be extended to the maximization of the global fidelity between $\mathcal{C}_U$ and $U^{\otimes M}$, just by replacing U with $U^{\otimes M}$ in all derivations. Indeed, the role of the target unitary U in our derivations is completely generic: we never used the fact that the unitary emulated by the machine was equal to the unitaries provided in the examples. Therefore, following the same proofs of Sec. III A it is immediate to see that also for the case of $M > 1$ copies with global fidelity the optimal strategy for storing consists in the parallel application of the examples on an input state of the form of Lemma 2 and that the optimal strategy for retrieving consists in measuring the optimal POVM $P_{\hat{U}}$ and in performing $\hat{U}^{\otimes M}$ conditionally on outcome $\hat{U}$. Therefore, also in this case optimal learning is equivalent to optimal estimation: precisely, the optimal learning is achieved by the estimation strategy that maximizes the expectation value of the goal function $f_M(U,\hat{U}) = (|{\rm Tr}[U^\dagger\hat{U}]|/d)^{2M}$, given by $\langle f_M\rangle = \int dU \int d\hat{U}\, f_M(U,\hat{U})\langle\varphi_U|P_{\hat{U}}|\varphi_U\rangle$. Note that in this case the coefficients $\{p_j\}$ in the optimal input state of Lemma 2) generally depend on M.

Remark (generalization to nonidentical group representations). Since we never used the fact that the N examples are identical, all the results of Sec. III A hold even when the input (output) uses are not identical copies $U^{\otimes N}$ ($U^{\otimes M}$), but generally N (M) different unitaries, each of them belonging to a different representation of the group G. For example, if $G = {\rm SO}(3)$ the N examples may correspond to rotations (of the same angle and around the same axis) of N quantum particles with different angular momenta. Of course, the same remark also holds when the M output copies.

IV. OPTIMAL INVERSION OF AN UNKNOWN UNITARY EVOLUTION

We now extend our results to the *optimal inversion* of an unknown unitary U: in this case the goal is not to produce M copies of U, but, instead M copies of its inverse $U^\dagger$. For this task the fidelity of the learning board is $F' = 1/d^2\int_G \langle\!\langle U^\dagger|^{\otimes M}\langle\!\langle U^*|^{\otimes N} L'|U^\dagger\rangle\!\rangle^{\otimes M}|U^*\rangle\!\rangle^{\otimes N} dU$, as obtained by substituting U with $U^{\dagger\otimes M}$ in the target of Eq. (8). From this expression it is easy to see that one can always assume $[L', V^{\otimes M}\otimes U^{*\otimes M}\otimes U_o^{*\otimes N}\otimes V_i^{\otimes N}] = 0$. Therefore, the optimal inversion is obtained from our derivations by simply substituting $U_{2N+3} \to V^{\otimes M}$ and $V_{2N+2} \to U^{\otimes M}$. Accordingly, the optimal inversion is achieved by measuring the optimal POVM $P_{\hat{U}}$ on the optimal state $|\varphi_U\rangle$ and by performing $\hat{U}^{\dagger\otimes M}$ conditionally on outcome $\hat{U}$. This provides the optimal approximate realignment of reference frames in the quantum communication scenario recently considered in Ref. [11], proving the optimality of the "measure-and-rotate" strategy conjectured therein. In that scenario, the state $|\varphi\rangle \in \widetilde{\mathcal{H}}$ serves as a token of Alice's reference frame and is sent to Bob along with a quantum message $|\psi\rangle \in \mathcal{H}^{\otimes M}$. Due to the mismatch of reference frames, Bob receives the decohered state $\sigma_\psi = \int_G |\varphi_U\rangle\langle\varphi_U|\otimes U|\psi\rangle\langle\psi|U^\dagger dU$, from which he tries to retrieve the message $|\psi\rangle$ with maximum fidelity $f = \int d\psi\,\langle\psi|\mathcal{R}'(\sigma_\psi)|\psi\rangle {\rm d}\psi$, where $\mathcal{R}'$ is the retrieving channel and $d\psi$ denotes the uniform probability measure over pure states.

The maximization of f is equivalent to the maximization of the channel fidelity $F' = \int_G \langle\langle U^\dagger | \langle \varphi_U^* | R' | U^\dagger \rangle\rangle | \varphi_U^* \rangle dU$, which is the figure of merit for optimal inversion. It is worth stressing that the state $|\varphi\rangle$ that maximizes the fidelity is not the state $|\varphi_{\rm lik}\rangle = \bigoplus_j \sqrt{d_j/L} |I_j\rangle\rangle$, $L = \sum_j d_j^2$ that maximizes the likelihood [15]. For $M = 1$ and $G = \mathrm{SU}(2)$, U(1) the state $|\varphi\rangle$ gives an average fidelity that approaches 1 as $1/N^2$, while for $|\varphi_{\rm lik}\rangle$ the scaling is $1/N$. On the other hand, Ref. [11] shows that for $M = 1$ $|\varphi_{\rm lik}\rangle$ allows a perfect correction of the misalignment errors with probability of success $p = 1 - 3/(N+1)$, which is not possible for $|\varphi\rangle$. The determination of the best input state to maximize the probability of success, and the study of the probability/fidelity trade-off remain open interesting problems for future research.

V. CONCLUSIONS

In conclusion, in this article we found the optimal storing-retrieving of an unknown group transformation with N input and M output copies, proving the optimality of the incoherent "measure-and-rotate" strategy, in strong contrast with the case of quantum cloning. The result has been extended to the optimal inversion of U, with application to the optimal approximate alignment of reference frames for quantum communication. An interesting development of this work is the analysis of optimal learning when the unknown unitaries do not form a group. This would be the case, for example, of the optimal learning of the unknown unitary transformation appearing in Grover's quantum search algorithm. The question whether coherent quantum strategies can lead to an improvement in these cases remains open and worth investigating.

ACKNOWLEDGMENTS

This work has been supported by by the Italian Ministry of Education through grant PRIN 2008 and by the EC through the projects COQUIT and CORNER. G.C. is grateful to R. Spekkens for useful discussions and to the Risk and Security Study Center of IUSS Pavia for financial support in the early stage of this work. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the province of Ontario through MRI.

[1] B. Julsgaard *et al.*, Nature **432**, 482 (2004).
[2] C. F. Roos *et al.*, Science **304**, 1478 (2004).
[3] P. Rabl, D. DeMille, J. M. Doyle, M. D. Lukin, R. J. Schoelkopf, and P. Zoller, Phys. Rev. Lett. **97**, 033003 (2006).
[4] G. Vidal, L. Masanes, and J. I. Cirac, Phys. Rev. Lett. **88**, 047905 (2002).
[5] M. Sasaki, A. Carlini, and R. Jozsa, Phys. Rev. A **64**, 022317 (2001).
[6] M. Sasaki and A. Carlini, Phys. Rev. A **66**, 022303 (2002).
[7] S. Gammelmark and K. Mølmer, New J. Phys. **11**, 033017 (2009).
[8] V. Scarani *et al.*, Rev. Mod. Phys. **77**, 1225 (2005).
[9] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008).
[10] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).
[11] S. D. Bartlett, T. Rudolph, R. W. Spekkens, and P. S. Turner, New J. Phys. **11**, 063013 (2009).
[12] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Phys. Rev. A **72**, 042338 (2005).
[13] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, Phys. Rev. Lett. **93**, 180503 (2004).
[14] V. Bužek, R. Derka, and S. Massar, Phys. Rev. Lett. **82**, 2207 (1999).
[15] G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, Phys. Rev. A **70**, 062105 (2004); Int. J. Quantum Inf. **4**, 453 (2006).