Optimal Quantum Learning of a Unitary Transformation

Alessandro Bisio,* Giulio Chiribella, Giacomo Mauro D'Ariano, Stefano Facchini, and Paolo Perinotti
QUIT group, Dipartimento di Fisica "A. Volta",
INFN Sezione di Pavia, via Bassi 6, 27100 Pavia, Italy.
(Dated: August 20, 2009)

We prove that the optimal strategy to store an unknown group transformation into a quantum memory is to apply the available uses in parallel on a suitable entangled state. The optimal retrieving strategy is the incoherent, "measure-and-rotate" strategy, in which the quantum memory is measured and a unitary depending on the outcome is performed. The same result holds for approximate re-alignment of reference frames for quantum communication.

PACS numbers:

A quantum memory would be an invaluable resource for Quantum Technology, and extensive experimental effort is in progress for its realization [1, 2, 3]. On a quantum memory one can store any unknown quantum state. Can we exploit a quantum memory to store an unknown quantum transformation, without keeping the device?

Consider the scenario in which a user can dispose of N uses of a black box implementing an unknown unitary transformation U. The user is allowed to exploit such uses today at his convenience, runnning an arbitrary quantum circuit that makes N calls to the black box. Tomorrow, however, the black box will not be available, and the user will be asked to reproduce U on a new input state $|\psi\rangle$ unknown to him. We refer to this scenario as to an instance of *quantum learning* of the unitary U from a finite set of examples. More generally, the user may be requied to reproduce U more than once, i. e. to produce $M \geq 1$ copies of U. In this case it is important to assess how the performance decays with the number of copies required, as in the case of quantum cloning [4].

Let us consider first the case of single input and output copies. Clearly, the only thing that the user can do today is to use the black box on a known (generally entangled) input state $|\varphi\rangle$. After that, what remains available is the output state $|\varphi_U\rangle = (U \otimes I)|\varphi\rangle$, which the user can store in a quantum memory. When the input state $|\psi\rangle$ will be available, the user will send $|\psi\rangle$ and $|\varphi_U\rangle$ to an optimal *retrieving machine*, which extracts U and applies it to $|\psi\rangle$. When $N > 1$ input copies are available, the user must also find the best storing strategy: he 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, generally alternated with other known unitaries, yielding $(UV_{N-1} \dots V_2 U V_1 U \otimes I)|\varphi\rangle$. The most general storing strategy is described by a *quantum circuit board*, i. e. a quantum network with open slots in which the input copies can be inserted [5]. In summary, finding the optimal quantum learning means finding the optimal storing board and the optimal retrieving machine.

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 strategy has the double advantage of avoiding the expensive use of a quantum memory, and of allowing one to reproduce U an unlimited number of times with constant quality. Incoherent strategies are suboptimal for the similar task of quantum cloning [4], and this would suggest that a coherent retrival achieves better performances. Surprisingly enough, we find that the incoherent strategies already achieve to ultimate performance of quantum learning. We analyze the case in which U is a completely unknown unitary in a group G, and we find that the performances of the optimal retrieving machine are equal to those of optimal estimation. For an unknown qubit unitary with N input copies the maximum fidelity approches unit asymptotically as $1/N^2$ and is achieved using N memory qubits. 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 the inverse $U^\dagger$. In this case, we provide the optimal approximate re-alignment of reference frames for the quantum communication scenario considered in Ref. [6].

To derive the optimal learning we use the method of *quantum combs* [5], briefly summarized here. A quantum channel $\mathcal{C}$ from $\mathcal{H}_i$ to $\mathcal{H}_j$ is described by its Choi operator $C = (\mathcal{C} \otimes \mathcal{I})(|I\rangle\rangle\langle\langle I|)$ on $\mathcal{H}_j \otimes \mathcal{H}_i$, where $|I\rangle\rangle$ is the maximally entangled vector $|I\rangle\rangle = \sum_n |n\rangle|n\rangle \in \mathcal{H}_i^{\otimes 2}$. We will use the one-to-one correspondence between bipartite vectors in $|A\rangle\rangle \in \mathcal{H}^{\otimes 2}$ and operators A on $\mathcal{H}$ given by $|A\rangle\rangle = \sum_{m,n} \langle m|A|n\rangle\ |m\rangle|n\rangle$, along with the property

$$(A \otimes I)|B\rangle\rangle = (I \otimes A^T)|B\rangle\rangle \ , \tag{1}$$

holding when $[A, B] = 0$, A^T denoting the transpose of A in the basis $\{|n\rangle\}$. If $\mathcal{D}$ is a channel from $\mathcal{H}_j$ to $\mathcal{H}_k$, the Choi operator of the channel $\mathcal{D} \circ \mathcal{C}$ resulting from the connection $\mathcal{C}$ and $\mathcal{D}$ is given by the *link product*

$$D * C = \mathrm{Tr}_1[(D \otimes I_i)(I_k \otimes C^{T_j})] \ , \tag{2}$$

T_j denoting partial transpose on $\mathcal{H}_j$. A *quantum comb* is the Choi operator associated to a quantum circuit board, and is obtained as the link product of all component circuits. Moreover, viewing states as a special kind of channels with one-dimensional input space, Eq. (2) yields

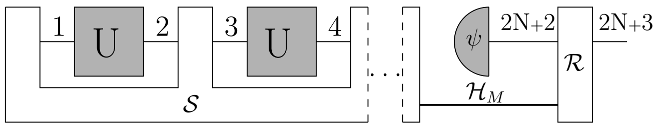

FIG. 1: The learning process is described by a quantum comb R (in white) representing the quantum circuit board, in which the N uses of an oracle U are plugged, along with the state $|\psi\rangle$ (in grey). The wires represent the input-output Hilbert spaces. The output of the first comb is stored in a quantum memory, later used in the retrieval stage when it interacts with the input of the second comb.

$\mathcal{C}(\rho) = C * \rho = \mathrm{Tr}[C(I_1 \otimes \rho^T)]$. A channel $\mathcal{C}$ from $\mathcal{H}_i$ to $\mathcal{H}_j$ is trace-preserving if and only if it satisfies the normalization condition $I_j * C \equiv \mathrm{Tr}_j[C] = I_i$.

We tackle the optimization of learning starting from the case $M = 1$. 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 a positive operator L on the tensor of all Hilbert spaces, and satisfies the normalization condition [5]:

$$\mathrm{Tr}_{2k+1}[L^{(k)}] = I_{2k} \otimes L^{(k-1)} \qquad k = 0, 1, \dots, N+1 \quad (3)$$

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 $C_U = L * |U\rangle\rangle\langle\langle U|^{\otimes N} = \mathrm{Tr}_{i,o}\left[L\left(I_{2N+3} \otimes I_{2N+2} \otimes (|U\rangle\rangle\langle\langle U|^{\otimes N})^T\right)\right]$, according to the definition of link product in Eq. (2).

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 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 average channel fidelity between $\mathcal{C}_U$ and the target unitary, which is nothing but the fidelity between the Choi states C_U/d and $|U\rangle\rangle\langle\langle U|/d$:

$$F = \frac{1}{d^2} \int_G \langle\langle U|\langle\langle U^*|^{\otimes N} L |U^*\rangle\rangle^{\otimes N}|U\rangle\rangle \, \mathrm{d}U \ , \quad (4)$$

U^* being the complex conjugate of U in the computational basis, and $\mathrm{d}U$ 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

$$[L, U_{2N+3} \otimes V^*_{2N+2} \otimes U_o^{*\otimes N} \otimes V_i^{\otimes N}] = 0 \ . \quad (5)$$

where U and V are arbitrary elements of G. Combining Eqs. (3) and (5) we then obtain

$$[L^{(N)}, U_o^{*\otimes N} \otimes V_i^{\otimes N}] = 0 \ . \quad (6)$$

Lemma 1 (Optimality of parallel storage) *The optimal storage of U can be achieved by applying $U^{\otimes N} \otimes I^{\otimes N}$ on a suitable input state $|\varphi\rangle \in \mathcal{H}_{o'} \otimes \mathcal{H}_{i'}$.*

Proof. The learning board $\mathcal{L}$ is obtained by connection of the storing board $\mathcal{S}$ with the retrieving channel $\mathcal{R}$, whence $L = R * S$. Denoting by $\mathcal{H}_M$ the Hilbert space of the quantum memory, $\mathcal{R}$ 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. (3) 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 [5], 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 $\mathcal{H}'_i = \mathcal{H}_i$, and assume $|\Phi\rangle\rangle = |L^{(N)\frac{1}{2}}\rangle\rangle$. Taking $V = I$ in Eq. (6) and using Eq. (1) we get $\left(U_o^{\otimes N} \otimes I_{i,o',i'}\right)|\Phi\rangle\rangle = \left(I_{o,i} \otimes U_{o'}^{T\otimes N} \otimes I_{i'}\right)|\Phi\rangle\rangle$. 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 $|\varphi\rangle = \langle\langle I^{\otimes N}|_{o,i}|\Phi\rangle\rangle \in \mathcal{H}_{o'} \otimes \mathcal{H}_{i'}$. This proves that every storing board gives the same output as a parallel scheme. ■

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. (4), and using the relation $\langle\langle U|\langle\langle U^*|^{\otimes N}(R * S)|U\rangle\rangle|U^*\rangle\rangle^{\otimes N} = \langle\langle U|R|U\rangle\rangle * \langle\langle U^*|^{\otimes N} S|U^*\rangle\rangle^{\otimes N} = \langle\langle U|R|U\rangle\rangle * \rho_U$, which gives

$$F = \frac{1}{d^2} \int_G \langle\langle U|\langle\varphi_U^*|R|U\rangle\rangle|\varphi_U^*\rangle \ \mathrm{d}U. \quad (7)$$

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

$$|\varphi\rangle = \bigoplus_j \sqrt{\frac{p_j}{d_j}} |I_j\rangle\rangle \in \widetilde{\mathcal{H}} \ , \quad (8)$$

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

Proof. Using Eqs. (1) and (6) it is possible to show that the local 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 arbitary 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

$$\mathrm{Tr}_{2N+3}[R] = I_{2N+2} \otimes I_{\widetilde{\mathcal{H}}} \ . \tag{9}$$

Combining the expression of the fidelity (4) with that of the input state (8), it is easy to see that one can always use a covariant retrieving channel, satisfying

$$\left[R, U_{2N+3} \otimes V^*_{2N+2} \otimes \widetilde{U}^* \widetilde{V}'\right] = 0, \tag{10}$$

where $\widetilde{V}' = \bigoplus_j (I_j \otimes V_j)$ acts on $\widetilde{\mathcal{H}}$. We now exploit the decompositions $U \otimes U_j^* = \bigoplus_K \left(U_K \otimes I_{m_K^{(j)}}\right)$ and $V^* \otimes V_j = \bigoplus_L \left(V_L^* \otimes I_{m_L^{(j)}}\right)$, which yield

$$U_{2N+3} \otimes V^*_{2N+2} \otimes \widetilde{U}^* \widetilde{V} = \bigoplus_{K,L} (U_K \otimes V_L^* \otimes I_{m_{KL}}) \ . \tag{11}$$

Here $I_{m_{KL}}$ is given by $I_{m_{KL}} = \bigoplus_{j \in \mathsf{P}_{KL}} \left(I_{m_K^{(j)}} \otimes I_{m_L^{(j)}}\right)$, 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 (10) and (11) then imply

$$R = \bigoplus_{K,L} (I_K \otimes I_L \otimes R_{KL}) \ , \tag{12}$$

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

$$|I\rangle\rangle|\varphi^*\rangle = \bigoplus_{j=0}^{N/2} \sqrt{\frac{p_j}{d_j}} |I\rangle\rangle|I_j\rangle\rangle = \bigoplus_K |I_K\rangle\rangle|\alpha_K\rangle \ , \tag{13}$$

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

$$|\alpha_K\rangle = \bigoplus_{j \in \mathsf{P}_{KK}} \sqrt{\frac{p_j}{d_j}}\ |I_{m_K^{(j)}}\rangle\rangle \ . \tag{14}$$

Exploiting Eqs. (12) and (13), the fidelity (7) can be rewritten as

$$F = \sum_K \frac{d_K}{d^2}\ \langle \alpha_K | R_{KK} | \alpha_K \rangle \ . \tag{15}$$

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 input sytem.*

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. (12) we obtain $\mathrm{Tr}_{2N+3}[R] = \sum_{K,L} \sum_{j \in \mathsf{P}_{KL}} \left(\frac{d_K}{d_j} I_j \otimes I_L \otimes \mathrm{Tr}_{m_K^{(j)}}[R_{KL}^{(j)}]\right)$. Eq. (9) 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

$$\mathrm{Tr}[R_{KK}^{(j)}] \leq \frac{d_j m_K^{(j)}}{d_K} \ . \tag{16}$$

For the fidelity (15) we then have the bound

$$F = \sum_K \frac{d_K}{d^2} \sum_{j,j' \in \mathsf{P}_{KK}} \sqrt{\frac{p_j p_{j'}}{d_j d_{j'}}} \langle\langle I_{m_K^{(j)}} | R_{KK} | I_{m_K^{(j')}} \rangle\rangle \tag{17}$$

$$\leq \sum_K \frac{d_K}{d^2} \left(\sum_{j \in \mathsf{P}_{KK}} \sqrt{\frac{p_j \langle\langle I_{m_K^{(j)}} | R_{KK}^{(j)} | I_{m_K^{(j)}} \rangle\rangle}{d_j}} \right)^2 \tag{18}$$

$$\leq \sum_K \frac{\left(\sum_{j \in \mathsf{P}_{KK}} m_K^{(j)} \sqrt{p_j}\right)^2}{d^2} = F_{\mathrm{est}} \ , \tag{19}$$

having used the positivity of R_{KK} for the first bound and Eq. (16) for the second. Regarding the last equality, it can be proved as follows. First, the Choi operator of the measure-and-prepare strategy is $R_{\mathrm{est}} = \int_G |\hat{U}\rangle\rangle\langle\langle\hat{U}| \otimes |\eta^*_{\hat{U}}\rangle\langle\eta^*_{\hat{U}}|\, \mathrm{d}\,\hat{U}$. Using Eq. (13) 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. (15) then gives $F_{\mathrm{est}} = \sum_K\ |\langle\alpha_K|\beta_K\rangle|^2/d^2$. ■

Using the above result it becomes easy to optimize the input state for storing. In fact, such a state is just the optimal state for the estimation of the unknown unitary U [7], whose expression is 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 [8]. For large number of copies, the optimal input state is given by $|\varphi\rangle \approx \sqrt{4/N}\ \sum_{j=j_{\min}}^{N/2} \frac{\sin(2\pi j/N)}{\sqrt{2j+1}}\ |I_j\rangle\rangle$, with $j_{\min} = 0(1/2)$ 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 [8]. 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 $F \approx 1 - 2\pi^2/(N+1)^2$ [9]. Again, the optimal state can be prepared using only N qubits.

Our result can be extended to the case where the user must reproduce $M > 1$ copies of the unknown unitary U. Indeed, let $\mathcal{C}_U$ be the M-partite channel obtained by the user, and $\mathcal{C}_U^{(1)}$ be the local channel $\mathcal{C}_U^{(1)}(\rho) = \mathrm{Tr}_{\bar{1}}[\mathcal{C}_U(\rho \otimes \Sigma)]$, where $\mathrm{Tr}_{\bar{1}}$ denotes the trace over all spaces except the first. The local channel $\mathcal{C}_U^{(1)}$ 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 $\mathcal{C}_U^{(1)}$ and the unitary U cannot be larger than the optimal fidelity F_{est} of Eq. (19), and the same holds for any local channel $\mathcal{C}_U^{(i)}$, in which all but the i-th input system are discarded. Therefore, the measure-and-prepare strategy is optimal also for the maximization of the single-copy fidelity of all local channels, and such fidelity does not decrease with increasing M. Moreover, our result 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. Again, the optimal retrieving is obtained by measuring the optimal POVM $P_{\hat{U}}$ and by performing $\hat{U}^{\otimes M}$ conditionally on outcome $\hat{U}$. Finally, we note that the same result holds 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.

We conclude by extending our result to the *optimal inversion* of an unknown unitary U. For this task the fidelity of the learning board is $F' = 1/d^2 \int_G \langle\langle U^\dagger|^{\otimes M} \langle\langle U^*|^{\otimes N}\ L'\ |U^\dagger\rangle\rangle^{\otimes M} |U^*\rangle\rangle^{\otimes N} \,\mathrm{d}\,U$, as obtained by substituting U with $U^{\dagger\otimes M}$ in the target of Eq. (4). From this expression it is easy to see that one can always assume $[L', V_{2N+3}^{\otimes M} \otimes U_{2N+2}^{*\otimes M} \otimes U_o^{*\otimes N} \otimes V_i^{\otimes N}] = 0$. 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 re-alignment of reference frames in the quantum communication scenario recently considered in Ref. [6]. In this scenario, the state $|\varphi\rangle \in \widetilde{\mathcal{H}}$ serves as a token of Alice's reference frame, and is sent to the 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 $\sigma_\psi = \int_G |\varphi_U\rangle\langle\varphi_U| \otimes U|\psi\rangle\langle\psi|U^\dagger \,\mathrm{d}\,U$, from which he tries to retrieve the message $|\psi\rangle$ with maximum fidelity $f = \int \mathrm{d}\,\psi\ \langle\psi|\mathcal{R}'(\sigma_\psi)|\psi\rangle \,\mathrm{d}\,\psi$, where $\mathcal{R}'$ is the retrieving channel and $\mathrm{d}\,\psi$ denotes the uniform probability measure over pure states. The maximization of f is equivalent to the maximization of the channel fidelity $F' = \int_G \langle\langle U^\dagger|\langle\varphi_U^*|R'|U^\dagger\rangle\rangle|\varphi_U^*\rangle \,\mathrm{d}\,U$, which is the figure of merit for optimal inversion. It is worth stressing that the state $|\varphi_{\mathrm{fid}}\rangle$ that maximizes the fidelity is not the state $|\varphi_{\mathrm{lik}}\rangle = \bigoplus_j \sqrt{d_j/L}|I_j\rangle\rangle$, $L = \sum_j d_j^2$ that maximizes the likelihood [10]. For $M = 1$ and $G = SU(2), U(1)$ the state $|\varphi_{\mathrm{fid}}\rangle$ gives an average fidelity that approaches 1 as $1/N^2$, while for $|\varphi_{\mathrm{lik}}\rangle$ the scaling is $1/N$. On the other hand, Ref. [6] shows that for $M = 1$ $|\varphi_{\mathrm{lik}}\rangle$ allows a prefect correction of the misalignment errors with probability of success $p = 1 - 3/(N+1)$, which is not possible for $|\varphi_{\mathrm{fid}}\rangle$. 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.

In conclusion, in this paper we found the optimal storing and retrieving of an unknown unitary transformation with N input and M output copies, proving the optimality of incoherent "measure-and-rotate" strategies under general hypotheses. The result has been extended to the optimal inversion of U, with application to the optimal approximate re-alignment of reference frames for quantum communication.

Acknowledgments.— This work has been supported by EC through the project CORNER. GC is grateful to R Spekkens for useful discussions, and to the Risk and Security Study Center of IUSS Pavia for financial support.

* Electronic address: bisio@fisicavolta.unipv.it

[1] B. Julsgaard, J. Sherson, J. I. Cirac, J. Fiurášek, and E. S. Polzik, Nature 432, 482 (2004);
[2] C. F. Roos, M. Riebe, H. Häffner,1 W. Hänsel,1 J. Benhelm, G. P. T. Lancaster, C. Becher, F. Schmidt-Kaler, R. Blatt, 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] V. Scarani *et al*, Rev. Mod. Phys. **77**, 1225 (2005).
[5] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008).
[6] S. D. Bartlett, T. Rudolph, R. W. Spekkens, and P. S. Turner, arXiv:0812.5040.
[7] G. Chiribella, G. M. D'Ariano and M. F. Sacchi, Phys. Rev. A 72 042338 (2005)
[8] G. Chiribella, G. M. D'Ariano, P. Perinotti, M. F. Sacchi, Phys. Rev. Lett **93**, 180503 (2004)
[9] V. Bužek, R. Derka, and S. Massar, Phys. Rev. Lett. **82**, 2207 (1999).
[10] 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).