Memory Effects in Quantum Channel Discrimination

Giulio Chiribella,[1] Giacomo M. D'Ariano,[1,2] and Paolo Perinotti[1]

[1]*QUIT-Quantum Information Theory Group, Dipartimento di Fisica "A. Volta" Università di Pavia, via A. Bassi 6, I-27100 Pavia, Italy*

[2]*Istituto Nazionale di Fisica Nucleare, Sez. di Pavia, via Bassi 6, I-27100 Pavia, Italy*

(Received 2 April 2008; published 27 October 2008)

We consider quantum-memory assisted protocols for discriminating quantum channels. We show that for optimal discrimination of memory channels, memory assisted protocols are needed. This leads to a new notion of distance for channels with memory, based on the general theory of quantum testers. For discrimination and estimation of sets of independent unitary channels, we prove optimality of parallel protocols among all possible architectures.

DOI: 10.1103/PhysRevLett.101.180501 PACS numbers: 03.67.Dd, 03.65.Ta, 03.67.Hk, 03.67.Lx

The problem of discrimination between quantum channels has been recently considered in quantum information [1–7]. For example, in Ref. [6] an application of discrimination of unitary channels as oracles in quantum algorithms is suggested. The optimal discrimination is achieved by applying the unknown channel locally on some bipartite input state of the system with an ancilla, and then performing some measurement at the output. A natural extension to multiple uses is obtained by applying the uses in *parallel* to a global input state. However, more generally, one can apply the uses partly in parallel and partly in series, even intercalated with other fixed transformations, as in Ref. [8]. Moreover, when the multiple uses are correlated—i.e., for *memory channels*—the uses can be applied either in parallel or in a causal fashion (see Fig. 1). In this Letter we show that this *causal* scheme is necessary, whereas it is not needed for independent uses of unitary channels (the case of nonunitary channels remains an open problem).

Memory channels [9–13] attracted increasing attention in the last years. They are quantum channels whose action on the input state at the n-th use can depend on the previous $n-1$ uses through a quantum ancilla. Optimal discrimination of two memory channels is crucial to assess whether a cryptographic protocol is concealing [14], to discriminate among different strategies of an opponent in a quantum game [15], for minimization of oracle calls in quantum algorithms, and for applications such as quantum illumination [7].

We will provide an example showing that a pair of memory channels can be perfectly discriminable, even though they never provide orthogonal output states when applied to the same global input state. This new causal setup provides the most general discrimination scheme for multiple quantum channels, and this fact leads to a new notion of distance between channels.

In the case of two unitary channels, optimal parallel discrimination with N uses was derived in Refs. [1,2], and in Ref. [5] a causal scheme without entanglement was proved to be equivalently optimal. In the following, we will prove the optimality of both schemes for discrimination of unitaries. We will generalize this result to discrimination of sequences of unitaries, and to estimation with multiple copies. Differently from the case of memory channels, we will prove that for all these examples causal schemes are not necessary.

It is convenient to represent a channel $\mathcal{C}$ by means of its Choi operator C defined as follows

$$C := (\mathcal{C} \otimes \mathcal{I})(|I\rangle\rangle\langle\langle I|), \qquad (1)$$

for a channel $\mathcal{C}$ with input or output states in $\mathcal{H}_{\text{in/out}}$, respectively, where $|I\rangle\rangle := \sum_n |n\rangle|n\rangle \in \mathcal{H}_{\text{in}}^{\otimes 2}$, $\{|n\rangle\}$ being an orthonormal basis for $\mathcal{H}_{\text{in}}$. In this representation complete positivity of $\mathcal{C}$ is simply $C \geq 0$) and the trace-preserving constraint is $\text{Tr}_{\text{out}}[C] = I_{\text{in}}$.

In a memory channel with N inputs and N outputs labeled as in Fig. 1, the causal independence of output $2n+1$ on input $2m$ with $m > n$ is translated to the following recursive property [8] of the Choi operator $C =: C^{(N)}$

$$\text{Tr}_{2n-1}[C^{(n)}] = I_{2n-2} \otimes C^{(n-1)}, \quad \forall\ 1 \leq n \leq N, \qquad (2)$$

where conventionally $C^{(0)} = 1$. A memory channel can be measured by a causal scheme as in Fig. 2, which is described by a *tester* [8], i.e., a set of positive operators $P_i \geq 0$ such that the probability of outcome i while testing the channel $\mathcal{C}$ is provided by the generalized Born rule

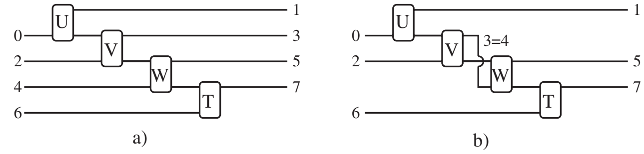

FIG. 1. Different usage schemes of a general memory channel, where the boxes U, V, W, T denote interactions of systems with ancillae. (a) *Parallel scheme* (a multipartite input state is evolved through the channel). (b) A particular case of *causal scheme* (the output of some use of the channel is fed into a successive use).

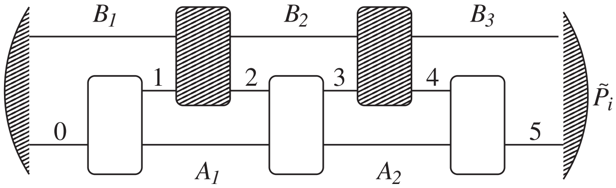

FIG. 2. The most general scheme for the connection of a memory channel to a quantum circuit corresponding to a *tester*. The memory channel is represented by its isometric gates (white boxes) which denote interaction of quantum systems (inputs are labeled by even integers and outputs by odd integers) with the ancillae A_1 and A_2. The tester is represented by dashed boxes, including the preparation phase (joint input state of system 0 and ancilla B_1) and the final measurement stage represented by the POVM $\{\tilde{P}_i\}$.

$$p(i|\mathcal{C}) := \mathrm{Tr}[P_i C]. \tag{3}$$

The notion of a tester is an extension of that of POVM, which describes the statistics of customary measurements on quantum states. The normalization of probabilities for testers on memory channels with N input-output systems is equivalent to the following recursive property, analogous to that in Eq. (2)

$$\begin{aligned} &\sum_i P_i = I_{2N-1} \otimes \Xi^{(N)}, \\ &\mathrm{Tr}_{2n-2}[\Xi^{(n)}] = I_{2n-3} \otimes \Xi^{(n-1)}, \quad \forall\ 2 \leq n \leq N, \\ &\mathrm{Tr}[\Xi^{(1)}] = 1, \end{aligned} \tag{4}$$

One can prove [8] that any tester can be realized by a concrete measurement scheme of the class represented in Fig. 2.

Mathematical structures analogous to Eqs. (2) and (4) have been introduced in Ref. [15] to describe strategies in a quantum game.

Every tester $\{P_i\}$ can be written in terms of a usual POVM $\{\tilde{P}_i\}$ and the *normalization operator* $\Xi^{(N)}$ as

$$P_i = (I \otimes \Xi^{(N)(1/2)})\tilde{P}_i(I \otimes \Xi^{(N)(1/2)}), \tag{5}$$

and for every memory channel $\mathcal{C}$ the generalized Born rule rewrites as the usual one in terms of the state

$$\tilde{C} := (I \otimes \Xi^{(N)(1/2)})C(I \otimes \Xi^{(N)(1/2)}). \tag{6}$$

The state $\tilde{C}$ corresponds to the output system-ancilla state in Fig. 2 after the evolution through all boxes of both the tester and the memory channel, on which the final POVM $\{\tilde{P}_i\}$ is performed [16].

The standard discriminability criterion for channels is the following. Two channels $\mathcal{C}_0$ and $\mathcal{C}_1$ on a d-dimensional system are perfectly discriminable if there exists a pure state $|\Psi\rangle\rangle$ in dimension d^2 such that $\rho_i' = \mathcal{C}_i \otimes \mathcal{I}(|\Psi\rangle\rangle \times \langle\langle\Psi|)$ with $i = 0, 1$ are orthogonal (every joint mixed state with an ancilla of any dimension can be purified with an ancilla of dimension d). Here we use the notation $|\Psi\rangle\rangle := \sum_{m,n}\Psi_{mn}|m\rangle|n\rangle$ which associates an operator Ψ to a bipartite vector. It is easy to show [17] that orthogonality between ρ_0' and ρ_1' is equivalent to

$$C_0(I \otimes \rho)C_1 = 0, \tag{7}$$

where $\rho := \Psi^*\Psi^T$, where Ψ^* and Ψ^T denote the complex conjugate and transpose of Ψ in the canonical basis $\{|n\rangle\}$, respectively. The criterion in Eq. (7), however, is too restrictive for memory channels. Indeed, the correct condition for perfect discriminability of two memory channels $\mathcal{C}_i$ with $i = 0, 1$ is equivalent to the existence of a tester $\{P_i\}$ with $i = 0, 1$, such that

$$\mathrm{Tr}[P_i C_j] = \delta_{ij}, \tag{8}$$

which means that the two channels can be perfectly discriminated by a measurement scheme as that of Fig. 2. Using Eqs. (5) and (6), Eq. (8) becomes $\mathrm{Tr}[\tilde{P}_i\tilde{C}_j] = \delta_{ij}$, whence the states $\tilde{C}_i$ with $i = 0, 1$ are orthogonal, and the same derivation as for Eq. (7) leads to

$$C_0(I \otimes \Xi^{(N)})C_1 = 0, \tag{9}$$

with $\Xi^{(N)}$ as in Eq. (4). In Eq. (9) the identity operator acts only on space $2N-1$, differently from Eq. (7) where it acts on all output spaces.

It is interesting to analyze the special case of memory channels made of sequences of independent channels $(\mathcal{C}_{ij})_{1\leq j\leq N}$ and $i = 0, 1$ (in Fig. 2 the memory channel is replaced by an array of channels without the ancillae A_1 and A_2). The condition for perfect discriminability is the same as Eq. (9) with C_0 and C_1 replaced by $\bigotimes_j C_{ij}$ for $i = 0, 1$, respectively. In terms of a Kraus form $\mathcal{C}_i = \sum_m K_{im}K_{im}^\dagger$ Eq. (9) becomes the orthogonality condition $\langle\langle K_{0m}|(I \otimes \Xi^{(N)})|K_{1n}\rangle\rangle = 0$, which for the sequences of maps becomes

$$\bigotimes_{j=1}^{N}\langle\langle K_{0m_j}^j|(I \otimes \Xi^{(N)})\bigotimes_{j'=1}^{N}|K_{1n_{j'}}^{j'}\rangle\rangle = 0 \tag{10}$$

for all choices of indices (m_j), $(n_{j'})$, where K_{im}^j are the Kraus operators for the channel $\mathcal{C}_{ij}$. For sets composed by single channels $\mathcal{C}_i$ with $i = 0, 1$, the condition becomes simply the existence of a state ρ such that

$$\mathrm{Tr}[\rho K_{0j}^\dagger K_{1k}] = 0, \quad \forall\ j, k, \tag{11}$$

and the minimum rank of such state ρ determines the amount of entanglement required for discrimination.

We now provide an example of memory channels $\mathcal{C}_0$ and $\mathcal{C}_1$ that cannot be discriminated by a parallel scheme, but can be discriminated with a tester. Their action on the joint state ρ on $\mathcal{H}_0 \otimes \mathcal{H}_1$ provides the output on $\mathcal{H}_1 \otimes \mathcal{H}_3$ as follows

$$\mathcal{C}_0(\rho) = \frac{1}{d^2}\sum_{p,q=0}^{d-1} |p,q\rangle\langle p,q| \otimes W_{p,q}\mathrm{Tr}_0[\rho]W_{p,q}^\dagger, \quad (12)$$

$$\mathcal{C}_1(\rho) = \frac{I}{d^2} \otimes |0\rangle\langle 0|. \quad (13)$$

$|p,q\rangle$ being an orthonormal basis in a d^2 dimensional Hilbert space, and the unitaries $W_{p,q} := Z^p U^q$ are the customary shift-and-multiply operators, with $Z|n\rangle = |n+1\rangle$ and $U|n\rangle = e^{(2\pi i/d)n}|n\rangle$. We will now show that the two channels are discriminable with a causal setup and not with a parallel one. Their Choi operators are

$$C_0 = \frac{1}{d^2}\sum_{p,q=1}^{d-1} |W_{p,q}\rangle\rangle\langle\langle W_{p,q}|_{32} \otimes |p,q\rangle\langle p,q|_1 \otimes I_0, \quad C_1 = \frac{1}{d^2}|0\rangle\langle 0|_3 \otimes I_{210}, \quad (14)$$

where the output spaces 1,3 have dimension d^2 and d, respectively. It is clear that the output states in Eqs. (12) and (13) are never orthogonal; hence, the channels are not discriminable by a parallel scheme. This is actually the same conclusion that can be drawn from the criterion in Eq. (7). Indeed, suppose that the channels are perfectly discriminable in parallel, then there exists ρ such that

$$C_0(I_{13} \otimes \rho_{02})C_1 = C_0C_1(I_{13} \otimes \rho_{02}) = 0, \quad (15)$$

where the second equality comes from the expression of C_1 in Eq. (14). Tracing both sides on the output spaces 1 and 3 one has $\mathrm{Tr}_{13}[C_0C_1]\rho = 0$. However,

$$\mathrm{Tr}_{13}[C_0C_1] = \frac{I}{d^2} \quad (16)$$

whence $\rho = 0$, which is absurd. We now show a simple causal scheme which allows perfect discrimination of the same channels. The first use of the channel is applied to any state $|\psi\rangle\langle\psi|$, then the measurement with POVM $\{|p,q\rangle \times \langle p,q|\}$ is performed at the output on system 1. Depending on the outcome $\bar{p}$, $\bar{q}$, the second use of the channel is applied to the state $W_{\bar{p},\bar{q}}^\dagger|1\rangle\langle 1|W_{\bar{p},\bar{q}}$. It is clear that the second output of channel $\mathcal{C}_0$ is the state $|1\rangle\langle 1|$, whereas the second output of channel $\mathcal{C}_1$ is $|0\rangle\langle 0|$. Indeed, this agrees with the criterion in Eq. (9), which is satisfied with

$$\Xi^{(2)} = \sum_{p,q} W_{p,q}^T|1\rangle\langle 1|W_{p,q}^* \otimes |p,q\rangle\langle p,q| \otimes |0\rangle\langle 0|. \quad (17)$$

This example highlights the need of using a causal scheme in order to discriminate between memory channels. In general, optimal causal discrimination implies a notion of distance between memory channels different from the usual distance between channels. Indeed, discriminability by parallel schemes is assessed by the usual cb-norm distance [18–20], which can be rewritten as follows (see, e.g., Ref. [3])

$$D_{cb}(\mathcal{C}_0, \mathcal{C}_1) = \max_\rho ||(I \otimes \rho^{(1/2)})\Delta(I \otimes \rho^{(1/2)})||_1, \quad \Delta := C_0 - C_1, \quad (18)$$

where the maximum is over all states ρ, and $||X||_1 := \mathrm{Tr}[\sqrt{X^\dagger X}]$ denotes the trace-norm. One has $D_{\mathrm{cb}}(\mathcal{C}_0, \mathcal{C}_1) \leq 2$, with the equality holding for perfectly discriminable channels, satisfying the criterion in Eq. (7). For memory channels the discriminability is equivalent to discriminability of states in Eq. (6), leading to the following definition

$$D(\mathcal{C}_0, \mathcal{C}_1) := \max_{\Xi^{(N)}} ||(I \otimes \Xi^{(N)(1/2)})\Delta(I \otimes \Xi^{(N)(1/2)})||_1, \quad (19)$$

where the maximum is over all $\Xi^{(N)}$ satisfying conditions (4). Clearly $D(\mathcal{C}_0, \mathcal{C}_1) \leq 2$, with equality holding if and only if Eq. (9) is satisfied. For $N = 1$ Eq. (19) reduces to (18), yielding $D(\mathcal{C}_0, \mathcal{C}_1) = D_{\mathrm{cb}}(\mathcal{C}_0, \mathcal{C}_1)$.

The easiest application of testers is the discrimination of sequences of unitary channels (T_j) and (V_j), with $j = 1, \ldots, N$. Without loss of generality we can always reduce to the discrimination of the sequence $(U_j) := (T_j^\dagger V_j)$ from the constant sequence (I). Let us first consider the case of sequences of two unitaries. By referring to the scheme in Fig. 2 we can restate the problem as the discrimination of $W^\dagger(U_1 \otimes I)W(U_2 \otimes I)$ from I on a bipartite system, where W describes the interaction with an ancillary system. It is well known that optimal discriminability of a unitary X from the identity is related to the angular spread $\Theta(X)$, defined as the maximum relative phase between two eigenvalues of X [1,2]. Apart from the degenerate case in which X has only two different eigenvalues, the discriminability of X from I is given by the quantity $\max\{0, \cos\Theta(X)/2\} \geq 0$, which is zero for $\Theta(X) \geq \pi$, corresponding to perfect discriminability [21]. Since unitary conjugation preseves $\Theta(X)$ and the angular spread of the product of two unitaries X, Y satisfies the following bound [22]

$$\Theta(XY) \leq \Theta(X) + \Theta(Y), \quad (20)$$

and finally $\Theta(X \otimes Y) = \Theta(X) + \Theta(Y)$, one has that $\Theta[W^\dagger(U_1 \otimes I)W(U_2 \otimes I)] \leq \Theta(U_1 \otimes U_2)$, then no causal scheme can outperform the parallel one. By induction, one can prove that this is true for sequences of any length N. Indeed, defining X_{N-1} as the product of the tester unitaries alternated with $U_j \otimes I$ for $1 \leq j < N$, if $\Theta(X_{N-1}) \leq \Theta(\bigotimes_{j=1}^{N-1} U_j)$ holds true, then it holds also for N, due to Eq. (20). By the same argument, one can also prove that the sequential scheme of Ref. [5] equals the performances of the parallel scheme, since there always exists T such that $\Theta(UTVT^\dagger) = \Theta(U \otimes V)$ (indeed it is sufficient that T transforms the eigenbasis of V into that of U, suitably matching the eigenvalues). Therefore, the schemes of Refs. [1,2,5] are optimal also for discriminating sequences of unitaries. Notice that this also includes the case of discrimination of two different permutations of a sequence of unitary transformations.

Another situation in which a parallel scheme already performs optimally is the case of estimation of unitary transformations U_g, $g \in G$ which make a unitary representation of the group G. For N uses of the unitary U_g the Choi operator is

$$R_g^{(N)} = R_g^{\otimes N}, \qquad R_g = (U_g \otimes I)|I\rangle\rangle\langle\langle I|(U_g^\dagger \otimes I). \quad (21)$$

The probability density of estimating h for actual element g is $p(h|g) = \mathrm{Tr}[P_h R_g^{\otimes N}]$. As a figure of merit for estimation one typically considers a cost function $c(h, g)$ averaged on h, with $c(h, g) = c(fh, fg)$ $\forall f \in G$ (the cost depends only on distance, not on specific location)

$$C_g(p) = \int_G \mu(dh)c(h, g)p(h|g), \quad (22)$$

where $\mu(dg)$ is the invariant Haar measure on G. The optimal density p is the one minimizing $\hat{C}(p) := \max_{g\in G} C_g(p)$. For every density $p(h|g)$ there exists a *covariant* one $p_c(h|g) = p_c(fh|fg)$ $\forall f \in G$ which can be obtained as the average $p_c(h|g) := \overline{p(fh|fg)}$ over $f \in G$ (practically this corresponds to randomly transforming the input before measuring and processing the output accordingly). Since $\hat{C}(p_c) = \bar{C}(p) \le \hat{C}(p)$, then the optimal density minimizing both costs $\hat{C}$ and $\bar{C}$ can be chosen as covariant. Now, since $p_c(h|g) = p_c(e|gh^{-1})$ (e denoting the identity element in G), this means that the optimal tester must be of the covariant form

$$P_h = (U_h \otimes I)^{\otimes N} P_e (U_h^\dagger \otimes I)^{\otimes N}. \quad (23)$$

For such P_h, the normalization $\int_G \mu(dh)P_h = I \otimes \Xi^{(N)}$ implies the commutation $[I \otimes \Xi^{(N)}, (U_h \otimes I)^{\otimes N}] = 0$, whence the POVM $\tilde{P}_h$ in Eq. (5) is itself covariant. The optimal tester problem is then equivalent to the optimal state estimation in the orbit $(I \otimes \Xi^{(N)(1/2)})R_g^{\otimes N}(I \otimes \Xi^{(N)(1/2)})$. This proves that the optimal estimation of U_g with $g \in G$ compact group can be reduced to a covariant state estimation problem, and the optimal parallel scheme of Ref. [23] is optimal among all possible architectures.

In conclusion, we considered the role of memory effects in the discrimination of memory channels and of customary channels with multiple uses. We used the new notion of *tester* [8], which describes any possible scheme with parallel, sequential, and combined setup of the tested channels. We provided an example of discrimination of memory channels which cannot be optimized by a parallel scheme, and for which the optimal discrimination is achieved by a sequential scheme. The new testing of memory channels corresponds to a new notion of distance between channels. Finally, using the theory of testers we showed that for the purpose of unitary channel discrimination and estimation with multiple uses, the optimal schemes are parallel.

This work has been supported by the EC through the project CORNER.

[1] G. M. D'Ariano, P. Lo Presti, and M. G. A. Paris, Phys. Rev. Lett. **87**, 270404 (2001).
[2] A. Acín, Phys. Rev. Lett. **87**, 177901 (2001).
[3] M. F. Sacchi, Phys. Rev. A **71**, 062340 (2005).
[4] G. M. D'Ariano, M. F. Sacchi, and J. Kahn, Phys. Rev. A **72**, 052302 (2005).
[5] R. Duan, Y. Feng, and M. Ying, Phys. Rev. Lett. **98**, 100503 (2007).
[6] A. Chefles, A. Kitagawa, M. Takeoka, M. Sasaki, and J. Twamley, J. Phys. A **40**, 10 183 (2007).
[7] S. Lloyd, Science **321**, 1463 (2008).
[8] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008).
[9] C. Macchiavello and G. M. Palma, Phys. Rev. A **65**, 050301(R) (2002).
[10] G. Bowen and S. Mancini, Phys. Rev. A **69**, 012306 (2004).
[11] D. Kretschmann and R. F. Werner, Phys. Rev. A **72**, 062323 (2005).
[12] N. Datta and T. C. Dorlas, J. Phys. A **40**, 8147 (2007).
[13] M. B. Plenio and S. Virmani, Phys. Rev. Lett. **99**, 120504 (2007).
[14] G. M. D'Ariano, D. Kretschmann, D. Schlingemann, and R. F. Werner, Phys. Rev. A **76**, 032328 (2007).
[15] G. Gutoski and J. Watrous, in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computation (STOC)* (2007), p. 565.
[16] This is the *delayed measurement principle*, stating that for any quantum setup involving measurements, there is an equivalent one in which all measurements are postponed at the final stage.
[17] The orthogonality condition is $0 = \rho_0' \rho_1' = (I \otimes \Psi^T)C_0(I \otimes \Psi^* \Psi^T)C_1(I \otimes \Psi^*) = H_0^\dagger H_0 H_1^\dagger H_1$, with $H_i = C_i^{(1/2)}(I \otimes \Psi^*)$, which holds iff $H_0 H_1^\dagger = 0$.
[18] V. I. Paulsen, *Completely Bounded Maps and Operator Algebras*, Pitman Research Notes in Math. 146. (Longman Scientific & Technical, Harlow, 1996).
[19] D. Aharonov, A. Kitaev, and N. Nisan, in *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computation (STOC)* (1997), p. 20.
[20] This distance is referred to in the literature as cb-norm distance, since it is induced by the norm of complete boundedness [18] (cb-norm for short, also defined as "diamond norm" in Ref. [19]).
[21] In the degenerate case, if $\Theta(U) + \Theta(V) > \pi$, it is always possible to find T such that $\Theta(UTVT^\dagger) = \pi$.
[22] A. M. Childs, J. Preskill, and J. Renes, J. Mod. Opt. **47**, 155 (2000).
[23] G. Chiribella, G. M. D'Ariano, and M. F. Sacchi, Phys. Rev. A **72**, 042338 (2005).