PHYSICAL REVIEW A **69**, 010303(R) (2004)

Transition behavior in the channel capacity of two-quibit channels with memory

Chiara Macchiavello,[1] G. Massimo Palma,[2] and S. Virmani[3]

[1]*Dipartimento di Fisica "A.Volta," and Istituto Nazionale per la Fisica della Materia (INFM), Via Bassi 6, I-27100 Pavia, Italy*

[2]*INFM-NEST, Dipartimento di Tecnologie dell'Informazione, Università degli Studi di Milano, via Bramante 65, I-26013 Crema (CR), Italy*

[3]*QOLS, Department of Physics, Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2BW, United Kingdom and Department of Computer Science, University of Hertfordshire, College Lane, Hatfield AL10 9AB, United Kingdom*

(Received 3 July 2003; published 15 January 2004)

We prove that a general upper bound on the maximal mutual information of quantum channels is saturated in the case of Pauli channels with an arbitrary degree of memory. For a subset of such channels we explicitly identify the optimal signal states. We show analytically that for such a class of channels entangled states are indeed optimal above a given memory threshold.

DOI: 10.1103/PhysRevA.69.010303 PACS number(s): 03.67.Hk

The study of the optimal coding and decoding of information in quantum systems has a long history [1]. The advent of present day quantum information theory [2,3] has not only revived interest in the subject but has also opened new problems. A key open question concerns the additivity of channel capacity when entangled states are used as signals. Although entanglement is a ubiquitous ingredient in nearly all quantum information processing protocols and algorithms, it is often regarded as being very fragile in the presence of environmental noise. This has led to the belief that in most circumstances the use of entanglement is not advantageous in the reliable transmission of classical information through quantum channels. For those memoryless channels (i.e., ones in which the noise acting on consecutive uses of the channel is uncorrelated) that have been studied so far, this is indeed the case. This was first proven analytically for two qubits in the case of the depolarizing channel [4], where isotropic noise acts on individual qubits, and then extended to a more general form of memoryless unital channel [5]. There has also been interesting recent work demonstrating that the potential additivity of channel capacities is equivalent to other well known additivity conjectures in quantum information theory [6]. The scenario changes when the channel is not memoryless, i.e., when the noise acting on consecutive uses is partially correlated. This phenomenon is not uncommon in physical situations, when the statistical properties of the physical source of noise can be time-dependent. The problem of quantum channels with memory was first introduced in Ref. [7], where, for the case of depolarizing channels with memory, it was shown that the use of entangled states enhances the mutual information. In Ref. [7] input states taken from a certain ansatz were considered, and it was shown that within this ansatz entangled states allow for the transmission of a larger amount of reliable information. However, it was not proved analytically that this ansatz is indeed optimal. Further results bounding the asymptotic capacities of noisy channels with memory have also recently been derived [8].

Here we prove the optimality of a set of entangled input signal states for a class of Pauli channels. To this end we will first obtain an upper bound on the channel capacity. We will then show that for the general case of Pauli channels with an arbitrary degree of memory this bound is saturated by states of minimal output entropy. For a class of Pauli channels we will derive these states explicitly. They turn out to be entangled above a given memory threshold and product states below it.

In order to set the scenario let us first consider a single qubit channel that is a random implementation of the Pauli transformations

$$\rho \to \sum_{i=0}^{3} q_i \sigma_i \rho \sigma_i , \tag{1}$$

where the q_i's give a probability distribution, and the σ_i's are the Pauli matrices according to the following convention:

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$
$$\sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{2}$$

We will sometimes refer to the Pauli matrices as the Pauli group, even though extra phases are required in order to make the matrices closed under matrix multiplication. However, since these phases cancel out when considering transformations of density operators, as

$$(e^{i\theta}\sigma_i)\rho(e^{i\theta}\sigma_i)^{\dagger} = \sigma_i \rho \sigma_i , \tag{3}$$

we will freely make this abuse of terminology.

In the typical memoryless channel scenario, an understanding of the action of an individual use, such as the one described in Eq. (1), is sufficient to fully describe the operation of the channel. However, the possibility of repeated access to the channel opens the question of optimizing the choice of signal states, including the ones that are entangled over many uses of the channel. This leads to questions concerning the additivity of channel capacities, and whether entangled inputs and output measurements can lead to improvements in information transmission. However, in the manner of Ref. [7] we would like to consider repeated applications of a single qubit channel that are *not* independent. In particular, we will consider a two qubit channel that is almost equivalent to two independent uses of the single qubit channel (1), aside from a memory factor $\mu\in[0,1]$ that introduces correlations, i.e.,

$$\rho\rightarrow\sum_{i,j=0}^{3} p_{ij}\sigma_i\otimes\sigma_j\rho\sigma_i\otimes\sigma_j, \qquad (4)$$

where

$$p_{ij}=(1-\mu)q_iq_j+\mu q_i\delta_{ij}. \qquad (5)$$

We can see that this evolution can be considered as two independent applications of Eq. (1), except for an additional effect due to the degree of memory μ, which with some probability forces the same Pauli transformation to be repeated in the second use of the channel.

We would like to compute the maximum amount of information that can be transmitted through a noisy channel of the form (5), and investigate how the use of entangled inputs in the two uses of the channel may improve its communication performance. To do this we will show that this is equivalent to finding the input pure state with minimal output entropy.

The maximum mutual information of a general quantum channel $\mathcal{E}$ is given by the Holevo-Schumacher-Westmoreland bound [9]

$$\chi(\mathcal{E})=\max_{\{p_i,\rho_i\}} S\left(\mathcal{E}\left(\sum_i p_i\rho_i\right)\right)-\sum_i p_iS(\mathcal{E}(\rho_i)), \qquad (6)$$

where $S(\omega)=-\mathrm{Tr}(\omega\log\omega)$ is the von Neumann entropy of the density operator ω and the maximization is performed over all input ensembles $\{p_i,\rho_i\}$ into the channel (ρ_i's are the input states on which classical information is encoded, and are transmitted with prior probabilities p_i). Note that this bound incorporates a maximization over all POVM (positive operator-valued measures) measurements at the receiver, including collective ones over multiple uses of the channel.

In our scenario the ρ_i's describe states of two qubits, and so we will refer to the maximum mutual information $\chi(\mathcal{E})$ as the *two-qubit* capacity of the channel. We will find it convenient to use the symbol $\rho_*(\mathcal{E})$ to denote a chosen input state that gives minimal output entropy when transmitted through the channel $\mathcal{E}$. As the maximally mixed state gives the largest possible entropy for any system, the formula (6) can clearly be bounded from above by

$$\chi(\mathcal{E})\leqslant\log_2(4)-S(\mathcal{E}(\rho_*))=2-S(\mathcal{E}(\rho_*)) \qquad (7)$$

for any two-qubit channel. We will now see that this upper bound can be achieved by any two-qubit channel whose action consists of random tensor products of Pauli transformations. The argument that we use to demonstrate this can be applied to any channel that is covariant with respect to an irreducible representation of a compact group, and has been also noted in Ref. [10]. The key ingredients will be the facts that the Pauli matrices (a) form an irreducible representation of a group, and (b) either commute or anticommute. Indeed, as these are essentially the only ingredients required, the same argument can easily be modified to multiqubit channels whose actions consist of random tensor products of Pauli matrices.

Let us consider an ensemble of input states given by the sixteen states defined by $\rho_{ij}:=\sigma_i\otimes\sigma_j\rho_*\sigma_i\otimes\sigma_j$, each with the same input probability 1/16. The commutation relations of the Pauli matrices imply that any channel $\mathcal{E}$ of the form (5) is covariant with respect to the Pauli rotations

$$\mathcal{E}(\sigma_i\otimes\sigma_j\rho_*\sigma_i\otimes\sigma_j)=\sigma_i\otimes\sigma_j\mathcal{E}(\rho_*)\sigma_i\otimes\sigma_j. \qquad (8)$$

As entropy is invariant under unitary transformations, we can immediately write

$$S(\mathcal{E}(\rho_*))=S(\mathcal{E}(\rho_{ij})), \qquad (9)$$

and therefore each of the states ρ_{ij} will also give the same minimal output entropy as ρ_*. Furthermore, the fact that the group of matrices $\{\sigma_i\otimes\sigma_j\}$ is an irreducible representation means that the ensemble will give an average output state that is maximally mixed [11]

$$\mathcal{E}\left(\sum_{ij}\frac{1}{16}\rho_{ij}\right)=\sum_{ij}\frac{1}{16}\sigma_i\otimes\sigma_j\mathcal{E}(\rho)\sigma_i\otimes\sigma_j=\frac{\mathbb{1}}{4}. \qquad (10)$$

Inserting Eqs. (9) and (10) into Eq. (6) we can see that the upper bound, Eq. (7) is attained by the input ensemble of states $\rho_{ij}:=\sigma_i\otimes\sigma_j\rho_*\sigma_i\otimes\sigma_j$ with equal prior probabilities. This means that to optimize the information transmission of our channel, we merely need to search for the input state that minimizes the output entropy. We will refer to any such state as an *optimal* input state.

In Ref. [7] a specific form of memory channel was investigated, where the weights in Eq. (5) were fixed by

$$q_0=x;\quad q_1=q_2=q_3=\frac{1-x}{3} \qquad (11)$$

and the degree of memory μ was allowed to take any value in the interval [0,1]. An ansatz for the form of the optimal input state was conjectured, but a full analytic proof is still lacking.

Consequently, here we will focus our attention on a kind of memory channel for which we can give an entirely analytic solution. The form of the channel is characterized by the following parameters in Eq. (5):

$$q_0=q_1=p,\quad q_2=q_3=q, \qquad (12)$$

where $q=(1-2p)/2$ and $0\leqslant p\leqslant 1$.

In order to identify the optimal input states we will first show that we can restrict our attention to input states that are invariant under the symmetry group $\{\sigma_0\otimes\sigma_0,\sigma_1\otimes\sigma_1\}$. The technique that we will use may be generalized to many other channels with a suitable structure [12]. Let us first consider the following modification of the channel $\mathcal{E}$: first rotate the input state by $\sigma_1\otimes\sigma_1$, and then act with $\mathcal{E}$. Let us call this new channel $\mathcal{E}':=\mathcal{E}\circ(\sigma_1\otimes\sigma_1)$. Using the standard relations for the Pauli group: $\sigma_0\sigma_1=\sigma_1$, $\sigma_1\sigma_1=\sigma_0$, $\sigma_2\sigma_1=-i\sigma_3$ and $\sigma_3\sigma_1=i\sigma_2$, and the fact that the Pauli matrices are Hermitian, we can see that preoperating with $\sigma_1 \otimes \sigma_1$ does not make any difference to the action of this channel, and therefore

$$\mathcal{E}' = \mathcal{E}. \tag{13}$$

We can also trivially say the same thing if we preoperate with the identity operation $\sigma_0 \otimes \sigma_0$. Let us now consider the following "averaging" preoperation:

$$\mathcal{F}(\rho) = \tfrac{1}{2}(\sigma_0 \otimes \sigma_0 \rho \sigma_0 \otimes \sigma_0 + \sigma_1 \otimes \sigma_1 \rho \sigma_1 \otimes \sigma_1). \tag{14}$$

From the arguments above follows immediately the equality

$$\mathcal{E} \circ \mathcal{F} = \mathcal{E}, \tag{15}$$

i.e., preoperating on our state with $\mathcal{F}$ does not affect the operation of the above channel. Since by construction $\mathcal{F}$ corresponds to averaging over the group $\{\sigma_0 \otimes \sigma_0, \sigma_1 \otimes \sigma_1\}$ we need only to consider input states that are invariant under it. Let us denote by R the whole set of two qubit density matrices. We are looking for the explicit form of an input state $\rho \in R$ which minimizes the output entropy. If we find such an optimal state ρ_*, then by the above arguments the input state $\mathcal{F}(\rho_*)$ will also give the same output entropy, and will therefore also be optimal. This means that instead of looking for the optimal state in R, we can instead restrict our search to finding an optimal state from the restricted set $\mathcal{F}(R)$. Since the optimal state $\rho_* \in \mathcal{F}(R)$ that minimizes the output entropy is by construction invariant under the group $\{\sigma_0 \otimes \sigma_0, \sigma_1 \otimes \sigma_1\}$ it can easily be checked that in the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ represented by the eigenvectors of $\sigma_1 \otimes \sigma_1$ it must take the form

$$\rho_* = \begin{pmatrix} a & 0 & 0 & c \\ 0 & d & f & 0 \\ 0 & f^* & e & 0 \\ c^* & 0 & 0 & b \end{pmatrix}. \tag{16}$$

From the form of ρ_* follows that it is a convex combination of pure states of the form

$$\alpha|00\rangle + \beta|11\rangle$$

or

$$\alpha|01\rangle + \beta|10\rangle. \tag{17}$$

We will now prove that to minimize the output von Neumann entropy we can restrict our attention to an input pure state of the form (17). Let us write ρ_* in terms of its pure state decomposition

$$\rho_* = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{18}$$

Then the action of the channel will give

$$\mathcal{E}(\rho_*) = \sum_i p_i \mathcal{E}(|\psi_i\rangle\langle\psi_i|) \tag{19}$$

and hence by the concavity of the von Neumann entropy [2], we have

$$S(\mathcal{E}(\rho_*)) \geqslant \sum_i p_i S(\mathcal{E}(|\psi_i\rangle\langle\psi_i|)). \tag{20}$$

In particular, suppose without loss of generality that $|\psi_1\rangle$ is the pure state in the decomposition of ρ_* that gives the lowest output entropy from all the eigenvectors of ρ_*. Then the above equation implies that

$$S(\mathcal{E}(\rho_*)) \geqslant S(\mathcal{E}(|\psi_1\rangle\langle\psi_1|)). \tag{21}$$

So indeed, as we have assumed that ρ_* is already optimal, this means that this last equation is actually a strict equality, and hence one of its eigenvectors will also be optimal, namely,

$$S(\mathcal{E}(\rho_*)) = S(\mathcal{E}(|\psi_1\rangle\langle\psi_1|)). \tag{22}$$

Therefore, we can restrict our attention to finding an input pure state of the form (17).

Let us rewrite without loss of generality the input state (17) as

$$|\psi_{\theta,\phi}\rangle = \cos\theta|00\rangle + e^{i\phi}\sin\theta|11\rangle. \tag{23}$$

The corresponding state at the output of the channel takes the form

$$\begin{aligned}\mathcal{E}(|\psi_{\theta,\phi}\rangle\langle\psi_{\theta,\phi}|) = \tfrac{1}{4}[&\sigma_0 \otimes \sigma_0 + \eta(\cos 2\theta)(\sigma_0 \otimes \sigma_1 + \sigma_1 \otimes \sigma_0) \\ &+ C\sigma_1 \otimes \sigma_1 + \mu(\sin 2\theta)(\cos\phi)(\sigma_2 \otimes \sigma_2 \\ &- \sigma_3 \otimes \sigma_3) + \mu\eta(\sin 2\theta)(\sin\phi)(\sigma_2 \otimes \sigma_3 \\ &+ \sigma_3 \otimes \sigma_2)],\end{aligned} \tag{24}$$

where $\eta = (4p-1)$ and $C = \mu + (1-\mu)\eta^2$. As we can easily verify, the above density operator has the following eigenvalues:

$$\lambda_{1,2} = \tfrac{1}{4}(1-C),$$

$$\lambda_{3,4} = \tfrac{1}{4}(1+C) \pm \tfrac{1}{2}\sqrt{\eta^2\cos^2 2\theta + \mu^2\sin^2 2\theta(\cos^2\phi + \eta^2\sin^2\phi)}. \tag{25}$$

As we can infer from the above form of the eigenvalues, the input state corresponding to the minimum entropy is given by $\phi = 0$. Moreover, when $\mu > \eta$, or equivalently $p < (\mu+1)/4$, the input state with minimum entropy is the maximally entangled state (17) with $\theta = \pi/4$. In the other case, when $\mu < 4p-1$, the input state corresponding to the minimum output entropy is a product state of the form $|00\rangle$.

The set of optimal 16 states discussed above, which maximizes the mutual information along the channel, reduces in these cases to a set of four equiprobable input orthogonal states. Therefore, similarly to the case of the depolarizing channel with memory [7], we can identify the onset of a threshold value $\mu_t = 4p - 1$, above which the mutual information along the channel is maximized by using equiprobable Bell states [13]. Below the threshold the use of entanglement does not bring any benefit since the information is optimized by transmitting product states, such as the set $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. It is noteworthy that the resulting channel capacity is a nondifferentiable function of the memory parameter μ, although it is continuous.

In conclusion, we have studied the performance of Pauli channels with memory effects for the transmission of classical information, and we have provided a complete proof that a certain class of Pauli channels exhibits the onset of a threshold on the degree of memory. We have shown that below this threshold the two-qubit capacity of the channels is achieved by input product states, while above it the capacity is achieved by maximally entangled input states. This is the first time that entanglement is rigorously proven to be a precious resource in the transmission of classical information in the presence of noise. Our results so far have covered a class of Pauli channels, characterized by a single noise parameter. However, we have numerical evidence that the onset of the threshold, and the corresponding enhancement of information transmission by using entangled states, are features of most two-qubit Pauli channels with correlated noise [14].

This work has been supported in part by the EC program QUPRODIS (Contract No. IST-2002-38877) and the US Army, Grant No. DAAD 19-02-0161.

[1] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic, New York, 1976).

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

[3] C. Macchiavello, G. M. Palma, and A. Zeilinger, *Quantum Computation and Quantum Information Theory* (World Scientific, Singapore, 2001).

[4] D. Bruß, L. Faoro, C. Macchiavello, and G.M. Palma, J. Mod. Opt. **47**, 325 (2000).

[5] C. King and M.B. Ruskai, IEEE Trans. Inf. Theory **47**, 192 (2001); C. King, e-print quant-ph/0103156.

[6] K. Matsumoto, T. Shimono, and A. Winter, e-print quant-ph/0206148; K.M.R. Audenaert and S.L. Braunstein, e-print quant-ph/0303045; P.W. Shor, e-print quant-ph/0305035.

[7] C. Macchiavello and G.M. Palma, Phys. Rev. A **65**, 050301(R) (2002).

[8] G. Bowen and S. Mancini, e-print quant-ph/0305010.

[9] B. Schumacher and M.D. Westmoreland, Phys. Rev. A **56**, 131 (1997); A.S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).

[10] J. Cortese, e-print quant-ph/0211093; A.S. Holevo, e-print quant-ph/0212025.

[11] This is a consequence of *Schur's Lemma*; see, for example, C. J. Isham, *Lectures on Groups and Vector Spaces for Physicists* (World Scientific, Singapore, 1989).

[12] The channel that we have chosen is a probabilistic application of unitary transformations drawn from a particular group, and we wish to force the input states to be invariant under one of its subgroups (namely, $\{\sigma_0 \otimes \sigma_0, \sigma_1 \otimes \sigma_1\}$). To ensure this we have picked the probabilities such that any elements drawn from the same left coset of the subgroup have identical weight. This is why the channel that we consider is invariant under preoperating with the ‘‘averaging over the subgroup’’ operation, and the same reasoning also applies to any subgroup/channel combination such that the weights are constant over each left coset.

[13] It might be argued that our derivation does not strictly imply that entanglement is necessary to achieve the capacity above the threshold, as our solution is not necessarily unique, and so it might be possible to achieve capacity using alternative ensembles of separable states. To show that this is not possible, we must show that there is no input separable state that gives minimal output entropy. As the preoperation $\mathcal{F}$ is not entangling, it is sufficient to consider input separable states of the form (16). Moreover, we have already seen that product pure states of this form, such as $|01\rangle$, etc., do not have minimal output entropy, so we should only be concerned with mixed separable states with the structure (16). It is not difficult (although slightly tedious) to verify that each of optimal input Bell states derived for memories above the threshold is taken to a distinct output state by the channel. Therefore, by the *strict* concavity of the entropy, no mixed separable state of the form (16) can have minimal output entropy either.

[14] C. Macchiavello, G. M. Palma, and S. Virmani (unpublished).