Exploiting entanglement in communication channels with correlated noise

Jonathan Ball,1 Andrzej Dragan,2 and Konrad Banaszek1

1Centre for Quantum Computation, Clarendon Laboratory, University of Oxford, Oxford OX1 3PU, United Kingdom
2Instytut Fizyki Teoretycznej, Uniwersytet Warszawski, Hoża 69, 00-681 Warszawa, Poland

We develop a model for a noisy communication channel in which the noise affecting consecutive transmissions is correlated. This model is motivated by fluctuating birefringence of fiber optic links. We analyze the role of entanglement of the input states in optimizing the classical capacity of such a channel. Assuming a general form of an ensemble for two consecutive transmissions, we derive tight bounds on the classical channel capacity depending on whether the input states used for communication are separable or entangled across different temporal slots. This result demonstrates that by an appropriate choice, the channel capacity may be notably enhanced by exploiting entanglement.

PACS numbers: 03.67.Hk, 03.65.Yz, 42.50.Dv

I. INTRODUCTION

Entanglement is a fragile feature of composite quantum systems that can easily diminish by uncontrollable interactions with the environment. At the same time however carefully crafted entangled states can protect quantum coherence from the deleterious effects of those random interactions. This idea underlies the principles of quantum error correcting codes that strengthen the optimism regarding the feasibility of implementing in practice complex quantum information processing tasks [1].

In this paper we demonstrate how quantum entanglement can help in the task of classical communication. To this end, we develop a simple model of a noisy communication channel, where the noise affecting consecutive transmissions is correlated. Within this model, we derive bounds on the classical channel capacity assuming either separable or entangled input states, and we show that using collective entangled states of transmitted particles leads to an enhanced capacity of the channel.

The motivation for our model comes from classical fiber optic communications [2]. In practice, light transmitted through a fiber optic link undergoes a random change of polarization induced by the birefringence of the fiber. The fiber birefringence usually fluctuates depending on the environmental conditions such as temperature and mechanical strain. At first sight, this makes the polarization degree of freedom unsuitable for encoding information, as the input polarization state gets scrambled on average to a completely mixed state. However, the birefringence fluctuations have a certain time constant which means that the transformation of the polarization state, though random, remains nearly the same on short time scales. Consider now sending a pair of photons whose temporal separation lies well within this time scale. Although the polarization state of each one of the photons when looked at separately becomes randomized, certain properties of the joint state remain preserved. For example, this is the case of the relative polarization of the second photon with respect to the first one. We can therefore try to decode from the output whether the input polarizations were mutually parallel or orthogonal. This property cannot be determined perfectly, as in general we cannot tell whether two general quantum states are identical or orthogonal if we do not know anything else about them [3], but even the ability of providing a partial answer establishes correlations between the channel input and output that can be used to encode information into the polarization degree of freedom. The situation becomes even more interesting when we allow for entangled quantum states. Then the singlet polarization state of the two photons, when sent as the input, remains invariant under such perfectly correlated depolarization, and it can be discriminated unambiguously against the triplet subspace. Therefore we can encode one bit of information into the polarization state of two photons by sending either a singlet state or any of the triplet states. We shall see that these simple observations will also emerge from our general analysis of the channel capacity.

The first example of entanglement-enhanced information transmission over a quantum channel with correlated noise has been recently analyzed by Macchiavello and Palma [4]. Our model assumes a different form of correlations, and its high degree of symmetries has allowed us to perform optimization of the channel capacity over arbitrary input ensembles. Although we analyze only zero- and one-photon signals, we define the action of the channel in terms of the transformations of the bosonic annihilation operators, which sets up a framework for possible generalizations, such as use of multiphoton signals. This application of entanglement in classical communication is a distinct problem from entanglement-assisted classical capacity of noisy quantum channels studied by Bennett et al. in Ref. [5], where it has been shown that prior entanglement shared between sender and receiver can increase the classical capacity. We also note that the non-zero time constant of phase and polarization fluctuations can be used in robust protocols for long-haul quantum key distribution [6, 7].

Before passing on to a detailed discussion of the prob-
lem in the subsequent sections, let us introduce some ba-

ic notation. The action of a channel is described by

d a completely positive map \( \Lambda(\cdot) \) that we will denote by

\( \Lambda(\hat{\rho}) \). The sender selects messages from an input ensemble \( \{p_i, \hat{\rho}_i\} \), where \( p_i \) is the probability of sending the

state \( \hat{\rho}_i \) through the channel. The capacity of the channel

is a function of the mutual information between the

input ensemble and measurement outcomes at the receiv-
ing stations: it characterizes the strength of correlations

between these two that are preserved by the channel. The

mutual information itself involves a specific measurement

scheme; however, it has a very useful upper bound in the

form of the Holevo quantity that depends only on the

output ensemble of states \( \{p_i, \Lambda(\hat{\rho}_i)\} \) emerging from the

channel \( \hat{\rho} \): \(^1\)

\[
\chi = S \left( \sum_i p_i \Lambda(\hat{\rho}_i) \right) - \sum_i p_i S(\Lambda(\hat{\rho}_i))
\]

(1)

where \( S \) is the von Neumann entropy \( S(\hat{\rho}) = -\text{Tr}(\hat{\rho} \log_2 \hat{\rho}) \). As we will see, in our model the Holevo

quantity will provide a tight bound on the mutual infor-

mation that could be achieved in practice using a simple

measurement scheme. The classical channel capacity is

obtained by assuming arbitrarily long sequences of possi-

bly entangled input systems, and calculating the average

capacity per single use of the channel. In our analysis,

we will perform a restricted optimization by considering

only two consecutive uses of the channel.

II. CHANNEL DECOMPOSITION

We will start our discussion by proving a rather gen-

eral lemma about channels that can be decomposed into

direct sum of maps acting on subspaces of the Hilbert

space of the input systems. In physical terms, such chan-

nels remove quantum coherence between the compo-

nents of the input state that belong to different subspaces,

by zeroing the respective off-diagonal blocks of the density

matrix characterizing the input state. This lemma will

greatly simplify our further calculations.

Lemma 1: Suppose that we can decompose the Hilbert

space \( \mathcal{H} \) of the system into a direct sum of subspaces

\[
\mathcal{H} = \bigoplus_k \mathcal{H}^{(k)}
\]

(2)

such that for an arbitrary input state \( \hat{\rho} \) the state emerging

from the channel \( \Lambda(\hat{\rho}) \) can be represented as

\[
\Lambda(\hat{\rho}) = \bigoplus_k \Lambda^{(k)}(\hat{\rho}^{(k)})
\]

(3)

where \( \hat{\rho}^{(k)} = \hat{\rho}|_{\mathcal{H}^{(k)}} \) is the input state \( \hat{\rho} \) truncated to the

subspace \( \mathcal{H}^{(k)} \), and each \( \Lambda^{(k)} \) is a certain trace-preserving

completely positive map acting in the corresponding sub-

space \( \mathcal{H}^{(k)} \). Then the optimal channel capacity can be

attained with an ensemble in which each state belongs to

one of the subspaces \( \mathcal{H}^{(k)} \).

Proof: Indeed, suppose that there is a state \( \hat{\rho} \) that
does not satisfy the above condition, i.e. it is defined

on more than one subspace \( \mathcal{H}^{(k)} \). We can replace it by a sub-

ensemble \( \{\text{Tr}(\hat{\rho}^{(k)}); \hat{\rho}^{(k)}/\text{Tr}(\hat{\rho}^{(k)})\} \), obtained by truncat-

ing the state \( \hat{\rho} \) to the subspaces \( \mathcal{H}^{(k)} \) and normalizing the

resulting density matrices. In other words, whenever

the sender is supposed to transmit \( \hat{\rho} \), she replaces it by

one of the normalized truncated states \( \hat{\rho}^{(k)}/\text{Tr}(\hat{\rho}^{(k)}) \) with

the corresponding probability \( \text{Tr}(\hat{\rho}^{(k)}) \). It is straightforward to verify that the average state obtained from such a subensemble is identical with \( \Lambda(\hat{\rho}) \).

The above observation has a useful consequence when

optimizing the Holevo bound on channel capacity. If the

input ensemble is of the form discussed above, then it

can be split into subensembles of states that belong to

separate subspaces \( \mathcal{H}^{(k)} \), with the probability distribu-

tions normalized to one within each subensemble, and \( p_k \)

denoting the probability of sending a state from the \( k \)

subensemble. It is then easy to check that the Holevo

quantity is given by the following expression:

\[
\chi = \sum_k p_k \chi^{(k)} - \sum_k p_k \log_2 p_k,
\]

(4)

where \( \chi^{(k)} \) is the Holevo quantity for the \( k \)th subensemb-

lare. Therefore, the maximization of the Holevo quantity

can be performed in two steps. The first one is the op-

timization of each of \( \chi^{(k)} \) separately, assuming an input

ensemble restricted to the subspace \( \mathcal{H}^{(k)} \). The second

step consists of optimizing the probability distribution

\( p_k \) with the normalization constraint \( \sum_k p_k = 1 \), and it

can be performed explicitly using the method of Lagrange

multipliers. Indeed, if we denote the Lagrange multiplier

as \( \lambda \), then differentiation over \( p_l \) yields:

\[
0 = \frac{\partial}{\partial p_l} \left( \chi - \lambda \sum_k p_k \right) = \chi^{(l)} - \log_2 p_l - \frac{1}{\ln 2} - \lambda.
\]

(5)

This formula allows us to express the probabilities \( p_l \) in
terms of the Lagrange multiplier \( \lambda \) as:

\[
p_l = 2^{\chi^{(l)}-\lambda} 1/\ln 2 - \lambda,
\]

(6)

and furthermore summation over \( l \) and using the fact

that \( \sum_l p_l = 1 \) gives the value of the Lagrange

multiplier as:

\[
\lambda = \log_2 \left( \sum_l 2^{\chi^{(l)}} \right) - \frac{1}{\ln 2}.
\]

(7)

Finally, inserting Eqs. (6) and (7) into Eq. (4) yields the

maximum value of the Holevo quantity equal to:

\[
\chi = \log_2 \left( \sum_k 2^{\chi^{(k)}} \right).
\]

(8)
belle by A and B, each comprising two orthogonal polarizations, are occupied by a joint state of radiation $\hat{\rho}_{AB}$, as shown schematically in Fig. 1. We will assume that the polarization transformation $\Omega_A$ affecting the slot A is completely random, but that the transformation $\Omega_B$ is correlated with the first one through a conditional probability distribution $p(\Omega_B|\Omega_A)$. The resulting transformation of the joint two-slot state is therefore given by the following completely positive map:

$$
\Lambda(\hat{\rho}_{AB}) = \int d\Omega_A \int d\Omega_B p(\Omega_B|\Omega_A) \times \hat{U}(\Omega_A) \otimes \hat{U}(\Omega_B) \hat{\rho}_{AB} \hat{U}^\dagger(\Omega_A) \otimes \hat{U}^\dagger(\Omega_B).
$$

(9)

Here $\hat{U}(\Omega)$ is a unitary matrix acting in the Hilbert space of one of the slots that represents the polarization transformation $\Omega$. We will assume that the conditional probability $p(\Omega_B|\Omega_A)$ depends only on the relative transformation between the slots $A$ and $B$ and that it can consequently be represented as $p(\Omega_B|\Omega_A) = p(\Omega_B \cdot \Omega_A^{-1})$. In such a case, we can substitute the integration variables in the second integral according to $\Omega_B = \Omega' \cdot \Omega_A$, and make use of the invariance of the integration measure $d\Omega_B = d\Omega'$. This procedure shows that the map $\Lambda$ can be represented as a composition of two maps: $\Lambda = (\mathbb{I} \otimes \Lambda_{\text{dep}}) \circ \Lambda_{\text{perf}}$. The first one of them, $\Lambda_{\text{perf}}$, acts on both the temporal slots and it depolarizes them in exactly the same way:

$$
\Lambda_{\text{perf}}(\hat{\rho}_{AB}) = \int d\Omega \hat{U}(\Omega) \otimes \hat{U}(\Omega) \hat{\rho}_{AB} \hat{U}^\dagger(\Omega) \otimes \hat{U}^\dagger(\Omega).
$$

(10)

The second map, $\Lambda_{\text{dep}}$, acts only on the slot $B$, and it introduces additional depolarization relative to the slot $A$ according to the probability distribution $p(\Omega')$:

$$
\Lambda_{\text{dep}}(\hat{\rho}_B) = \int d\Omega' p(\Omega') \hat{U}(\Omega') \hat{\rho}_B \hat{U}^\dagger(\Omega').
$$

(11)

We will assume later that the distribution $p(\Omega')$ has sufficient symmetry to describe the action of the map $\Lambda_{\text{dep}}$ in the relevant Hilbert space with the help of two simple parameters.

We now introduce a further simplification by imposing a condition that each temporal slot may contain at most one photon. Therefore the relevant Hilbert space for each slot is spanned by three states: the zero-photon state $|0\rangle$, and horizontally and vertically polarized one-photon states $|\leftrightarrow\rangle$ and $|\downarrow\rangle$. We can conveniently write the explicit form of the unitary transformation $\hat{U}(\Omega)$ using the irreducible unitary representations of the group SU(2). We will denote by $\mathcal{D}^j(\Omega)$ a $(2j + 1) \times (2j + 1)$ matrix that is a $(2j + 1)$-dimensional representation of an SU(2) element obtained from $\Omega$ by fixing the overall phase factor to one. These matrices are well known in the quantum theory of angular momentum as describing transformations of a spin-$j$ particle under the rotation group SU(2). We will also denote by $\alpha(\Omega)$ the overall phase
of the element $\Omega$. Then the unitary transformation of the input state corresponding to the polarization rotation $\Omega$ is given by the matrix:

$$
\hat{U}(\Omega) = \left( \begin{array}{ccc} \hat{D}^0(\Omega) & 0 & 0 \\ 0 & e^{i\alpha(\Omega)}\hat{D}^{1/2}(\Omega) \end{array} \right). \quad (12)
$$

In this formula, the one-dimensional representation $\hat{D}^0(\Omega)$ is identically equal to one, and $e^{i\alpha(\Omega)}\hat{D}^{1/2}(\Omega)$ is a $2 \times 2$ unitary matrix itself; however, we will keep this more general notation in order to be able to use results from the theory of group representations. In particular, the following property of the rotation matrix elements will allow us to evaluate directly a number of expressions:

$$
\int d\Omega [\hat{D}^j_{mn}(\Omega)]^* \hat{D}^j_{m'n'}(\Omega) = \frac{1}{2j+1} \delta_{jj'} \delta_{mm'} \delta_{nn'}. \quad (13)
$$

The action of the map $\Lambda_{\text{perf}}$ on a joint two-slot state can be analyzed most easily if we decompose the complete Hilbert space into a direct sum of subspaces with a fixed number of photons: $\mathcal{H} = \mathcal{H}^{(0)} \oplus \mathcal{H}^{(1)} \oplus \mathcal{H}^{(2)}$, where the upper index labels the number of photons. The zero-photon subspace is spanned by a single state $|0_A 0_B\rangle$. The one-photon space has a basis formed by four vectors: $|\uparrow_A 0_B\rangle$, $|\downarrow_A 0_B\rangle$, $|0_A \uparrow_B\rangle$, and $|0_A \downarrow_B\rangle$. Finally, in the two-photon subspace $\mathcal{H}^{(2)}$ we will introduce a basis that consists of the singlet state $|\Psi_\text{-}\rangle = (|\uparrow_A 0_B\rangle - |\downarrow_A 0_B\rangle)/\sqrt{2}$ and the three triplet states $|\Psi_\text{+}\rangle = (|\uparrow_A 0_B\rangle + |\downarrow_A 0_B\rangle)/\sqrt{2}$, and $|0_A 0_B\rangle_B$. The reason for this choice is that then the action of the tensor product $\hat{D}^{1/2}(\Omega) \otimes \hat{D}^{1/2}(\Omega)$ on a two-photon state can be decomposed into the sum: $\hat{D}^{1/2}(\Omega) \otimes \hat{D}^{1/2}(\Omega) = \hat{D}^0(\Omega) \oplus \hat{D}^1(\Omega)$ where $\hat{D}^0(\Omega)$ acts on the singlet state $|\Psi_\text{-}\rangle$, and $\hat{D}^1(\Omega)$ is a three-dimensional matrix acting in the triplet subspace. Using our decomposition of the complete Hilbert space, the action of the tensor product $\hat{U}(\Omega) \otimes \hat{U}(\Omega)$ on a general two-slot state in the basis specified above is given by:

$$
\hat{U}(\Omega) \otimes \hat{U}(\Omega) = \hat{D}^0(\Omega) \oplus e^{i\alpha(\Omega)} \left( \begin{array}{ccc} \hat{D}^{1/2}(\Omega) & 0 & 0 \\ 0 & 0 \hat{D}^{1/2}(\Omega) & 0 \\ 0 & 0 \hat{D}^1(\Omega) \end{array} \right) \quad (14)
$$

If we now insert this formula into Eq. (13), it can be easily seen that the invariant integration over the overall phase factor $\alpha(\Omega)$ kills all the off-block diagonal elements of the density matrix that link different subspaces $\mathcal{H}^{(k)}$. In other words, all the coherence between states with different photon numbers is completely removed by the phase fluctuations. Furthermore, the operation $\Lambda_{\text{dep}}$, acting only on the second slot, does not mix subspaces with different photon numbers. Therefore the conditions of our lemma are satisfied and we can consider only states with a definite number of photons as elements of the input ensemble. Thus we need to calculate are three corresponding Holevo quantities $\chi^{(0)}$, $\chi^{(1)}$, and $\chi^{(2)}$ that can be combined into a Holevo bound for the overall channel capacity according to Eq. (5). This calculation forms the contents of the next section.

IV. CHANNEL CAPACITY

The communication capacity $\chi^{(0)}$ of the zero-photon subspace itself $\mathcal{H}^{(0)}$ is naturally zero, as we have only a single state $|0_A 0_B\rangle$ at our disposal. This state can of course be used as an element of a larger ensemble thus contributing to the overall capacity. This fact is reflected in the form of Eq. (5), where $\chi^{(0)} = 0$ indeed does increase the total value of $\chi$.

A. One-photon subspace

A less trivial problem to calculate is the capacity of the one-photon subspace. If we assume a normalized input state $\hat{\varrho}_\text{in}$ from the subspace $\mathcal{H}^{(1)}$, then the action of the channel $\Lambda_{\text{perf}}$ restricted to this subspace is given by:

$$
\Lambda_{\text{perf}}^{(1)}(\hat{\varrho}_\text{in}) = \frac{1}{2} \left( \begin{array}{cccc} a & 0 & b & 0 \\ 0 & a & 0 & b \\ b^* & 0 & 1 - a & 0 \\ 0 & b^* & 0 & 1 - a \end{array} \right) \quad (15)
$$

where the parameters $a$ and $b$ are defined in terms if the input density matrix as:

$$
\begin{align*}
a &= \langle \uparrow_A 0_B | \hat{\varrho}_\text{in} | \downarrow_A 0_B \rangle + \langle \downarrow_A 0_B | \hat{\varrho}_\text{in} | \uparrow_A 0_B \rangle \\
b &= \langle \uparrow_A 0_B | \hat{\varrho}_\text{in} | 0_A \uparrow_B \rangle + \langle \downarrow_A 0_B | \hat{\varrho}_\text{in} | 0_A \downarrow_B \rangle & (16)
\end{align*}
$$

For the form of density matrix given in Eq. (15), the depolarizing channel $\Lambda_{\text{dep}}$ affects only the off-diagonal elements $b$ and $b^*$. We will assume that the symmetry of the distribution $p(\Omega')$ is such that the effect of $\Lambda_{\text{dep}}$ is a rescaling of these elements by a real parameter $\eta'$ bounded between 0 and 1. It is now easy to check that the entropy of the one-photon state emerging from the channel can be written as

$$
S(\Lambda(\hat{\varrho}_\text{in})) = 1 + S \left( \frac{\eta'}{\eta'^2 + 1 - a} \right) \quad (17)
$$
where the $2 \times 2$ matrix appearing in the second term can be interpreted as a state of a qubit. Therefore, the second term is bounded by 0 and 1, and consequently $1 \leq S(\Lambda(\hat{\varrho}_{in})) \leq 2$. It is a straightforward observation that the Holevo quantity is bound from above by the difference between the maximum and the minimum possible entropies of states emerging from the channel. Therefore we obtain that $\chi^{(1)} \leq 1$. This inequality can be saturated simply by taking a one-photon state confined either to the first or to the second temporal slot, with an arbitrary polarization. Thus, the channel capacity is not enhanced in the one-photon sector.

**B. Two-photon subspace**

The most interesting regime is when both the temporal slots are occupied by photons. As we will see below, in this case quantum correlations can then enhance the capacity of the channel. If we take a normalized input state $\hat{\varrho}_{in}$ from the two-photon subspace $\mathcal{H}^{(2)}$, then the map $\Lambda_{\text{perf}}$ produces a Werner state $\hat{W}_c$:

$$\Lambda_{\text{perf}}^{(2)}(\hat{\varrho}_{in}) = \hat{W}_c,$$

where we have introduced the following notation:

$$\hat{W}_c = -c|\Psi_-\rangle\langle\Psi_-| + (1 + c)\frac{\mathbb{1}}{4}$$

and we will use for $c$ the name of the Werner parameter of the input state $\hat{\varrho}_{in}$, defined as:

$$c = \frac{1}{3} - \frac{4}{3}\langle\Psi_-|\hat{\varrho}_{in}|\Psi_-\rangle.$$  \hfill (20)

This result, derived previously in Ref. [13], can be verified independently using the property given in Eq. (13).

The second operation affecting the input state is the partially depolarizing channel $\hat{1} \otimes \Lambda_{\text{dep}}$. We will assume that the action of the map $\Lambda_{\text{dep}}$ acting on the photon in the second temporal slot is simply isotropic depolarization shrinking the length of the Bloch vector by a factor $\eta$ satisfying $0 \leq \eta \leq 1$. Such an operation preserves the Werner form of the transmitted state, and its only effect is the multiplication of the parameter $c$ by the factor $\eta$. Thus, the state emerging from the channel is given by:

$$\Lambda^{(2)}(\hat{\varrho}_{in}) = \hat{W}_{\eta c},$$

with the parameter $c$ defined by the input state $\hat{\varrho}_{in}$ according to Eq. (20).

At this point the possibility of enhanced communication capacity by exploiting entanglement manifests itself. The difference between the separable and entangled alphabets can be seen by comparing the allowed ranges of the parameter $c$. The positivity of the input density matrix $\hat{\varrho}_{in}$ requires that

$$-1 \leq c \leq 1/3$$

and this is the only condition if we consider the most general, possibly entangled input states. However, if the input states are restricted to separable ones, then as shown by Horodeccy [14], the allowed range for the parameter $c$ is reduced to

$$-1/3 \leq c \leq 1/3.$$  \hfill (23)

This limitation will underlie the reduced channel capacity in the case of separable states.

As the two-photon states emerging from the channel are fully characterized by the Werner parameters of the respective input states, optimization of the Holevo quantity can be carried out over the ensemble $\{q_j; c_j\}$ of the probabilities $q_j$ of sending the $j$th state with the Werner parameter equal to $-c_j$. The output states emerging from the channel is therefore given by an ensemble of Werner states $\{q_j; \hat{W}_{\eta c_j}\}$. Because a statistical mixture of Werner states is also a Werner state with the average parameter:

$$\sum_j q_j \hat{W}_{\eta c_j} = \hat{W}_{\sum_j q_j \eta c_j},$$

the Holevo quantity can be expressed with the help of a single real-valued function $f(c)$:

$$\chi^{(2)} = S\left(\sum_j q_j \hat{W}_{\eta c_j}\right) - \sum_j q_j S(\hat{W}_{\eta c_j})$$

$$= f\left(\sum_j q_j \eta c_j\right) - \sum_j q_j f(\eta c_j)$$

where the explicit form of the function $f(c)$ is given by:

$$f(c) = 2 - \frac{3}{4}(1+c)\log_2(1+c) - \frac{1}{4}(1-3c)\log_2(1-3c).$$  \hfill (26)

The optimization of the Holevo quantity, which in principle needs to be performed over an arbitrarily large input ensemble of permitted quantum states, can be greatly simplified using the following observation.

**Lemma 2.** Let $f(\gamma)$ be a concave function defined on a closed interval $[\alpha, \beta]$, and let $q_j$ be a probability distribution for a set $\gamma_j$ of real numbers taken from the range $\alpha \leq \gamma_j \leq \beta$. Then the following inequality holds:

$$f\left(\sum_j q_j \gamma_j\right) - \sum_j q_j f(\gamma_j) \leq \sup_{\alpha \leq \gamma \leq \beta} \left(f(\gamma) - \frac{\beta - \gamma}{\beta - \alpha} f(\alpha) - \frac{\gamma - \alpha}{\beta - \alpha} f(\beta)\right).$$  \hfill (27)

**Proof:** The concavity of the function $f(c)$ implies that for every $j$ we have:

$$f(\gamma_j) \geq \frac{\beta - \gamma_j}{\beta - \alpha} f(\alpha) + \frac{\gamma_j - \alpha}{\beta - \alpha} f(\beta).$$  \hfill (28)
If we now multiply the above equation by \(-q_j\), perform the summation over \(j\), and add a term \(\sum_j f(q_j \gamma_j)\) to both sides of the equation, we will obtain an inequality whose left hand side is identical with that of Eq. (27) and the right hand side is exactly the argument of the supremum for \(\gamma = \sum_j q_j \gamma_j\). Obviously, this value of \(\gamma\) lies between \(\alpha\) and \(\beta\), and consequently the supremum may only exceed the value obtained from this calculation. This confirms that Eq. (27) is indeed satisfied.

The above lemma reduces the whole problem of optimizing the Holevo bound to maximizing a one-parameter real-valued function that is the argument of the supremum on the right hand side of Eq. (27). Inserting the explicit form of the function \(f(\gamma)\) given in Eq. (29) and differentiating the resulting expression over \(\gamma\) shows that the supremum in the right hand side of Eq. (27) is attained for

\[
\gamma_{\text{opt}} = \frac{1 - 2^{\mu/\beta}}{3 + 2^{\mu/\beta}} \quad (29)
\]

where \(\mu = [f(\beta) - f(\alpha)]/|\beta - \alpha|\).

As we have seen, the permitted range of the parameters \(c_j\) characterizing the states belonging to the input ensemble depends on whether we allow most general, possibly entangled states, or rather restrict the input to separable states only. If we assume that this range spans from \(c_{\text{min}}\) to \(c_{\text{max}}\):

\[
c_{\text{min}} \leq c_j \leq c_{\text{max}} \quad (30)
\]

then we can easily apply Lemma 2 to the expression of the Holevo quantity \(\chi^{(2)}\) in terms of the function \(f(c)\) that has been given in the second line of Eq. (27). Taking \(\alpha = \eta c_{\text{min}}\) and \(\beta = \eta c_{\text{max}}\) and using the explicit value of the turning point derived in Eq. (28) yields the following bound:

\[
\chi^{(2)} \leq \log_2(3 + 2^{\mu/\beta}) - f(\eta c_{\text{min}}) + \mu(\eta c_{\text{min}} - 1/3) - 2 \quad (31)
\]

where \(\mu\) is given in terms of the input ensemble characteristics as:

\[
\mu = \frac{f(\eta c_{\text{max}}) - f(\eta c_{\text{min}})}{\eta (c_{\text{max}} - c_{\text{min}})}. \quad (32)
\]

We will analyze in detail numerical values of the channel capacity in the next section. Before doing so, we will close this section by describing a simple intuitive picture of Lemma 2 that gives an additional insight into the form of the input ensemble.

**C. Graphical interpretation**

The result of Lemma 2 can be visualized using the following geometrical reasoning depicted in Fig. 2. Consider a graph of the function \(f(\gamma)\) versus its argument \(\gamma\). The numbers \(\gamma_j\) and the corresponding values of the function \(f\) are given by a set of points \(G_j = (\gamma_j, f(\gamma_j))\) in the plane of the graph. The probability distribution \(q_j\) for the arguments \(\gamma_j\) defines an average

\[
\bar{G} = \left( \sum_j q_j \gamma_j, \sum_j q_j f(\gamma_j) \right) \quad (33)
\]

that can be interpreted as a center of gravity for the system of points \(G_j\) that have been assigned respective masses \(q_j\). Obviously, if the probability distribution is arbitrary, then this average can lie anywhere within the convex polygon spanned by the points \(G_j\). Since the function \(f\) is strictly concave over the range considered, the whole polygon lies within the area bounded by the graph of the function \(f(\gamma)\) on one side, and a straight line connecting the points \((\alpha, f(\alpha))\) and \((\beta, f(\beta))\) on the other side. This straight line is given by a function \(g\) defined as:

\[
g(\gamma) = \frac{\beta - \gamma}{\beta - \alpha} f(\alpha) + \frac{\gamma - \alpha}{\beta - \alpha} f(\beta). \quad (34)
\]

The left hand side of Eq. (27) is now given by the length of a vertical line connecting \(G\) with the point \(H' = (\bar{\gamma}, f(\bar{\gamma}))\) on the graph of the function \(f(\gamma)\), where \(\bar{\gamma} = \sum_j q_j \gamma_j\). Clearly, the line \(GH'\) will be always equal in length or shorter than the line \(H'H''\) where the point \(H'' = (\bar{\gamma}, g(\bar{\gamma}))\) lies on the graph of the function \(g(\gamma)\). Furthermore, in order to find the maximum possible length of the line \(H'H''\), it is clear from this geometrical construction that we need to maximize the difference \(f(\gamma) - g(\gamma)\) over \(\gamma\) belonging to the interval \([\alpha, \beta]\). This procedure is expressed explicitly in the right hand side of Eq. (27) and the parameter \(\mu\) introduced in the previous subsection is simply the gradient of the function \(g(\gamma)\).

It is clearly seen from this geometric construction that enlarging the interval \([\alpha, \beta]\) can only increase the value...
The Holevo quantity $\chi$ is only an upper bound on the channel capacity and therefore is not necessarily attainable. Users of a communication channel need two relevant pieces of information. The first one is the optimal form of the input ensemble that should be used by the sender. The second one is a measurement scheme that should be employed at the output of the channel in order to optimize the capacity.

Let us start by summarizing the results of the preceding section and specifying the input ensemble implied by these considerations. We have seen that in the zero- and one-photon subspaces the channel capacity cannot be enhanced by exploiting the polarization degree of freedom. Therefore as the elements of the input ensemble we can take for example states $|0_A 0_B\rangle$, $|1_A 0_B\rangle$, and $|0_A 1_B\rangle$, where for concreteness we have fixed the polarization of single-photon states to vertical. The polarization degree of freedom starts to play a nontrivial role when both the temporal slots are occupied by photons. In this subspace, we need to select two input states characterized by the Werner parameters that are as distant as it is allowed by the constraints on the input ensemble. If we restrict ourselves to separable states, then according to Eq. \((23)\) we need to take one separable state with $c_{\text{min}} = -1/3$ and another one with $c_{\text{max}} = 1/3$. It is easy to verify using Eq. \((20)\) that the pair of separable states satisfying this condition can be taken as $|A^{-}B\rangle$ and $|A_{+}B\rangle$.

We thus see that in agreement with the simple picture developed in the introduction to this paper, the relevant quantity is the relative polarization of the photons occupying consecutive slots. If we allow for entangled input, then the lower limit for the Werner parameters of the input states shifts down to $c_{\text{min}} = -1$. This value can be of course attained by taking the singlet state $|\Psi_{-}\rangle$ itself as one element of the input ensemble, and any state with $c_{\text{max}} = 1/3$, for example again $|A_{+}B\rangle$ as the second one.

In order to complete the description of the communication protocol, we need to specify the measurement applied to the states emerging from the channel. This task can be decomposed into two steps. The first one is the determination of the total number of photons contained in the two slots and it can in principle be accomplished by a collective quantum non-demolition measurement \[15\] on all the modes involved that would determine the total photon number without destroying coherence between the modes. Depending on the outcome, the second step needs to be either finding the temporal slot occupied by a photon in the one-photon subspace which can be realized by direct temporally resolved detection, or discriminating between the states used to encode information in the two-photon subspace. It is easy to see that this discrimination takes a simple form in the case of perfectly correlated noise and entangled input states: we need to determine whether the received states belong to the singlet or the triplet subspace, which corresponds to a two-element projective measurement:

$$
\hat{O}_S = |\Psi_{-}\rangle \langle \Psi_{-}|,
$$
$$
\hat{O}_T = \hat{1} - |\Psi_{-}\rangle \langle \Psi_{-}|
$$

(35)

It turns out that the same measurement saturates the Holevo bound also in the general case of any value of the parameter $\eta$ with either entangled or separable input states. In Fig. 4(a) we depict conditional probabilities of obtaining the singlet or the triplet outcomes for a
two-element input ensemble characterized by Werner parameters $c_{\text{min}}$ and $c_{\text{max}}$. A lengthy but straightforward calculation shows that if we take as the input probabilities the values discussed in the preceding section, the mutual information is given exactly by the right hand side of Eq. (31). Thus the described procedure indeed maximizes the channel capacity in the two-photon subspace.

It is instructive to compare the above diagram for optimal entangled and separable input ensembles in the case of perfect correlations $\eta = 1$. For the optimal entangled ensemble, shown in Fig. 4(b) we can distinguish perfectly between the two inputs as they belong to orthogonal subspaces even after the transmission. For the separable ensemble, the emerging states can no longer be perfectly discriminated as seen in Fig. 4(c).

The complete channel capacity obtained by combining Eq. (8) with the results of Sec. IV is shown as a function of $\eta$ in Fig. 5. It is seen that using an entangled input ensemble gives a clear advantage over the separable states over the complete range of the correlation parameter $\eta$.

We note that the measurement discriminating between the singlet and the triplet subspaces can be implemented using the Braunstein-Mann scheme based on linear optics.

VI. CONCLUSIONS

We have introduced a model of a communication channel with correlated noise motivated by random birefringence fluctuations in a fiber optic link. Within this model, we have demonstrated that introducing quantum correlations between consecutive uses of the channel increases its capacity. This demonstrates how specifically quantum phenomena such as entanglement can be helpful in the task of transferring classical information. Making use of entanglement requires more complex preparation procedures that provide joint input states extending over a number of temporal slots. A related question is the role of collective quantum measurements on the output of the channel rather than detecting radiation in each of the slots individually and combining classical outcomes of separate measurements.

The action of the channel has been defined in terms of transformations of the bosonic field operators. This opens up a route towards interesting generalizations of the present work, for example including arbitrary multi-photon states. Another direction would be extending the model to an arbitrary number of temporal slots rather than just allowing for correlations between pairs of consecutive slots as in our example. It is easy to give a simple protocol showing that in this case the channel capacity
can be enhanced even further. Suppose that the sender generates a train of zero- and one-photon states with the same probabilities equal to one half. The first time she is to transmit a photon, she sends half of maximally entangled pair. In the second instance when a one photon should be transmitted, she sends the remaining member of the pair transforming it in such a way that the joint two-photon polarization state belongs either to the singlet or the triplet subspace. The receiver implements a polarization-independent quantum non-demolition measurement on each temporal slot. When a photon is detected, it needs to be stored until the arrival of the second member of a pair, when the discrimination between the singlet and the triplet subspaces can be performed with the help of a joint measurement. If the fluctuations in random birefringence can be neglected over the temporal separation between the photons in a pair, this procedure allows one to encode one extra bit of information into each pair of transmitted photons. This gives the average channel capacity equal to 2.5 per a pair of temporal slots, enhancing further the optimal value shown in Fig. 5.

ACKNOWLEDGEMENTS

This research was supported by EPSRC and Polish KBN.