# Chapter 15: Quantum Information Theory and Photonic Communications

**Farhan Rana (MIT)**

March, 2002

**Abstract**

The fundamental physical limits of optical communications are discussed in the light of the recent developments in quantum information theory. The limits on channel capacities and bit error rates are described. Classical information theory is reviewed, and its relationship with the quantum information theory is examined. Recent developments in quantum information theory have opened new possibilities for transmitting classical and quantum information over optical channels. The current status of the field and the open questions are discussed.

## Contents

# 1 Introduction

The fundamental limits of communication has been an interesting subject for many years. The capacity $C$ of a communication channel is the maximum rate at which information can be transmitted without error from the channel's input to its output. The limits on capacities for transmitting classical information over classical channels were first described in the seminal work by Shannon [1]. Since then a vast body of literature has been published in this field [2]. The transmission of classical and quantum information over quantum channels is described by quantum information theory. The effects of the quantum nature of electromagnetic waves on the capacities of optical channels have been discussed in the literature by many authors (see [3] and [4] and references therein). More recently, transmission of quantum information over quantum channels has also been investigated (see [5] and references therein). At a more fundamental level it has become clear that an information theory based on quantum principles extends and completes classical information theory. The new theory not only provides quantum generalizations of classical concepts such as sources, channels, and codes but also introduces ingredients that are distinctively quantum mechanical in nature such as uncertainty and entanglement. The goal of this Chapter is to explore the limits of communication over optical channels in the light of the recent developments in quantum information theory.

# 2 Classical Information Theory

In 1948 Claude Shannon in his landmark paper established two main results of classical information theory [1]. The two central problems he addressed were:

1. **Data Compression:** How much can a *message* be compressed (Shannon's noiseless coding theorem).

2. **Data Transmission:** At what rate can reliable communication occur over a noisy channel (Shannon's noisy channel coding theorem).

One of Shannon's key insights was that *entropy* provides a suitable way of quantifying the information content of a message. In the next three sections the main results of classical information theory are reviewed.

## 2.1 Data Compression and Shannon's Noiseless Channel Coding Theorem

In this Chapter, a *message* would be considered to be a long string of letters with which two people communicate with each other. Each letter in a message is chosen from an alphabet of $k$ letters,

$$A = \{a_1, a_2, a_3, \ldots, a_k\} \tag{1}$$

Each letter in a message is statistically independent and occurs with an a priori probability $P(a_i)$. Now suppose we want to send messages, each containing $n$ letters, over a noiseless channel. We can assign a binary codeword to each letter. Since there are $k$ different letters, each binary codeword has $\log_2 k$ bits. Each message sent over the channel will therefore contain $n \log_2 k$ bits, or $\log_2 k$ bits per letter. Now we ask: is it possible to compress the messages before sending them over the channel to less than $\log_2 k$ bits per letter but still convey essentially the same information? The answer, as shown below, is yes.

The probability that a message contains letters $a_{i_1}, a_{i_2}, \ldots, a_{i_n}$ is $P(a_{i_1}, a_{i_2}, \ldots, a_{i_n})$, where,

$$-\frac{1}{n}\log_2 P(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) = -\frac{1}{n}\sum_{j=1}^{n}\log_2 P(a_{i_j}) \tag{2}$$

From the law of large numbers,

$$\lim_{n\to\infty} -\frac{1}{n}\log_2 P(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) = \lim_{n\to\infty} -\frac{1}{n}\sum_{j=1}^{n}\log_2 P(a_{i_j}) \to -\sum_{i=1}^{k} P(a_i)\log_2 P(a_i) = H(A) \tag{3}$$

This implies that for any $\delta > 0$, one can choose n sufficiently large such that,

$$2^{-n(H(A)+\delta)} \leq P(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) \leq 2^{-n(H(A)-\delta)} \tag{4}$$

The set of messages for which the above relation holds is called the typical set, an its members are called typical messages. In a typical message letter $a_i$ occurs approximately $nP(a_i)$ times. Not all messages belong to the typical set. For example, a message containing $n$ letters all of which are $a_1$ is not a typical message (unless $P(a_1)$ equals 1). The law of large number shows that for large $n$ *almost all* the messages are typical messages. The probability of every member in the typical set is approximately the same, and it is $2^{-nH(A)}$. This implies that there must be approximately $2^{nH(A)}$ typical messages in the typical set. If one assigns binary codewords only to the typical messages, and map all the non-typical messages to the binary codeword containing all zeros, then, with a probability of error approaching zero, one can transmit a message across the noiseless channel using only $nH(A)$ bits, or only $H(A)$ bits per letter. It follows from the definition of $H(A)$ that if all the letters are equally likely then $H(A)$ equals $\log_2 k$ and no compression is possible.

The quantity $H(A)$ is called the entropy. In general the entropy $H(X)$ of a random variable $X$, which can take the possible value $x$ with probability $P_X(x)$, is given by the relation,

$$H(X) = -\sum_{x} P_X(x)\log_2 P_X(x) \tag{5}$$

From the above discussion it follows that the entropy may be interpreted as the number of bits required *on the average* to represent the outcome of the random variable. In other words, entropy is the information in bits acquired, *on the average*, by gaining knowledge of an outcome of the random variable $X$.

## 2.2 Data Transmission and Shannon's Noisy Channel Coding Theorem

A communication channel is shown in Fig. 1. The input to the channel is a random variable $X$ which can take any value from the input alphabet $A = \{a_1, a_2, \ldots, a_k\}$ with a priori probability $P_X(x)$. Channel output is described by the random variable $Y$ which can take any value from the output alphabet $B = \{b_1, b_2, \ldots, b_k\}$. The channel is described by the transition probabilities $P_{Y|X}(y|x)$. In this section the maximum achievable rate, at which information in bits per letter can be transmitted across this channel, is computed. A rate is achievable if the probability of error approaches zero as the size of the message approaches infinity.

If the channel were noiseless (i.e. $P_{Y|X}(y|x) = \delta_{y,x}$), then as shown in the previous section, one can transmit information at the rate of $H(X)$ $(= H(A))$ bits per letter (provided one uses messages that are long enough). If the channel is noisy one would expect the information transmission rate to be less than $H(X)$ bits per letter. If the channel is noisy each letter observed at the output of the channel would convey less than $H(X)$ bits of information. But less by how much? The conditional entropy $H(X|Y)$ quantifies this reduction in information in bits, and it is defined as,

$$H(X|Y) = -\sum_{y} P_Y(y)\sum_{x} P_{X|Y}(x|y)\log_2 P_{X|Y}(x|y) \tag{6}$$

Thus, only $H(X) - H(X|Y)$ bits of information are obtained by observing a letter at the output of the channel. This quantity is called the mutual information $I(X:Y)$,

$$I(X:Y) = H(X) - H(X|Y) \tag{7}$$

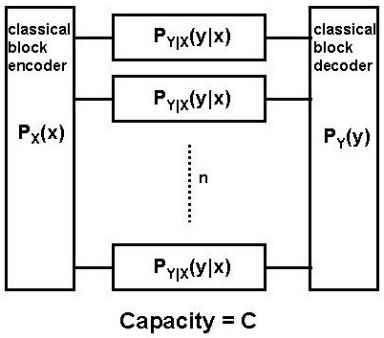


Figure 1: A classical communication channel

Defining the joint entropy $H(Y,X)$ between $X$ and $Y$ as,

$$H(Y,X) = \sum_{y,x} P_{Y,X}(y,x) \log_2 P_{Y,X}(y,x) \tag{8}$$

one obtains the following identities,

$$\begin{aligned} I(X:Y) &= H(X) - H(X|Y) && (9)\\ &= H(X) + H(Y) - H(Y,X) && (10)\\ &= H(Y) - H(Y|X) && (11)\\ &= I(Y:X) && (12) \end{aligned}$$

One may also interpret the above result in the following way. There are approximately $2^{nH(X)}$ typical input messages each containing $n$ letters. There are also approximately $2^{nH(Y)}$ typical output messages. But for each message at the input there are only $2^{nH(Y|X)}$ typical output messages. In order to ensure that no two input messages produce the same output message we cannot use more than $2^{nH(Y)}/2^{nH(Y|X)}$ typical input messages. The information in bits conveyed by an input message cannot therefore exceed,

$$\log_2 \left( 2^{nH(Y)}/2^{nH(Y|X)} \right) = n(H(Y) - H(Y|X)) = nI(Y:X)$$

It follows that the information transmitted per letter is $I(Y:X)$ bits. One can increase the information transmission rate by using those letters more frequently that are less likely to be affected by the channel noise. The channel capacity $C$ is the maximum rate at which information in bits per letter can be transmitted across the channel, and is obtained by maximizing $I(Y:X)$ over the probabilities $P_X(x)$ for the letters in the input alphabet $A$,

$$C = \max_{P_X(.)} I(Y:X) \tag{13}$$

Shannon's greatest achievement was to prove that the rate $C$ is achievable. It can be shown that any rate less than $C$ is achievable, and any rate above $C$ is not achievable [6].

## 2.3 The Gaussian Channel

In this section we compute the capacity of the classical gaussian channel. For the gaussian channel the input alphabet is the set of real numbers with zero mean. A message consisting of $n$ letters input to the channel must also satisfy the average power constraint,

$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 \leq P \tag{14}$$

The channel adds zero mean gaussian noise to the input letter. If $x$ is the input letter and $y$ is the output letter then,

$$y = x + z \tag{15}$$

where the noise $Z$ has the normal distribution $\mathcal{N}(0, N)$. The capacity $C$ is,

$$\begin{aligned} C &= \max I(Y:X) && (16)\\ &= \max\left[H(Y) - H(Y|X)\right] = \max H(Y) - H(Z) && (17)\\ &= \max H(Y) - \frac{1}{2}\log_2 2\pi e N && (18)\\ & && (19) \end{aligned}$$

Equations (14) and (15) imply that the message at the output of the channel must satisfy the constraint,

$$\frac{1}{n}\sum_{i=1}^{n} y_i^2 \leq P + N \tag{20}$$

Subject to the above constraint the entropy $H(Y)$ can be maximized if $Y$ has the normal distribution $\mathcal{N}(0, P+N)$. A normal distribution at the channel output can be achieved if the channel input $X$ also has the normal distribution $\mathcal{N}(0, P)$. The channel capacity is therefore,

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{N}\right) \tag{21}$$

The above result is of fundamental importance in classical information theory. It has a simple visual interpretation. If we represent each input message of $n$ letters as a point in an n-dimensional space, then the channel output can be represented by drawing small $n$-spheres of radius $\sqrt{N}$ around each of these points. The $n$-spheres represent regions around each input point within which the channel output can wander as a result of the channel noise. The number of distinct messages containing $n$ letters each that can be sent across the channel is the number of $n$-spheres of radius $\sqrt{N}$ that can be packed in a $n$-space of radius $\sqrt{P+N}$. This number is,

$$\frac{(\sqrt{P+N})^n}{(\sqrt{N})^n} = \left(1 + \frac{P}{N}\right)^{n/2} \tag{22}$$

The amount of information conveyed per letter is therefore $\frac{1}{2}\log_2(1 + P/N)$ bits. If the channel has a bandwidth $B$, then from Nyquist's theorem $2B$ letters can be sent across the channel per second. Assuming the noise spectral density to be $S_n$, the channel capacity in bits per second becomes,

$$C = B\log_2\left(1 + \frac{P}{BS_n}\right) = B\log_2\left(1 + SNR\right) \tag{23}$$

where we have recognized $P/BS_n$ to be the signal to noise ratio (SNR) at the channel output. It must be emphasized here that in order to achieve capacity messages must be block encoded and block decoded.

# 3 Quantum Information Theory

## 3.1 The *Bit* and the *Qubit*

The *bit* is the fundamental unit of classical information. The fundamental unit of quantum information is the *quantum bit*, or the *qubit* for short. Two possible states for a qubit are $|0\rangle$ and $|1\rangle$, which correspond to the states 0 and 1 of a classical bit. But unlike the classical bit, the qubit can also be in a *linear superposition* of states $|0\rangle$ and $|1\rangle$, for example,

$$|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$$

where $\alpha$ and $\beta$ are complex numbers. In other words, the state of a qubit is a vector in a two-dimensional complex vector space.

## 3.2 The Density Operator and the Von Neumann Entropy

The state of a quantum system is completely described by its density operator $\rho$, which is a hermitian positive semi-definite operator with trace equal to unity. The information content of a quantum state is given by the Von Neumann entropy $S(\rho)$, which is the quantum analog of the Shannon entropy,

$$S(\rho) = -\mathrm{Tr}\left(\rho \log_2 \rho\right) \tag{24}$$

Von Neumann's entropy plays a dual role. As clarified later in this paper, it quantifies not only the quantum information content in qubits of the quantum state (i.e. the minimum number of qubits needed to reliably encode the quantum state), but also its classical information content (i.e. the maximum information in bits that can be gained about the quantum state by making the best possible measurement). Below we consider three important cases :

1. Pure States: If $\rho = |\phi\rangle\langle\phi|$ then $S(\rho) = 0$.

2. Mixed States: If the density operator represents an ensemble of pure orthogonal states then,

$$\rho = \sum_{i=1}^{k} P_i |\phi_i\rangle\langle\phi_i| \tag{25}$$

   where $\langle\phi_i|\phi_j\rangle = \delta_{i,j}$. The entropy is,

$$S(\rho) = -\sum_{i=1}^{k} P_i \log_2 P_i = H(P) = \text{Shannon entropy of the ensemble} \tag{26}$$

   In this case the quantum source is essentially classical. If the states $|\phi_i\rangle$ were not orthogonal then $S(\rho)$ would be less than $H(\rho)$.

3. Ensemble of Mixed States: In this most general case,

$$\rho = \sum_{i=1}^{k} P_i \rho_i \tag{27}$$

   There is no simple expression for $S(\rho)$. However, if the density operators $\rho_i$ have support on orthogonal spaces then $S(\rho)$ will equal $H(P)$.

## 3.3 Quantum Data Compression and Schumacher's Quantum Noiseless Channel Coding Theorem

A quantum message consists of quantum letters (or quantum states) chosen from an alphabet $Q$ of $k$ letters,

$$Q = \{\rho_1, \rho_2, \rho_3, \ldots, \rho_k\} \tag{28}$$

The $j$-th letter occurs with an a priori probability $P_j$. The density operator for each letter is therefore,

$$\rho = \sum_{i=1}^{k} P_i \rho_i \tag{29}$$

The density operator for the entire message of $n$ letters is,

$$\rho^n = \rho \otimes \rho \otimes \rho \otimes \ldots \otimes \rho \tag{30}$$

We have assumed above that each letter in the message is i.i.d. Now we ask: is it possible to compress the message to a smaller Hilbert space without compromising the fidelity of the message? In 1995 Schumacher [7] showed that if the density operator for each letter represented an ensemble of pure (not necessarily orthogonal) states then with appropriate quantum block encoding operations the message can be compressed at best to a $2^{nS(\rho)}$ dimensional Hilbert space. The proof of Schumacher's theorem

closely resembles the proof of Shannon's noiseless channel coding theorem. For large $n$ the density operator $\rho^n$ in Equation (30) has all of its support on the typical subspace of the full Hilbert space of the messages. This conclusion follows from the corresponding classical statement if we consider the orthonormal basis in which $\rho$ is diagonal. Therefore, one can encode the typical subspace and ignore its orthogonal compliment without much loss of fidelity. The dimension of the typical subspace is $2^{nS(\rho)}$. Therefore, the quantum information in the message is only $S(\rho)$ qubits per letter.

If the density operator of each letter represents an ensemble of mixed quantum states, as in Equation (29), then the maximum achievable compressibility is not firmly established and is the subject of current research. Although Schumacher's theorem guarantees that compression to $S(\rho)$ qubits per letter can still be achieved, one can generally do better. In this most general case a lower bound on the compressibility is known. It can be shown that high fidelity compression to less than $I(\rho)$ qubits per letter is not possible, where $I(\rho)$ is given by the expression [8],

$$I(\rho) = S(\rho) - \sum_{i=1}^{k} P_i S(\rho_i) \tag{31}$$

$I(\rho)$ is also called the Holevo information associated with the density operator $\rho$. For ensembles of pure states or mixed orthogonal states it can be shown that compression to $I(\rho)$ qubits per letter can be achieved. Evidently, $I(\rho)$ depends not just on the density operator $\rho$, but also on the particular way that $\rho$ is realized as an ensemble of mixed states. It has been conjectured [8] that if the encoder is aware of the source alphabet and the a priori probabilities of each letter then compression to $I(\rho)$ qubits per letter may be asymptotically achievable.

## 3.4 Classical Information Content of Quantum States

The previous section was devoted to quantifying the quantum information content in qubits of messages constructed from an alphabet of quantum states. This section will focus on the classical information in bits that can be extracted from such messages by making appropriate measurements. The Von Neumann entropy and the Holevo information will also play a central role in what follows.

### 3.4.1 Generalized Quantum Measurements

The most general quantum measurements can be described in terms of a complete set of positive hermitian operators $\{F_j\}$ which provide a resolution of the identity operator,

$$\sum_j F_j = 1 \tag{32}$$

These generalized measurements constitute a positive operator valued measure (POVM). The probability $P_i$ that the outcome of a measurement will be $i$ is given as,

$$P_i = \mathrm{Tr}\,(\rho\, F_i) \tag{33}$$

### 3.4.2 The Holevo Bound

A theorem, stated by Gordon [9] (without proof), and proved by Holevo [10], gives an upper bound on the amount of classical information $I(M:P)$ that can be gained about the preparation of a quantum state $\rho$ (as given in Equation (29)) by making the best possible measurement,

$$\max_F I(M:P) \leq I(\rho) = S(\rho) - \sum_{i=1}^{k} P_i S(\rho_i) \tag{34}$$

The upper limit in Holevo's theorem can be achieved if and only if the quantum states of all the letters commute, i.e. $[\rho_i, \rho_j] = 0$. If all the letters commute then they can all be diagonalized in a common orthonormal basis, say $|\phi_\alpha\rangle$,

$$\rho_i = \sum_\alpha P(\alpha|i)|\phi_\alpha\rangle\langle\phi_\alpha| \tag{35}$$

If measurement is made using the POVM $\{F_\alpha\}$, where $F_\alpha = |\phi_\alpha\rangle\langle\phi_\alpha|$, then the probability $P_\alpha$ of measuring $\alpha$ is,

$$P_\alpha = \mathrm{Tr}\,(\rho\, F_\alpha) = \sum_i P(\alpha|i) P_i \tag{36}$$

The Shannon entropy $H(M)$ in the measurement outcome is equal to the Von Neumann entropy of the quantum state $\rho$, i.e.,

$$H(M) = -\sum_\alpha P_\alpha \log_2 P_\alpha = S(\rho) \tag{37}$$

The mutual information $I(M:P)$, which is the maximum classical information that can be gained about the preparation of the quantum state by making the measurement, is,

$$\begin{aligned} I(M:P) &= H(M) - H(M|P) && (38)\\ &= H(M) + \sum_i P_i \sum_\alpha P(\alpha|i) \log_2 P(\alpha|i) && (39)\\ &= S(\rho) - \sum_i P_i S(\rho_i) && (40)\\ &= I(\rho) && (41) \end{aligned}$$

Therefore, if the letters commute then $I(\rho)$ bits of classical information can be obtained about the preparation of the quantum state $\rho$ by making the best possible measurement. The similarity between the Holevo information and the classical mutual information is evident. If $\rho$ is an ensemble of pure or mixed orthogonal states then mutual information $I(M:P)$ will equal the source entropy $H(P)$. The essential message here is that information about the preparation is lost when non-orthogonal (pure or mixed) quantum states are chosen as letters. One can easily deduce the following conclusions :

1. If non-commuting quantum states are chosen as letters (which are necessarily non-orthogonal),

$$\max_F I(M:P) < I(\rho) < H(P) \tag{42}$$

2. For non-orthogonal, but commuting, letters (which cannot be pure states),

$$\max_F I(M:P) = I(\rho) < H(P) \tag{43}$$

3. For orthogonal letters (which are necessarily commuting),

$$\max_F I(M:P) = I(\rho) = H(P) \tag{44}$$

# 4 Channel Capacity Limits for Transmitting Classical Information Over Quantum Channels

Perhaps the oldest branch of quantum information theory deals with the capacity limits for transmitting classical information over quantum channels [11]. As shown in Fig. 2 through Fig. 5, there are four different ways in which classical information can be transmitted over quantum channels:

1. **Quantum Block Encoding and Quantum Block Decoding:** In this scheme the message is block encoded so that the density operator for the entire message is not necessarily a tensor product of the density operators of individual letters. The decoding operation is carried out by making the best possible measurement on the density operator of the entire message. The capacity for this channel is $C_{QQ}$.

2. **Classical Encoding and Quantum Block Decoding:** In this scheme the density operator for the message is a tensor product of the density operators of individual letters, but the decoding is still carried out by using a quantum block decoder. The capacity is $C_{CQ}$.

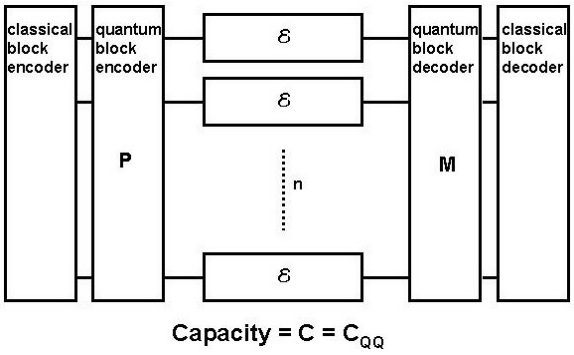


Figure 2: Channel with capacity $C_{QQ}$

3. **Quantum Block Encoding and Classical Decoding:** A classical decoder is used and separate measurements are made on each letter. However, the encoding is performed by a quantum block encoder. The capacity is $C_{QC}$.

4. **Classical Encoding and Classical Decoding:** A classical encoder is used and also separate measurements are made on each letter. The capacity is $C_{CC}$.

The capacities $C_{QC}$, $C_{QQ}$, and $C_{CC}$ are not known. But it is obvious from their definitions that $C_{CC} \leq \{C_{CQ}\,, C_{QC}\} \leq C_{QQ}$. The capacity $C_{CQ}$ is known [12]. Suppose the $n$ letter message at the input to the channel is described by the density operator $\rho^n$, where,

$$\rho^n = \rho \otimes \rho \otimes \rho \otimes \ldots \otimes \rho \tag{45}$$

and,

$$\rho = \sum_{i=1}^{k} P_i \rho_i \tag{46}$$

The channel is described by a trace preserving linear quantum operation $\mathcal{E}$ such that the density operator $\mathcal{E}(\rho)$ of each letter at the output of the channel is related to the density operator at the input to the channel by the relation,

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger \quad \text{where} \quad \sum_k E_k^\dagger E_k = 1 \tag{47}$$

The capacity $C_{CQ}$ of the channel in bits per letter is given by the relation [12],

$$C_{CQ} = \max_{P_i}\, I(\mathcal{E}(\rho)) = \max_{P_i}\, S(\mathcal{E}(\rho)) - \sum_{i=1}^{k} P_i S(\mathcal{E}(\rho_i)) \tag{48}$$

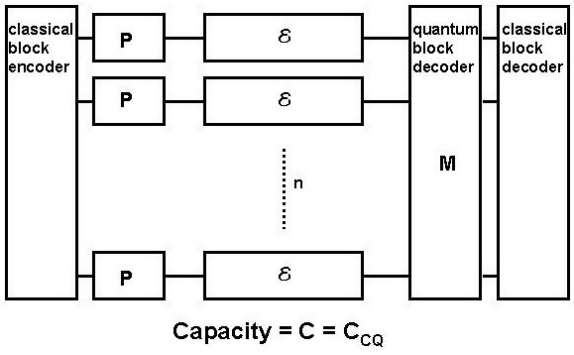


Figure 3: Channel with capacity $C_{CQ}$

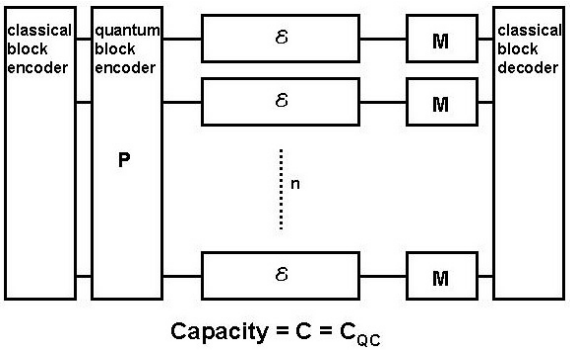


Figure 4: Channel with capacity $C_{QC}$

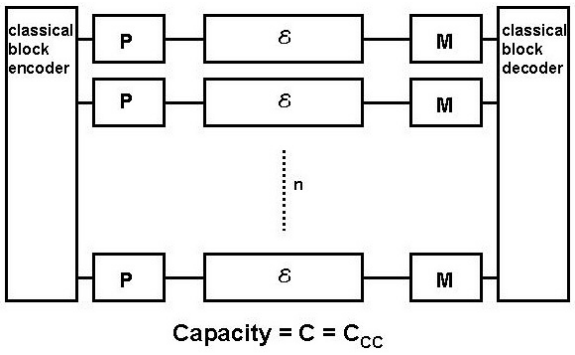


Figure 5: Channel with capacity $C_{CC}$

The above equation shows that $C_{CQ}$ is related to the Holevo information of the density operator at the output of the channel. This should not come as a surprise. In the last section it was shown that the maximum amount of classical information that can be obtained by performing the most suitable measurement on a quantum state is limited by the Holevo information. This upper limit is achievable if the letters of the alphabet commute. However, the result in Equation (48) holds even if the alphabet consists of non-commuting letters. The proof of Equation (48) closely resembles the proof of Shannon's noisy channel coding theorem. The typical subspace of the entire output message has $2^{nS(\mathcal{E}(\rho))}$ dimensions. If the input message is known to be,

$$\underline{\rho}^n = \rho_{i_1} \otimes \rho_{i_2} \otimes \ldots \otimes \rho_{i_n} \tag{49}$$

then the entropy of the output message would be $S(\mathcal{E}(\underline{\rho}^n))$. The mean entropy of the output message for a known input message is,

$$\begin{aligned} \langle S(\mathcal{E}(\underline{\rho}^n))\rangle &= \sum_{i_1,i_2,\ldots,i_n} P_{i_1} P_{i_2} \ldots P_{i_n}\, S(\mathcal{E}(\rho_{i_1} \otimes \rho_{i_2} \otimes \ldots \otimes \rho_{i_n})) && (50)\\ &= n \sum_i P_i S(\mathcal{E}(\rho_i)) && (51)\\ &= n\langle S(\mathcal{E}(\rho))\rangle && (52) \end{aligned}$$

For each typical input message the typical subspace of the output message would have $2^{n\langle S(\mathcal{E}(\rho))\rangle}$ dimensions. Thus, the entire $2^{nS(\mathcal{E}(\rho))}$ dimensional typical subspace can be divided into smaller $2^{n\langle S(\mathcal{E}(\rho))\rangle}$ dimensional sub-subspaces corresponding to each typical input message. At the channel output a measurement can be performed on the entire received message to determine the typical sub-subspace of the received message, and thereby determine the input message. This scheme would work provided the number of different input messages do not exceed,

$$\frac{2^{nS(\mathcal{E}(\rho))}}{2^{n\langle S(\mathcal{E}(\rho))\rangle}} = 2^{nI(\mathcal{E}(\rho))} \tag{53}$$

Therefore, the maximum information that can be conveyed by such a scheme is $I(\mathcal{E}(\rho))$ bits per letter. As in the classical case, the channel capacity $C_{CQ}$ is determined by maximizing $I(\mathcal{E}(\rho))$ with respect to the probabilities of the input letters. This capacity is also called the fixed alphabet product state capacity, since the optimization is not performed over the choice of input letters and the input letters are not assumed to be entangled over multiple uses of the channel.

# 5 Channel Capacities in Photonic Communications

The maximum amount of classical information that can be transmitted by electromagnetic waves is discussed in this section. Since the measurement schemes discussed below are not all optimum, the capacities computed below should be taken as the capacities achievable with a given measurement scheme (and a given input alphabet).

## 5.1 The Number State Channel with Photon Number Detection

For the number state channel the input alphabet consists of photon number eigenstates $\{|n\rangle\langle n|\}$. The input density operator $\rho$ for each letter is $\rho = \sum_n P_P(n)|n\rangle\langle n|$. The measurements are performed using an ideal photon counter and are described by the POVM $\{F_n = |n\rangle\langle n|\}$.

### 5.1.1 Noiseless Channel

In this case the POVM $\{F_n\}$ is optimum, and the capacity follows from Holevo's theorem,

$$C = \max_{P_P} I(\rho) = \max_{P_P} S(\rho) \tag{54}$$

The maximization is performed with the contraint that the average photon number $\sum_n nP_P(n)$ equals $n_o$. The resulting probability distribution turns out to be the thermal distribution,

$$P_P(n) = \frac{1}{1+n_o}\left(\frac{n_o}{1+n_o}\right)^n \tag{55}$$

and the capacity in bits per letter comes out to be,

$$C = \log_2(1+n_o) + n_o \log_2\left(1+\frac{1}{n_o}\right) \tag{56}$$

Assuming a channel bandwidth B, and average power $P = B\hbar\omega n_o$, the channel capacity in bits per second can be written as,

$$C = B\log_2\left(1+\frac{P}{B\hbar\omega}\right) + \frac{P}{\hbar\omega}\log_2\left(1+\frac{B\hbar\omega}{P}\right) \tag{57}$$

For signal powers much larger than $B\hbar\omega$ the above expression is identical to the classical expression given in Equation (23) for the capacity of the gaussian channel, provided it is assumed that the noise spectral density $S_n$ equals $\hbar\omega$. For signal powers much smaller than $B\hbar\omega$ the capacity equals,

$$C = \frac{P}{\hbar\omega}\log_2\left(\frac{B\hbar\omega}{P}\right) \tag{58}$$

This result can be understood as follows. For small signal powers one can choose a transmission time $T$ long enough such that one photon gets transmitted in time $T$. This would require choosing $T$ to equal $\hbar\omega/P$. If the channel bandwidth is $B$ then the transmission time $T$ can be divided into $BT$ time slots. The transmitted photon can occupy any one of these time slots. The channel capacity becomes,

$$C = \frac{\log_2(BT)}{T} = \frac{P}{\hbar\omega}\log_2\left(\frac{B\hbar\omega}{P}\right) \tag{59}$$

This type of coding technique is called pulse position modulation (PPM).

Equation (56) holds for a narrow band number state channel. To find the capacity of a wide band number state channel one needs to maximize,

$$C = \int \frac{d\omega}{2\pi}\, \log_2(1+n_\omega) + n_\omega \log_2\left(1+\frac{1}{n_\omega}\right) \tag{60}$$

subject to the power constraint,

$$P = \int \frac{d\omega}{2\pi}\, \hbar\omega n_\omega \tag{61}$$

which yields the wide band capacity [4],

$$C = \log_2(e)\sqrt{\frac{\pi}{3}}\sqrt{\frac{P}{\hbar}} \tag{62}$$

### 5.1.2 Channel with Thermal Noise

A channel with thermal noise can be described by the quantum operation $\mathcal{E}_{th}$. The action of $\mathcal{E}_{th}$ on the number state $|n\rangle\langle n|$ can be described as,

$$\mathcal{E}_{th}(|n\rangle\langle n|) = \sum_m P_{th}(m)|n+m\rangle\langle n+m| \tag{63}$$

where $P_{th}(m)$ denotes the thermal distribution,

$$P_{th}(m) = \frac{1}{1+n_{th}}\left(\frac{n_{th}}{1+n_{th}}\right)^m \tag{64}$$

The POVM $\{F_n\}$ is optimum even in the presence of thermal noise, and the capacity as before can be found from Holevo's theorem,

$$C = \max_{P_P(n)|\sum_n nP_P(n)=n_o} S(\mathcal{E}(\rho)) - \sum_n P_P(n) S(\mathcal{E}(|n\rangle\langle n|)) \tag{65}$$

$$= \max_{P_P(n)|\sum_n nP_P(n)=n_o} S(\mathcal{E}(\rho)) - S_{th}(n_{th}) \tag{66}$$

$$= \max_{P_P(n)|\sum_n nP_P(n)=n_o+n_{th}} \sum_n P_P(n)|n\rangle\langle n| - S_{th}(n_{th}) \tag{67}$$

$$= S_{th}(n_o + n_{th}) - S_{th}(n_{th}) \tag{68}$$

$$= \log_2\left(1+\frac{n_o}{1+n_{th}}\right) + (n_o+n_{th})\log_2\left(1+\frac{1}{n_o+n_{th}}\right) - n_{th}\log_2\left(1+\frac{1}{n_{th}}\right) \tag{69}$$

$S_{th}(n_{th})$ above denotes the entropy of a thermal state with an average number of photons equal to $n_{th}$. From the above equation it can be deduced that the minimum energy required to transmit one bit of information (at a vanishingly small rate) at temperature $T$ is $K_B T \ln(2)$.

## 5.2 The Coherent State Channel with Balanced Heterodyne Detection

For the coherent state channel the input alphabet consists of coherent states $\{|\alpha\rangle\langle\alpha|\}$. The input density operator $\rho$ for each letter is,

$$\rho = \sum_\alpha P_P(\alpha)|\alpha\rangle\langle\alpha| = \int d^2\alpha\, P_P(\alpha)|\alpha\rangle\langle\alpha| \tag{70}$$

The POVM for balanced heterodyne detection is $\{F_\beta\}$, where,

$$F_\beta = \frac{1}{\pi}\,|\beta\rangle\langle\beta| \tag{71}$$

This POVM implies that both the field quadratures are measured simultaneously. The channel capacity can easily be found from standard classical arguments. The conditional probability $P_{M|P}(\beta|\alpha)$ that coherent state $\beta$ is measured when $\alpha$ is prepared equals,

$$P_{M|P}(\beta|\alpha) = \frac{1}{\pi}\,|\langle\beta|\alpha\rangle|^2 = \frac{1}{\pi}\,e^{-|\beta-\alpha|^2} \tag{72}$$

The channel capacity can be found by maximizing the mutual information $I(M:P)$ over the input probability distribution $P_P(\alpha)$ subject to the average power constraint,

$$n_o = \int d^2\alpha\,|\alpha|^2\,P_P(\alpha) \tag{73}$$

which yields,

$$C = \log_2(1+n_o) \tag{74}$$

and the capacity is achieved with a gaussian distribution at the input,

$$P_P(\alpha) = \frac{1}{\pi n_o}\,\exp\left(-\frac{|\alpha|^2}{n_o}\right) \tag{75}$$

It should be emphasized here that the operators for the field quadratures don't commute. Therefore, simultaneous measurement of both the field quadratures results in noise in excess of that dictated by the uncertainty principle [3, 13]. To avoid this excess noise balanced homodyne detection can be used with coherent states, as discussed next.

## 5.3 The Coherent State Channel with Balanced Homodyne Detection

The input alphabet in this case are coherent states $\{|\alpha_1\rangle\langle\alpha_1|\}$. Only one of the field quadratures carries information. This quadrature is assumed to be $x_1 = (a+a^\dagger)/2$. The other quadrature is left unexcited. The input density operator $\rho$ for each letter is,

$$\rho = \int d\alpha_1\, P_P(\alpha_1)|\alpha_1\rangle\langle\alpha_1| \tag{76}$$

The measurement at the channel output is performed by the projection operators $\{F_{x_1} = |x_1\rangle\langle x_1|\}$, which project the state at the channel output on the eigenstates of the $x_1$ quadrature. The conditional probability $P_{M|P}(x_1|\alpha_1)$ that $x_1$ is measured when $\alpha_1$ is prepared equals,

$$P_{M|P}(x_1|\alpha_1) = |\langle x_1|\alpha_1\rangle|^2 = \frac{1}{\sqrt{2\pi\,(1/4)}} \exp\left(-\frac{(x_1-\alpha_1)^2}{2\,(1/4)}\right) \tag{77}$$

The capacity is found by maximizing the mutual information $I(M:P)$ subject to the power constraint,

$$n_o = \int d\alpha_1\, \alpha_1^2\, P_P(\alpha_1) \tag{78}$$

which gives,

$$C = \frac{1}{2}\log_2\left(1 + 4n_o\right) \tag{79}$$

The capacity achieving input distribution is,

$$P_P(\alpha_1) = \frac{1}{\sqrt{2\pi n_o}}\, \exp\left(-\frac{\alpha_1^2}{2n_o}\right) \tag{80}$$

For average photon numbers much larger than unity this scheme provides one-half the capacity achieved with balanced heterodyne detection. But, as shown later in the paper, balanced homodyne detection yields smaller bit error rates than balanced heterodyne detection.

### 5.4 The Coherent State Channel with Photon Number Detection

In this case the input alphabet consists of coherent states but the measurement is made using the POVM $\{F_n = |n\rangle\langle n|\}$. The conditional probability $P_{M|P}(n|\alpha)$ that number state $|n\rangle$ is measured when coherent state $|\alpha\rangle$ is prepared equals,

$$P_{M|P}(n|\alpha) = e^{-|\alpha|^2}\, \frac{|\alpha|^{2n}}{n!} \tag{81}$$

The channel capacity as before can be found by maximizing the mutual information $I(M:P)$ over the input probability distribution $P_P(\alpha)$ subject to the average power constraint,

$$n_o = \int d^2\alpha\, |\alpha|^2\, P_P(\alpha) \tag{82}$$

The maximization procedure turns out to be analytically cumbersome. However, in the limit $n_o << 1$ the capacity is,

$$C \approx n_o \log_2\left(\frac{1}{n_o}\right) \tag{83}$$

and in the limit $n_o >> 1$ the capacity for this channel turns out to be [11],

$$C \approx \frac{1}{2}\, \log_2(n_o) \tag{84}$$

The above results are not surprising. When $n_o << 1$ the capacity equals the capacity of the number state channel. When $n_o >> 1$ the capacity is one-half the capacity of a coherent state channel with balanced heterodyne detection. This is because when $n_o$ is much larger than unity half of the information is in the photon number and the other half is in the phase. By performing photon number measurements the information in the phase is thrown away. On the other hand when $n_o$ is much smaller than unity the phase is not well defined, and all the information is carried in the photon number.

### 5.5 The Quadrature Squeezed State Channel with Balanced Homodyne Detection

In the quadrature squeezed state channel the input alphabet consists of quadrature squeezed states $\{|\alpha_1, r\rangle\langle r, \alpha_1|\}$, where the displacement of the squeezed quadrature is $\alpha_1$ and the squeezing parameter is $r$. The squeezed quadrature is assumed to be $x_1 = (a + a^\dagger)/2$. The squeezed states are produced by first squeezing the vacuum and then displacing it [14],

$$|\alpha, r\rangle = \exp\left(\alpha_1 a^\dagger - \alpha_1 a\right) \exp\left(\frac{1}{2} r \left(a^2 - a^{2\dagger}\right)\right) |0\rangle \tag{85}$$

The resulting variance of the squeezed quadrature is $e^{-2r}/4$, and the variance of the amplified quadrature is $e^{2r}/4$. The displacement of the amplified quadrature is chosen to be zero in Equation (85), since information will be encoded only in the squeezed quadrature and unnecessary excitation of the amplified quadrature would waste power. The input density operator for each letter is,

$$\rho = \int d\alpha_1 \, P_P(\alpha_1) |\alpha_1, r\rangle\langle r, \alpha_1| \tag{86}$$

The measurement at the channel output is performed by the operators $\{F_{x_1} = |x_1\rangle\langle x_1|\}$. The conditional probability $P_{M|P}(x_1|\alpha_1)$ that $x_1$ is measured when $\alpha_1$ is prepared equals,

$$P_{M|P}(x_1|\alpha_1) = |\langle x_1|\alpha_1, r\rangle|^2 = \frac{1}{\sqrt{2\pi \left(e^{-2r}/4\right)}} \exp\left(-\frac{(x_1 - \alpha_1)^2}{2\left(e^{-2r}/4\right)}\right) \tag{87}$$

The capacity is found as usual by maximizing the mutual information $I(M : P)$ subject to the power constraint,

$$n_o = \sinh^2(r) + \int d\alpha_1 \, \alpha_1^2 \, P_P(\alpha_1) \tag{88}$$

which gives,

$$C = \frac{1}{2} \log_2 \left[1 + 4e^{2r}(n_o - \sinh^2(r))\right] \tag{89}$$

and the capacity maximizing input distribution is gaussian,

$$P_P(\alpha_1) = \frac{1}{\sqrt{2\pi(n_o - \sinh^2(r))}} \exp\left(-\frac{\alpha_1^2}{2(n_o - \sinh^2(r))}\right) \tag{90}$$

A further optimization with respect to the squeezing parameter $r$ yields the maximum capacity,

$$C = \log_2(1 + 2n_o) \tag{91}$$

which is achieved when,

$$e^{2r} = 2n_o + 1 \tag{92}$$

Comparing Equations (91) and (74) it is seen that for large average photon numbers the quadrature squeezed states transmit only one more bit per letter compared to coherent states. This does not seem all that attractive given the complexity required in generating squeezed states. But, as is shown in the next section, squeezed states do much better than coherent states in reducing the bit error rates.

## 6 Bit Error Rates in Photonic Communications

The channel capacity is not the only important figure of merit relevant to photonic communication channels. Bit error rates are equally, if not more, important. In this section we compute bit error rates for different photonic channels.

## 6.1 The Binary Coherent State Channel with Balanced Heterodyne Detection (Binary PSK)

The input alphabet for this channel consists of coherent state letters $|\alpha\rangle$ and $|-\alpha\rangle$. The two field quadratures carry equal amount of power. The error probability $Pr(E)$ for this channel can easily be found from the analysis done earlier,

$$Pr(E) \quad = \quad Q(\sqrt{2}|\alpha|) \leq \frac{1}{4}\exp\left(-|\alpha|^2\right) = \frac{1}{4}\exp\left(-n_o\right) \tag{93}$$

This scheme would require 20 photons per letter to get the probability of error below $10^{-9}$.

## 6.2 The Binary Coherent State Channel with Balanced Homodyne Detection (Binary PSK)

This is similar to the channel above except that all the power is placed in only one of the field quadratures. By doing so one avoids the excess noise associated with the simultaneous measurement of two non-commuting observables. The probability of error becomes,

$$Pr(E) = Q(2|\alpha|) \leq \frac{1}{4}\exp\left(-2|\alpha|^2\right) = \frac{1}{4}\exp\left(-2n_o\right) \tag{94}$$

Only 10 photons per letter would be required in this case to get the probability of error below $10^{-9}$. In a recent experiment by Kahn [15] 45 photons per bit were required to achieve a probability of error less than $10^{-9}$ using Binary PSK with balanced homodyne detection. Given the non-idealities in the experimental setup this result is impressive.

## 6.3 The Binary Coherent State Channel with Direct Photon Number Detection (Binary ASK)

In this case the letters used to communicate are the coherent states $|\alpha\rangle$ and $|0\rangle$. The photon detector is assumed to be ideal with no dark current. Thus, the probability of error when $|0\rangle$ is used is 0. The probability of error when $|\alpha\rangle$ is used equals $e^{-|\alpha|^2}$. The average probability of error becomes,

$$Pr(E) = \frac{1}{2}\exp\left(-|\alpha|^2\right) = \frac{1}{2}\exp\left(-n_o\right) \tag{95}$$

In this case also 10 photons per letter would be required on the average to achieve a probability of error less than $10^{-9}$.

## 6.4 The Binary Quadrature Squeezed State Channel with Balanced Homodyne Detection

The input letters in this case are the quadrature squeezed states $|\alpha_1, r\rangle$ and $|r, -\alpha_1\rangle$. The probability of error is,

$$Pr(E) = Q(\sqrt{4e^{2r}(n_o - \sinh^2(r))}) \tag{96}$$

Maximizing the probability of error with respect to the squeezing parameter $r$ yields the lowest achievable probability of error,

$$Pr(E) = Q(\sqrt{4n_o(n_o+1)}) \leq \frac{1}{4}\exp\left(-2n_o(n_o+1)\right) \tag{97}$$

Only 3 photons per letter can yield a probability of error less than $10^{-9}$ when quadrature squeezed states are used to communicate. Although squeezed states offer modest improvement in channel capacity over coherent states, they offer large improvements in bit error rates. However, error correcting codes can also easily reduce the probability of error by significant margins (at the expense of bit rates), and with the availability of relatively cheap high speed electronics forward error correction seems more attractive than using squeezed states.

# 7 Conclusion

In this paper principles from quantum information theory were used to study the limits of photonic communications. The Holevo bound represents the maximum amount of classical information that can be extracted from quantum states by performing suitable measurements. It plays a central role in determining the capacity of photonic channels in the presence of noise. The quantum information capacity in qubits of quantum channels has been left out from the discussion above for two reasons. First, the quantum capacity of channels is less well understood and most of the results available are not firmly established [5]. Second, a full discussion of quantum capacity would have shifted the focus away from the discussion of the most commonly used optical channels. For the same reasons other related topics, including the entanglement assisted classical capacity of quantum channels [16] and super dense coding, have also been left out. Photons can be used to convey quantum information in qubits and, therefore, it is likely that one day channels carrying quantum information may also become a part of the *most commonly used* optical channels.

## References


[1] C. E. Shannon, W. Weaver, *Mathematical Theory of Communication*, U. of Illinois Press (1963).

[2] S. Verdu, S. W. McLaughlin, *Information Theory: 50 Years of Discovery*, IEEE Press (1999).

[3] Y. Yamamoto, H. A. Haus, Rev. Mod. Phys. **58**, 1001 (1986).

[4] C. M. Caves, P. D. Drummond, Rev. Mod. Phys. **66**, 481 (1994).

[5] C. H. Bennet, P. W. Schor, IEEE Transactions on Information Theory **44**, 2724 (1998).

[6] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons (1991).

[7] B. Schumacher, Phys. Rev. **A 51**, 2738 (1995).

[8] M. Horodecki, Phys. Rev. **A 57**, 3364 (1998).

[9] J. P. Gordon, in *Quantum Electronics and Coherent Light*, edited by P. A. Miles, Academic Press (1964).

[10] A. S. KHolevo, Probl. Peredachi Inf. **9**, 177 (1973).

[11] J. P. Gordon, Proc. IRE **50**, 1898 (1962).

[12] B. Schumacher, M. D. Westmoreland, Phys. Rev. **A 56**, 131 (1997).

[13] E. Arthurs, J. L. Kelly, Bell Syst. Tech. J. **44**, 725 (1965).

[14] H. A. Haus, *Electromagnetic Noise and Quantum Optical Measurements*, Springer Verlag (2001).

[15] J. M. Kahn, IEEE Photon. Tech. Lett. **1**, 340 (1989).

[16] C. H. Bennet, P. W. Shor, J. A. Smolin, A. V. Thapliyal, Phys. Rev. Lett. **83**, 3081 (1999).