6.962 Week 12 Summary
Quantum Information Theory

Presenter: Yonina Eldar

May 5, 2001

1 Introduction

Quantum information theory refers to the distinctive information processing properties of quantum systems which arise when information is stored in or retrieved from quantum states. Previous attempts to describe quantum information have generally relied on the formulas of classical information theory supplemented with probabilities derived from quantum rules. A new approach is to treat information as a quantum concept and to ask what new insight can be gained by encoding this information in individual quantum systems. Instead of simply applying classical information theory to probabilities derived from quantum rules, we can adopt notions of coding and measures of information that are themselves distinctly quantum mechanical. Schumacher, in his landmark paper [1], introduced for the first time the notion of a quantum bit (*qubit*) together with quantum mechanical coding techniques and measures of information; this launched the field of quantum information.

Many of the central results of classical information theory have quantum analogs that have been discovered and developed recently. These include quantum mechanical notions of coding and channel fidelity [1, 2], compression of quantum information [1], bounds on classical information encoded in quantum systems [3], and bounds on quantum information sent reliably over a noisy quantum channel [4].

In this presentation we will start with an introduction to quantum states and measurements. We will then discuss two basic topics that quantum information theory deals with: Transmission of classical information over quantum channels and transmission of quantum information over quantum channels. We will focus primarily on the problem of detecting classical information conveyed by quantum states, and the problem of determining the minimal resources needed to transmit a quantum state over a noiseless quantum channel, namely the problem of quantum coding.

2 States and Measurement

Quantum theory is in essence a set of rules allowing the computation of probabilities for the outcomes of measurements performed on a quantum system. The new feature introduced by quantum theory is that the probabilistic nature of a measurement outcome is inherent to quantum physics, and is not due to imperfections in the measurement process. A quantum system is described by a *state* where a pure state is a vector in a Hilbert (inner-product) space. In this summary, and in the presentation, we will use the quantum "bra-ket" notation, in which $|\psi\rangle$ denotes a column vector ("ket") and the corresponding row representation $\langle\psi|$ ("bra") is obtained by transposing and taking the complex conjugate. The inner product between two vectors is denoted by $\langle\psi|\phi\rangle$. (Although possibly confusing at first, this notation is actually very convenient once you get a hang of it – maybe we can start a campaign to introduce it into the signal processing/information theory/communication literature ...)

A measurement of a quantum system consists of a complete set of orthogonal projection operators $\{P_n\}$, *i.e.,* a set of operators satisfying

$$P_n = P_n^*, \quad P_n^2 = P_n, \quad P_n P_m = \delta_{mn} P_m, \quad \sum_n P_n = I. \tag{1}$$

The measurement process selects one of the projectors, where the probability of obtaining outcome k given that the system is in the state $|\psi\rangle$ is

$$p(k) = \langle \psi | P_k \psi \rangle. \tag{2}$$

Right after the measurement the quantum state collapses (is projected) onto a state that is consistent with the measurement outcome. Thus if output k is obtained, then the state after the measurement is a normalized vector in the direction of $P_k|\psi\rangle$[1].

The simplest type of measurement is a measurement consisting of rank-one projections of the form $P_k = |h_k\rangle\langle h_k|$, where the vectors $|h_k\rangle$ are mutually orthonormal. In this case (2) reduces to

$$p(k) = |\langle \psi | h_k \rangle|^2, \tag{3}$$

and if output k is obtained, then the state after the measurement is $|h_k\rangle$.

As an example, suppose the state of a system is given by $|\psi\rangle = \frac{1}{2}|h_1\rangle + \frac{\sqrt{3}}{2}|h_2\rangle$ where $|h_1\rangle$ and $|h_2\rangle$ are orthonormal vectors in a two-dimensional Hilbert space. For our measurement we choose the orthogonal projections $P_1 = |h_1\rangle\langle h_1|$ and $P_2 = |h_2\rangle\langle h_2|$. Then the probability of obtaining outcome 1 is $p(1) = \langle\psi|P_1\psi\rangle = |\langle\psi|h_1\rangle|^2 = 1/4$. Similarly, $p(2) = |\langle\psi|h_2\rangle|^2 = 3/4$. If *e.g.,* outcome 1 is obtained, then the state of the system right after the measurement is $|h_1\rangle$. The measurement process is illustrated in Figure 1.

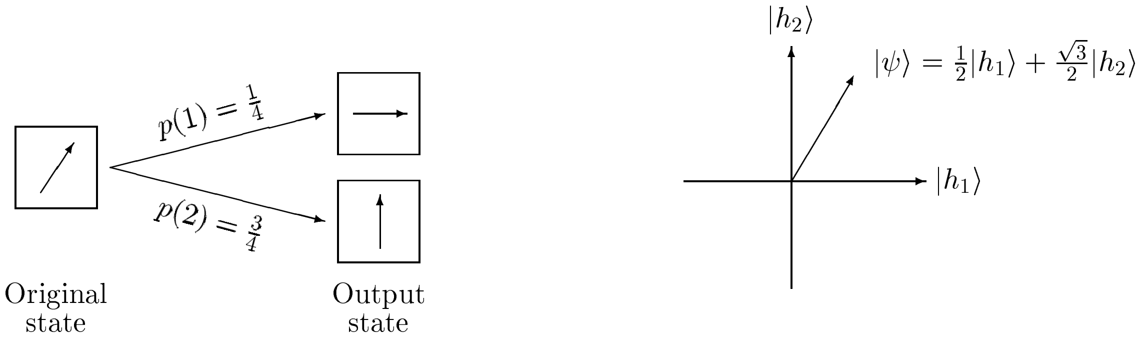

Figure 1: Illustration of a quantum measurement.

In general two quantum states are reliably distinguishable if and only if their vector representations are orthogonal, as demonstrated in the following example. Assume we have a quantum

[1]By adding an auxiliary system and performing orthogonal measurements on the combined system, we can implement *generalized measurements* [5], where a generalized measurement consists of a set of positive Hermitian operators A_n that are not restricted to be projection operators but satisfy $\sum_n A_n = I$. Such a set of operators is termed a *positive operator-valued measure (POVM).* In this presentation we will focus our attention on orthogonal measurements.

system that is prepared in one of the states $|0\rangle = (1\ 0)^T$ or $|1\rangle = (0\ 1)^T$ with given probabilities. To distinguish between the two preparations we measure $P_0 = |0\rangle\langle 0|$ and $P_1 = |1\rangle\langle 1|$. Using (3) we can calculate the probabilities of the different outputs given that the system was prepared in one of the two states. If the system was prepared in the state $|0\rangle$, then the measurement P_0 will yield $|0\rangle$ with probability $|\langle 0|0\rangle|^2 = 1$. If the system was prepared in the state $|1\rangle$, then the measurement P_1 will yield $|1\rangle$ with probability $|\langle 1|1\rangle|^2 = 1$. Therefore, we can determine the state of the system with probability one. On the other hand, if the system was prepared in one of the states $|0\rangle = (1\ 0)^T$ or $|+\rangle = 1/\sqrt{2}(1\ 1)^T$, then no measurement can perfectly determine the state. Indeed, if $|+\rangle$ was prepared then the measurement P_0 will yield $|0\rangle$ with probability $|\langle +|0\rangle|^2 = 1/2$. Thus if the outcome $|0\rangle$ is obtained, then we cannot perfectly determine which state was prepared.

As we said in the introduction, quantum states can be used to convey both classical information and quantum information. Thus, two main topics that quantum information theory deals with are:

- Transmission of classical information over quantum channels;
- Transmission of quantum information over quantum channels.

The first topic includes the problem of detecting classical information conveyed by quantum states, and the problem of determining the number of bits that can be sent reliably over a quantum channel. In the presentation we consider only the first problem; the second problem is discussed in [3, 4].

The second topic includes the problem of quantum coding, *i.e.,* determining the minimal resources needed to transmit a quantum state over a noiseless quantum channel. The problem of transmitting quantum information over a noisy quantum channel is tied in with quantum error correction, which is a very broad topic – and beyond our scope!

3 Classical Information via Quantum States

Suppose that a sender, Alice, wants to transfer classical information to a receiver, Bob, using a quantum mechanical channel (for some reason the characters in quantum information are always called Alice and Bob). Alice sends information by sending one of a possible set of quantum states $\{|\psi_k\rangle, 1 \leq k \leq m\}$. This setup is very similar to a classical communication problem but instead of Alice sending bits or waveforms to represent the different messages, Alice is now sending quantum states. For example, to send a $'0'$ Alice can send the state $\uparrow$ and to send a $'1'$ Alice can send the state $\nearrow$. Bob receives the quantum system which is in one of the possible states. In order for Bob to detect the information, he needs to measure the system to decide which state was prepared. If the quantum states are orthogonal, then Bob can determine the state correctly with probability one [6] by performing a measurement that consists of projections onto the given states. However, if the given states are not orthogonal, then no measurement will allow Bob to distinguish perfectly between them. Bob's problem is therefore to construct a measurement optimized to distinguish between non-orthogonal quantum states. If we restrict our attention to measurements consisting of rank-one projections of the form $\{P_k = |h_k\rangle\langle h_k|, 1 \leq k \leq m\}$ for some set of orthonormal measurement vectors $\{|h_k\rangle\}$, then Bob's problem is to choose a 'good' set of orthonormal vectors. This is the quantum detection problem in its simplest form.

A straightforward approach is to seek the orthonormal measurement vectors $|h_k\rangle$ that maximize the probability of detection, which is given by

$$P_d = \sum_{k=1}^{m} p(\psi_k)p(h_k|\psi_k) = \sum_{k=1}^{m} p(\psi_k)|\langle\psi_k|h_k\rangle|^2, \tag{4}$$

where $p(\psi_k)$ is the prior probability of $|\psi_k\rangle$. Necessary and sufficient conditions for an optimum measurement maximizing (4) have been derived [7, 8, 9]. However, except in some particular cases [9, 10, 11, 12], obtaining a closed-form analytical expression for the optimal measurement directly from these conditions is a difficult and unsolved problem. Thus in practice, iterative procedures maximizing the probability of detection [13] or *ad-hoc* suboptimal measurements are used.

An alternative approach is to choose a different criterion; if you can't solve it ... change it! Instead of using a probability of detection criterion we can use a different criterion that makes sense in this problem and that has a simple closed-form solution. Such a criterion is a least-squares criterion where we choose a set of orthonormal measurement vectors $|\hat{h}_k\rangle$ that are closest to the states $|\psi_k\rangle$ in a least-squares sense, as illustrated in Figure 2. Thus, the measurement vectors $|h_k\rangle$ are chosen to be orthonormal and to minimize the squared error E, defined by

$$E = \sum_{k=1}^{m} \langle e_k|e_k\rangle, \tag{5}$$

where $|e_k\rangle = |\psi_k\rangle - |h_k\rangle$. The minimizing vectors are given by the columns of $\widehat{H} = \Psi(\Psi^*\Psi)^{-1/2}$ where Ψ is the matrix of columns $|\psi_k\rangle$. This measurement is referred to as the least-squares measurement (LSM) [14].

It turns out that this measurement has many desirable properties. Its construction is relatively simple; it can be determined directly from the given states; it minimizes the probability of a detection error when the states exhibit certain symmetries [14, 12]; it is "pretty good" when the states to be distinguished are equally likely and almost orthogonal [15]; and it is asymptotically optimal [3].

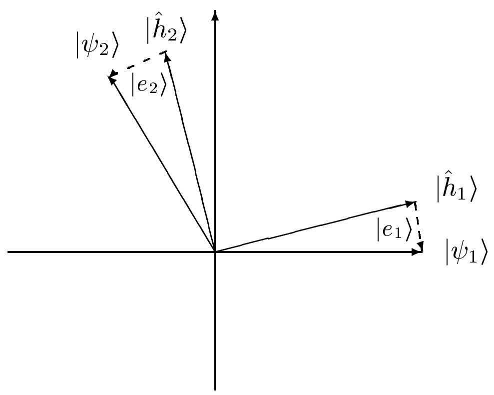

Figure 2: 2-dimensional example of the LSM.

4 Quantum Information

We now turn to discuss the problem of using quantum states to convey quantum information. In particular, we consider coding techniques and measures of information that are distinctly quantum mechanical.

4.1 The Qubit

The elementary quantity of classical information is the bit, which can take on one of two values, 0 or 1. The corresponding unit of quantum information is the quantum bit or *qubit*. (This term was first suggested by Schumacher in [1].) The states 0 and 1 are represented by a pair of orthogonal vectors corresponding to reliably distinguishable states of the qubit. In correspondence with the classical bit we can call these states $|0\rangle$ and $|1\rangle$. The novel feature of quantum information is that a qubit can exist in a continuum of intermediate states ("superpositions") represented mathematically as unit vectors in a two-dimensional inner-product space (the Hilbert space $\mathcal{H}_2$). Any normalized vector in the space can be represented as:

$$|\psi\rangle = a|0\rangle + b|1\rangle, \quad |a|^2 + |b|^2 = 1, \tag{6}$$

where $a, b \in \mathcal{C}$. Intermediate states (*i.e.*, $a, b \neq 0$) cannot be reliably distinguished from the bases states. With respect to any measurement that distinguishes the states $|0\rangle$ and $|1\rangle$ the superposition $|\psi\rangle = a|0\rangle + b|1\rangle$ behaves like $|0\rangle$ with probability $|a|^2$ and like $|1\rangle$ with probability $|b|^2$. The measurement irrecoverably disturbs the state (unless $a = 0$ or $b = 0$), since it projects the state onto the output state. If the value of the qubit is initially unknown then there is no way to determine a and b with any conceivable measurement. However, after the measurement the qubit has been prepared in a known state - either $|0\rangle$ or $|1\rangle$ - that differs (in general) from its previous state. In this respect a classical bit differs from a qubit. We can measure a classical bit without disturbing it and we can extract all the information it encodes.

The state of n qubits is represented by a complex unit vector in the 2^n-dimensional Hilbert space defined as the tensor product of n copies of the 2-dimensional Hilbert space representing states of a single qubit[2]. Thus, a pair of qubits can exist in four bases states, *e.g.*, $|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle$, as well as all possible superpositions of them (where *e.g.*, $|0\rangle \otimes |1\rangle$ represents a pair of qubits for which the state of the first qubit is $|0\rangle$ and the state of the second qubit is $|1\rangle$). The Hilbert space of two qubits contains *separable states*, which can be interpreted in terms of individual states for the two qubits, *e.g.*, $|\psi\rangle = 1/\sqrt{2}(|0\rangle \otimes |0\rangle + |0\rangle \otimes |1\rangle) = 1/\sqrt{2}|0\rangle \otimes (|0\rangle + |1\rangle)$, as well as *entangled states*, where neither qubit on its own has a definite state although the pair together does, *e.g.*, $|\psi\rangle = 1/\sqrt{2}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle)$.

4.2 The Quantum Source

A quantum source, illustrated in Figure 3, prepares messages of n letters where each letter is chosen independently from an ensemble $\varepsilon = \{p_i, |\psi_i\rangle\}$ consisting of states $\{|\psi_i\rangle\}$ (not necessarily

[2]The tensor product of two matrices $A(k \times l)$ and $B(m \times n)$ is defined as the $(mk \times nl)$ matrix $A \otimes B = \begin{bmatrix} a_{00}B & a_{01}B & \dots & a_{0,l-1}B \\ a_{10}B & a_{11}B & \dots & a_{1,l-1}B \\ \vdots & \vdots & & \vdots \\ a_{k-1,0}B & a_{k-1,1}B & \dots & a_{k-1,l-1}B \end{bmatrix}$.

orthogonal) emitted with probabilities $\{p_i\}$.

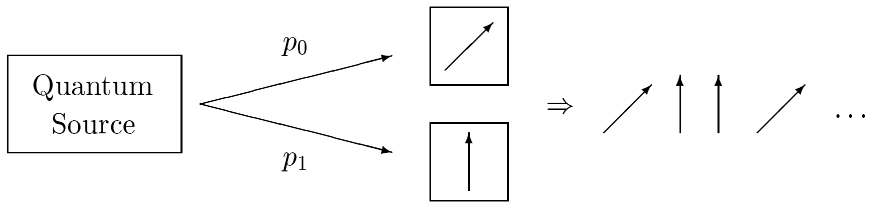

Figure 3: A quantum source.

How do we characterize the sate of a letter emitted by the source? Suppose that we perform a measurement consisting of the projections $\{P_i\}$, to determine the letter. The probability of obtaining *e.g.,* the outcome 0 is given by

$$p(0) = \sum_i p_i p(0|\psi_i) = \sum_i p_i \langle \psi_i | P_0 \psi_i \rangle = \mathrm{Tr}(\rho P_0), \tag{7}$$

where ρ is the density matrix defined as

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{8}$$

Thus, the sate of a letter emitted by the source is characterized by the density matrix given by (8). Since the letters are drawn independently, the entire message of length n is described by the density matrix

$$\rho^n = \rho \otimes \ldots \otimes \rho. \tag{9}$$

From its definition, the density matrix is Hermitian, non-negative definite, and $\mathrm{Tr}(\rho) = 1$. We will see that these properties are very important in proving the quantum coding theorem.

4.3 von Neumann Entropy

We now want to associate an entropy with the quantum source we just described. We have seen that the state of a letter emitted by the source is given by the density matrix ρ. Therefore we expect the measure of entropy to depend on this matrix. Indeed the expression for entropy in quantum systems, first proposed by von Neumann [16] thus termed the von Neumann entropy, is given by

$$S(\rho) = -\mathrm{tr}(\rho \log \rho). \tag{10}$$

If we choose an orthonormal basis $\{|a\rangle\}$ that diagonalizes ρ, $\rho = \sum_a \lambda_a |a\rangle\langle a|$, then $S(\rho) = H(A)$ where $H(A)$ is the Shannon entropy of the ensemble $A = \{\lambda_a, a\}$. As discussed in Section 2, when the source ensemble consists of mutually orthogonal pure states, all of the states can be perfectly distinguished. Given the output of the source we can determine the state the source prepared with probability one. Thus, the quantum source reduces to a classical one and $S(\rho) = H(\{p_m\}) = -\sum_m p_m \log p_m$. The quantum source is more interesting when the states are not orthogonal. In this case the states are not fully distinguishable as physical states and the ensemble's von Neumann entropy is less than the Shannon entropy of the probabilities, $S(\rho) < H(\{p_m\})$ [17]. Intuitively, the entropy of a source represents the information gain when the output of the source is revealed. If the states of the quantum source are not orthogonal, no measurement of the output can fully determine its state. Thus, the information gain when we are given the output is less than the uncertainty associated with the prior probabilities.

5 Quantum Coding

In [1] Schumacher proves a quantum coding theorem that is analogous to the noiseless coding theorem of classical information theory. Classical data compression allows information from a redundant source, *e.g.,* a binary source emitting 0 and 1 with unequal probability, to be compressed to the Shannon entropy of the source when coding large blocks of data and allowing for a small probability of error. Similarly, quantum data compression allows qubits from a redundant quantum source to be compressed to the von Neumann entropy of the quantum source with arbitrarily high fidelity, when coding large blocks.

Assume we have a quantum source that prepares messages of n letters where each letter is chosen independently from an ensemble $\varepsilon = \{p_a, |a\rangle\}$. Each letter is described by the density matrix $\rho = \sum_a p_a |a\rangle\langle a|$, and the entire message of length n is described by the density matrix $\rho^n = \rho \otimes \ldots \otimes \rho$.

We would like to devise a quantum code that will result in compression of the message to a smaller Hilbert space without comprising the fidelity of the message. Schumacher [1] proves that the best possible compression with arbitrarily good fidelity as $n \to \infty$ is compression to a Hilbert space $\mathcal{H}$ of dimension $\dim(\mathcal{H}) = 2^{nS(\rho)}$, which is equivalent to $nS(\rho)$ qubits. In this sense the von Neumann entropy is the number of qubits of quantum information carried per letter. The proof of this theorem relies heavily on the noiseless coding theorem, which we now briefly discuss. (The proof in [1] is quite obscure; a more coherent proof of this theorem is given in [18]).

5.1 Classical Noiseless Coding Theorem

We begin by defining a message as a string of n statistically independent letters chosen from an ensemble $A = \{p(a_m), a_m\}$, where the letter a_m occurs with probability $p(a_m)$. We want to represent this message by a sequence of bits with minimal length. A particular n-letter message $a_1 a_2 \ldots a_n$ occurs with probability $p(a_1 a_2 \ldots a_n) = p(a_1)p(a_2)\ldots p(a_n)$, thus $\log p(a_1 a_2 \ldots a_n) = \sum_{m=1}^{n} \log p(a_m)$. If n is large we can apply the strong law of large numbers to $-\frac{1}{n}\sum_{m=1}^{n} \log p(a_m)$ and conclude that

$$-\frac{1}{n}\log p(a_1 a_2 \ldots a_n) = -\frac{1}{n}\sum_{m=1}^{n} \log p(a_m) \sim E(-\log p(a)) \triangleq H(A), \tag{11}$$

where $H(A)$ is the Shannon entropy of the ensemble defined as

$$H(A) = -\sum_m p(a_m)\log p(a_m) = -E(\log p(a)) \tag{12}$$

(we use base-2 logarithms). We define a *typical message* as a message for which $p(a_1 a_2 \ldots a_n) \sim 2^{-nH(A)}$. Since each typical message has a probability of approximately $2^{-nH(A)}$, and the probability that a message is typical is almost 1, there are approximately $2^{nH(A)}$ typical messages. Therefore, with a block code of length $n(H(A))$ bits we can encode all typical messages. Since an atypical message will occur with negligible probability, it doesn't matter what we do in that case.

5.2 Quantum Data Compression

In the classical noiseless theorem there are three central features. The first is the elementary quantity of information - the bit. Second, instead of encoding individual letters we encode messages consisting of n independent letters. Third, we allow for an arbitrarily small probability of error. Each of these three essential ideas has a quantum counterpart. The elementary quantity of information is the qubit. All messages are encoded into sequences of qubits. The "closeness" of the input and output states is measured by the quantum *fidelity*. The fidelity F is defined as the overall probability that the decoded letter passes a "validation test", *i.e.,* the probability that the decoded letter, represented by the density matrix w_a, is equal to the original letter $|a\rangle$. Thus,

$$F = \sum_a p_a \langle a|w_a a\rangle. \tag{13}$$

From (13), $0 \leq F \leq 1$ and $F = 1$ only in the case of perfect decoding of all letters.

The key to Shannon's noiseless coding theorem is that we can code the typical messages and ignore the rest, without much loss of fidelity. In quantum coding we introduce the notion of a *typical subspace*. The key to Schumacher's noiseless quantum coding theorem is that we can code the typical subspace and ignore its orthogonal complement without much loss of fidelity.

The density matrix for the letter ensemble can be written in terms of its eigenvectors and eigenvalues as $\rho = \sum_{i=1}^{2} \lambda_i |\psi_i\rangle\langle\psi_i|$. Since ρ is Hermitian, non-negative definite, and $\mathrm{Tr}(\rho) = 1$ we have that

$$\lambda_1, \lambda_2 \geq 0,\ \ \lambda_1 + \lambda_2 = 1. \tag{14}$$

The density matrix of the n letter message has 2^n eigenvectors of the form $|\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle$, where $|\psi_m\rangle$ $m = 1, \ldots n$ can be either of the two eigenvectors $|\psi_1\rangle, |\psi_2\rangle$. The corresponding eigenvalue is

$$\lambda_1 \lambda_2 \ldots \lambda_n. \tag{15}$$

These 2^n eigenvectors are orthonormal and span the Hilbert space $\mathcal{H} = \mathcal{H}_2 \otimes \ldots \otimes \mathcal{H}_2$ where $\mathcal{H}_2$ is the two dimensional Hilbert space. Thus, every vector in $\mathcal{H}$ can be expressed as a linear combination of the eigenvectors of ρ^n. The conditions of (14) imply that we can regard the eigenvalues of ρ as a probability distribution, *i.e.,* we can imagine a classical source that emits two letters with probabilities λ_1, λ_2 and has Shannon entropy $H(\{\lambda_i\}) = S(\rho)$. The eigenvalue of (15) then represents the probability of a sequence of n independent letters emitted by the classical source. We can now use Shannon's results regarding typical messages to conclude that there are no more than $2^{n(S(\rho)+\epsilon)}$ eigenvalues of the form of (15) that have a total sum greater than $1 - \epsilon$. We define the typical subspace as the subspace spanned by the eigenvectors corresponding to these eigenvalues. Intuitively, the density matrix ρ^n has most of its support on the typical subspace. Denoting the typical subspace by Λ and the projection onto this subspace by P_Λ it follows that $\mathrm{Tr}(\rho^n P_\Lambda) > 1 - \epsilon$ (since the trace is equal to the sum of the "typical" eigenvalues).

We can now use the following coding scheme: Perform the measurement P_Λ on the message. The measurement projects the message onto either Λ or $\Lambda^\perp$ (the orthogonal complement of Λ).

The probability of projecting onto the typical subspace is given by $p(\Lambda) = \mathrm{Tr}(\rho^n P_\Lambda) > 1-\epsilon$. The projected state can be encoded using $n(S(\rho)+\epsilon)$ qubits since it lies in a subspace of $\mathcal{H}$ of dimension $2^{n(S(\rho)+\epsilon)}$. Since the probability of projecting onto $\Lambda^\perp$ can be made arbitrarily small for large n it matters little what we do in that case. Using this coding scheme we can represent an n letter message using $nS(\rho)$ qubits, or $S(\rho)$ qubits per letter. A mathematical proof of the argument outlined above as well as a converse are given in [18].

We could imagine a situation where we prepare a message using the quantum source by first drawing a sequence of n numbers $i_1, i_2, \ldots i_n$ independently from the ensemble $\varepsilon = \{p_i, i\}$, and then prepare the quantum message $|a_{i_1}\rangle \otimes \ldots \otimes |a_{i_n}\rangle$. We can encode this message by encoding the classical information $i_1, i_2, \ldots i_n$ in mutually orthogonal quantum states. We can then achieve high fidelity compression to $H(\{p_i\})$ bits or qubits per letter. However, if the letters are drawn from nonorthogonal pure states, this procedure is not optimal. Some of the classical information regarding the preparation of the state is redundant since nonorthogonal quantum states cannot be perfectly distinguished. Using Schumacher's scheme we can compress the message to $S(\rho) < H(\{p_i\})$ qubits per letter. In this case the decoded qubits will be in (almost) the same quantum state as the original qubits, however there is no way to fully determine this state by measuring the qubits. So achieving optimal compression comes with a price – we don't know what state we have and any attempt to determine the state will irrecoverably disturb it.

To summarize, there is a close analogy between the classical noiseless coding theorem and the quantum noiseless coding theorem. Classical compression is achieved by noting that with high probability a message emitted from the source is typical. We can code only the typical messages and still have a small probability of error. Compression in the quantum case is achieved by noting that with high probability a message emitted from the source will lie in the typical subspace. We can encode only the typical subspace and still achieve high fidelity.

References

[1] B. Schumacher, "Quantum coding," *Phys. Rev. A*, vol. 51, no. 4, pp. 2738–2747, 1995.

[2] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher, "General fidelity limit for quantum channels," *Phys. Rev. A*, vol. 54, no. 6, pp. 4707–4711, 1996.

[3] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wooters, "Classical information capacity of a quantum channel," *Phys. Rev. A*, vol. 54, no. 3, pp. 1869–1876, 1996.

[4] B. Schumacher and M. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, pp. 131–138, 1997.

[5] A. Peres, "Newmark's theorem and quantum inseparability," *Found. Phys.*, vol. 20, no. 12, pp. 1441–1453, 1990.

[6] A. Peres, *Quantum Theory: Concepts and Methods*, Kluwer, Boston, 1995.

[7] A. S. Holevo, "Statistical decisions in quantum theory," *J. Multivar. Anal.*, vol. 3, pp. 337–394, Dec. 1973.

[8] R. S. Kennedy H. P. Yuen and M. Lax, "Optimum testing of multiple hypotheses in quantum detection theory," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 125–134, Mar. 1975.

[9] C. W. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press, Inc. , 1976.

[10] C. Bendjaballah M. Charbit and C. W. Helstrom, "Cutoff rate for the m-ary psk modulation channel with optimal quantum detection," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1131–1133, Sep. 1989.

[11] M. Ban M. Osaki and O. Hirota, "Derivation and physical interpretation of the optimum detection operators for coherent-state signals," *Phys. Rev. A*, vol. 54, pp. 1691–1701, Aug. 1996.

[12] R. Momose M. Ban, K. Kurukowa and O. Hirota, "Optimum measurements for discrimination among symmetric quantum states and parameter estimation," *Int. J. Theor. Phys.*, vol. 36, pp. 1269–1288, 1997.

[13] C. W. Helstrom, "Bayes-cost reduction algorithm in quantum hypothesis testing," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 359–366, Mar. 1982.

[14] Y. C. Eldar and G. D. Forney, Jr., "On quantum detection and the square-root measurement," *IEEE Trans. Inform. Theory*, vol. 47, pp. 858–872, Mar. 2001.

[15] P. Hausladen and W. K. Wootters, "A 'pretty good' measurement for distinguishing quantum states," *J. Mod. Opt.*, vol. 41, pp. 2385–2390, 1994.

[16] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, 1955.

[17] L. P. Hughston, R. Jozsa, and W. K. Wooters, "A complete classification of quantum ensembles having a given density matrix," *Phys. Letters A*, vol. 183, pp. 14–18, 1993.

[18] B. Schumacher and R. Jozsa, "A new proof of the quantum noiseless coding theorem," *J. Mod. Opt.*, vol. 41, pp. 2343, 1994.