A New Proof of the Quantum Noiseless Coding Theorem

Richard Jozsa,[1] Benjamin Schumacher[2]

[1] *Département IRO, Université de Montréal, C.P. 6128, Succursale "A", Montréal (Québec), Canada H3C 3J7*

[2] *Department of Physics, Kenyon College, Gambier, OH 43022, USA*

Abstract

We give an account of the quantum noiseless coding theorem, including a new proof based on a simplified block coding scheme. We also discuss an illustrative example of quantum coding.

Introduction

Suppose that a quantum source produces a sequence of signals, each of which is a pure state from the list $|a_1\rangle, \ldots, |a_m\rangle$, occurring with *a priori* probabilities $p_1, \ldots, p_m$ and the states are assumed to lie in an n-dimensional Hilbert space $\mathcal{H}_n$. We may associate the density matrix

$$\rho = \sum_{i=1}^{n} p_i |a_i\rangle\langle a_i|$$

to the source.

In general the states $|a_i\rangle$ are not mutually orthogonal so even though we know the list of possible states and their *a priori* probabilities, there is no reliable way of identifying the signals with certainty.[2]

Suppose that we wish to store or code the signals with a minimum of resources, in the sense that we want to use the least possible number of Hilbert space dimensions per signal for the coding or storage. Equivalently we may consider the problem of transmitting the signals faithfully over a quantum channel of fixed dimension, asking for the best compression of the signal stream, i.e. the best transmission rate.

Clearly a storage capacity of n dimensions per signal suffices for perfect storage. However we may improve on this if we allow two reasonble concessions:

(a) We allow "block coding" i.e. we code strings of signals rather than individual signals separately.

(b) We tolerate a small error in the signals reconstituted from the coded version. This error can be made *arbitrarily* small without increasing the number of dimensions per signal (at the expense of coding longer blocks).

With regard to (a), we introduce the notion of K-blocking of the source. Suppose for example, that a source A emits the three states $|a_i\rangle$, $i = 1, 2, 3$ with equal probability. Then we may view the source alternatively as the source A_2 of strings of signals of length two i.e. A_2 emits the nine states $|a_i\rangle|a_j\rangle$ with equal probability. Similarly given any source A we may consider the K-blocked version A_K of A. If A has m distinct signals in $\mathcal{H}_n$, then A_K will have m^K signals in $\mathcal{H}_{n^K}$.

Suppose now that after coding and reconstitution, the signal $|a_i\rangle$ (or more generally a K-block of signals) is reconstituted as a mixed state with density matrix W_i. (This is the most general possible description.) As the number of dimensions is reduced in coding, typically part of a system will be discarded in the process giving an evolution from a pure state to a mixed state. For perfect reconstitution we would have $W_i = |a_i\rangle\langle a_i|$ but in view of (b) this will not generally be the case. We define the fidelity F of the coding scheme to be [1]

$$F = \sum_i p_i \langle a_i | W_i | a_i \rangle \qquad (1)$$

Recall that $\langle a_i|W_i|a_i\rangle$ is the probability that the state W_i passes the yes/no test of being the state $|a_i\rangle$ (where the test is the measurement of the observable $proj_{|a_i\rangle}$). A perfect coding scheme has $W_i = |a_i\rangle\langle a_i|$ and $F = 1$, but in general $0 \leq F \leq 1$. Note that the operation of reconstitution or decoding must be a fixed unitary operation on the coded signal, independent of the signal state, since the coded signal constitutes the entire information we have about the original state. If d dimensions per signal were used to code the states, then the W_i's will be density matrices all supported in some fixed d dimensional subspace of $\mathcal{H}_n$. The quantum noiseless coding theorem, first introduced and proved in [1], solves the problem of the minimal resources necessary for (block) coding with arbitrarily high fidelity.

By analogy with the classical measure of information, the bit, as a 2-state classical system, we use the term "qubit" [1] to refer to the quantum state storage capacity of a two dimensional Hilbert space. The polarization of a single photon, for example, has a storage capacity of one qubit. We also introduce the von Neumann entropy

$$S(\rho) = -\ trace\ \rho \log_2 \rho \tag{2}$$

We may now state the quantum noiseless coding theorem [1]:

For any quantum source with von Neumann entropy $S(\rho)$ and any given ϵ, δ (however small)

(a) If $S(\rho) + \delta$ qubits are available per signal then for each sufficiently large N there exists a coding scheme, having fidelity $F > 1 - \epsilon$, for signal strings of length N.

(b) If $S(\rho) - \delta$ qubits are available per signal then for any coding scheme, strings of length N will be decoded with fidelity $F < \epsilon$ for all sufficiently large N.

This theorem is clearly a quantum analogue of Shannon's noiseless coding theorem (in the form given in [4] §5.6 or [5] §3.2). Shannon's theorem may be viewed as a special case of the above, in which the signal states are all mutually orthogonal. The more general structure of the qubit compared to the bit provides no advantage in this orthogonal situation. As in Shannon's theorem, if δ is fixed, ϵ can be made arbitrarily small in (a) at the expense of coding longer blocks of signals.

Proof of the Theorem

Let A be a quantum source emitting states $|a_1\rangle, \ldots, |a_m\rangle$ in $\mathcal{H}_n$ with *a priori* probabilities $p_1, \ldots, p_m$. (The $|a_i\rangle$'s here will actually be K-blocks of signals when the following lemmas are used in the proof of the main theorem.) As before we set

$$\rho = \sum_{i=1}^{m} p_i |a_i\rangle\langle a_i|$$

Lemma 1. Consider any arbitrary association of signal states to density matrices:

$$|a_i\rangle\langle a_i| \longleftrightarrow W_i \qquad i = 1, \ldots, m \tag{3}$$

where each W_i is a density matrix supported on some fixed d-dimensional subspace D of $\mathcal{H}_n$. Let the sum of the d largest eigenvalues of ρ be η. Then the fidelity of the association (3) is no greater than η.

Proof. Write $\Pi_i = |a_i\rangle\langle a_i|$ and let Γ be the projection into D. Since W_1 is supported in D there is an orthonormal basis $|\phi_1\rangle, \ldots, |\phi_d\rangle$ of D consisting of eigenvectors of W_1. Thus we can write

$$\Gamma = \sum |\phi_i\rangle\langle\phi_i|$$

$$W_1 = \sum q_i\, |\phi_i\rangle\langle\phi_i| \quad where \quad 0 \le q_i \le 1$$

Then

$$\langle a_1|W_1|a_1\rangle = \sum q_i\, \langle a_1|\phi_i\rangle\langle\phi_i|a_1\rangle \le \sum \langle a_1|\phi_i\rangle\langle\phi_i|a_1\rangle = \ trace\, \Pi_1\Gamma$$

Similarly for each W_i. Hence the fidelity

$$F \le \sum p_i\, trace\, \Pi_i\Gamma = trace\, \rho\Gamma$$

Now introduce $|e_1\rangle, \ldots, |e_n\rangle$, an orthonormal basis of eigenvectors of ρ, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

$$trace\, \rho\Gamma = \sum \lambda_i \langle e_i|\Gamma|e_i\rangle$$

where $0 \le \langle e_i|\Gamma|e_i\rangle \le 1$ and $\sum\langle e_i|\Gamma|e_i\rangle = trace\, \Gamma = d$. Hence $trace\, \rho\Gamma$ is no larger than the sum of the d largest eigenvalues, proving the lemma.

This upper bound for $trace\, \rho\Gamma$ is attained if D is the subspace spanned by the d eigenvectors corresponding to the d largest eigenvalues.□

We are thinking here of (3) as arising from a coding-decoding prescription using d dimensions per signal. In general, not every association (3) will be physically realisable within the laws of quantum mechanics. Nevertheless the lemma applies to arbitrary associations.

The proof of lemma 1 suggests that, within the restriction of d dimensions, the highest fidelity is offered by D being the span of the d eigenvectors of ρ corresponding to the d largest eigenvalues. Let Λ denote this subspace. Intuitively Λ is the d dimensional subspace on which the ensemble of signals has the largest "weight".

We now introduce an explicit coding scheme based on $D = \Lambda$. Let $|0\rangle$ denote any fixed chosen state in Λ and let $\Lambda^\perp$ denote the orthogonal complement of Λ in $\mathcal{H}_n$. For each signal $|a_i\rangle$ we construct W_i as follows. Write $|a_i\rangle$ in terms of its components in Λ and $\Lambda^\perp$:

$$|a_i\rangle = \alpha_i|l_i\rangle + \beta_i|m_i\rangle \tag{4}$$

where $|l_i\rangle \in \Lambda$ and $|m_i\rangle \in \Lambda^\perp$ are normalised. Then W_i consists of a mixture of two states $|l_i\rangle$ and $|0\rangle$ with probabilities $|\alpha_i|^2$ and $|\beta_i|^2$ respectively i.e.

$$W_i = |\alpha_i|^2 |l_i\rangle\langle l_i| + |\beta_i|^2 |0\rangle\langle 0| \tag{5}$$

Thus W_i is obtained by measuring the observable $proj_\Lambda$ on $|a_i\rangle$ and if the result 0 is obtained (i.e. if $|a_i\rangle$ projects to $\Lambda^\perp$) then $|0\rangle \in \Lambda$ is substituted for the post-measurement state.

Lemma 2. Suppose that the sum of the d largest eigenvalues of ρ is greater than $1-\xi$. Then the association

$$|a_i\rangle \longleftrightarrow W_i$$

defined by (5) has fidelity $F > 1 - 2\xi$.

Proof. To compute the fidelity we use (5) to get

$$\langle a_i|W_i|a_i\rangle = |\alpha_i|^2 |\langle a_i|l_i\rangle|^2 + |\beta_i|^2 |\langle a_i|0\rangle|^2 \;\geq\; |\alpha_i|^2 |\langle a_i|l_i\rangle|^2 = |\alpha_i|^4 \;\geq\; 2|\alpha_i|^2 - 1$$

Hence

$$F = \sum p_i \langle a_i|W_i|a_i\rangle \geq 2\sum (p_i|\alpha_i|^2) - 1 = 2\, trace\, \rho\Lambda - 1 \geq 1 - 2\xi$$

where we have used $\sum p_i |\alpha_i|^2 = trace\, \rho\Lambda =$ sum of d largest eigenvalues$> 1-\xi$.□

The final ingredient for the proof of the main theorem is a result from classical information theory - the existence of "typical sequences"[3, 4, 5]. Let $\mathcal{P} = \{\lambda_1, \ldots, \lambda_m\}$ be a probability distribution. For any K let $\mathcal{P}_K$ denote the set of products of K-strings of the λ_i's

$$\mathcal{P}_K = \{\lambda_{i_1} \ldots \lambda_{i_K} : 1 \leq i_j \leq m\}$$

(Here strings which multiply up to the same value are kept distinct.) i.e. $\mathcal{P}_K$ is the probability distribution for a sequence of K independent trials of $\mathcal{P}$. Define the probability of any subset $T \subseteq \mathcal{P}_K$ to be the sum of the members of T. Let

$$H(\mathcal{P}) = -\sum \lambda_i \, \log_2 \lambda_i$$

be the Shannon entropy of $\mathcal{P}$. Then we can state:

Theorem of Typical Sequences. For any ϵ and δ (however small), and for all sufficiently large K

(a) The set $\mathcal{P}_K$ may be partitioned into a subset $\mathcal{L}$ of "likely" sequences containing at most $2^{K(H+\delta)}$ members, and its complement $\mathcal{L}^c$ of "unlikely" sequences. Furthermore $\mathcal{L}$ has probability greater than $1-\epsilon$.

(b) Any subset of $\mathcal{P}_K$ having less than $2^{K(H-\delta)}$ members, has probability less than ϵ.

Outline of Proof. (a) is a standard result (c.f. theorem 1.3.1 of [3], theorem 5.3.1 of [4] or theorem 3.1.2 of [5]). It is remarkable that $\mathcal{L}$ is generally an exponentally small part of $\mathcal{P}_K$ yet has arbitrarily large probability. For (b) we use the parameters $\epsilon/2$ and $\delta/2$ to obtain a set of "unlikely" sequences of total probability less than $\epsilon/2$. Also, for each sufficiently large K, each likely sequence has probability $\leq 2^{-K(H-\delta/2)}$ (this fact is part of the above quoted theorems in [3, 4, 5]). Now choose any $2^{K(H-\delta)}$ sequences. In general some are likely and some are unlikely. Hence their total probability p is smaller than $\epsilon/2$ (all unlikely sequences) plus the probability of $2^{K(H-\delta)}$ likely sequences:

$$p \leq \epsilon/2 + 2^{K(H-\delta)}2^{-K(H-\delta/2)}$$

which is $\leq \epsilon$ if K is large enough to ensure $2^{-K\delta/2} \leq \epsilon/2$. □

Proof of the quantum noiseless coding theorem. Consider a quantum source A emitting states $|a_1\rangle, \ldots, |a_m\rangle$ in $\mathcal{H}_n$ with *a priori* probabilities $p_1, \ldots, p_m$. Let the density matrix of A be ρ with eigenvalues $\lambda_1, \ldots, \lambda_n$ forming a probability distribution $\mathcal{P}$ as above. Note that the Shannon entropy $H(\mathcal{P})$ is just $S(\rho)$. Also the K-blocked version A_K of A has density matrix $\rho_K = \bigotimes^K \rho$ with eigenvalues given by the set $\mathcal{P}_K$.

Suppose that ϵ and δ are given. From the theorem of typical sequences part (a), it follows that for all sufficiently large K, there is a set of $2^{K(S+\delta)}$ eigenvalues of ρ_K having sum greater that $1-\epsilon/2$. Hence the sum of the $2^{K(S+\delta)}$ *largest* eigenvalues has sum greater than $1-\epsilon/2$. By lemma 2 there exists a coding scheme for A_K using $K(S+\delta)$ qubits per signal of A_K, i.e. $S+\delta$ qubits per signal of A, and having fidelity $F > 1-\epsilon$ for K-strings. This proves part (a) of the main theorem.

From the theorem of typical sequences part (b), it follows that for all sufficiently large K, *any* subset of $\mathcal{P}_K$ of size less than $2^{K(S-\delta)}$ has probability less than ϵ i.e. the sum of the largest $2^{K(S-\delta)}$ eigenvalues of ρ_K has sum less than ϵ. By lemma 1, every coding scheme for A_K using $K(S-\delta)$ qubits per signal of A_K, i.e. $S-\delta$ qubits per signal of A, will have fidelity less than ϵ for K-strings. This holds for all sufficiently large K and proves part (b) of the main theorem.□

An Example

We can illustrate some of these ideas by means of a simple example. Suppose a quantum source generates two non-orthogonal signals $|+\rangle$ and $|-\rangle$ with equal probability. With respect to a particular pair of orthogonal states $|a\rangle$ and $|b\rangle$ these signal states are

$$|\pm\rangle = \alpha\,|a\rangle \pm \beta\,|b\rangle$$

where $\alpha = \sqrt{0.9}$ and $\beta = \sqrt{0.1}$. The density matrix $\rho = \frac{1}{2}(|+\rangle\langle+| + |-\rangle\langle-|)$ is diagonal in the $|a\rangle$, $|b\rangle$ basis and has eigenvalues 0.9 and 0.1, yielding an entropy of 0.4690 qubits. We could code the output of this signal source perfectly using one qubit per signal, by mapping $|a\rangle \to |0\rangle$ and $|b\rangle \to |1\rangle$. The quantum noiseless coding theorem says that, in the long run, we should be able to represent the output of the signal source with arbitrarily high fidelity using less than one-half of a qubit per signal.

For this example, we will set ourselves the less difficult task of using two-thirds of a qubit per signal—that is, using an average of two qubits for every three signals generated by the source. One rather trivial way to achieve this, at some sacrifice of fidelity, would be to break up the stream of independent signals into blocks of three, then code only the first two signals of each block into individual qubits as described above. The receiver, by prearrangement, generates the state $|\,a\rangle$ for the omitted signals. Thus, in each 3-block of signals, two signals are transmitted perfectly, and the third signal would pass a fidelity test measurement with probability $\alpha^2 = 0.9$. The overall fidelity of a given 3-block is thus 0.9.

This fidelity can be improved considerably by a block coding scheme more closely resembling the one we have described. We note that $|\,a\rangle$ has a much greater "weight" in the signal ensemble than $|\,b\rangle$. Once again we break the signal stream up into 3-blocks. For each 3-block, let Λ be the subspace spanned by vectors in the $|\,a\rangle$, $|\,b\rangle$ product basis which include a majority of $|\,a\rangle$'s:

$$\Lambda = \text{Span}\,\{|\,a\rangle\,|\,a\rangle\,|\,a\rangle\,,|\,a\rangle\,|\,a\rangle\,|\,b\rangle\,,|\,a\rangle\,|\,b\rangle\,|\,a\rangle\,,|\,b\rangle\,|\,a\rangle\,|\,a\rangle\}\,.$$

Our coding scheme is as follows: We first make a measurement of the 3-block to see if the state lies in the (four-dimensional) subspace Λ. If so, we can encode the state unitarily into the state of a pair of qubits via

$$\begin{array}{lcl} |\,a\rangle\,|\,a\rangle\,|\,a\rangle & \to & |\,0\rangle\,|\,0\rangle \\ |\,a\rangle\,|\,a\rangle\,|\,b\rangle & \to & |\,0\rangle\,|\,1\rangle \\ |\,a\rangle\,|\,b\rangle\,|\,a\rangle & \to & |\,1\rangle\,|\,0\rangle \\ |\,b\rangle\,|\,a\rangle\,|\,a\rangle & \to & |\,1\rangle\,|\,1\rangle\,. \end{array}$$

If the state of the 3-block is not found to be in Λ, we put the qubits into the state $|\,0\rangle\,|\,0\rangle$. The decoding is accomplished by inverting the unitary mapping, yielding a decoded signal in Λ.

In this coding scheme, none of the actual signal states of the 3-block are coded perfectly, since none of them lie in Λ. Nevertheless, the subspace Λ contains most of the "weight" of the signal ensemble. Let Γ be the projection onto Λ and ρ_3 is the density matrix of the 3-block. Then $\mathit{trace}\,\rho_3\Gamma = (0.9)^3 + 3(0.9)^2(0.1) = 0.972$. By lemma 2, the fidelity of the coding scheme is at least 0.944.

This is already an improvement over the simple "two out of three" coding scheme, but in fact the actual fidelity is even higher. We have coded those signals not found to be in Λ into the state $|\,0\rangle\,|\,0\rangle$, which is then decoded as $|\,a\rangle\,|\,a\rangle\,|\,a\rangle$. This state is itself not very far from the initial state. That is, even the "failures" of the coding scheme have a substantial probability of passing a fidelity test measurement. The fidelity may be calculated to be $F = 0.9924$. We can therefore encode a 3-block of signals from our quantum source into two qubits with a fidelity above 99%.

Acknowledgments

We would like to thank C. H. Bennett and W. K. Wootters for several helpful discussions. We are also very grateful to the Rank Prize Funds for their sponsorship of the 1993 Mini-Symposium on Quantum Communication and Cryptography, Broadway, England which provided the opportunity for this collaboration.

References

[1] B. Schumacher, "On Quantum Coding", *Phys. Rev.* A, (to appear 1993)

[2] A. Peres, *Phys. Lett. A* **128**, 19, (1988)

[3] R. B. Ash,"Information Theory",(Dover Publications 1990).

[4] D. Welsh, "Codes and Cryptography",(Clarendon Press Oxford, 1989).

[5] T. M. Cover, J. A. Thomas, "Information Theory", (John Wiley and sons Inc. 1991) Chapter 3.