Correspondence

The Capacity of the Quantum Channel with General Signal States

A. S. Holevo, *Member, IEEE*

***Abstract*—It is shown that the capacity of a classical-quantum channel with arbitrary (possibly mixed) states equals the maximum of the entropy bound with respect to all *a priori* distributions. This completes the recent result of Hausladen, Jozsa, Schumacher, Westmoreland, and Wootters, who proved the equality for the pure state channel.**

***Index Terms*— Average error probability, entropy bound, quantum coding theorem, typical subspace.**

I. Introduction

The fundamental limitations on the quality and rate of information transmission due to the quantum-mechanical nature of the information carrier are the subject of quantum communication theory (see [4], [10], and [15] for survey and further references). Initiated in the 1960's and developed in the 1970's, it is now getting new impetus from such fields as quantum optics, precision experiments, quantum cryptography and computing. Not only have ideas and methods of quantum communication theory been successfully applied in these fields, but also new light is shed on some old issues of quantum communication. One of such issues—the coding theorem for quantum communication channels—was substantially clarified in recent publications [9], [17] on noiseless quantum channels. The present correspondence treats the case of noisy quantum channel.

A theoretical framework for quantum communication channels is developed in [11]–[14], and more recently [2], [18]. For convenience, we start by repeating some definitions and results from [14]. Let $\mathcal{H}$ be a d-dimensional Hilbert space describing quantum-mechanically the physical carrier of the information. A simple quantum communication channel (*classical-quantum channel* in the terminology of [13]) consists of the input alphabet $A = \{1, \cdots, a\}$ and a mapping $i \to S_i$ from the input alphabet to the set of quantum states in $\mathcal{H}$. A quantum state is a density operator (d.o.), i.e., a positive operator S in $\mathcal{H}$ with unit trace, $\mathrm{Tr}\, S = 1$. Sending a letter i results in producing the quantum state S_i of the information carrier.

Like in the classical case, the input is described by an *a priori probability distribution* $\pi = \{\pi_i\}$ on A. At the receiving end of the channel a quantum measurement in the sense of [15] is performed. Mathematically it is described by a *resolution of identity* in $\mathcal{H}$, that is, by a family $X = \{X_j\}$ of positive operators in $\mathcal{H}$ satisfying $\sum_j X_j = I$, where I is the unit operator in $\mathcal{H}$. The index j runs through some finite output alphabet, which is not fixed here. The conditional probability of the output j, given the input was i, equals $P(j|i) = \mathrm{Tr}\, S_i X_j$. The Shannon information is given by the classical formula

$$I_1(\pi, X) = \sum_j \sum_i \pi_i P(j|i) \log \left(\frac{P(j|i)}{\sum_k \pi_k P(j|k)} \right) \tag{1}$$

(in what follows we use the binary logarithms).

In the same way we can consider the product channel in the tensor product Hilbert space $\mathcal{H}^{\otimes n} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}$ with the input alphabet A^n consisting of words $u = (i_1, \cdots, i_n)$ of length n, with the d.o.

$$S_u = S_{i_1} \otimes \cdots \otimes S_{i_n} \tag{2}$$

corresponding to the word u. If π is a probability distribution on A^n and X is a resolution of identity in $\mathcal{H}^{\otimes n}$, we define the information quantity $I_n(\pi, X)$ by a formula similar to (1). Defining

$$C_n = \sup_{\pi, X} I_n(\pi, X)$$

we have the property of superadditivity $C_n + C_m \le C_{n+m}$, hence the following limit exists:

$$C = \lim_{n \to \infty} C_n / n \tag{3}$$

and is called the *capacity* of the initial channel [14]. This definition is justified by the fact easily deduced from the classical Shannon coding theorem, that C is the least upper bound of rate (bits per symbol) of information which can be transmitted with asymptotically vanishing error. More precisely, we call by *code of size* N a sequence $(u_1, X_1), \cdots, (u_N, X_N)$, where u_k are words of length n, and $\{X_k\}$ is a family of positive operators in $\mathcal{H}^{\otimes n}$, satisfying

$$\sum_{j=1}^{N} X_j \le I.$$

Defining

$$X_0 = I - \sum_{j=1}^{N} X_j$$

we have a resolution of identity in $\mathcal{H}^{\otimes n}$. An output $k (1 \le k \le N)$ means the decision that the word u_k was transmitted, and the output 0 is interpreted as evasion of any decision. The average error probability for such a code is

$$\mathsf{P}_{\mathrm{er}} = \frac{1}{N} \sum_{k=1}^{N} [1 - \mathrm{Tr}\, S_{u_k} X_k].$$

Let us denote $p(n, N)$ the minimum of this error probability with respect to all codes of the size N with words of length n. Then

$$p(n, 2^{n(C-\delta)}) \to 0 \text{ and } p(n, 2^{n(C+\delta)}) \not\to 0 \tag{4}$$

where $\delta > 0$, if $n \to \infty$. The same holds for the minimum of the maximal (with respect to k) error probability, which does not presume any *a priori* probabilities for the words (see [6] and [14]).

Manuscript received December 20, 1996; revised May 30, 1997. This work is based on Report quant-ph/9611023, Nov. 14, 1996, Los Alamos Electronic Archive for Quantum Physics (http://xxx.lanl.gov/archive/quant-ph). It was supported by Tamagawa University, JSPS and RFBR under Grant 96-01-01709. The material in this correspondence was presented in part at the Information Physics Workshop, University of New Mexico, Albuquerque, April 30, 1997.

The author is with the Steklov Mathematical Institute, 117966 Moscow, Russia.

Publisher Item Identifier S 0018-9448(98)00070-4.

II. The Entropy Bound

The main result of [14] was a demonstration of the possibility that the inequality $C > C_1$ implies strict superadditivity of the sequence C_n. This is in sharp contrast with the situation for the corresponding classical memoryless channel, for which $C_n = nC_1$ and hence $C = C_1$, and is just another manifestation of the quantum nonseparability. This fact is in a sense dual to the existence of EPR (Einstein–Podolsky–Rosen) correlations [3]: the latter are due to entangled states and hold for disentangled measurements while the superadditivity is due to entangled measurements and holds for disentangled states (see also [19]). The inequality $C \neq C_1$ raised the problem of the actual value of the capacity C.

In what follows we take $d \leq \infty$. Denoting by $H(S) = -\operatorname{Tr} S \log S$ the von Neumann entropy of a d.o. S, we *assume that* $H(S_i) < \infty$. If $\pi = \{\pi_i\}$ is an *a priori* distribution on A, we denote

$$\overline{S} = \sum_{i \in A} \pi_i S_i \qquad \overline{H}(S_{(\cdot)}) = \sum_{i \in A} \pi_i H(S_i)$$

(we use the symbol $(\cdot)$ to indicate dependence on the variable, with respect to which the averaging is performed), and

$$\Delta H(\pi) = H(\overline{S}) - \overline{H}(S_{(\cdot)}).$$

According to Lemma 1 proved in the Appendix, the quantity $\Delta H(\pi)$ is well-defined and is continuous in π.

The entropy bound [12] says that

$$\sup_X I_1(\pi, X) \leq \Delta H(\pi)$$

where the equality is achieved if and only if all operators $\pi_i S_i$ commute (see [22] for a detailed discussion of history and some generalizations of the entropy bound, including the case $d = \infty$). Applied to $I_n(\pi, X)$ and combined with the additivity and continuity properties of $\Delta H(\pi)$ proved in the Appendix this inequality implies

$$C \leq \max_\pi \Delta H(\pi).$$

In [14] a conjecture was made that in fact this might be an equality. Recently, Hausladen, Jozsa, Schumacher, Westmoreland, and Wootters [9] proved this in the case of pure states S_i. The problem for the case of general (possibly mixed) states was left open in [9] and is the subject of our present work. The main result is the estimate for the error probability implying converse inequality $C \geq \max_\pi \Delta H(\pi)$. Thus we have the following quantum coding theorem, confirming the physical wisdom according to which the entropy bound was used to evaluate the quantum capacity [8].

Theorem: The capacity of the quantum communication channel with arbitrary signal states S_i, having finite entropy $H(S_i)$, is given by

$$C = \max_\pi \left[H\left(\sum_{i \in A} \pi_i S_i \right) - \sum_{i \in A} \pi_i H(S_i) \right]. \tag{5}$$

The key points of the proof are the estimate for the error probability, which is substantially more complicated than the estimate for pure states given already in [14] and a similar estimate from [9], and the idea of projection onto the typical subspace due to [9] and [17], modified here for the case of mixed states.

Note Added in Proof: A similar result in finite-dimensional case has been obtained in the recent paper by B. Schumacher and M. D. Westmoreland [20]. Recently, the theorem of the present paper was generalized in several directions. In [5], exponential upper bounds for the error probability in pure-state quantum channel were given, providing lower bounds for the quantum reliability function, and implying, in particular, an alternative proof of the coding theorem, not using projection onto the typical subspace. In [16], the coding theorem was extended to quantum channels with constrained inputs and infinite (continuous) alphabets, and applied to calculate the capacity of quantum Gaussian channel. In [7], the channels with the property $C = C_1$ are characterized.

III. Typical Subspaces of Density Operators

We denote $D = \{1, \cdots, d\}$ if $d < \infty$ and $D = \{1, \cdots\}$ if $d = \infty$. Let

$$\overline{S} = \sum_{j \in D} \lambda_j |e_j\rangle\langle e_j|$$

be the spectral decomposition of the d.o. $\overline{S}$, where λ_j are the eigenvalues, $|e_j\rangle$ are the eigenvectors of $\overline{S}$ (we are using the Dirac "braket" notations, see. e.g., [10], [15]). Then the spectral decomposition of $\overline{S}^{\otimes n} = \overline{S} \otimes \cdots \otimes \overline{S}$ is

$$\overline{S}^{\otimes n} = \sum_{J \in D^n} \lambda_J |e_J\rangle\langle e_J|$$

where $J = (j_1, \cdots, j_n)$, $\lambda_J = \lambda_{j_1} \cdot \cdots \cdot \lambda_{j_n}$, $|e_J\rangle = |e_{j_1}\rangle \otimes \cdots \otimes |e_{j_n}\rangle$. Following [9] we introduce the spectral projector onto the *typical subspace* of the d.o. $\overline{S}^{\otimes n}$ as

$$P = \sum_{J \in B} |e_J\rangle\langle e_J| \tag{6}$$

where

$$B = \{J \colon 2^{-n[H(\overline{S})+\delta]} < \lambda_J < 2^{-n[H(\overline{S})-\delta]}\} \subset D^n.$$

A sequence $J \in B$ is "typical" for a probability distribution on D^n given by eigenvalues λ_J of the d.o. $\overline{S}^{\otimes n}$ in the sense of classical information theory (see, e.g., [6]). It follows that for fixed small positive ϵ, δ and all $n \geq n_1(\pi, \epsilon, \delta)$

$$\operatorname{Tr} \overline{S}^{\otimes n}(I - P) \leq \epsilon. \tag{7}$$

Indeed, $\operatorname{Tr} \overline{S}^{\otimes n} P$ is equal to the probability

$$\begin{aligned} \mathsf{P}\{J \in B\} &= \mathsf{P}\{n[H(\overline{S}) - \delta] < -\log \lambda_J < n[H(\overline{S}) + \delta]\} \\ &= \mathsf{P}\{|n^{-1} \sum_{l=1}^{n} \log \lambda_{j_l} + H(\overline{S})| < \delta\} \end{aligned}$$

which tends to 1 as $n \to \infty$, according to the Law of Large Numbers, since $H(\overline{S}) = -\mathsf{M} \log \lambda_{(\cdot)}$ (we denote by M the expectation of a random variable throughout the correspondence).

The next step is a development of this idea necessary to prove the Theorem for mixed states. Let

$$S_i = \sum_{j \in D} \lambda_j^i |e_j^i\rangle\langle e_j^i|$$

be the spectral decomposition of the d.o. S_i. Let $u = (i_1, \cdots, i_n)$ be a word of the input alphabet and $S_u = S_{i_1} \otimes \cdots \otimes S_{i_n}$ be the corresponding d.o. Its spectral decomposition is

$$S_u = \sum_{J \in D^n} \lambda_J^u |e_J^u\rangle\langle e_J^u|$$

where

$$\lambda_J^u = \lambda_{j_1}^{i_1} \cdot \cdots \cdot \lambda_{j_n}^{i_n}, \quad |e_J^u\rangle = |e_{j_1}^{i_1}\rangle \otimes \cdots \otimes |e_{j_n}^{i_n}\rangle.$$

We introduce the spectral projector onto the typical subspace of S_u as

$$P_u = \sum_{J \in B_u} |e_J^u\rangle\langle e_J^u| \tag{8}$$

where

$$B_u = \{J \colon 2^{-n[\overline{H}(S_{(\cdot)})+\delta]} < \lambda_J^u < 2^{-n[\overline{H}(S_{(\cdot)})-\delta]}\}.$$

Let on the set of all words A^n the following probability distribution be defined:

$$\mathsf{P}\{u = (i_1, \cdots, i_n)\} = \pi_{i_1} \cdot \cdots \cdot \pi_{i_n}. \tag{9}$$

Then for fixed small positive ϵ, δ and all $n \geq n_2(\pi, \epsilon, \delta)$

$$\mathsf{M}\,\mathrm{Tr}\, S_u(I - P_u) \leq \epsilon. \tag{10}$$

Indeed, consider the sequence of independent trials with the outcomes i_l, j_l; $l = 1, \cdots, n$, where the probability of the outcome (i, j) in each trial is equal to $\pi_i \lambda_j^i$. Then

$$\begin{aligned}\mathsf{M}\,\mathrm{Tr}\, S_u P_u &= \mathsf{P}\{J \in B_u\} \\ &= \mathsf{P}\{n[\overline{H}(S_{(\cdot)}) - \delta] < -\log\, \lambda_J^u < n[\overline{H}(S_{(\cdot)}) + \delta]\} \\ &= \mathsf{P}\left\{\left| n^{-1} \sum_{l=1}^{n} \log\, \lambda_{j_l}^{i_l} + \overline{H}(S_{(\cdot)}) \right| < \delta \right\}\end{aligned}$$

which tends to 1 as $n \to \infty$, according to the Law of Large Numbers, since $\overline{H}(S_{(\cdot)}) = -\mathsf{M} \log\, \lambda_{(\cdot)}^{(\cdot)}$. In what follows we put

$$n(\pi, \epsilon, \delta) = \max\{n_1(\pi, \epsilon, \delta), n_2(\pi, \epsilon, \delta)\}.$$

IV. The Choice of the Suboptimal Decision Rule

In quantum detection theory the optimal decision rule minimizing the average (or more general Bayes) error probability not always can be found explicitly; however, just as in the classical coding theorem, it is not necessary to use a precisely optimal decision rule, and it is possible to design a decision rule asymptotically optimal in the limit of "almost orthogonal" states and hence sufficient for the proof of the coding theorem.

Let $u_1, \cdots, u_N$ be a sequence of words. To simplify notations we denote the words by their numbers $1, \cdots, N$. Put

$$X_u = \left(\sum_{u'=1}^{N} P P_{u'} P\right)^{-(1/2)} P P_u P \left(\sum_{u'=1}^{N} P P_{u'} P\right)^{-(1/2)} \tag{11}$$

where $X^{-(1/2)}$ denotes the generalized inverse of the operator $X^{1/2}$, i.e., operator equal 0 on the null subspace $\mathcal{N}$ of X and $(X^{1/2})^{-1}$ on $\mathcal{N}^{\perp}$. Then

$$\sum_{u=1}^{N} X_u \leq I.$$

The idea behind this choice is a noncommutative analog of the idea of using "jointly typical" sequences in the proof of the classical Shannon coding theorem [6]: projection P_u selects the typical subspace for the signal state S_u and P does the same for the mixture of the signal states. The normalizing factors on both sides of (11) are the source of major analytical difficulties in evaluating the average error probability.

Put $|\hat{e}_J^u\rangle = P|e_J^u\rangle$ where P is defined by (6), then

$$\begin{aligned}X_u = &\left(\sum_{u'=1}^{N} \sum_{J \in B_{u'}} |\hat{e}_J^{u'}\rangle\langle\hat{e}_J^{u'}|\right)^{-(1/2)} \sum_{J \in B_u} |\hat{e}_J^u\rangle\langle\hat{e}_J^u| \\ &\cdot \left(\sum_{u'=1}^{N} \sum_{J \in B_{u'}} |\hat{e}_J^{u'}\rangle\langle\hat{e}_J^{u'}|\right)^{-(1/2)}.\end{aligned}$$

By denoting

$$\alpha_{(u,\,J),\,(u',\,J')} = \langle\hat{e}_J^u| \left(\sum_{v=1}^{N} \sum_{K \in B_v} |\hat{e}_K^v\rangle\langle\hat{e}_K^v|\right)^{-(1/2)} \hat{e}_{J'}^{u'}\rangle$$

and taking into account that $X_u = PX_uP$, the average error probability corresponding to the choice (11) can be written as

$$\mathsf{P}_{\mathrm{er}} = \frac{1}{N} \sum_{u=1}^{N} \left[1 - \sum_{J \in D^n} \sum_{J' \in B_u} \lambda_J^u |\alpha_{(u,\,J),\,(u,\,J')}|^2\right]. \tag{12}$$

V. The Estimate for the Error Probability

Taking into account that $\sum_{J \in D^n} \lambda_J^u = 1$ and omitting some nonpositive terms, we see that

$$\mathsf{P}_{\mathrm{er}} \leq \frac{1}{N} \sum_{u=1}^{N} \left[\sum_{J \in B_u} \lambda_J^u (1 - \alpha_{(u,\,J),\,(u,\,J)}^2) + \sum_{J \notin B_u} \lambda_J^u\right]. \tag{13}$$

Let us denote

$$\gamma_{(u,\,J),\,(u',\,J')} = \langle\hat{e}_J^u|\hat{e}_{J'}^{u'}\rangle = \langle e_J^u|P e_{J'}^{u'}\rangle \tag{14}$$

and introduce the Gram matrix

$$\Gamma = [\gamma_{(u,\,J),\,(u',\,J')}]$$

where $J \in B_u$, $J' \in B_{u'}$ and $u, u' = 1, \cdots, N$. Then

$$\Gamma^{1/2} = [\alpha_{(u,\,J),\,(u',\,J')}].$$

In particular,

$$\alpha_{(u,\,J),\,(u,\,J)}^2 \leq \gamma_{(u,\,J),\,(u\,J)} \leq 1.$$

Then from (13)

$$\mathsf{P}_{\mathrm{er}} \leq \frac{1}{N} \sum_{u=1}^{N} \left[2 \sum_{J \in B_u} \lambda_J^u (1 - \alpha_{(u,\,J),\,(u,\,J)}) + \sum_{J \notin B_u} \lambda_J^u\right]. \tag{15}$$

By introducing the diagonal matrix $\Lambda = \mathrm{diag}[\lambda_J^u]$ and denoting by E the unit matrix and the trace of matrices by Sp as distinct from the trace of operators in Hilbert space, we have

$$\begin{aligned}&2 \sum_{u=1}^{N} \sum_{J \in B_u} \lambda_J^u (1 - \alpha_{(u,\,J),\,(u,\,J)}) \\ &\quad = 2\,\mathrm{Sp}\,\Lambda(E - \Gamma^{1/2}) \\ &\quad = \mathrm{Sp}\,\Lambda(E - \Gamma^{1/2})^2 + \mathrm{Sp}\,\Lambda(E - \Gamma) \\ &\quad \leq \mathrm{Sp}\,\Lambda(E - \Gamma)^2 + \mathrm{Sp}\,\Lambda(E - \Gamma)\end{aligned} \tag{16}$$

since $(E - \Gamma^{1/2})^2 = (E - \Gamma)^2(E + \Gamma^{1/2})^{-2} \leq (E - \Gamma)^2$ [14]. Calculating the traces, we obtain the right-hand side of (16) as

$$\begin{aligned}\sum_{u=1}^{N} \sum_{J \in B_u} \lambda_J^u \Bigg[&2 - 3\gamma_{(u,\,J),\,(u,\,J)} + \gamma_{(u,\,J),\,(u,\,J)}^2 \\ &+ \sum_{J':\, J' \neq J} |\gamma_{(u,\,J),\,(u,\,J')}|^2 \\ &+ \sum_{u':\, u' \neq u} \sum_{J' \in B_{u'}} |\gamma_{(u,\,J),\,(u',\,J')}|^2\Bigg].\end{aligned}$$

This quantity will not decrease if the range of J is enlarged to the full range D^n and if $2 - 3\gamma_{(u,\,J),\,(u,\,J)} + \gamma_{(u,\,J),\,(u,\,J)}^2$ is replaced with $2 - 2\gamma_{(u,\,J),\,(u,\,J)}$. Then we obtain

$$\begin{aligned}\mathsf{P}_{er} \leq \frac{1}{N} \sum_{u=1}^{N} \Bigg\{\sum_{J \in D^n} \lambda_J^u \Bigg[&2 - 2\gamma_{(u,\,J),\,(u,\,J)} \\ &+ \sum_{J':\, J' \neq J} |\gamma_{(u,\,J),\,(u,\,J')}|^2 \\ &+ \sum_{u':\, u' \neq u} \sum_{J' \in B_{u'}} |\gamma_{(u,\,J),\,(u',\,J')}|^2\Bigg] + \sum_{J \notin B_u} \lambda_J^u\Bigg\}.\end{aligned}$$

Taking into account the definition (14) of $\gamma_{(u,\,J),\,(u',\,J')}$ and the fact that $\langle e_J^u | e_{J'}^u \rangle = 0$ for $J \neq J'$, we can write the last inequality as

$$\mathsf{P}_{\rm er} \leq \frac{1}{N} \sum_{u=1}^{N} \Bigg\{ 2 \operatorname{Tr} S_u (I-P) + \operatorname{Tr} S_u (I-P) P_u (I-P) + \sum_{u' :\, u' \neq u} \operatorname{Tr} P S_u P P_{u'} + \operatorname{Tr} S_u (I - P_u) \Bigg\}.$$

The second term is less or equal than $\operatorname{Tr} S_u (I-P)$. Thus finally

$$\mathsf{P}_{\rm er} \leq \frac{1}{N} \sum_{u=1}^{N} \Bigg\{ 3 \operatorname{Tr} S_u (I-P) + \sum_{u' :\, u' \neq u} \operatorname{Tr} P S_u P P_{u'} + \operatorname{Tr} S_u (I - P_u) \Bigg\}. \tag{17}$$

VI. Random Coding

Just as in the classical Shannon random coding scheme, assume that the words $u_1, \cdots, u_N$ are chosen at random, independently, and with the probability distribution (9) for each word. Then $\mathsf{M} S_u = \overline{S}^{\otimes n}$ [9] and from (17), by independence of S_u, $P_{u'}$

$$\mathsf{M} \mathsf{P}_{\rm er} \leq 3 \operatorname{Tr} \overline{S}^{\otimes n} (I-P) + (N-1) \operatorname{Tr} P \overline{S}^{\otimes n} P \mathsf{M} P_{u'} + \mathsf{M} \operatorname{Tr} S_u (I - P_u).$$

By the inequalities (7), (10), and by the properties of trace

$$\mathsf{M} \mathsf{P}_{\rm er} \leq 4\epsilon + (N-1) \| P \overline{S}^{\otimes n} P \| \operatorname{Tr} \mathsf{M} P_{u'}$$

for $n \geq n(\pi,\, \epsilon,\, \delta)$. By the definition of P

$$\| P \overline{S}^{\otimes n} P \| \leq 2^{-n[H(\overline{S}) - \delta]}$$

and by the definition of P_u

$$\operatorname{Tr} \mathsf{M} P_{u'} = \mathsf{M} \operatorname{Tr} P_{u'} \leq \mathsf{M} \operatorname{Tr} S_{u'} \cdot 2^{n[\overline{H}(S_{(\cdot)}) + \delta]} = 2^{n[\overline{H}(S_{(\cdot)}) + \delta]}.$$

Thus

$$\mathsf{M} \mathsf{P}_{\rm er} \leq 4\epsilon + (N-1) 2^{-n[H(\overline{S}) - \overline{H}(S_{(\cdot)}) - 2\delta]}. \tag{18}$$

Let us choose the distribution $\pi = \pi^0$ maximizing the entropy bound $\Delta H(\pi)$. Then (18) implies

$$p(n,\, N) \leq 4\epsilon + (N-1) 2^{-n[\Delta H(\pi^0) - 2\delta]} \tag{19}$$

for $n \geq n(\pi^0,\, \epsilon,\, \delta)$. Thus $p(n,\, 2^{n[\Delta H(\pi^0) - 3\delta]}) \to 0$ as $n \to \infty$, whence $\Delta H(\pi^0) - 3\delta \leq C$ by (4) for arbitrary δ, and (5) follows.

Appendix

Lemma 1: Let $H(S_i) < \infty$, then $\Delta H(\pi)$ is well defined for all π and is a continuous function of π.

Proof: By an inequality of Lanford and Robinson (see [21, eq. (2.3)])

$$H\left(\sum_i \pi_i S_i \right) \leq \sum_i \pi_i H(S_i) - \sum_i \pi_i \log \pi_i$$

hence $\Delta H(\pi)$ is well defined. It follows also that

$$H\left(\sum_i S_i \right) = -\operatorname{Tr} \sum_i S_i \log \sum_i S_i = a \left[H\left(a^{-1} \sum_i S_i \right) - \log a \right] < \infty$$

while

$$\sum_i \pi_i S_i \leq \sum_i S_i$$

for all π. Since $\sum_i \pi_i S_i$ is weakly continuous in π, by the dominated convergence theorem of Simon (see [21, Sec. D]), $H(\sum_i \pi_i S_i)$ and hence $\Delta H(\pi)$ is continuous in π.

Let A_k, $k = 1, 2$, be finite alphabets and let $\{S_i^k,\, i \in A^k\}$ be families of d.o. in Hilbert spaces $\mathcal{H}_k$. Let $\{\pi_{ij}\}$ be a probability distribution on $A^1 \times A^2$ and denote

$$\Delta H(\{\pi_{ij}\}) = H\left(\sum_{ij} \pi_{ij} S_i^1 \otimes S_j^2 \right) - \sum_{ij} \pi_{ij} H(S_i^1 \otimes S_j^2).$$

Lemma 2: (Additivity of the entropy bound)

$$\max_{\pi_{ij}} \Delta H(\{\pi_{ij}\}) = \max_{\pi_i^1} \Delta H(\{\pi_i^1\}) + \max_{\pi_i^2} \Delta H(\{\pi_i^2\}).$$

Proof: By subadditivity of entropy (see [1] and [21, Sec. F])

$$H(S) \leq H(\operatorname{Tr}_2 S) + H(\operatorname{Tr}_1 S)$$

where S is a d.o. in $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $\operatorname{Tr}_k S$, $k = 1, 2$, is partial trace with respect to $\mathcal{H}_k$, we have

$$H\left(\sum_{ij} \pi_{ij} S_i^1 \otimes S_j^2 \right) \leq H\left(\sum_i \pi_i^1 S_i^1 \right) + H\left(\sum_j \pi_j^2 S_j^2 \right)$$

where $\{\pi_i^1\}$, $\{\pi_i^2\}$ are the marginal distributions of $\{\pi_{ij}\}$. It follows that

$$\max_{\pi_{ij}} \Delta H(\{\pi_{ij}\}) \leq \max_{\pi_i^1} \Delta H(\{\pi_i^1\}) + \max_{\pi_i^2} \Delta H(\{\pi_i^2\}).$$

The converse inequality follows by restricting to $\pi_{ij} = \pi_i^1 \times \pi_i^2$ and using the additivity of quantum entropy for product states.

Acknowledgment

The work was stimulated by discussions with Prof. R. Jozsa and Prof. A. Yu. Kitaev during the 3rd Conference on Quantum Communication and Measurement in Hakone, Japan, September 1996, where the result of [9] was reported. The author is grateful to Prof. O. Hirota, Dr. M. Osaki, and Dr. M. Sasaki (Tamagawa University) for their hospitality and stimulating discussion.

References

[1] H. Araki and E. H. Lieb, "Entropy inequalities," *Commun. Math. Phys.*, vol. 18, no. 2, pp. 160–170, 1970.
[2] M. Ban, M. Osaki, and O. Hirota, "Upper bound of the accessible information and lower bound of the Bayes cost in quantum signal detection processes," *Phys. Rev. A*, vol. 54, no. 4, pp. 2718–2727, 1996.
[3] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics.* New York: Cambridge Univ. Press, 1987.
[4] C. Benjaballah, *Introduction to Photon Communication* (Lecture Notes in Physics, vol. 29). Heidelberg, Germany: Springer-Verlag, 1995.
[5] M. V. Burnashev and A. S. Holevo, "On reliability function of quantum communication channel," LANL Rep, quant-ph/9703013, Mar. 10, 1997.
[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.
[7] A. Fujivara and H. Nagaoka, "Operational capacity and semi-classicality of quantum channel," *IEEE Trans. Inform. Theory*, submitted.
[8] J. P. Gordon, "Quantum electronics and coherent light," in *Proc. Int. School Phys. "Enrico Fermi"*, Course XXXI, P. A. Miles, Ed. New York: Academic, 1964, pp. 156–181.
[9] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. Wootters, "Classical information capacity of a quantum channel," *Phys. Rev. A*, vol. 54, no. 3, pp. 1869–1876, Sept. 1996.
[10] C. W. Helstrom, *Quantum Detection and Estimation Theory.* New York: Academic, 1976.

[11] A. S. Holevo, "Towards the mathematical theory of quantum communication channels," *Probl. Pered. Inform.*, vol. 8, no. 1, pp. 63–71, 1972. (English translation: *Probl. Inform. Transm.*, vol. 8, no. 1, pp. 47–56).
[12] ____, "Some estimates of the information transmitted by quantum communication channel," *Probl. Pered. Inform.*, vol. 9, no. 3, pp. 3–11, 1973. (English translation: *Probl. Inform. Transm.*, vol. 9, no. 3, pp. 177–183).
[13] ____, "Problems in the mathematical theory of quantum communication channels," *Rep. Math. Phys.*, vol. 12, no. 2, pp. 273–278, 1977.
[14] ____, "On the capacity of quantum communication channel," *Probl. Pered. Inform.*, vol. 15, no. 4, pp. 3–11, 1979. (English translation: *Probl. Inform. Transm.*, vol. 15, no. 4, pp. 247–253).
[15] ____, *Probabilistic and Statistical Aspects of Quantum Theory*. Amsterdam, The Netherlands: North Holland, 1982.
[16] ____, "On quantum communication channels with constrained inputs," LANL Rep. quant-ph/9705054, May 30, 1997.
[17] R. Jozsa and B. Schumacher, "A new proof of the quantum noiseless coding theorem," *J. Mod. Opt.*, vol. 41, no. 12, pp. 2343–2349, 1994.
[18] M. Ohya, D. Petz, and N. Watanabe, "On capacity of quantum channels," *IEEE Trans. Inform. Theory*. to be published.
[19] A. Peres and W. K. Wootters, "Optimal detection of quantum information," *Phys. Rev. Lett.*, vol. 66, no. 9, pp. 1119–1122, 1991.
[20] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channel," *Phys. Rev. A*, 1997, to be published.
[21] A. Wehrl, "General properties of entropy," *Rev. Mod. Phys.*, vol. 50, no. 2, pp. 221–260, 1978.
[22] H. P. Yuen and M. Ozawa, "Ultimate information carrying limit of quantum systems," *Phys. Rev. Lett.*, vol. 70, no. 4, pp. 363–366, Jan. 1993.

Lattice Codes Can Achieve Capacity on the AWGN Channel

Rüdiger Urbanke and Bixio Rimoldi, *Member, IEEE*

***Abstract*— It is shown that lattice codes can achieve capacity on the additive white Gaussian noise channel. More precisely, for any rate R less than capacity and $\epsilon > 0$, there exists a lattice code with rate no less than R and average error probability upper-bounded by ϵ. These lattice codes include *all* points of the (translated) lattice within the spherical bounding region (not just the ones inside a thin spherical shell).**

***Index Terms*— AWGN channel, Blichfeldt's principle, lattice codes, Minkowski–Hlawka theorem.**

I. Introduction

Consider the additive white Gaussian noise (AWGN) channel with peak signal-power constraint S, i.e., each codeword x of an n-dimensional code for this channel must satisfy

$$||x||^2 \leq nS$$

where $||\cdot||$ is the Euclidean norm. It is well known [1] that the capacity of this channel is

$$C = \frac{1}{2}\log_2\left(1 + \frac{S}{N}\right) \text{ bits/channel use}$$

where N is the variance of the independent and identically distributed (i.i.d.) Gaussian noise. The proof in [1] is based on a *random coding* argument and, hence, codes that achieve capacity may exhibit little or no structure, making them ill-suited for practical applications. It is, hence, of interest to investigate the maximal reliable transmission rates achievable by structured ensembles of codes.

An important class of structured codes is the class of lattice codes. For the purpose of this correspondence we define a lattice code C_n as the intersection of a (possibly translated) n-dimensional full-rank lattice Λ_n with a region B_n of bounded support. What is the maximal reliable transmission rate achievable by means of lattice codes?

To take full advantage of the underlying lattice structure we would like to neglect the effects of the bounding region B_n and simply decode to the nearest *lattice point* (which may or may not be a code point). This decoding procedure is often referred to as *lattice decoding*. The optimum decoding procedure, on the other hand, is *minimum-distance* decoding which maps the received point into the closest *code point*. Lattice decoding is, in general, significantly less complex and it results in a uniform probability of error among all codewords, i.e., the average and the maximum probability of error are equal. Minimum-distance decoding, on the other hand, minimizes the average probability of error but a "good" code with respect to an average probability of error criterion may contain some "bad" codewords.

The known facts about lattice codes for the AWGN channel can be summarized as follows.

1) For any rate $R < \frac{1}{2}\log_2(S/N)$, there exists a lattice code C_n with arbitrarily small (maximum) probability of error when used with lattice decoding [2]. Moreover, the bounding region B_n can be chosen to be the n-dimensional ball of radius $\sqrt{nS}$.
2) By choosing B_n to be a "thin" spherical shell centered at the origin, rates up to capacity can be achieved with arbitrarily low average probability of error using a minimum distance decoder [3], [4].

To prove the first result, de Buda used the Minkowski–Hlawka Theorem. Loeliger [5], [6] derived the same result using a standard averaging argument for linear codes applied to lattices. He also conjectured that $\frac{1}{2}\log_2(S/N)$ is indeed the highest achievable rate under lattice decoding. Regarding the second result, in [4] it was pointed out that because of the "thin" spherical bounding region these codes loose most of their structure and resemble more "random" codes.

In this correspondence we will use [3] and [4] to close one of the remaining gaps by showing that lattice codes with a bounding region equal to an n-dimensional ball as opposed to a spherical shell also achieve capacity under minimum-distance decoding. Compared to the codes used in [3] and [4], these codes exhibit more structure which should facilitate their encoding and decoding process. Given the already known results this result is not surprising since in high dimensions most of the volume within a ball lies in a thin spherical shell and, hence, one might expect that most code points also lie in this spherical shell. This intuitive idea is made precise in Lemmas 1 and 3.

The main result of this correspondence is summarized by

Theorem 1: Let S, N, and $\epsilon > 0$ be given. If

$$R < \frac{1}{2}\log_2\left(1 + \frac{S}{N}\right)$$

then there exists a lattice code C_n for the additive white Gaussian noise channel with peak power constraint S and noise variance N, where B_n is the n-dimensional ball of radius $\sqrt{nS}$, such that C_n

Manuscript received October 23, 1994; revised May 28, 1997. This work was supported by the National Science Foundation under Grants NCR-9357689 and NCR-9304763. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Whistler, BC, Canada, Sept. 17–22, 1995.

R. Urbanke is with Bell Laboratories, Murray Hill, NJ 07974 USA.
B. Rimoldi is with the Department of Electrical Engineering, ESSRL, Washington University, St. Louis, MO 63130 USA.

Publisher Item Identifier S 0018-9448(98)00005-4.