A lower bound on the information capacity of a quantum narrow-band free-space link

1981 J. Phys. A: Math. Gen. 14 2625

(http://iopscience.iop.org/0305-4470/14/10/017)

J. Phys. A: Math. Gen. 14 (1981) 2625–2630. Printed in Great Britain

A lower bound on the information capacity of a quantum narrow-band free-space link

W G Chambers

Department of Mathematics, Westfield College, (University of London), Kidderpore Avenue, London, NW3 7ST

Received 21 November 1980

Abstract. A method suggested by Holevo is used to obtain a lower bound for the information capacity of a quantum narrow-band free-space link without extraneous noise. At high photon rates the bound is better than those previously proposed and comes close to a fundamental upper bound. It is not so good at low rates, where it is beaten by photon-detecting systems. The normal modes are not treated as independent channels. A system is describeu where the normal modes are used in pairs.

1. Introduction

In a previous article (Chambers 1981, henceforth referred to as I) the author attempted to find upper and lower bounds for the information capacity of a narrow-band free-space link. The upper bound is due to the finite number of orthogonal states available to the electromagnetic field. The lower bounds were obtained by considering the information rates of hypothetical systems, in which the normal modes of the field are used as independent channels. Since then the author's attention has been drawn to an article by Holevo (1979) in which it is proved that if the modes are used cooperatively it is possible to achieve higher rates. In this article Holevo's techniques are adapted for the narrow-band free-space link to obtain a better lower bound.

It may be useful to picture a narrow-band free-space link of bandwidth B as a system where the receiver exposes (say by means of a synchronised switch) a number B of simple harmonic oscillators in every second. The transmitter then puts each oscillator into a predetermined 'coherent state' (Helstrom 1976). The mean number of photons per oscillator $\bar{n}$ is equal to $P/(Bh\nu)$ where ν is the central frequency and P the received power. We define J as the information rate divided by B, so that it is in effect the information per oscillator.

It was argued in I on simple grounds that the information J could not exceed a value

$$C_{\max} = (\bar{n}+1)\ln(\bar{n}+1) - \bar{n}\ln\bar{n}. \tag{1}$$

It was also suggested that this bound was probably unattainable. For small values of $\bar{n}$ it is found that $C_{\max} \approx \bar{n}\ln(1/\bar{n}) + \bar{n} + O(\bar{n}^2)$, so that the information per photon is roughly equal to $\ln(1/\bar{n})$. A photon-detecting system also showed this logarithmic behaviour and gave a value of J which compared favourably with $C_{\max}$. At high values of $\bar{n}$ it is found that $C_{\max} = \ln(\bar{n}) + 1 + \frac{1}{2}\bar{n}^{-1} + O(\bar{n}^{-2})$. A 'lattice system' was described with

$$J = \ln \bar{n} + 1 - D + O(\bar{n}^{-1}) \tag{2}$$

with the 'defect' D equal to approximately 0.71.

It will be shown that there is a lower bound for the capacity of

$$C_{\min} = \ln\{\tfrac{1}{2}[1 + (1 + 4\bar{n}^2)^{1/2}]\} + 4\bar{n}/[1 + 2\bar{n} + (1 + 4\bar{n}^2)^{1/2}]. \tag{3}$$

For large $\bar{n}$ this gives $C_{\min} = \ln(\bar{n}) + 1 + \frac{1}{4}\bar{n}^{-1} + O(\bar{n}^{-2})$, which is remarkably close to $C_{\max}$. Unfortunately for small $\bar{n}$ we find that $C_{\min} \approx 2\bar{n}$, without the logarithmic behaviour. Thus (3) is presumably not the last result in this subject.

Equation (3) is derived in § 3 after some preliminaries in § 2. In § 4 it is shown that the cooperative use of pairs of oscillators in a 'lattice system' can indeed improve on the author's previous results, though not as strikingly as equation (3).

2. A bound on the error

Suppose that a signalling system uses an alphabet of M symbols over a discrete 'memoryless' channel, and that there is a probability q_r that the rth symbol is incorrectly received. The mean probability of error is defined as $\lambda = M^{-1}\Sigma_r q_r$, for which Holevo (1979) derived a bound. This bound does not directly guarantee a bound on the maximum probability of error, but it is easy to show that there is a sub-alphabet with at least $\frac{1}{2}M$ symbols with a maximum probability of error not exceeding 2λ. For suppose we throw out those symbols whose probability of error exceeds 2λ. Then the ensuing sum S of the probabilities of error must satisfy $S < M\lambda - 2n\lambda$, where n is the number of symbols thrown out. Since $S \geq 0$ we must have $n < \frac{1}{2}M$.

We imagine that when the rth symbol is sent, the receiver is put into a quantum state represented by the normalised state vector ψ_r. These vectors are not necessarily orthogonal or even linearly independent. We attempt to determine which symbol was sent by measuring some dynamical variable (Dirac 1958) with orthonormal eigenstates e_r. If the system is then found to be in the state e_l we assume that the lth symbol was sent. The probability of finding e_l when ψ_r was used is just $|(\psi_r, e_l)|^2$. Thus the mean probability of error is $\lambda = M^{-1}\Sigma_r[1 - |(\psi_r, e_r)|^2]$. Holevo has proved that the minimum value $\lambda_{\min}$ of this expression obtained by varying the orthonormal set e_r satisfies

$$\lambda_{\min} \leq M^{-1}\Sigma'_{r,s}|(\psi_r, \psi_s)|^2 \tag{4}$$

where the prime denotes that the terms with $r = s$ are omitted. The proof assumes that the ψ_r are linearly independent, but the extension to the case when this is not so is easily made. Since the proof is brief and perhaps not immediately accessible to all readers, a version is given here.

From the inequality $|1 - \alpha|^2 \geq 0$ we find $2 - 2\,\mathrm{Re}(\alpha) \geq 1 - |\alpha|^2$, and so it follows that $\lambda M \leq \Sigma_r[2 - 2\,\mathrm{Re}(\psi_r, e_r)]$. We now wish to choose the orthonormal set e_r to minimise the right-hand side. We define the Hermitian matrix $\Gamma_{rs} = (\psi_r, \psi_s)$ whose diagonal elements are equal to 1. We may then find a unitary matrix C_{rk} such that $\Sigma_s \Gamma_{rs} C_{sk} = C_{rk}\gamma_k$ where the γ_k are the eigenvalues of $\mathbf{\Gamma}$, which are positive if the ψ_r are linearly independent. Thus we find

$$\Sigma_r(\psi_r, e_r) = \Sigma_{krs} C^*_{rk}(\psi_r, e_s) C_{sk} = \Sigma_{ks}\gamma_k^{1/2}(\varphi_k, e_s) C_{sk}$$

where the vectors $\varphi_k = \gamma_k^{-1/2} \Sigma_r \psi_r C_{rk}$ form an orthonormal set. Hence the matrix $U_{ks} = (\varphi_k, e_s)$ is unitary, and so the above expression is $\Sigma_k \gamma_k^{1/2} (\boldsymbol{UC})_{kk}$ where the matrix $\boldsymbol{UC}$ is also unitary. Since $\gamma_k^{1/2} > 0$ the real part is maximised by setting $\boldsymbol{UC} = \boldsymbol{I}$, which maximises (to unity) the real parts of every diagonal element of $\boldsymbol{UC}$. Thus we find that

$$\begin{aligned} M\lambda_{\min} &\leqslant \Sigma_k (2 - 2\gamma_k^{1/2}) = \mathrm{Tr}(2 - 2\Gamma^{1/2}) = \mathrm{Tr}(1 + \Gamma - 2\Gamma^{1/2}) \\ &= \mathrm{Tr}(1 - \Gamma^{1/2})^2 = \mathrm{Tr}[(1 - \Gamma)/(1 + \Gamma^{1/2})]^2 \leqslant \mathrm{Tr}(1 - \Gamma)^2 \\ &= \Sigma'_{rs} |(\psi_r, \psi_s)|^2 \end{aligned}$$

which is the result (4).

If the ψ_r are linearly dependent, then we may assume that there is a linearly independent subset on which the others depend. We introduce a positive parameter η and an orthonormal set of vectors f_r which are also orthogonal to the space spanned by the ψ_r. (Thus the state-space is extended.) We define $\eta_r = 0$ if ψ_r belongs to the linearly independent subset, and $\eta_r = \eta$ otherwise. Then it is easy to see that the vectors $\chi_r = (1 + \eta_r^2)^{-1/2} (\psi_r + \eta_r f_r)$ form a set of linearly independent normalised vectors, for which the preceding analysis may be used. It is then readily shown by substituting for ψ_r that

$$1 - |(\psi_r, e_r)|^2 \leqslant 1 - |(\chi_r, e_r)|^2 + 2\eta_r (1 + \eta_r^2)^{1/2}$$

and by substituting for χ_r that

$$|(\chi_r, \chi_s)|^2 \leqslant |(\psi_r, \psi_s)|^2 + \delta_{rs} (2\eta_r^2 + \eta_r^4)$$

and so overall it is found that

$$M\lambda_{\min} \leqslant \Sigma'_{rs} |(\psi_r, \psi_s)|^2 + F(\eta)$$

where $F(\eta) \to 0+$ as $\eta \to 0+$. Since the other terms are not dependent on η we reobtain (4).

3. A lower bound

The lower bound (3) can be obtained as follows. We imagine that N normal modes or oscillators are used together in some way. Each receiver state is then a direct product of N coherent states $|\alpha_1, \ldots, \alpha_N\rangle = |\alpha_1\rangle \ldots |\alpha_N\rangle$ where the α_j are complex numbers. We may represent such a state as a point $\boldsymbol{\alpha}$ in a $2N$-dimensional real space. The quantity $|\boldsymbol{\alpha}|^2 = \Sigma |\alpha_j|^2$ is then the expectation value of the total photon number which is $\bar{n}N$, $\bar{n}$ being the average number of photons per oscillator. Thus with a given value of $\bar{n}$, the points are confined to the surface of a $2N$-dimensional sphere of radius $a = (N\bar{n})^{1/2}$ (see figure 1). We now use the technique of random encoding (Holevo 1979, Shannon 1948). The M input symbols are represented by M points on this sphere. We then average over all such encodings, so that any given input symbol may be represented by any point on the sphere with uniform probability distribution. The sum in (4) involves $M(M-1)$ pairs of points. For two states represented by the points P and Q in figure 1 the square of the matrix element is simply $\exp(-q^2)$ where q is the distance PQ (Helstrom 1976). Without any loss of generality we may take P as fixed, and then average $\exp(-q^2)$ over all positions of Q on the surface. This bounds the average of

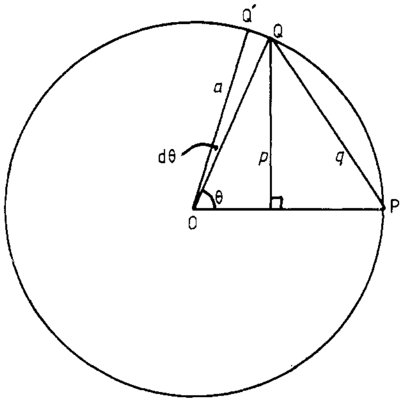

Figure 1. The geometry used for evaluating the average of $\exp(-q^2)$ over the surface of a sphere of radius $a=(\bar{n}N)^{1/2}$ in $2N$ dimensions.

$\lambda_{\min}$, which we denote by $\bar{\lambda}$. Thus we find

$$M\bar{\lambda}\leqslant M(M-1)\int \exp(-q^2)\,\mathrm{d}S\Big/\int \mathrm{d}S$$

where the integrals are taken over the sphere in figure 1. The belt produced by rotating QQ′ about the axis OP has an 'area' $K_{2N-1}p^{2N-2}a\,\mathrm{d}\theta$, where $K_{2N-1}p^{2N-2}$ is the 'surface area' of a sphere of radius p in $2N-1$ dimensions. By the geometry illustrated in figure 1 we have $p=a\sin\theta$, $q=2a\sin(\frac{1}{2}\theta)$, with $a=(\bar{n}N)^{1/2}$, so that

$$\bar{\lambda}\leqslant(M-1)\int_0^{\pi}\exp[-4\bar{n}N\sin^2(\tfrac{1}{2}\theta)](a\sin\theta)^{2N-2}a\,\mathrm{d}\theta\Big/\int_0^{\pi}(a\sin\theta)^{2N-2}a\,\mathrm{d}\theta.$$

We require a bound for $\ln\bar{\lambda}$ in the limit of large N, and for this purpose we may use the method of stationary phase in a fairly cavalier way. We replace the exponents $2N-2$ by $2N$, the factor $M-1$ by M and we evaluate the logarithm of each integral as the logarithm of the greatest value of the corresponding integrand. The integrand in the numerator has a maximum when $\theta=\cos^{-1}\{2\bar{n}/[1+(1+4\bar{n}^2)^{1/2}]\}$, and so it is found that $\ln\bar{\lambda}\leqslant\ln M-NC_{\min}$ where $C_{\min}$ is given by (3). The information sent per oscillator J is given by $JN=\ln M$ and so if J is less than $C_{\min}$ and independent of N we find that the bound on $\bar{\lambda}$ falls exponentially with N. Now $\bar{\lambda}$ is the probability of error averaged over all encodings. Thus there is an encoding whose value of $\lambda_{\min}$ is less than this. Overall, provided that $J<C_{\min}$ we can certainly find an encoding with a probability of error as small as we please, by making N large enough. This establishes a lower bound for the information capacity.

A plot of $C_{\min}$ and $C_{\max}$ against $\bar{n}$ is shown in figure 2. It is evident that the bounds are fairly close, although the ratio falls off for small values of $\bar{n}$.

4. Lattice system with paired oscillators

Proofs based on random encoding are not constructive, so it is worth considering a definite situation to show how the cooperative use of oscillators (or normal modes) can increase the rate of information. A lattice model was set up in I as a hypothetical

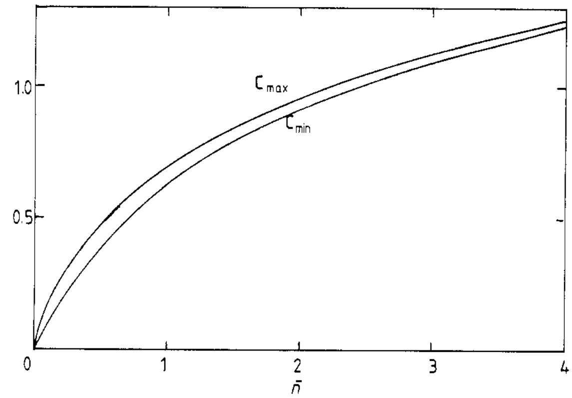

Figure 2. Plots of the bounds C_{max} (equation(1)) and C_{min} (equation (3)) against $\bar{n}$ on the information per oscillator (in natural units).

transmission system with independently used oscillators (for the case of large $\bar{n}$). It was shown that if the values of the coherent-state parameter α were permitted to form a square lattice of spacing $\pi^{1/2}$ then the displacement operators which translated the coherent states into one another formed an abelian group. Thus in vector notation we may write $\alpha = \pi^{1/2}(l_1, l_2)$ with l_1 and l_2 integral. Hence solid-state techniques can be used to form orthonormal Wannier functions (Ziman 1964). (Incidentally for this system these form the optimal set e_r as described in § 2, provided the coherent states are linearly independent.) Moreover it was shown that if the coherent states had been orthogonal, so that there need have been no error in the measuring process, then the information would have been given by (2) with $D = 0$. Unfortunately this square lattice could not be directly used, since these coherent states are linearly dependent. So a larger lattice spacing was used, which is equivalent to restricting the values of l_1 and l_2 by $l_1 + l_2 \equiv 0 \bmod 2$. Since this effectively halves the number of points for each normal mode it reduces the information rate by ln 2 plus a little more for the 'equivocation' due to the probability of error. Hence it was found that $D = 0.71$ in equation (2).

An alternative is to pair off the oscillators in some way, and specify the direct product of pairs of coherent states by two complex numbers $\alpha = \pi^{1/2}(l_1, l_2)$ and $\beta = \pi^{1/2}(l_3, l_4)$ which form a 'cubic' lattice in four dimensions if the l_i are integers. We then introduce a simple coding by demanding that the l_i add up to a multiple of 2. The effect of this is to halve the density of lattice points and thus to reduce the information per pair of oscillators by ln 2. Rather than use the method of quadratures in $\boldsymbol{k}$-space suggested in I to compute the equivocation (a rather daunting prospect in a four-dimensional non-cubic Brillouin zone!), the following perturbation method can be used. The amplitudes corresponding to (ψ_r, e_l) in this system turn out to be the Fourier components of a function $\gamma_k^{1/2}$ in $\boldsymbol{k}$-space where $\gamma_k = 1 + \Sigma'_{\boldsymbol{R}} M_{\boldsymbol{R}} \exp(\mathrm{i}\boldsymbol{k} \cdot \boldsymbol{R})$. Here the $M_{\boldsymbol{R}}$ are overlap integrals for coherent states spaced by $\boldsymbol{R}$ (in vector notation) and the prime denotes the absence of the term with $\boldsymbol{R} = 0$. The Fourier components of $\gamma_k^{1/2}$ can then be picked out from the ordinary binomial expansion for the square root. Using this method we obtain a value for the 'defect' D in equation (2) of about 0.40 after halving the computed equivocation to allow for the pairing of the oscillators.

Presumably it is possible to group oscillators into larger blocks, and 'thin out' the multi-dimensional lattice by imposing several parity constraints on the l_i. Whether this makes D tend to zero has not been investigated.

5. Concluding remarks

The author has no serious suggestions for obtaining better answers at the other limit, when $\bar{n}$ is very small. It may be that in this case the photons are acting in a classical manner, as particles, and so the old theorem may apply that for classical links the optimum rate can be obtained by using the normal modes independently (McEliece 1977). It seems unlikely that the upper bound $C_{\max}$ can be attained in this way. If the upper bound cannot be attained for small $\bar{n}$, then it may not be attainable for large $\bar{n}$ either, although the method of 'random coding' comes remarkably close.

Acknowledgments

The author would like to thank Professor C W Helstrom for drawing his attention to the article by A S Holevo.

References

Chambers W G 1981 *J. Phys. A: Math. Gen.* **14** 133
Dirac P A M 1958 *Principles of Quantum Mechanics* (Oxford: Clarendon) 4th edn, ch 2
Helstrom C W 1976 *Quantum Estimation and Detection Theory* (New York: Academic) §§ V.2, V.5
Holevo A S 1979 *Problemi peredachi informatsii* **15** 3–11 (in Russian)
McEliece R J 1977 *The Theory of Information and Coding* (Encyclopedia of Mathematics and its Applications) § 1.2 (Reading, Mass: Addison-Wesley)
Shannon C 1948 *Bell Syst. Tech. J.* **27** 379–423, 623–55
Ziman J M 1964 *Principles of the Theory of Solids* § 6.2 (Cambridge: University Press)