Exceeding the Classical Capacity Limit in a Quantum Optical Channel

Mikio Fujiwara, Masahiro Takeoka, Jun Mizuno, and Masahide Sasaki*

Communications Research Laboratory, Koganei, Tokyo 184-8795, Japan

(Received 13 November 2002; revised manuscript received 21 January 2003; published 24 April 2003)

The amount of information transmissible through a communications channel is determined by the noise characteristics of the channel and by the quantities of available transmission resources. In classical information theory, the amount of transmissible information can be increased twice at most when the transmission resource is doubled for fixed noise characteristics. In quantum information theory, however, the amount of information transmitted can increase even more than twice. We present a proof-of-principle demonstration of this superadditivity of classical capacity of a quantum channel by using the ternary symmetric states of a single photon, and by event selection from a weak coherent light source. We also show how the superadditive coding gain, even in a small code length, can boost the communication performance of the conventional coding technique.

DOI: 10.1103/PhysRevLett.90.167906 PACS numbers: 03.67.Hk, 03.65.Ta, 42.50.Dv

In any transmission of signals at the quantum level, such as a long-haul optical communication where the signals at the receiving end are weak coherent pulses, ambiguity among signals may be more a matter of noncommutativity of quantum states, i.e., $\hat{\rho}_0\hat{\rho}_1 \neq \hat{\rho}_1\hat{\rho}_0$, rather than any classical noise. Such states can never be distinguished perfectly even in principle. This imposes an inevitable error in signal detection even in an ideal communications system. It was only recently that communication theory was extended into quantum domain to include this aspect of ambiguity, and the expressions of channel capacity were finally obtained [1]. Classical communication theory [2] describes the special case of the signals prepared in commuting density matrices.

For reliable transmission in the presence of noise, redundancy must be introduced in representing messages by letters, such as $\{0, 1\}$, so as to correct errors at the receiving side. The capacity is associated with the functional meaning of this channel coding. Messages of k (bit) are encoded into block sequences of given letters in length n ($>k$). The $n - k$ (bit) redundancy allows one to correct errors at the receiving side. For a channel with a capacity C (bit/letter), it is possible [2] with the rate $R = k/n < C$ to reproduce k bit messages with an error probability as small as desired by appropriate encoding and decoding in the limit $n \to \infty$.

In extending the theory of capacity into quantum domain, primary concern is decoding of code words made of noncommuting density matrices of letters. The optimal decoding essentially uses a process of entangling letter states constituting code words prior to the measurement to enhance the distinguishability of signals. Such a process is nothing but a quantum computation on code word states. This is a new aspect, not found in conventional coding techniques, and leads to a larger capacity. A significant consequence of this so-called quantum collective decoding is that the capacity can increase even more than twice when the code length is doubled. In classical information theory, on the contrary, the capacity can be increased twice at most. This feature, the superadditive quantum coding gain [3–7], will be an important design rule for communications at the quantum level.

The theory of capacity, however, generally gives no guidance on how to construct codes that approach the capacity. The practical problem is then to find good codes for small blocks. Although several coding schemes have been proposed to exhibit superadditive coding gain [4–7], little attention has been paid to this topic so far, and no experimental work has been reported yet. In fact, putting these theoretical predictions into practice has been considered as a formidable task with present technologies. In this Letter, we experimentally demonstrate the superadditive coding gain by designing a coding circuit for a quantum channel consisting of the ternary symmetric states in a two-state system (qubit) of a single photon.

For binary nonorthogonal pure states, the most basic signals, the superadditive coding gain is predicted [5] for the minimum length, $n = 2$. The amount of gain, however, is so small to be observed experimentally, that is, 5.2×10^{-4} (bit) as the net increase of retrievable information per letter from the classical limit. For $n = 3$, the net gain of 0.009 (bit) is predicted [6]; however, this requires quantum gating more than ten steps with high precision, which is something hard to do. Therefore we consider the letter-state set that shows the largest amount of the coding gain with the minimum code length, $n = 2$, among the known codes [5–7].

Let us consider the set of the ternary letters $\{0, 1, 2\}$ conveyed by the symmetric states of a qubit system, $\hat{\rho}_x = |\psi_x\rangle\langle\psi_x|$ with $|\psi_0\rangle = |0\rangle$, $|\psi_1\rangle = -\frac{1}{2}|0\rangle - \sqrt{3}/2|1\rangle$, $|\psi_2\rangle = -\frac{1}{2}|0\rangle + \sqrt{3}/2|1\rangle$. Here $\{|0\rangle, |1\rangle\}$ is the orthonormal basis set. We assume that these states arrive at the receiver's hand through a noiseless transmission line. If the letter states were prepared in commuting density matrices, they could be distinguished perfectly, and $\log_2 3$ (bit) of information (the maximum Shannon entropy of the set $\{0, 1, 2\}$) could be faithfully retrieved per letter, meaning that the capacity would be $\log_2 3$ (bit/letter). However, the

0031-9007/03/90(16)/167906(4)$20.00

states $\hat{\rho}_x$ here are noncommuting, and distinguishing them is always associated with finite errors. In fact, the average error probability can never be lower than 1/3 when they are used with equal prior probabilities [8].

The capacity is mathematically given [2] based on the mutual information $I(X{:}Y)$ which is defined from the input variable $X = \{x\}$, the output variable $Y = \{y\}$, the prior distribution $\{P(x)\}$ of X, and the conditional probability $\{P(y|x)\}$ of Y for given X. For the given channel model $[P(y|x)]$, the capacity is defined by $C = \max_{\{P(x)\}} I(X{:}Y)$. In the quantum context, on the other hand, only the input variable X and the corresponding set of quantum states $\{\hat{\rho}_x\}$ at the receiver's hand are given. The output variable Y is to be sought for the best quantum measurement, i.e., a positive operator valued measure $\{\hat{\Pi}_y\}$. The channel matrix elements are now given by $P(y|x) = \mathrm{Tr}(\hat{\Pi}_y \hat{\rho}_x)$, and one is to find the quantity [9]

$$C_1 = \max_{\{P(x)\}} \max_{\{\hat{\Pi}_y\}} I(X{:}Y). \tag{1}$$

For the ternary set $\{|\psi_x\rangle\}$, the C_1 was evaluated as 0.6454 (bit/letter) which is attained by using only two of the three letters, say $\{|\psi_0\rangle, |\psi_1\rangle\}$, with equal probability 1/2 and by applying the binary measurement to form a binary symmetric channel [10]. The quantity C_1 is, however, not the ultimate capacity allowed by quantum mechanics. In fact, C_1 specifies the classical capacity limit when the given initial channel is used with classical channel coding [11]. It is this quantity that limits the performance of all conventional communications systems.

Now let us consider a quantum channel coding of length two. There are nine possible sequences in length two coding of three letters. Peres and Wootters showed [4] that $I(X^2{:}Y^2) = 1.3690$ (bit) of information can be retrieved in principle, which is greater than twice the classical limit $2C_1 = 1.2908$ (bit). This can be achieved in the following way; only three sequences $|\Psi_{xx}\rangle = |\psi_x\rangle \otimes |\psi_x\rangle$ $(x = 0, 1, 2)$ are used as the code words with equal probability 1/3, and they are decoded by the measurement represented by the elements $\hat{\Pi}_{yy} = |\Pi_{yy}\rangle \times \langle \Pi_{yy}|$ $(y = 0, 1, 2)$ which compose the orthonormal basis expanding $\{|\Psi_{xx}\rangle\}$; that is,

$$|\Psi_{00}\rangle = c|\Pi_{00}\rangle - \frac{s}{\sqrt{2}}|\Pi_{11}\rangle - \frac{s}{\sqrt{2}}|\Pi_{22}\rangle, \tag{2a}$$

$$|\Psi_{11}\rangle = -\frac{s}{\sqrt{2}}|\Pi_{00}\rangle + c|\Pi_{11}\rangle - \frac{s}{\sqrt{2}}|\Pi_{22}\rangle, \tag{2b}$$

$$|\Psi_{22}\rangle = -\frac{s}{\sqrt{2}}|\Pi_{00}\rangle - \frac{s}{\sqrt{2}}|\Pi_{11}\rangle + c|\Pi_{22}\rangle, \tag{2c}$$

where $c = \cos\frac{\gamma}{2} = (\sqrt{2} + 1)/\sqrt{6}$, and $s = \sin\frac{\gamma}{2} = (\sqrt{2} - 1)/\sqrt{6}$ $(\gamma \simeq 19.47°)$. Figure 1 shows a geometrical representation of Eq. (2). The superadditive coding gain is $I(X^2{:}Y^2)/2 - C_1 = 0.0391$ (bit/letter).

To demonstrate this gain, we must be able to entangle two letter states at the receiver's hand prior to a measurement. Unfortunately quantum gating operations demonstrated to date are not precise enough to observe the small quantum coding gain. Therefore our method for proof-of-principle demonstration is based on the use of two physically different kinds of qubits of a single photon. The first and second letters of a code word are drawn from the ternary letter-state sets made of a polarization and a location qubit, respectively. Then entangling the polarization and location degrees of freedom of a photon can be performed by linear optical components with very high accuracy. The polarization qubit consists of the horizontal $|H\rangle$ and vertical $|V\rangle$ polarization states of a single photon. The location qubit for the second letter is realized by guiding the polarization qubit into two different paths A and B through a polarizing beam splitter (PBS) which reflects the vertical polarization and transmits the horizontal polarization. Thus, the length two coding space is spanned by the two orthonormal basis sets [12]

$$|00\rangle = |0\rangle_P \otimes |0\rangle_L = |H\rangle_A \otimes |\mathrm{vacuum}\rangle_B, \tag{3a}$$

$$|01\rangle = |0\rangle_P \otimes |1\rangle_L = |\mathrm{vacuum}\rangle_A \otimes |H\rangle_B, \tag{3b}$$

$$|10\rangle = |1\rangle_P \otimes |0\rangle_L = |V\rangle_A \otimes |\mathrm{vacuum}\rangle_B, \tag{3c}$$

$$|11\rangle = |1\rangle_P \otimes |1\rangle_L = |\mathrm{vacuum}\rangle_A \otimes |V\rangle_B. \tag{3d}$$

The code word states $|\Psi_{xx}\rangle = |\psi_x\rangle_P \otimes |\psi_x\rangle_L$ are represented in this product space. Thus the increase in resources in our coding format is due to doubling the spatial resource which is analogous to doubling the transmission bandwidth, as opposed to doubling the number of polarized photons. We want then to observe the increase more than double the amount of information transmitted.

An optical circuit for this coding is shown in Fig. 2(a). The angles of the three half wave plates (HWPs) θ's are chosen as $(\theta_0, \theta_1, \theta_2) = (0°, 0°, 0°)$, $(30°, -30°, -15°)$, and $(30°, 30°, 15°)$ for $|\Psi_{00}\rangle$, $|\Psi_{11}\rangle$, and $|\Psi_{22}\rangle$, respectively. This decoding circuit is derived with slight modifications from a general circuit design of Fig. 2(b), which can be applied to any physical qubits. The received code word is decided to be either of $|\Psi_{00}\rangle$, $|\Psi_{11}\rangle$, or $|\Psi_{22}\rangle$ according to the detection of a photon by the avalanche photodetector APD0, APD1, or APD2, respectively.

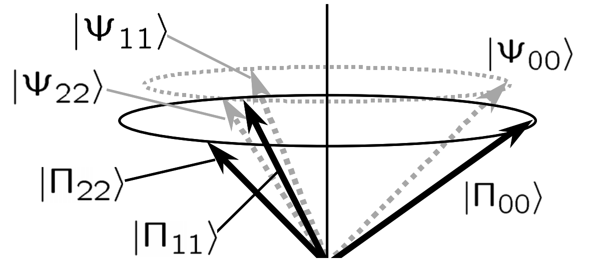

FIG. 1. Geometrical representation of the code word (dotted arrows) and decoding (solid arrows) state vectors in a real three dimensional space.

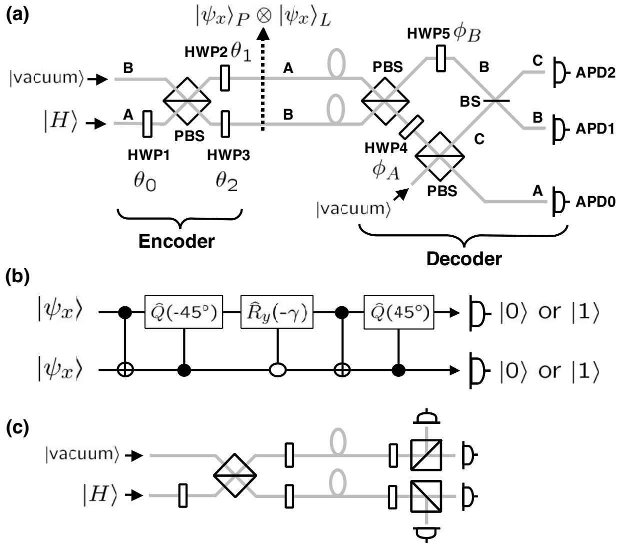

FIG. 2. (a) Encoding and decoding circuits. The angles of the HWPs, θ_0, θ_1, and θ_2 are chosen as described in the text, and $\phi_A = -\gamma/2 = -9.74°$ and $\phi_B = -45°$. (b) Quantum circuit to realize the collective decoding by $\{|\Pi_{yy}\rangle\}$, which can be applied to any physical qubits. A received code word state is first transformed by the five controlled gates and is then detected by a standard von Neumann measurement on each letter. The open circle indicates conditioning on the control qubit being set to zero, and $\hat{Q}(\varphi) = \hat{R}_y(\varphi)\hat{\sigma}_z$, and $\gamma = 19.47°$. Other nomenclature follows Ref. [13]. (c) Circuit for separable (classical) decoding for C_1.

In our experiment, the cw light from a He-Ne laser at the wavelength of 632.8 nm with 1 mW power is strongly attenuated such that about 10^{-2} photons exist on average in the whole circuit. The signal photons are guided to the Si APDs whose quantum efficiency and dark count are typically 70% and 100 (count/sec), respectively, through a multimode optical fiber with coupling efficiency of about 80%. In this setup, the mutual information is evaluated by constructing the 3-by-3 channel matrix $[P(yy|xx) \equiv |\langle\Pi_{yy}|\Psi_{xx}\rangle|^2]$ from a statistical data of single-photon events detected by either of the three APDs conditioned on an input code word $|\Psi_{xx}\rangle$. The mutual information thus obtained measures the ratio of number of bits retrieved per number of total photon counts. This allows us to simulate communications of "pure" code word states of two letters by sending and detecting the photons one by one through the channel. The error performance is then determined only by the noncommutativity of the signal states, imperfect alignment of the whole interferometer, and the dark count of the APDs.

Each polarization Mach-Zehnder interferometer must be adjusted simultaneously at a proper operating point. This is done by using a bright reference beam and Piezo transducers with low noise voltage sources. The visibility of the whole interferometer is typically 98%. Once the circuit is adjusted, the reference beam is shut off. The signal light is then guided into the encoder and decoder. Photon counts are measured for 5-sec duration. This procedure is repeated for each code word, composing a full sequence of measuring the channel matrix. The temporal stability in this sampling mode corresponds to the relative path length change within 3 nm for at least more than 200 sec, which causes the error in mutual information within ±0.005 (bit).

An example of the channel matrix measured is shown as a histogram in Fig. 3. Ideally, the diagonal and off-diagonal elements must be $c^2 = 0.9714$ and $s^2/2 = 0.0143$, respectively. The total events counted for 1 sec is of order 10^6, while the average count for the off-diagonal elements is about 1.9×10^4. The background photons amount to about 300 (count/sec). Including dark counts, the total background photon count is 2% of the average count for the off-diagonal elements. The mutual information is evaluated as $I(X^2{:}Y^2) = 1.312 \pm 0.005$ (bit). For experimental clarity, we measured the variation of the mutual information when the code word state set $\{|\Psi_{xx}\rangle\}$ is rotated with respect to the decoder state set $\{|\Pi_{yy}\rangle\}$ around the vertical axis in Fig. 1. The result is shown in Fig. 3. The gap between the data points (diamonds) and the ideal curve (solid curve) is mainly attributed to the imperfection of the PBSs. Fluctuation of the data points is mainly due to thermal drifts. The

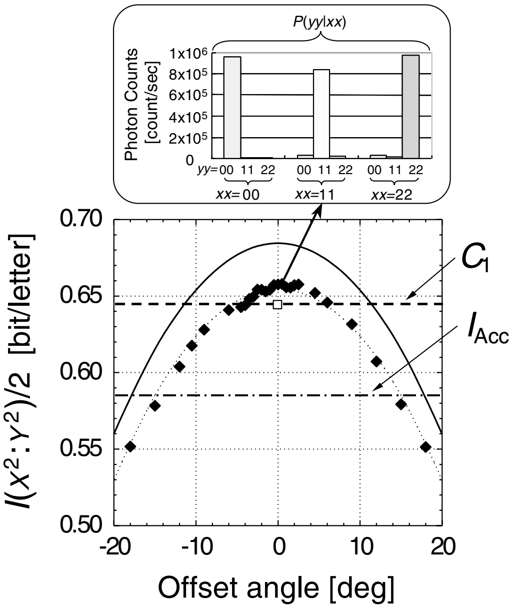

FIG. 3. Top: histogram of photon counts for the channel matrix elements $P(yy|xx)$ corresponding to the maximum mutual information. Bottom: measured (diamonds) and theoretical (solid curve) mutual information as a function of the offset angle of the code word state set $\{|\Psi_{xx}\rangle\}$ from the decoder state set $\{|\Pi_{yy}\rangle\}$ around the vertical axis in Fig. 1. The dotted curve is just a guide for the eyes. The theoretical C_1 and accessible information I_{Acc} are shown by the dashed and one-dotted lines, respectively [10]. The square represents the experimental C_1.

corresponding error bars [~ ±0.005 (bit)] are about the same size of the diamonds. The measured mutual information per letter, 0.656 ± 0.003 (bit/letter), is clearly greater than the classical theoretical limit $C_1 = 0.6454$ (bit/letter), which is the level shown by the dashed lines. The white square represents the experimental C_1, 0.644 ± 0.001 (bit/letter). This is measured by the circuit for classical decoding of Fig. 2(c), which does not entangle the polarization and location qubits. The retrieved information can never exceed $2C_1$. Our results clearly show that when an appropriate quantum circuit for entangling the letter states is inserted in front of the separable decoding, one can retrieve information more than twice per letter.

The superadditive coding gain in small blocks is not only valuable as a proof-of-principle demonstration but also of practical importance in quantum-limited communications. Even a two-qubit quantum circuit like Fig. 2(b) is useful in boosting the performance of a classical decoder. It can be shown that the decoding error can be greatly reduced by inserting the quantum circuit in front of the classical decoder. The quantum circuit processes a received code word state quantum collectively prior to converting it into a classical signal. We call such a scheme quantum-classical hybrid coding (QCHC). The theoretical error exponents [2] of QCHC and all-classical coding (ACC) in the ternary letter-state case are listed in Table I for low and high transmission rates R. The improvement is more drastic in the higher rate limit. For the rate $R = 0.62$ (bit/letter) (96% of C_1), it is possible to reduce the decoding error as $P_e = 2^{-(n/2)E(R)} = 2^{-0.0488n}$ by an appropriate classical coding with the composite letters $\{00, 11, 22\}$ assisted by the pairwise quantum decoding. To achieve a standard error-free criterion $P_e = 10^{-9}$, QCHC requires the code length $n = 614$ (307 composite letter pairs), whereas ACC typically needs $n = 57\,300$. As codes get longer, the complexity of the decoder, such as the total number of arithmetic operations, increases and eventually limits the effective transmission speed. For some asymptotically good codes, the total number of arithmetic operations is evaluated [14] to be typically of order $(n \log n)^2$. Then the reduction of code length attained by QCHC will be practically significant in the trade-off between performance and decoding complexity. This suggests a useful application of small scale quantum computation.

The superadditive quantum coding gain should eventually be applied to more practical resources such as optical pulses of coherent states, especially, heavily attenuated coherent states $\{|\alpha_k\rangle\}$ of phase-shift and/or amplitude-shift keying. Unlike our channel model of single-photon polarization and location modes, one must be able to entangle weak coherent pulses with respect to the degrees of phase and/or amplitude. This is another big challenge.

TABLE I. Error exponent $E(R)$ of QCHC and ACC.

R (bit/letter)	$E(R)$ of QCHC	$E(R)$ of ACC
0.1	0.842	0.315
0.62	9.753×10^{-2}	5.218×10^{-4}

The authors acknowledge J. A. Vaccaro for his valuable comments. This work was supported by the CREST of JST.

*Electronic address: psasaki@crl.go.jp

[1] P. Hausladen *et al.*, Phys. Rev. A **54**, 1869 (1996); B. Schumacher and M. Westmoreland, Phys. Rev. A **56**, 131 (1997); A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).

[2] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948), Pt. I; **27**, 623 (1948), Pt. II; R. G. Gallager, *Information Theory and Reliable Communication* (John Wiley & Sons, New York, 1968).

[3] A. S. Holevo, Prob. Peredachi Inf. **15**, 3 (1979).

[4] A. Peres and W. K. Wootters, Phys. Rev. Lett. **66**, 1119 (1991).

[5] J. R. Buck *et al.*, Phys. Rev. A **61**, 032309 (2000).

[6] M. Sasaki *et al.*, Phys. Rev. A **58**, 146 (1998).

[7] S. Usami *et al.*, *Quantum Communication, Computing, and Measurement 3*, edited by P. Tombesi and O. Hirota (Plenum, New York, 2001), p. 35.

[8] C.W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[9] For a brief review of the mutual information and the capacity, see Appendix in J. Mizuno *et al.*, Phys. Rev. A **65**, 012315 (2001).

[10] M. Osaki *et al.*, *Quantum Communication, Computing, and Measurement 2*, edited by P. Kumar *et al.* (Plenum, New York, 2000), p. 17; P.W. Shor, quant-ph/0206058. If the three letters are used with equal probabilities 1/3, the maximum mutual information remains 0.5850 (bit/letter), which is called the accessible information for a fixed *a priori* probability; see M. Sasaki *et al.*, Phys. Rev. A **59**, 3325 (1999); R. B. M. Clarke *et al.*, Phys. Rev. A **64**, 012303 (2001).

[11] A. Fujiwara and H. Nagaoka, IEEE Trans. Inf. Theory **44**, 1071 (1998).

[12] N. J. Cerf *et al.*, Phys. Rev. A **57**, R1477 (1998); R. J. C. Spreeuw, Found. Phys. **28**, 361 (1998); S. Takeuchi, in *Proceedings of PhysComp96*, edited by T. Toffoli (New England Complex System Institute, Boston, 1996), p. 299.

[13] A. Barenco *et al.*, Phys. Rev. A **52**, 3457 (1995).

[14] S. Hirasawa *et al.*, IEEE Trans. Inf. Theory **26**, 527 (1980).