7 Quantum Communication Technologies

7-1 Quantum detection circuit for quantum channel coding

Masahide SASAKI, Jun MIZUNO, and Mikio FUJIWARA

Light, main carrier in the present information technology, is electromagnetic wave, and also an ensemble of energy quanta, photon as well. At present only the fact is used that light propagagtes as an energy flux, and the wave nature of light is never used any more so far. Conventional information theory is readily capable of designing communication system based on the wave nature of light, and providing its performance limit. However the ulitimate performace limit of optical communication is eventually given by the law of quantum mechanics that governs photon dynamics.

Quantum channel capacity is then determined by the distinguishability of optical quantum states. In this article we consider one of the most basic quantum signal system, symmetric states of a single photon polarization, which is often used for quantum key distributiontion. It is clarified what is the maximum amount of information that can be extracted from that source and how one can implement the optimal detector for attaing it. Conventionally detection of polarization modulation signals is made by using a polarizer. For a single photon state, the binary output detection based on a polarizer is the standerd mesurement, which is called the von Neumann measurement. On the other hand the optimal solution in quantum information theory is given by the detector with three outputs at most regardless of the number of the signal components. This type of detector can be implemented by the polarization interferometer. We developed such a detector and could demondtrate the 96% of the predected limit which is superior to the conventional von Neumann limit.

1 Introduction

Information technology stands on the modeling of information processing in terms of the binary symbols 0 and 1. That is, messages are represented by sequences of 0 and 1, and their transitions correspond to transmission and processing of information. Noise effect disturbing these tasks is modeled as probabilistic transitions between 0 and 1. Physical entities conveying {0, 1} are macroscopic ensembles of electrons and photons. In the ideal noiseless limit such 0- and 1-state can be in principle distinguished from each other. In addtion such states can be copied and amplified. Under these basic assumptions, information theory quantifies information as the uncertainty associated with probabilistic events, and turn out to be a powerful tool to design the optimal information trans-

mission and processing systems.

Information technology has now started to control directly quantum objects such as an atom, electron, and photon. {0, 1} conveyed by these quantum entities can neither be in general copied nor amplified[1][2]. Perfect discrimination between quantum states is prohibited in principle[3][4]. While this fact imposes the new performance limit on information technology, this provides us the new principle for quantum cryptography that ensures the perfect security. More importantly, in transitions among sequences of 0 and 1, there appears quantum mechanical interference effect inherent in a quantum system, and by controling this effect optimally we are led to the new performance limit that is much superior to the one extraporated by direct application of conventional theory. Information theory must include these limits and be modeled in terms of quantum mechanics. On the other hand recent progress in quantum optics enable us to extend quantum *entanglement* seen typically in atomic scale to a few 10km range, and it is clarified that this new resource opens the new paradigm such as quantum teleportation, quantum cryptography and netork quantum computing[5]. Namely conventional information theory is just one of the possibilities for modeling and definition of information. Much wider definition of information and new information tasks are definitly possible. The guiding principle for information technology should now be unified with quantum mechanics.

The whole view of information technology predicted by this new theory, quantum information theory, is hardly seen yet. We have just started to study all kinds of possibilities in information technology[6]. Once we turn to the experimental research area, only a few works have been made toward realization of the theoretical predictions. One of the most important problems might be quantum detection problem because distinguishing quantum states is at the root of any information tasks. In this article, we focus on this problem from the view point of maximization of the mutual information, which plays an essensial role in coding technology toward the ultimate channel capacity. In particular we consider one of the simplest quantum signals based on a single photon system, and review our recent progress.

2 Basics of quantum detection and mutual information

Primary concerns of information theory are *how to represent messages as effectively as possible* and *how to transmit messages as precisely as possible*. In communications systems a sender, Alice, has a source of messages S and selects one of a known set $\{a, b, ..., z\}$ with given prior probabilities $\{P(a), P(b), ..., P(z)\}$. This source may be characterized by the random variable $S = \{a, b, ..., z; P(a), P(b), ..., P(z)\}$. The sender represents each of these messages by a sequence of a given set of *letters* $\{x_i\}$ such as {0, 1}. These are the symbols running through the transmission channel. Each message is then represented by a *codeword* formed from a sequence of letters. This is *source coding*. Information theory tells us that the effectiveness of source coding can be measured by the minimum of the average length required for a codeword and that it is given by the Shannon entropy

$$H(S) = -\sum_{A=a,b,\ldots} P(A)\log_2 P(A)\,. \qquad (1)$$

This is a measure of uncertainty in the random variable S. It takes its maximum value when all elements appear with equal probability, that is, when we know nothing better than a random guess for each element. This measure of uncertainty is regarded as the amount of information required to represent S.

A channel is usually subject to various types of noise disturbances. Information theory provides means and limits for reliable information transmission with such noisy channels. The key idea is to introduce some redundancy in the codeword representation prior to transmission so as to allow the correction of errors at the receiving side. This entails adding some

redundant letters to the codewords and hence increases their length. This is *channel coding*. The mutual information quantifies how much redundancy is required for error-free transmission.

The output from the source encoder is a sequence of the letters forming the codewords representing the messages. For such sequences one can find the frequencies of appearance $P(x_i)$ for each letter x_i. Thus we can de-fine a random variable $X = \{x_i; P(x_i)\}$ for the outputs from the source encoder. This is the set of inputs to the channel. A mathematical model for the channel is specified by the set of possible outputs $\{y_i\}$ and the conditional probability $P(y_i \mid x_i)$ for each input. Given X, $\{y_i\}$, and *channel matrix* $[P(y_i \mid x_i)]$, we can determine the existence or nonexistence of encoders and decoders that achieve a given level of transmission performance.

The mutual information is defined between the input and output random variables X and $Y = \{y_i ; P(y_i)\}$. Here

$$P(y_j) \equiv \sum_{x_i} P(y_j|x_i)P(x_i) \tag{2}$$

is the probability of having y_i . The uncertainty of the input random variable X is measured by the Shannon entropy

$$H(X) = -\sum_i P(x_i) \log P(x_i) \tag{3}$$

defined in a similar way to Eq. (1).

If the receiver detects the output signal y_i , then he is now more certain about X. The new probability distribution conditioned by y_i is given as

$$P(x_i|y_j) = \frac{P(y_j|x_i)P(x_i)}{P(y_j)} . \tag{4}$$

One can then define the average conditional entropy by

$$(X|Y) = -\sum_{y_j} P(y_j) \sum_{x_i} P(x_i|y_j) \log P(x_i|y_j). \tag{5}$$

This quantifies the remaining uncertainty of X after having the knowledge on the conditioning variable Y. The information extracted by the receiver is naturally defined by the reduction of the uncertainty,

$$I(X:Y) = H(X) - H(X|Y)$$
$$= \sum_{x_i, y_j} P(x_i)P(y_j|x_i) \log \left[\frac{P(y_j|x_i)}{\sum_{x_k} P(x_k)P(y_j|x_k)} \right] . \tag{6}$$

This $I(X : Y)$ is the mutual information between X and Y.

Now let us consider a block coding of length n. The output from the source encoder is a letter sequence, which is devided into blocks (*message blocks*) of length k ($<n$). Each block is supplemented by an additional block (*correction block*) of $n-k$ letters to compose a transmission codeword $\{x^p\}$:

$$\mathbf{x}^p = \overbrace{x^p{}_1 x^p{}_2 \cdots x^p{}_k}^{\text{message block}} \overbrace{x^p{}_{k+1} x^p{}_{k+2} \cdots x^p{}_n}^{\text{correction block}} \tag{7}$$
$$(\text{for } p = 1, 2, \ldots, L^k) ,$$

where each $x^p{}_l$ (l = 1, ..., n) is an element of possible letters $\{x_i; \text{ i = 0, 1, ..., } L-1\}$. Note that although there are L^n possible sequences of length n in total, only part of them, i.e. L^k sequences, are used as codewords. This redundancy, together with appropriate choice of correction blocks, allows us to recover the possible errors in transmission.

The input codeword $\mathbf{x}^p$ will be disturbed in the channel so as to come out as a different sequence $\mathbf{y}^q = y^q{}_1 y^q{}_2 \ldots y^q{}_n$. The channel decoder processes this output codeword to assign an appropriate sequence which should be the correct input codeword. The average error in this decoding should be as small as possible, while the redundancy $n-k$ should also be as small as possible. In other words, keeping the ratio $R = k/n$, so-called the *transmission rate*, as large as possible, we wish to attain a small error in decoding.

Let us suppose that encoding is made under the constraint that the frequency of x_i's occurring in the set of codewords $\{\mathbf{x}^p\}$ is $P(x_i)$. Information theory says that by an appropriate design of the coding scheme it is possible to transmit the messages with an error probability as small as desired if $R < \text{I}(X : Y)$ is satisfied. For the fixed channel model $[P(y_i \mid x_i)]$, one may further adjust prior probabilities

$\{P(x_i)\}$ to maximize the mutual information. The maximum value

$$C_c = \max_{\{P(x_i)\}} I(X:Y) \tag{8}$$

is called the *channel capacity*. Then the channel coding theorem tells us[7][8][9] that if $R < C$ holds there exists a coding scheme which transmits messages with an error probability as small as desired. Thus the mutual information is related to the ultimate use of the channel.

The basic frameworks described above also apply to a quantum limited channel. However a new ingredient comes into play, which is a quantum effect in the detection process. A detection process is represented mathematically by the probability operator measure (POM), which consists of nonnegative (generally not normalized) Hermitian operators satisfying the resolution of the identity[3][10][11]:

$$\hat{\Pi}_j^\dagger = \hat{\Pi}_j, \quad \hat{\Pi}_j \geq 0 \quad {}^\forall j, \quad \sum_j \hat{\Pi}_j = \hat{I}. \tag{9}$$

Each element $\hat{\Pi}_j$ is associated with the measurement outcome j and hence implies the output letter y_i . Let us consider the simplest case where the letter set $\{x_i\}$ is conveyed by a set of pure quantum states $\{|\psi_i\rangle\}$, *letter state*, possibly a nonorthogonal set, through a noiseless channel. Then the channel model is specified by a POM $\{\hat{\Pi}_j\}$ and the channel matrix $P(y_i\,|x_i) = \langle\psi_i|\hat{\Pi}_j|\psi_i\rangle$.

In the conventional (classical) context, the channel matrix $[P(y_i\,|x_i)]$ is given and fixed. In quantum domain, however, one may ask what is the best possible POM for the given set of letter states $\{|\psi_i\rangle\}$. This is actually a nontrivial problem. The problem can be decomposed into several steps. First we can consider the maximization of the mutual information with respect to a POM $\{\hat{\Pi}_j\}$ for the fixed $\{|\psi_i\rangle\}$ and prior probabilities $\{P(x_i)\}$. The maximum value

$$I_{\mathrm{Acc}}(\{|\psi_i\rangle; P(x_i)\}) = \max_{\{\hat{\Pi}_j\}} I(\{|\psi_i\rangle; P(x_i)\}:Y) \tag{10}$$

is called the *accessible information* of $\{|\psi_i\rangle; P(x_i)\}$. We can then consider the maximization of the accessible information over prior probabilities $\{P(x_i)\}$, and may define the quantity C_1 as

$$C_1 = \max_{\{P(x_i)\}} I_{\mathrm{Acc}}(\{|\psi_i\rangle; P(x_i)\}). \tag{11}$$

This would be a natural extension from the conventional idea. However, this C_1 is not in general the maximum bound for the transmission rate for error-free communication, and hence it is not the channel capacity. In fact, there is the peculiar quantum interference effect in quantum detection of codeword states, which was not taken into account in the conventional theory. The true capacity for a pure state channel is given by Hausladen *et al*[12]. The general theory for a mixed state channel is given by Holevo[13] and by Schumacher and Westmoreland[14].

To realize reliable transmission ensured by quantum theory of the channel capacity, one may need quantum computation for the decoding process[15][16]. This is, however, far beyond present technologies. If only a quantum detection on each letter state is available, then I_{Acc} and C_1 practically specify the limit of communication ability. Let us suppose again that encoding is made such that x_i (i.e.$|\psi_i\rangle$) occurs in the set of codewords $\{\mathbf{x}^p\}$ (i.e.$\{|\psi_{i1}\rangle \otimes \cdots \otimes |\psi_{in}\rangle\}$) with the probability $P(x_i)$. We further suppose that $\{\hat{\Pi}_j\}$ is the POM attaining the accessible information for $X = \{|\psi_i\rangle; P(x_i)\}$ and the receiver applies this detection *separately* on each letter states to get output sequences $\{y_{i1}y_{i2} \ldots y_{in}\}$. If $R < I_{Acc}$ holds, then a reliable transmission of the letters with an arbitrarily small error is possible by an appropriate *classical* coding. The optimum POM for the accessible information is thus an important concern for devising a good code for a quantum limited channel.

To find the optimal solution for the accessible information closely related to the other kinds of optimization tasks. The simplest requirement is that Bob wants to decide which letter state he has received among the set $\{|\psi_i\rangle\}$ with the smallest error. This usually

means minimizing the average error probability, or bit error rate P_e[17][18]. A second possibility is for Bob to eliminate all errors by allowing the possibility of inconclusive results by means of unambiguous state discrimination [19]~[26]. The optimum strategy in this case will be the one that minimizes the probability P_i of inconclusive outcomes. This type of detection has been proposed for quantum key distribution[27].

For the communication of messages, however, Bob does best by devising a detection strategy so as to retrieve Alice's message with the greatest probability. This does not necessarily mean minimizing either P_e or P_i, but instead means reducing the uncertainty in some *random variable* $X = \{x_i, P(x_i)\}$. Such a detection strategy is directly related to reliable communication by coding technique and is actually used as a basic building block for effective decoding procedures of codeword states formed from the letter states $\{|\psi_i\rangle\}$. (A more detailed explanation of this point is given in Appendix.)

The reduction of the uncertainty caused by a detection is quantified by the Shannon mutual information $I(X : Y)$ between the input (Alice's) and output (Bob's) random variables, $X = \{x_i, P(x_i)\}$ and $Y = \{y_i , P(y_i)\}$. This mutual information $I(X : Y)$ can be regarded as the amount of information extracted from X. Bob's optimum strategy will be the one that maximizes $I(X : Y)$. Other figures of merit have also been considered and these include the fidelity[28][29].

The optimum conditions are already known for minimizing the error probability [17][18]. It is not an easy task, however, to find the optimum detection strategies from these conditions. In fact, optimum strategies are only known for some special cases such as the set of binary states, sets of symmetric states [15][17][18][30][31] and multiply symmetric states[32]. Unambiguous state discrimination is possible if and only if the letter states are linearly independent and an explicit method for constructing the optimum strategy has been given in this case[25][26]. Finding optimum solutions for $I(X : Y)$ is much more difficult than those for P_e and P_i due to the nonlinearity of logarithmic function of $I(X : Y)$ with respect to a POM. Optimum solutions are known only for the set of binary pure states [33][34] and for sets of real symmetric qubit states with equal prior probabilities[35][36].

It seems intuitively reasonable that we might obtain most information by minimizing either the average error probability P_e or the probability of inconclusive outcomes P_i. In fact, the maximum mutual information for binary states is attained by the same strategy that realizes the minimum average error probability. There are, however, cases where the maximum information must be obtained neither by minimizing P_e nor P_i[35][37].

Devices capable of demonstrating near optimum detection at the single photon level have been demonstrated in the laboratory. The simplest of these is discrimination between the set of binary photon polarization states with the minimum allowed average error probability[38]. Unambiguous discrimination between two non-orthogonal polarization states has also been demonstrated[39][40]. A set of more than three polarization states is linearly dependent and hence it is not possible to carry out unambiguous state discrimination. Clarke *et al.* have demonstrated state discrimination with near minimum error probability for both the trine and tetrad polarization states[41]. They have also demonstrated the ability to extract more information than is possible by the best, standard von Neumann measurement (a projection onto binary orthogonal polarization states).

In this paper we describe our experimental implementation of a class of optimum strategies for maximizing the mutual information, as predicted by Ref[36]. One of these is the ternary or trine set of states discussed by Clarke *et al.*[41]. We have improved upon the information yield obtained by these authors and have also measured the information obtained from signals formed from five and seven possible polarizations. Our letter states are implemented physically as single photon

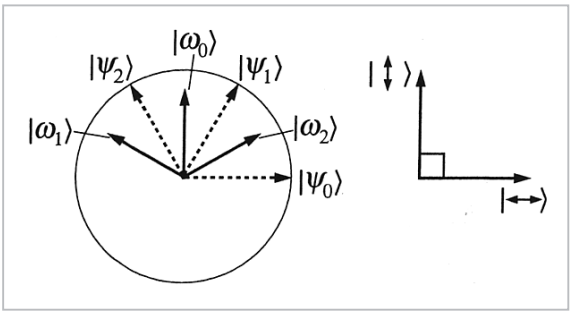


Fig.1 *The measurement state vectors for the optimum strategy (solid line) and the signal state vectors in the case of the ternary (trine) signals*

polarizations. The required equiprobable real symmetric qubit states are then states of linear polarization. Such sets of states have previously found application in quantum key distribution[42][43]. From the view point of fundamental interests, they might be the simplest system with which to test the peculiar effect predicted by Davies' theorem. According to the theorem, there must exist at least one solution, that maximizes the mutual information, which has N possible outputs, where N is bounded by $d \leq N \leq d^2$ with d being the dimension of the Hilbert space $\mathcal{H}_s$ supported by Alice's set[35]. For real state sets, this bounding inequality becomes $d \leq N \leq d(d+1)/2$[36]. Thus for a single photon polarization system, one can always optimize the mutual information by constructing a device with just three possible outputs. This is true regardless of the number of letter states. In the case of ternary or trine signals, the optimum measurement consists of three symmetric state vectors with the length less than the unity, and has been demonstrated experimentally in Ref.[41]. In the cases of quinary and septenary signals, the optimum strategies consist of three nonorthogonal state vectors with different lengths. In the septenary case, there are two different configurations of measurement state vectors. We study how each of these strategies work and the extent to which they allow us to access the theoretical maximum amount of mutual information.

3 Real symmetric qubit sets and optimum detection

Let $\{|\leftrightarrow\rangle, |\updownarrow\rangle\}$ be the orthogonal basis of linear polarizationstates of a single photon. Then the real symmetric qubit states are defined as

$$|\psi_i\rangle = \cos\frac{i\pi}{M}|\leftrightarrow\rangle + \sin\frac{i\pi}{M}|\updownarrow\rangle \quad (i = 0, \ldots, M-1). \tag{12}$$

We assume that each state is selected with equal prior probability $1/M$. This set is one of the few quantum state sets for which optimum strategies for the accessible information are explicitly known[33]~[36].

For $M > 2$, the signal states cannot be distinguished perfectly, thus $P_i = 1$. The minimum average error probability is

$$P_e = 1 - \frac{2}{M}, \tag{13}$$

Table 1 *The channel matrix of the optimum POM for the ternary signals*

	$\|\psi_0\rangle$	$\|\psi_1\rangle$	$\|\psi_2\rangle$
$\|\omega_0\rangle$	0	0.5	0.5
$\|\omega_1\rangle$	0.5	0	0.5
$\|\omega_2\rangle$	0.5	0.5	0

which is attained by the POM $\{\hat{\Pi}_j\}$[3][31]

$$\hat{\Pi}_j = |a_j\rangle\langle a_j| \quad \text{with} \tag{14}$$

$$|a_j\rangle = \sqrt{\frac{2}{M}}\left(\cos\frac{j\pi}{M}|\leftrightarrow\rangle + \sin\frac{j\pi}{M}|\updownarrow\rangle\right) \quad (j = 0, \ldots, M-1). \tag{15}$$

This POM is unique in leading to the minimum error probability and has the same number of POM elements, corresponding to the measurement outcomes, as the letter states.

In contrast, maximizing the mutual information requires a POM with three rank-one elements at most, corresponding to just three measurement outcomes[36]. Although it is also possible to construct optimum POMs with elements more than three, a strategy with minimum outputs is often the one desired in practice.

If M is even, a von Neumann measurement, i.e. a pair of orthogonal projectors, can

be the optimum strategy with minimum outputs. If M is odd, then at least three outputs are required and a standard von Neumann measurement fails in maximizing the mutual information. The three rank-one elements required for the optimum POM $\{\hat{\Pi}_j\}$ are specified as follows:

$$\hat{\Pi}_j = |\omega_j\rangle\langle\omega_j| \tag{16}$$

$$\text{with} \quad \begin{cases} |\omega_0\rangle = -\sin\frac{\gamma}{2}|\updownarrow\rangle \\ |\omega_1\rangle = \frac{1}{\sqrt{2}}\left(-|\leftrightarrow\rangle + \cos\frac{\gamma}{2}|\updownarrow\rangle\right) \\ |\omega_2\rangle = \frac{1}{\sqrt{2}}\left(|\leftrightarrow\rangle + \cos\frac{\gamma}{2}|\updownarrow\rangle\right) \end{cases} \tag{17}$$

where γ is determined from

$$\cos\frac{\gamma}{2} \equiv \cot\frac{m\pi}{M}, \quad \sin\frac{\gamma}{2} \equiv -\sqrt{1-\cot^2\frac{m\pi}{M}} \tag{18}$$

for an integer parameter m within the range $\frac{M}{4} < m < \frac{M}{2}$. We will refer to the unnormalized vectors given in Eq. (17) as measurement state-vectors.

In the case of M = 3 (ternary or trine), the optimum POM is given by m = 1 which results in the set of three measurement state-vectors with equal norms. The signal and measurement state-vectors are schematically shown in Fig.1. In this figure, each arrow represents the polarization direction where the horizontal and the vertical directions correspond to the two unit bases $|\leftrightarrow\rangle$ and $|\updownarrow\rangle$, respectively. The length of each arrow represents the norm of the associated state vector, e.g. $|\psi_i\rangle$ or $|\omega_j\rangle$.

The optimum measurement in this case means that the state vectors $|\psi_j\rangle$ and $|\omega_i\rangle$ are orthogonal, and thus

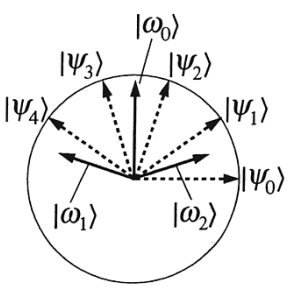


Fig.2 *The measurement state vectors for the optimum strategy (solid line) and the signal state vectors in the case of the quinary signals (M = 5)*

Table 2 *The channel matrix of the optimum POM for the quinary signals*

	$\|\psi_0\rangle$	$\|\psi_1\rangle$	$\|\psi_2\rangle$	$\|\psi_3\rangle$	$\|\psi_4\rangle$
$\|\omega_0\rangle$	0	0.309	0.809	0.809	0.309
$\|\omega_1\rangle$	0.5	0.191	0	0.191	0.5
$\|\omega_2\rangle$	0.5	0.5	0.191	0	0.191

$$P(y_j|x_j) = \langle\psi_j|\hat{\Pi}_j|\psi_j\rangle = 0. \tag{19}$$

The other two possible measurement outcomes occur with equal probabilities. This situation is summarized in Table 1.

In the cases of M = 5 (quinary) and M = 7 (septenary), Eq. (17) results in the three measurement statevector with two distinct norms. The relationship between the quinary letter states and the three measurement state-vectors (with m = 2) is depicted in Fig.2. The channel matrix in this case is summarized in Table 2. In the septenary case, there are two different POMs with three elements given by Eq. (17), with m = 2 and m = 3 in Eq. (18) respectively. They are depicted in Fig. 3 and summarized in Tables 3 and 4. In either case, there are combinations of (i,j) that give $P(y_j|x_j) = 0$, although j is not necessarily equal to i (a difference from the ternary case).

The method to implement the optimum POM with minimum outputs, as given in Eq. (17), is prescribed in detail in Ref.[36]. In short, the nonorthogonal measurement basis $\{|\omega_j\rangle\}$ is considered as the projection of a three-dimensional orthonormal basis in an enlarged space. Such an enlarged space is achieved by introducing another independent binary basis.

In practice, the concept described above is realized as the polarization Mach–Zehnder interferometer shown in Fig.4. The four-dimensional space is composed of $\{|\leftrightarrow\rangle_a, |\updownarrow\rangle_a, |\leftrightarrow\rangle_b, |\updownarrow\rangle_b\}$, where subscripts represent the optical paths (a, b) indicated in Fig.4. Our letter states present in the subspace spanned by the first two of these vectors. The additional port (at b in Fig.4) with an input of vacuum state $|0\rangle$ enlarges the space.

The unitary operation of the Mach–Zehnder part (indicated as $\hat{U}$ in Fig.4) can be written as

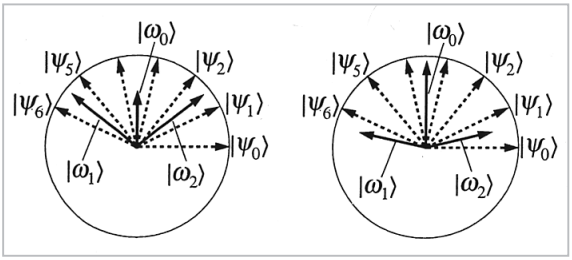


Fig.3 *The two optimum strategies in the case of the septenary signals (M = 7)*

The left figure corresponds to the choice m = 2, while the right one corresponds to the other choice m = 3.

Table 3 *The channel matrix of the optimum POM with m=2 (see Eq. (18)) for the septenary signals*

	$\|\psi_0\rangle$	$\|\psi_1\rangle$	$\|\psi_2\rangle$	$\|\psi_3\rangle$	$\|\psi_4\rangle$	$\|\psi_5\rangle$	$\|\psi_6\rangle$
$\|\omega_0\rangle$	0	0.069	0.223	0.346	0.346	0.223	0.069
$\|\omega_1\rangle$	0.5	0.154	0	0.154	0.5	0.777	0.777
$\|\omega_2\rangle$	0.5	0.777	0.777	0.5	0.154	0	0.154

Table 4 *The channel matrix of the optimum POM with m=3 (see Eq. (18)) for the septenary signals*

	$\|\psi_0\rangle$	$\|\psi_1\rangle$	$\|\psi_2\rangle$	$\|\psi_3\rangle$	$\|\psi_4\rangle$	$\|\psi_5\rangle$	$\|\psi_6\rangle$
$\|\omega_0\rangle$	0	0.178	0.579	0.901	0.901	0.579	0.178
$\|\omega_1\rangle$	0.5	0.322	0.099	0	0.099	0.322	0.5
$\|\omega_2\rangle$	0.5	0.5	0.322	0.099	0	0.099	0.322

$$A'_{\mathrm{H}}|\leftrightarrow\rangle_a + A'_{\mathrm{V}}|\updownarrow\rangle_a + B'_{\mathrm{H}}|\leftrightarrow\rangle_b + B'_{\mathrm{V}}|\updownarrow\rangle_b$$
$$= \hat{U}\left(A_{\mathrm{H}}|\leftrightarrow\rangle_a + A_{\mathrm{V}}|\updownarrow\rangle_a + B_{\mathrm{H}}|\leftrightarrow\rangle_b + B_{\mathrm{V}}|\updownarrow\rangle_b\right) \quad (20)$$

$$\text{with} \quad \begin{bmatrix} A'_{\mathrm{H}} \\ A'_{\mathrm{V}} \\ B'_{\mathrm{H}} \\ B'_{\mathrm{V}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\gamma/2 & \sin\gamma/2 & 0 \\ 0 & -\sin\gamma/2 & \cos\gamma/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A_{\mathrm{H}} \\ A_{\mathrm{V}} \\ B_{\mathrm{H}} \\ B_{\mathrm{V}} \end{bmatrix}$$

where $\gamma/2$ is twice the angle of HWP1. (This $\gamma/2$ represents the angle of one of the unit basis in the enlarged space relative to the signal plane.) In our setup, the inputs are $B_{\mathrm{H}} = B_{\mathrm{V}} = 0$ and hence $B'_{\mathrm{V}} = 0$. Thus the apparatus of Fig.4 actually couples a three-dimensional state space.

PD0 detects $|\leftrightarrow\rangle_b$ components whose amplitude is given by

$$B'_{\mathrm{H}} = -\sin(\gamma/2)A_{\mathrm{V}}\,. \quad (21)$$

Its null result guarantees that the signal was not $|\psi_o\rangle$. On the other hand, $|\leftrightarrow\rangle_a$ and $|\updownarrow\rangle_a$ components are further mixed at HWP2 and

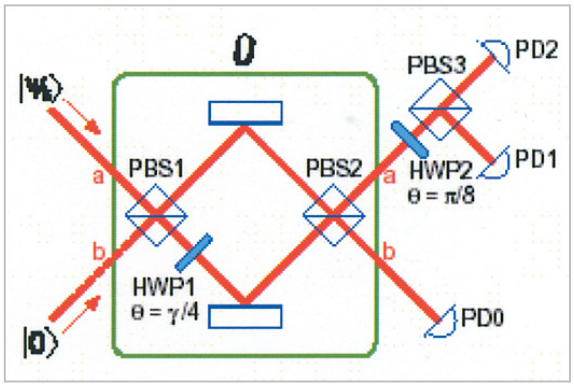


Fig.4 *Principle of the detector that realizes the optimum POM*

Here PBS stands for a polarizing beam splitter, HWP for a half waveplate whose axis is rotated by θ, and PD for a photodetector.

PBS3, resulting in amplitudes of

$$\frac{1}{\sqrt{2}}(A'_{\mathrm{H}} \pm A'_{\mathrm{V}}) = \frac{1}{\sqrt{2}}\left[A_{\mathrm{H}} \pm \cos\frac{\gamma}{2}A_{\mathrm{V}}\right] \quad (22)$$

which are then detected at PD1 and PD2. By inspecting Eqs. (21) and (22), it can be seen that $|\omega_j\rangle$ given in Eq. (17) were reproduced. When the condition Eq. (18) is satisfied, the null result at PD1 or PD2 excludes one of the possible signals ($|\psi_{k_\pm}\rangle$ with $k_+ = \mathrm{M}-m$ and $k_- = m$).

4 Experiment

The principle described in the previous section is realized in an actual setup to confirm the theoretical results. In the experiment, the polarization basis $\{|\leftrightarrow\rangle, |\updownarrow\rangle\}$ correspond to P- (within the paper plane in Fig.5) and S- (perpendicular to the paper plane) polarizations, respectively.

The light source is a He–Ne laser (Spectra–Physics, model 117A) operating at the wavelength of 632.8 nm. The laser light of 1mW is first attenuated by the attenuator ATN1 by a factor of 10^{-6}, purified to the horizontally polarized state by the polarizing beam splitter PBS0. The half waveplate HWP0, driven by a stepping motor, works as a modulator to produce the set $\{|\psi_i\rangle\}$. Then the beam is further attenuated by ATN2 by a factor of 10^{-4}. At the input of the Mach–Zehnder inter-

ferometer, the light power is of order 100fW ($\approx 3 \cdot 10^5$ photon/sec). In other words, the beam contains about 10^{-3} photons in one meter, whereas our detecting circuit is shorter than that.

The polarization Mach–Zehnder interferometer is composed of two PBSs, PBS1 and PBS2. Each PBS is carefully mounted so as to operate with an extinction ratio of 1 : 1000 (see below and Ref.[44]). Each path of the Mach–Zehnder contains one half waveplate, HWP1 and HWP1 ′. The angle of HWP1 is adjusted to a quarter of γ in Eq. (18) so that the polarization of the light is rotated by $\gamma/2$, whereas HWP1 ′is inserted for symmetry and thus adjusted not to affect the polarization state.

The beams from the two paths are superimposed at PBS2, resulting in two output beams from the Mach–Zehnder. The one corresponds to path *b* in Fig.4 is detected directly at Port 0. The beam in path a in Fig.4 is delivered to HWP2 at an angle of π /=8 and then to PBS3, in order to visualize the interference of the beams from the two paths. The two outputs from PBS3 are detected at Ports 1 and 2.

The relative path length of the Mach–Zehnder is adjusted to be a proper operating point (which is the minimum at either of Port 1 or 2) by a PZT actuator through a feedback system utilizing the modulation-demodulation method. Once the relative path length is adjusted, a sample-and-hold circuit keeps the mirror position fixed during a measurement sequence (see below) which lasts typically 20–30 seconds.

There are two photodetectors at each port, a silicon photodiode and an APD (avalanche photodiode, EG & G, SPCM–AQ–141–FC) guided through a multimode optical fiber. The former is for alignment purpose (with increased light) and the photon counting process is carried out with the latter, by mechanically switching the beam between them. The coupling efficiency of the fiber is measured to be 0.75–0.8, including the coupling lens and the connectors before the APD. The output from each APD is sent to a pulse counter (EG & G ORTEC, model 995) to count the number of photon-induced pulses.

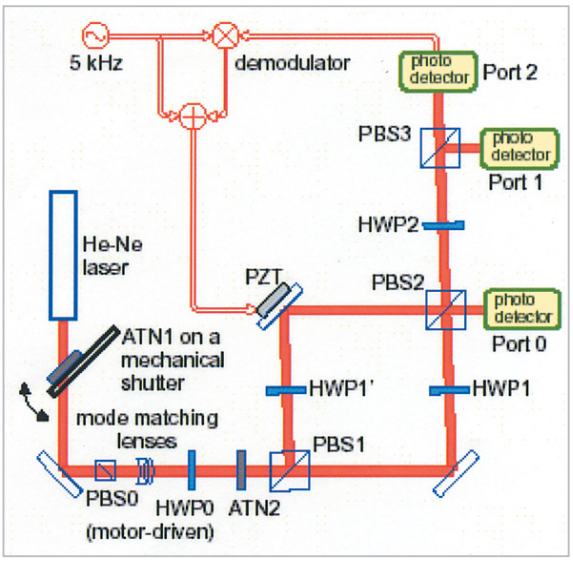


Fig.5 *Experimental configuration*

The same symbols as in Fig.4 and ATN for an attenuator are used. Each of Ports 0, 1, and 2 contains an APD and a silicon photodiode with a mechanical shutter to switch the beam between them. All PBS are adjusted for the maximum separation of two polarization, resulting in a slightly (≈ 0.02 rad) slanted parallelogram arrangement for the Mach–Zehnder.

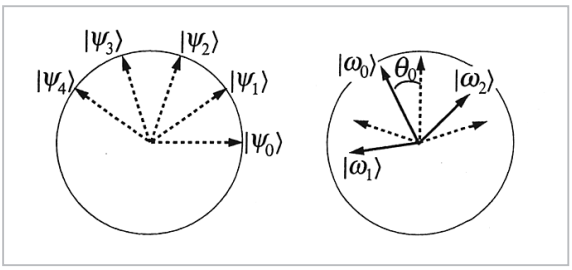


Fig.6 *The relation between the measurement state vectors (left, quinary case M = 5 in this example) and the signal state vectors (right) with an initial offset angle* θ_0

The counters are activated simultaneously by a common trigger, typically of one-second duration and fivetime repetition. The numbers of counts in each duration are read by a computer from all counters, so that we can analyze statistical errors. This procedure is repeated for each signal $|\psi_i\rangle$ with i = 0, ..., M–1, composing a full sequence of measuring the mutual information. The ratio of counts in the three APDs provides the channel matrix $P(y_j|$

x_i) from which the mutual information is derived.

As is discussed in Section **3**, in the optimum detection scheme proposed, the mutual information is increased by excluding one of the possible signals. Thus, realizing zero probabilities at the output ports is essential in achieving a high mutual information. In practice, however, there are several causes that increase the probability at the output where ideally zero is expected. Among them, the most pronounced ones are the pulses from an APD without any light (APD error), the finite extinction ratio of a PBS (PBS error), and the finite contrast of interference (interferometer error).

Without any light at all, the average dark counts of the APDs were measured to be slightly less than 100 count/sec. Although the whole interferometer is enclosed in a box, the environmental light increases the number of counts to around 300 count/sec, even if no laser light is injected. When the laser light is injected, the leak light due to the imperfection of the interferometer is added, and was measured to the average count of around 1000 count/sec for the output port at which no count is expected ideally (see Tables 1–4). The last increment is considered as the contributions from the PBS errors and the interferometer error. At the ports for which finite counts are expected, we had the counts of order 10^5 count/sec at most, which is within the linear range of APDs.

In general, a PBS has an angular-dependent separation of two polarization components. In our case, it turned out to be possible to achieve the separation better than 1 : 1000 for both polarization components, by carefully aligning the angle of incidence slightly (≈ 0.02 rad) different from the standard value $\frac{\pi}{4}$. Then the expected contrast is ≈ 0.998, which we thought sufficient for our experiment. We adopted this angle in our polarization Mach–Zehnder interferometer, resulting in a parallelogram arrangement (see Fig.5).

The actual contrast obtained with this interferometer can be as high as

$$\frac{P_{\max} - P_{\min}}{P_{\max} + P_{\min}} \approx 0.98\,, \tag{23}$$

though the typical values under normal experimental conditions were slightly lower than this. Thus, this is limited not by the PBS imperfection but by, e.g., the spatial mode mismatch of the two beams.

In order to analyze the performance of our detecter circuit, we measured not only the mutual information of the optimum detection scheme but also its dependence on the relative angle between the signal set $\{|\psi_i\rangle\}$ and the measurement state vectors $\{|\omega_j\rangle\}$. This is relevant to, for example, the possible rotation of polarization in the transmitting fiber. We measured the mutual information against the signal set $\{|\psi_i'(\theta_0)\rangle\}$ where

$$|\psi_i'(\theta_0)\rangle = \cos\left(\frac{i\pi}{M} + \theta_0\right)|\leftrightarrow\rangle + \sin\left(\frac{i\pi}{M} + \theta_0\right)|\updownarrow\rangle \quad (i = 0, \ldots, M-1), \tag{24}$$

as a function of the initial offset angle θ_0 (the optimum detection corresponds to $\theta_0 = 0$). The relation between $|\psi_i'(\theta_0)\rangle$ and $|\omega_j\rangle$ is depicted in Fig.6 for the case of quinary signals. In the experiment, θ_0 was changed in steps of $\frac{\pi}{90}$ radian (two degrees).

5 Results

We carried out the optimum measurements described in Section **3** for the sources comprising the ternary (trine), quinary and septenary states. For the septenary signal states, both of the two optimum detection schemes (with $m = 2$ and $m = 3$ in Eq. (18)) were tested.

Fig.7 shows the relative output counts at the three detectors as the polarization of the input light is varied in the ternary case. This relative power corresponds to the probability for the measurement outcome to occur for a single input photon.

For the polarization angles $\left\{-\frac{\pi}{6}, \frac{\pi}{6}, \frac{\pi}{2}\right\}$ we are performing the state discrimination with the minimum error probability, while for the angles $\left\{-\frac{\pi}{3}, 0, \frac{\pi}{3}\right\}$ we are realizing a

measurement that allows unambiguous elimination of one possibility among the three letter states. These measurements were referred to as the trine and anti-trine measurements in Ref.[41]. These authors found an *rms* deviation of 3.8% from the theoretical value given in Table 1. Our results indicate a lower value of 1.1 %. The reason for our lower value is that we have been able to achieve a smaller PBS error.

The data depicted in Fig.7 leads to the mutual information presented in Fig.8. At the optimum operating point, corresponding to the best detection strategy, we clearly find that the mutual information exceeds that attainable with the best von Neumann measurement. Our value also exceeds that obtained earlier by Clarke *et al.*[41] represented as triangles in our figure. The reason for this is again the smaller PBS error. Our experimental value is slightly lower than the theoretical maximum and this is due mainly to a residual PBS error of approximately 0.1% and also to the imperfect contrast of interference. It was found[44] that despite the PBS error is not the limiting factor of the interference contrast, it has non-negligible effects on the mutual information.

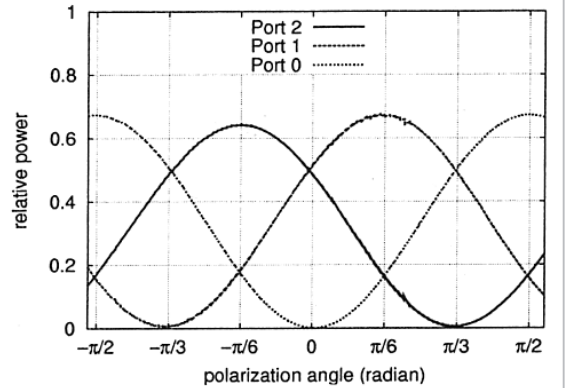


Fig.7 *The dependence of the relative outputs at the three APDs on the polarization angle of the injected beam in the ternary experiment*

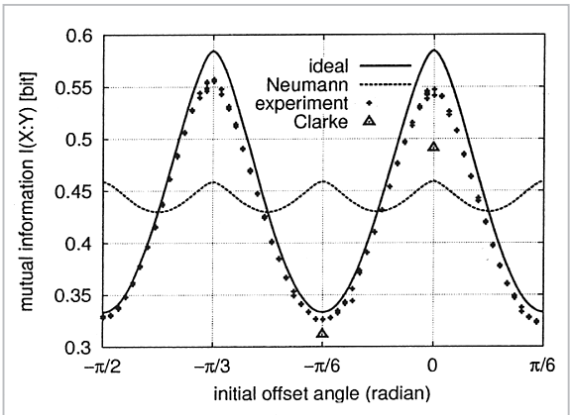


Fig.8 *The dependence of the mutual information on the initial offset angle θ_0 in the ternary experiment ("experiment", pluses)*

The ideal case ("ideal", solid curve) and the ideal von Neumann case ("Neumann", dashed curve) are shown for comparison. The values in an earlier experiment[41] ("Clarke", triangles at $\theta_0 = 0$ and $-\pi/6$) are also shown.

Fig.9 shows the relative output counts at our three detectors for the quinary case. These provide the data with which to calculate the mutual information depicted in Fig.10. Our data show a marginal increase in the mutual information beyond the value that may be attained with the best von Neumann measurement. The diffierence between our experimental result and the theoretical value is again principally attributable to the PBS error and the imperfect contrast.

As mentioned earlier, the optimum detection scheme increases the amount of the mutual information by excluding one of the possible signals. With three detectors, only three signals can be excluded at most, and the remaining signals do not contribute the mutual information very much. This fact reduces the maximum mutual information in quinary case (and in septenary case as well) from that in ternary case. Although the absolute difference (of ≈ 0.02) between the experimental and ideal values in the quinary case is similar to that in the ternary case, the excess from the von Neumann measurement became only marginal.

Fig.11 shows the mutual informations derived with the two possible optimum detection schemes for the septenary case. Even in an ideal case, the increase in the attainable mutual information over that found using the best von Neumann measurement is quite small. In both cases our experimental values failed to reach even the value attainable by means of the best von Neumann measurement.

The result with m = 3 shows a higher

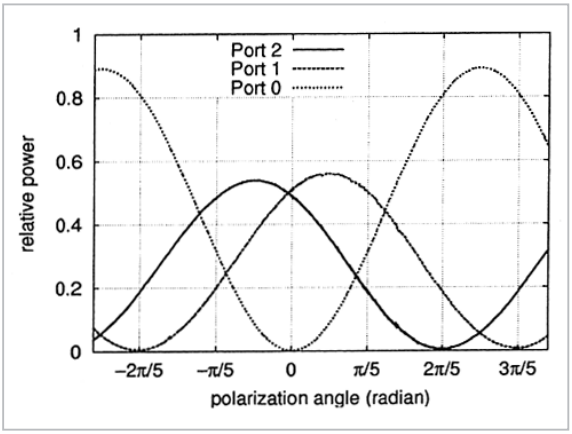


Fig.9 *The dependence of the relative outputs at the three APDs on the polarization angle of the injected beam in the quinary experiment*

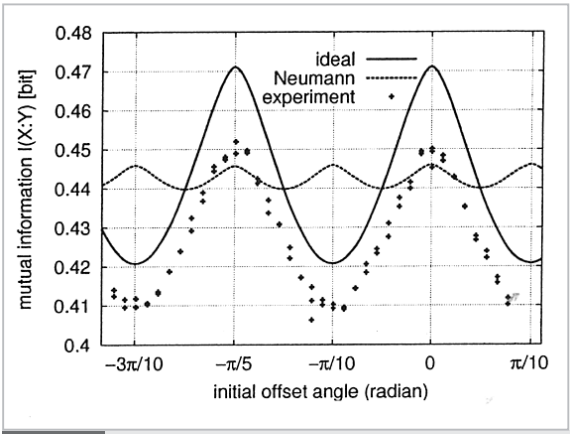


Fig.10 *The dependence of the mutual information on the initial offset angle θ_0 in the quinary experiment*

The symbols are the same as in Fig.8.

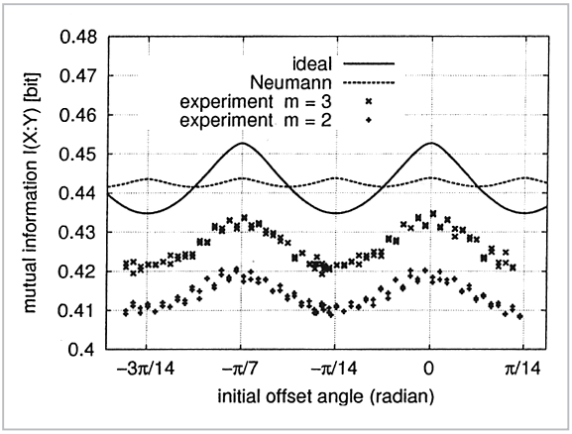


Fig.11 *The dependence of the mutual information on the initial offset angle θ_0 in the septenary experiment with m = 2 (crosses) and m = 3 (pluses). Other symbols are the same as in Fig.8*

mutual information than that with $m = 2$. This difference is not by an experimental failure, but due to the difference in the influences of inperfect contrast between the two cases. The reduced contrast increases the light leaking towards the port where ideally no light is expected, which in turn reduces the mutual information. The absolute amount of leak light is proportional to the amount of light interfering. This qualitatively explains the difference of experimental results with $m = 2$ and $m = 3$. In the former case the interfering light is greater than the latter, thus the influence on the mutual information is larger.

6 Discussion and Concluding Remarks

Our ability to communicate classical information by means of a quantum channel is limited by the existence of non-orthogonal quantum states and the associated restrictions in discriminating among them. These factors are fundamental to quanta as distinct from classical information theory and make quantum key distribution possible[42][43].

The optimum use of a quantum communication channel is closely related to the maximization of mutual information, as discussed in Appendix. The accessible information is obtained by maximizing the mutual information through the selection of the detection process. There are only a very few examples of signal states for which the accessible information is known[33]~[36]. One such example is that of the real symmetric qubit states[36].

In this paper we have described our polarization Mach–Zehnder interferometer that was designed to extract the accessible information from signals formed from symmetric polarization states. For the ternary (trine) states, our results proved an amount of information close (96 %) to the theoretical limit. Our value for the mutual information exceeds that reported in an earlier experiment[41]. The difference between our measured value for the mutual information and the theoretical limit is due principally to the leakage of the 'wrong' polarization through our polarizing beam splitters and also to the imperfect contrast. The effect

of this leakage is more pronounced when we consider the quinary and septenary signal states. Our experiments suggest that optimum quantum communication based on the ternary (trine) polarization states, for example the quantum key distribution by the Phoenix–Barnett–Chefles protocol[43], should be feasible. Schemes based on the quinary and septenary states will present a greater challenge.

In the light of fundamental interests, the quinary and septenary states meet with the simplest cases where the maximum amount of information can be extracted by a detection in which the number of possible outputs is less than that of input states. Davies' theorem predicted that a device with three possible outputs suffices for any real polarization system of a single photon. In our experiment, Davies' theorem has been tested within the PBS error. For the complete confirmation, further study might be necessary, e.g. comparing the minimum-output optimum detection with the one corresponding the group covariant optimal solution which consists of the same number of outputs as inputs.

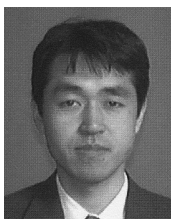
Acknowledgments

We are grateful to Dr. Izutsu, Professor Hirota, Dr. Riis, and Dr. Clarke for discussion and encouragement. This work was supported, in part, by the British Council, the Royal Society of Edinburgh, and by the Scottish Executive Education and Lifelong Learning Department.

References

1 W. K. Wootters and W.H. Zurek, Nature 299, 802 (1982).
2 H. P. Yuen, Phys. Lett. A, 113, 405 (1986).
3 C.W. Helstrom, Quantum Detection and Estimation Theory (Academic Press, New York, 1976).
4 O. Hirota : Theory of Optical Communication in Japanese (Morikita, Tokyo, 1985).
5 C. H. Bennett and P. Shor, IEEE Trans. Inform. Theory, IT-44, No. 6, pp2724-2742 (1998).
6 M. Sasaki and M. Ban, Reviews, BUTSURI, 57(1) Jan. (2002).
7 C.E. Shannon, Bell System Tech. J. 27, 379 (PartⅠ) and 623 (PartⅡ) (1948).
8 R.G. Gallager: Information Theory and Reliable Communication (John Wiley and Sons, New York, 1968).
9 T. Cover and J. Thomas: Elements of Information Theory (John Wiley and Sons, New York, 1991).
10 A.S. Holevo : Probabilistic and Statistical Aspects of Quantum Theory (North-Holland, Amsterdam, 1982).
11 A. Peres: Quantum Theory: concepts and methods, 279 (Kluwer Academic Publishers, Dortrecht, 1993).
12 P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W.K. Wootters, Phys. Rev. A54, 1869 (1996).
13 A.S. Holevo, IEEE Trans. Inf. Theory IT-44, 269 (1998).
14 B. Schumacher and M. Westmoreland, Phys. Rev. A56, 131 (1997).
15 M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, Phys. Rev. A58, 146 (1998).
16 M. Sasaki, T.S. Usuda, M. Izutsu, and O. Hirota, Phys. Rev. A58, 159 (1998).
17 A.S. Holevo, J. Multivar. Anal. 3, 337 (1973).
18 H.P. Yuen, R.S. Kennedy, and M. Lax, IEEE Trans. Inf. Theory 21(2), 125 (1975).
19 I.D. Ivanovic, Phys. Lett. A123, 257 (1987).
20 D. Dieks, Phys. Lett. A126, 303 (1988).

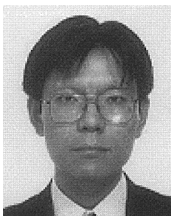
21 A. Peres, Phys. Lett. A128, 19 (1988).
22 G. Jaeger and A. Shimony, Phys. Lett. A197, 83 (1995).
23 S.M. Barnett, Phil. Trans. R. Soc. Lond. A255, 2279 (1997).
24 A. Chefles and S.M. Barnett, J. Mod. Opt., 45, 1295 (1998).
25 A. Chefles, Phys. Lett. A239, 339 (1998).
26 A. Chefles, Contemp. Phys. 41, 401 (2000) and references therein.
27 B. Huttner and A. Peres, J. Mod. Opt. 41, 2397 (1994).
28 B. Schumacher, Phys. Rev. A51, 2738 (1995).
29 S.M. Barnett, C.R. Gilson, and M. Sasaki, J. Phys. A: Math. Gen. (in press).
30 M. Osaki, M. Ban, and O. Hirota, Phys. Rev. A54, 1691 (1996).
31 M. Ban, K. Kurokawa, R. Momose, and O. Hirota, Inter. J. Theor. Phys. 36, 1269 (1997).
32 S.M. Barnett, Phys. Rev. A 64, 030303(R) (2001).
33 L.B. Levitin, Quantum Communication, and Measurement (Eds. V.P. Belavkin, O. Hirota, and R.L. Hudson, Prenum, New York, 1995), 439.
34 M. Osaki, M. Ban, and O. Hirota, Quantum Communication, Computing, and Measurement 2 (Eds. P. Kumar, G M. D'fAriano, and O. Hirota, Kluwer academic/Prenum publishers, New York, 2000) 17.
35 E.B. Davies, IEEE Trans. Inf. Theory IT-24, 596 (1978).
36 M. Sasaki, S.M. Barnett, R. Jozsa, M. Osaki, and O. Hirota, Phys. Rev. A59, 3325 (1999).
37 P.W. Shor, "On the Number of Elements Needed in a POVM Attaining the Accessible Information", to appear in Quantum Communication, Computing, and Measurement 3 (Eds. O. Hirota and P. Tombesi, Kluwer, Dordrecht, 2001). Also available as ArXiv:quant-ph/0009077.
38 S.M. Barnett and E. Riis, J. Mod. Opt. 44, 1061 (1997).
39 B. Huttner, A. Muller, J.D. Gautier, H. Zbinden, and N. Gisin, Phys. Rev. A54, 3783 (1996).
40 R.B.M. Clarke, A. Chefles, S.M. Barnett, and E. Riis, Phys. Rev. A63, 040305 (2001).
41 R.B.M. Clarke, V.M. Kendon, A. Chefles, S.M. Barnett, E. Riis, and M. Sasaki, Phys. Rev. A64, 012303 (2001).
42 S.J.D. Phoenix and P.D. Townsend, Comtemp. Phys. 36, 165 (1995) and references therein.
43 S.J.D. Phoenix, S.M. Barnett, and A. Chefles, J. Mod. Opt. 47, 507 (2000).
44 J. Mizuno et al., paper in preparation (2001).



***Masahide SASAKI**, Ph. D.*

Leader, Quantum Information Technology Group, Basic and Advanced Research Division

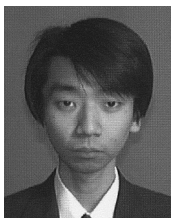
Quantum information theory



***Jun MIZUNO**, Ph. D.*

Research Fellow, Quantum Information Technology Group, Basic and Advanced Research Division

Optical sensing



***Mikio FUJIWARA**, Ph. D.*

Senior Researcher, Quantum Information Technology Group, Basic and Advanced Research Division

Photodetection technology