Theoretical Studies on the Performance of Lossy Photon Channels

FUMIO KANAYA, MEMBER, IEEE, AND KENJI NAKAGAWA, MEMBER, IEEE

Abstract—**Since light in a number state is free from inherent quantum statistical fluctuations associated with an ideal photon counting measurement, it is promising as an information bearing light in direct-detection optical communication systems. However, after undergoing independent random photon deletions due to the lossy effect of the channel, even number-state light gives rise to a random photon counting process at the output of a perfectly performing photon counting device. Therefore, even in noiseless optical channels, the lossy effect must be taken into account as an important degradation factor which limits the channel performance.**

The objective of this paper is first to study how channel losses affect the capacity of noiseless optical communication systems utilizing the number state; then, based on these findings, to evaluate photon efficiency of these systems in the presence of losses. Then, by comparing the photon efficiency of the number-state PPM and the coherent-state PPM under certain realistic conditions, it is found that the former is always superior to the latter from the perspective of photon efficiency regardless of losses.

I. INTRODUCTION

IT is a familiar fact that light emitted by a laser operating in a pure coherent state exhibits quantum mechanical photon-flux fluctuations. In other words, the number of photons contained in one propagation mode or light pulse must be considered to be a random variable whose statistical distribution is Poisson. This results in the possibility of nonzero decoding error, even in a noiseless optical channel where the background optical noise and thermal noise of the receiving system are negligibly small.

Recently, considerable interest has emerged in quantum states of light other than the coherent state. One example of such a quantum state is the squeezed state (or two-photon coherent state), whose quantum-mechanical properties have been extensively studied theoretically (see, e.g., [1]) and generation has been observed experimentally [2]. Applied to optical signals, this state features reduced quadrature amplitude noise and optical communications systems which exploit the properties of squeezed light have already been proposed [3], [4].

Another example of a nonclassical light field is a number state, whose generation has been the target of a recent study [5]. This state features zero uncertainty for the photon number, that is, the number of photons contained in one light pulse is a single, well-defined value rather than a Poisson distributed random variable. Since there can be no quantum statistical fluctuation present in the signal at the output of a photon counting device, an optical communications system utilizing direct photon detection could conceivably benefit from the unique properties of the number state. This concept has been confirmed by computing the channel capacity [6] for a lossless optical channel using the number state where every photon of the transmitted light beam reaches the photon detector and properly adds to the photon count.

However, if the channel is lossy, the benefit of utilizing the number state is thought to be lessened because after undergoing independent random photon deletions due to the lossy effect of the channel, even the number state light would yield a random photon counting process at the output of the photon detector, resulting quite naturally in a reduced channel capacity. Therefore, even in noiseless optical channels, transmission loss must be taken into account as an important degradation factor. Thus, it is essential in a lossy optical communications context to understand how loss affects the performance of an optical communications system utilizing the number state.

In this paper, first the capacity of an optical channel with loss is calculated for a number of modulation schemes utilizing the number state, and closed-form capacity formulae are obtained in terms of the transmittance coefficient. Then based on these formulae the photon efficiency is derived, which is defined as the capacity per transmitted photon. This is optimized for a PPM scheme subject to the constraint that both the channel capacity per second and the channel bandwidth are fixed at certain values. Then, the optimized photon efficiency is compared between the number state and the coherent state PPM scheme. From this it is concluded that the photon efficiency in the number state is always superior to that in the coherent state regardless of the transmittance coefficient.

II. CALCULATION OF CHANNEL CAPACITY

In order to characterize the channel matrix of various photon channels, it is first necessary to refer to the problem of photon counting statistics.

Let the lossy effect of the optical channel be represented in terms of real-valued transmittance coefficient $\eta \in [0, 1]$. Then, the probability that the number state of m photons transmitted over this lossy channel produces an output of n photons to the receiver is given by a familiar binomial distribution [7]

$$p(n)=\binom{m}{n}\eta^n(1-\eta)^{m-n} \quad n=0, 1, \cdots, m. \tag{1}$$

Concerning the above equation, it is noted that since photon counting statistics specific to lossy photon channels utilizing the number state are not Poisson, any capacity formula ever derived for Poisson-type channels (see, e.g., [13]) is not directly applicable to the former case.

In order to focus on the lossy effect of the channel, it is hereafter assumed that the background radiation noise and the thermal noise of the receiving system are negligibly small and that photon counting is executed perfectly.

A. PAM Scheme

In the M-ary PAM scheme, M different number states 0, 1, 2, $\cdots$, $M-1$ are prepared at the sending side for optical pulse transmission, and at the receiving side the actually transmitted state is estimated based on direct photon counting.

Paper approved by the Editor for Modulation Theory and Nonlinear Channels of the IEEE Communications Society. Manuscript received February 23, 1987; revised November 1, 1987. This paper was presented at the 9th Symposium on Information Theory and Its Applications, Akakura, Japan, October 29–31, 1986 (in Japanese).

The authors are with NTT Laboratories, Yokosuka-shi, Kanagawa 238-03, Japan.

IEEE Log Number 8824900.

0090-6778/89/0100-0039$01.00

Thus, from (1) the channel matrix for this scheme can be characterized by the following transition probability:

$$Q(j|i)=\begin{cases}\binom{i}{j}\eta^j(1-\eta)^{i-j}, & i=0, 1, \cdots, M-1 \\ & j=0, 1, \cdots, i \\ 0 & \text{otherwise}\end{cases} \tag{2}$$

In the event that $M = 2$, the channel matrix is

$$Q=\begin{pmatrix}1 & 0\\ 1-\eta & \eta\end{pmatrix} \tag{3}$$

and thus, using a general capacity formula for a 2×2 channel matrix [8], the channel capacity is given by

$$C=\log(1+\exp(-H_1/\eta)) \tag{4}$$

where

$$H_1 \triangleq H(\eta, 1-\eta). \tag{5}$$

Throughout the paper, the channel capacity C is measured by nats per channel use and an ordinary entropy function is denoted as H, that is, for the probability distribution $(q_1, q_2, \cdots, q_n)$

$$H(q_1, q_2, \cdots, q_n) \triangleq -\sum_{i=1}^{n} q_i \log q_i. \tag{6}$$

In the event that $M = 3$, the channel matrix is

$$Q=\begin{pmatrix}1 & 0 & 0\\ 1-\eta & \eta & 0\\ (1-\eta)^2 & 2\eta(1-\eta) & \eta^2\end{pmatrix} \tag{7}$$

and, again using a general capacity formula for a 3×3 channel matrix [8], the channel capacity is given by

$$C=\begin{cases}\log(1+\exp(-H_1/\eta) \\ \quad +\exp(-(H_2-2(1-\eta)H_1)/\eta^2)), & \eta_0 \leqq \eta \leqq 1, \\ \log(1+\exp(-\theta_1)), & 0 \leqq \eta \leqq \eta_0\end{cases} \tag{8}$$

where

$$H_2 \triangleq H((1-\eta)^2, 2\eta(1-\eta), \eta^2), \tag{9}$$

$$\theta_1 \triangleq H((1-\eta)^2, 1-(1-\eta)^2)/(1-(1-\eta)^2), \tag{10}$$

and η_0 is the unique solution to the equation

$$(2-\eta)\log 2+\log(1-\eta)=0. \tag{11}$$

Fig. 1 shows the channel capacity C plotted against transmittance coefficient η where $M = 2$ and 3.

In the event that $M \geqq 4$, it is not a simple task to derive the closed-form expression of C which is valid for all $\eta \in [0, 1]$. However, for η in the vicinity of one or zero, using the geometric method [8], it is possible to obtain

$$C=\begin{cases}\log\left(1+\sum_{k=1}^{M-1}\exp\left(-\sum_{l=1}^{k} w_{kl}H_l\right)\right), & \eta \fallingdotseq 1, \quad (12) \\ \log(1+\exp(-\theta_2)), & \eta \fallingdotseq 0 \quad (13)\end{cases}$$

where the inverse matrix of Q is introduced by $W = (w_{kl})$. H_l is defined as the entropy of the lth row probability vector of Q, i.e.,

$$H_l \triangleq H\left((1-\eta)^l, \binom{l}{1}\eta(1-\eta)^{l-1}, \cdots, \eta^l\right) \tag{14}$$

and θ_2 is defined as

$$\theta_2=H((1-\eta)^{M-1}, 1-(1-\eta)^{M-1})/(1-(1-\eta)^{M-1}). \tag{15}$$

For the details of the calculation, see Appendix A. Fig. 2 shows the channel capacity C plotted against η for $M \geqq 4$. In Fig. 2, the thin curves are obtained by using a well-known numerical capacity computation method [9]. It can be concluded from Fig. 2 that (12) and (13) are valid for a rather wide range of η.

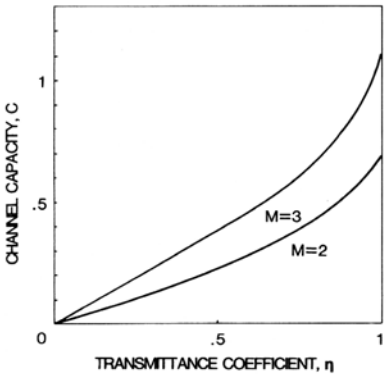

Fig. 1. Channel capacity versus transmittance coefficient for PAM scheme.

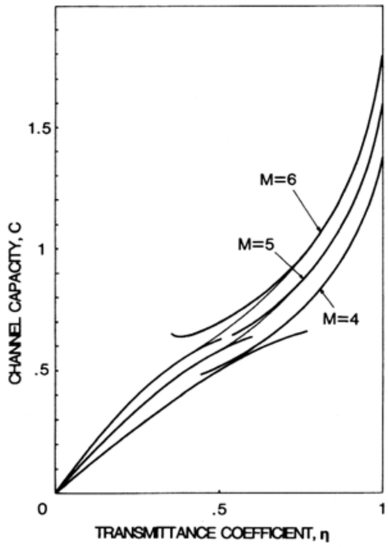

Fig. 2. Channel capacity versus transmittance coefficient for PAM scheme. Thin curves are obtained by Arimoto's numerical capacity computing algorithm [9]. Thick curves are obtained from (12) and (13).

B. PPM Scheme

In the M-ary PPM scheme, a total of M symbols are prepared at the transmitting side. The kth symbol is shown in Fig. 3. Here it is assumed that a number state of m photons is needed to transmit each symbol; that is, exactly m photons are assigned to each optical pulse having a time duration of τ seconds. At the receiving side, the kth symbol is estimated to be actually transmitted only if direct photon counting indicates at least one photon existing at the kth time slot. Therefore, symbol erasure occurs if no photons are detected at the time slot. Thus, the channel matrix for this case is $M \times (M + 1)$

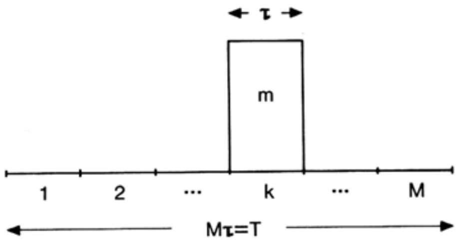

Fig. 3. The kth symbol of PPM scheme.

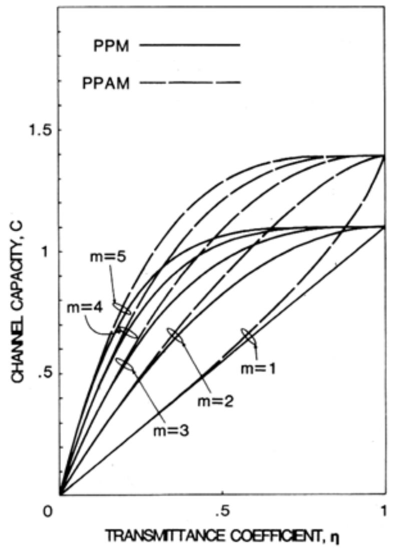

Fig. 4. Channel capacity versus transmittance coefficient for PPM and PPAM schemes where $M = 3$.

and given by

$$Q=\begin{pmatrix} 1-\theta & \theta & 0 & 0 & \cdots & 0 \\ 1-\theta & 0 & \theta & 0 & & 0 \\ & \vdots & & & & \\ 1-\theta & 0 & 0 & 0 & \cdots & \theta \end{pmatrix} \tag{16}$$

where $1 - \theta$ is the symbol erasure probability, i.e.,

$$1-\theta=(1-\eta)^m. \tag{17}$$

This represents a so-called symmetric channel, whose capacity is given by the well-known formula

$$C=\theta \log M. \tag{18}$$

Fig. 4 shows the capacity plotted against η for several values of m with M fixed at 3. It is clear that a larger m is useful in combating loss, if only disregarding a sacrifice in photon efficiency. This issue will be discussed in greater detail later in this paper.

C. PPAM Scheme

The PPAM scheme is a sort of combination of the PPM and PAM schemes. That is, in the PPAM scheme an extra symbol ϕ meaning no photons at all time slots is added to the ordinary M-ary PPM symbols. Thus, the channel matrix in this case is $(M + 1) \times (M + 1)$, and given by

$$Q=\begin{pmatrix} 1 & 0 & 0 & 0 & & 0 \\ 1-\theta & \theta & 0 & 0 & \cdots & 0 \\ 1-\theta & 0 & \theta & 0 & & 0 \\ & \vdots & & & & \\ 1-\theta & 0 & 0 & 0 & \cdots & \theta \end{pmatrix} \tag{19}$$

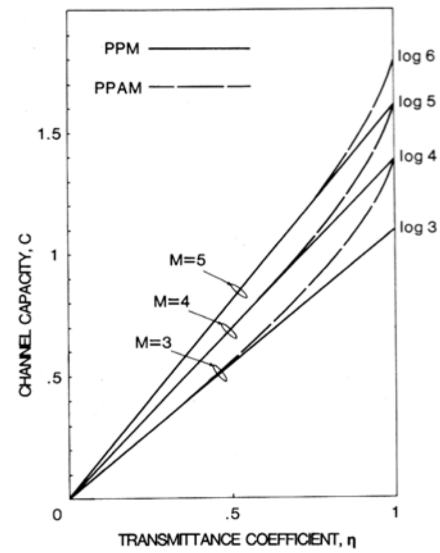

Fig. 5. Channel capacity versus transmittance coefficient for PPM and PPAM schemes where $m = 1$.

where θ is the same as defined in (17). It is obvious that when M is set equal to one, the PPAM scheme can be reduced to the binary PAM. On the other hand, it is assumed that the PPAM scheme begins to look like the PPM scheme as the probability of symbol ϕ becomes smaller. Actually, the channel capacity formulae derived below are something like mix of those of PAM and PPM schemes. For $M = 1$ and 2, utilizing a general capacity formulae for a 2×2 and 3×3 channel matrix, [8] readily yields

$$C=\log(1+M\exp(-\hat{H}/\theta)), \qquad 0 \leqq \theta \leqq 1. \tag{20}$$

For $M \geqq 3$, the geometric computation method in [8] can also be used profitably to yield

$$C=\begin{cases} \log(1+M\exp(-\hat{H}/\theta)), & \theta_0 \leqq \theta \leqq 1 \quad (21) \\ \theta \log M, & 0 \leqq \theta \leqq \theta_0 \quad (22) \end{cases}$$

where

$$\hat{H} \triangleq H(\theta, 1-\theta) \tag{23}$$

and θ_0 is the unique nonzero solution to the equation

$$\theta \log M + \log(1-\theta)=0. \tag{24}$$

For the detailed calculations, see Appendix B. In Fig. 4, the dashed curves show the channel capacity of the PPAM scheme. It is evident that the PPAM scheme is superior to the PPM scheme, especially for η in the vicinity of one. Fig. 5 shows channel capacity curves plotted against η for various M values with photon number m fixed at one.

III. Comparison of Photon Efficiency

As mentioned previously, it is possible to combat the effect of loss and increase channel capacity by increasing the number of photons transmitted per symbol. However, this leads to a decrease in photon efficiency and thus, to excessive consumption of optical power. Here photon efficiency is measured by nats per transmitted photon. That is, let the average number of photon transmitted per symbol be denoted by N; photon efficiency C_p can then be defined as

$$C_p \triangleq C/N \quad \text{(nats/photon)}. \tag{25}$$

Initially, for the sake of simplicity, consider a lossless

TABLE I
MAXIMUM ATTAINABLE PHOTON EFFICIENCY IN LOSSLESS CASE

Alphabet size	C_P nat/photon		
	PAM	PPM	PPAM
2	2log 2	log 2	———
3	log 3	log 3	(3/2)log 3
4	(2/3)log 4	log 4	(4/3)log 4
5	(1/2)log 5	log 5	(5/4)log 5

channel. In this case, the photon efficiency of each modulation scheme is as follows.

For the PAM scheme

$$C_p = (2/m) \log (1+m), \tag{26}$$

for the PPM scheme

$$C_p = (1/m) \log M, \tag{27}$$

and for the PPAM scheme

$$C_p = ((1+M)/mM) \log (1+M) \tag{28}$$

where it is assumed that for each modulation scheme, peak transmitter power is limited to the common value of m photons per symbol and quite naturally each symbol is equiprobable. Table I shows the maximum attainable photon efficiency of each scheme as an alphabet size, that is, the maximum number of transmitted symbols, is varied. It is evident from Table I that 1) among binary modulation schemes, the PAM scheme yields the optimum photon efficiency, 2) for an alphabet size larger than two, the PPAM scheme is always the best with respect to the photon efficiency, and 3) the photon efficiency values for the PPM and PPAM schemes asymptotically become equal as the alphabet size increases. Hereafter, let us focus on the PPM scheme. The objective of the following discussion is to compare the photon efficiency between PPM schemes using the number state and the coherent state.

It has been known for some time that the photon efficiency of the noiseless photon channel without loss can be infinite even for the coherent-state PPM scheme only if either the channel bandwidth is allowed to exponentially approach infinity or the channel capacity per second approaches zero [10]. However, neither of these conditions is realistic. Even in optical communications systems, the bandwidth actually available is limited and the need to realize some given nonzero capacity measured in nats per second using as little optical power as possible is also a real constraint. Therefore, prior to comparing the photon efficiency, the PPM scheme must be optimized for both the coherent state and the number state subject to the constraint that both the channel capacity per second and the channel bandwidth are fixed at certain prescribed values.

For the coherent state PPM scheme, such an optimization has been already carried out for a lossless photon channel in [11]. Thus, by using our notation and taking into consideration lossy effect on the average number of received photons, it follows from [11, eqs. (2) and (3)] that for both the channel capacity per second $\hat{C}_T$ and the photon efficiency $\hat{C}_p$ in the presence of channel losses

$$\hat{C}_T = ((1-\exp(-\eta m))/\tau M) \log M \quad \text{(nats/s)}, \tag{29}$$

$$\hat{C}_P = ((1-\exp(-\eta m))/m) \log M \quad \text{(nats/photon)}. \tag{30}$$

Since the parameter τ appearing in (29) is, as mentioned earlier, the pulse width and thus directly related to the inverse of the channel bandwidth, it is acceptable to consider a constraint on τ instead of on the bandwidth itself. At this point, $\hat{C}_P$ must be maximized, subject to the condition that $\hat{C}_T$ and τ are fixed at specific values c_0 and τ_0, respectively. It can be easily shown that this is equivalent to maximizing $\hat{C}_P$ as expressed by the following function of M:

$$\hat{C}_P = -\eta c_0 \tau_0 M / \log (1 - c_0 \tau_0 M/\log M). \tag{31}$$

Since M is an integer larger than one and is required to satisfy the constraint that the argument of the logarithmic function in the denominator of (31) must be positive, only a finite number of M values are permissible. Thus, it assured that $\hat{C}_P$ actually attains a maximum.

For the number-state PPM scheme, considering (17) and (18), the channel capacity per second C_T and the photon efficiency C_P are given by

$$C_T = ((1-(1-\eta)^m)/\tau M) \log M \quad \text{(nats/s)}, \tag{32}$$

$$C_P = ((1-(1-\eta)^m)/m) \log M \quad \text{(nats/photon)}. \tag{33}$$

Now the problem is to maximize C_P with respect to m and M which can be varied as the optimizing parameters subject to the constraint that $C_T = c_0$ and $\tau = \tau_0$. Because of the additional constraint that both m and M must be positive integers, it is difficult to obtain the analytic solution to this maximization problem. However, since some mathematical consideration ensures that C_P actually attains a maximum at specific finite values of m and M, the numerical computation method can be profitably utilized.

Tables II, III, and IV show a comparison of the maximum attainable photon efficiency between the number state and the coherent-state PPM given optimizing parameters m and M as indicated where η is varied from 1.0 to 0.1 and $c_0\tau_0$ is set to

TABLE II
COMPARISON OF THE PHOTON EFFICIENCY BETWEEN THE NUMBER STATE AND THE COHERENT-STATE PPM SCHEMES WHEN $c_0\tau_0 = 10^{-1}$

η	C_P	m	M	$\hat{C}_P$	$\hat{m}$	$\hat{M}$	$C_P/\hat{C}_P$
1.0	3.56	1	35	1.86	0.86	16	1.91
.9	3.06	1	30	1.67	0.96	16	1.83
.8	2.61	1	26	1.49	1.08	16	1.75
.7	2.13	1	21	1.30	1.23	16	1.64
.6	1.66	1	16	1.12	1.43	16	1.49
.5	1.24	1	12	0.93	1.72	16	1.34
.4	0.93	2	18	0.74	2.15	16	1.24
.3	0.65	2	13	0.56	2.87	16	1.17
.2	0.41	4	16	0.37	4.30	16	1.10
.1	0.19	8	15	0.19	8.61	16	1.04

TABLE III
COMPARISON WHEN $c_0\tau_0 = 10^{-2}$

η	C_P	m	M	$\hat{C}_P$	$\hat{m}$	$\hat{M}$	$C_P/\hat{C}_P$
1.0	6.47	1	647	4.23	0.42	176	1.53
.9	5.71	1	571	3.81	0.46	176	1.50
.8	4.96	1	496	3.38	0.52	176	1.47
.7	4.23	1	423	2.96	0.59	176	1.43
.6	3.52	1	351	2.54	0.69	176	1.39
.5	2.82	1	282	2.12	0.83	176	1.33
.4	2.15	1	214	1.69	1.04	176	1.27
.3	1.50	1	150	1.27	1.39	176	1.19
.2	0.94	2	188	0.85	2.08	176	1.11
.1	0.45	4	178	0.42	4.16	176	1.05

TABLE IV
COMPARISON WHEN $c_0\tau_0 = 10^{-3}$

η	Cp	m	M	$\hat{C}$p	$\hat{m}$	$\hat{M}$	Cp/$\hat{C}$p
1.0	9.12	1	9118	6.55	0.28	1830	1.39
.9	8.10	1	8099	5.90	0.31	1830	1.37
.8	7.09	1	7093	5.24	0.35	1830	1.35
.7	6.10	1	6101	4.59	0.40	1830	1.33
.6	5.13	1	5125	3.93	0.47	1830	1.30
.5	4.17	1	4167	3.28	0.56	1830	1.27
.4	3.23	1	3232	2.62	0.70	1830	1.23
.3	2.33	1	2325	1.97	0.93	1830	1.18
.2	1.46	1	1456	1.31	1.40	1830	1.11
.1	0.69	3	2069	0.66	2.79	1830	1.05

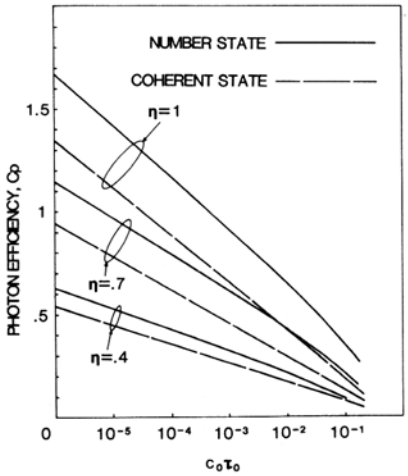

Fig. 6. Photon efficiency of the number state and the coherent-state PPM schemes as a function of the product $c_0\tau_0$.

10^{-1}, 10^{-2}, and 10^{-3} nats, respectively. It is quite evident from Tables II, III, and IV that the number state PPM scheme is always superior to the coherent state from the view point of photon efficiency irrespective of η.

The maximal photon efficiency plotted against $c_0\tau_0$ for several values of η is shown in Fig. 6. It can be decisively concluded from Fig. 6 that irrespective of η, the optimal photon efficiency of the PPM scheme increases monotonically when either the channel capacity per second or the pulse width is reduced.

It is important to note in Tables II, III, and IV, that even in lossless case where $\eta = 1$, the photon efficiency of the number-state PPM scheme is at most twice as large as that of the coherent-state PPM scheme. It should be noted, however, that such high photon efficiency is attainable in the coherent-state PPM scheme only by the use of a very complex and often impractical error correcting code, which is entirely unnecessary for the number-state PPM scheme.

IV. Conclusion

First, the capacity of lossy photon channels utilizing the number state, is calculated for PAM, PPM, and PPAM modulation schemes. As a result, regardless of losses, it is shown that the PPAM scheme is superior to the PPM scheme from the capacity point of view. Then through an evaluation of photon efficiency, measured in nats per transmitted photon, it is revealed that the number-state PPM and PPAM schemes asymptotically become equal in photon efficiency as the alphabet size increases. Finally, photon efficiency of the number state and the coherent-state PPM scheme are compared under certain realistic conditions. It is concluded that the former is always superior to the latter from the perspective of photon efficiency regardless of losses.

Appendix A

Calculation of C in (12) and (13)

Necessary definitions are given first.

Probability distribution q is defined as a nonnegative row vector

$$q=(q_1, q_2, \cdots, q_n), \quad \sum_{j=1}^{n} q_j=1, \quad \text{all } q_j \geqq 0. \quad \text{(A1)}$$

Let $\bar{\Delta}^n$ be the set of all probability distributions, i.e.,

$$\bar{\Delta}^n=\left\{(q_1, q_2, \cdots, q_n) \middle| \sum_{j=1}^{n} q_j=1, \quad \text{all } q_j \geqq 0\right\}. \quad \text{(A2)}$$

The Kullback–Leibler information $D(q^1 \| q^2)$ between two probability distributions q^1 and q^2 is defined by

$$D(q^1 \| q^2) \triangleq \sum_{j=1}^{n} q_j^1 \log (q_j^1/q_j^2). \quad \text{(A3)}$$

For k points $q^1, \cdots, q^k$ in n-dimensional Euclidean space, let $C(q^1, \cdots, q^k)$ be the convex hull of $q^1, \cdots, q^k$. Now, the calculation of C in (12) can proceed as follows. Let the ith row vector of the channel matrix characterized by (2) be denoted as the probability distribution given by Q^i

$$Q^i \triangleq (Q(0|i), Q(1|i), \cdots, Q(M-1|i)) \quad \text{(A4)}$$

and let V be defined as

$$V \triangleq C(Q^0, Q^1, \cdots, Q^{M-1}). \quad \text{(A5)}$$

It is then obvious that each Q^i is an extreme point of the convex hull V and considering (2), it is also evident that $V \fallingdotseq \bar{\Delta}^M$ if $\eta \fallingdotseq 1$. Therefore, letting q^0 denote an equidistant point from all extreme points of V, i.e.,

$$D(Q^0 \| q^0)=D(Q^1 \| q^0)=\cdots=D(Q^{M-1} \| q^0) \quad \text{(A6)}$$

it is evident that q^0 must be contained in V if $\eta \fallingdotseq 1$. Thus, according to the geometric characterization of channel capacity [8], it can be shown that C is given by the Kullback-Leibler information satisfying (A6).

Then, the next problem is solving (A6). Letting $q^0 = (q_0, q_1, \cdots, q_{M-1})$ and

$$C=D(Q^0 \| q^0) \quad \text{(A7)}$$

the following equations concerning q_i's are obtained from (A6)

$$Q\begin{pmatrix} \log q_0+C \\ \log q_1+C \\ \vdots \\ \log q_{M-1}+C \end{pmatrix} = -\begin{pmatrix} H_0 \\ H_1 \\ \vdots \\ H_{M-1} \end{pmatrix} \quad \text{(A8)}$$

where for $l = 0, 1, \cdots, M-1$, H_l is defined in (14). Using the inverse matrix $W = (w_{kl})$ of Q, it follows that solutions of (A8) are given by

$$q_k=\exp\left(-C-\sum_{l=1}^{k} w_{kl}H_l\right), \quad k=0, 1, \cdots, M-1. \quad \text{(A9)}$$

Consequently, channel capacity C in (12) is readily obtained by combining the last equations with the normalization equation $\Sigma_{k=0}^{M-1} q_k = 1$.

As η becomes smaller than one, it is no longer possible to postulate that $V \doteqdot \bar{\Delta}^M$. As a result, the equidistant point q^0 is no longer contained in V. Geometrical consideration of the channel capacity [8] shows that in this case, C in (12) is no longer able to represent the actual channel capacity and that the correct channel capacity in this case must be given by the projection point of q^0 onto V. However, this projection problem proves very difficult to solve, in general.

We are now in a position to calculate C in (13). For the case $\eta \doteqdot 0$, it can be demonstrated by direct calculation that among points $Q^0, Q^1, \cdots, Q^{M-1}$, Q^0, and Q^{M-1} are the two remotest points measured by the Kullback-Leibler information $D(Q^0 \| Q^i)$. Therefore, in this case it is presupposed that the channel capacity is attained only if the two symbols 0 and $M-1$ are used with nonzero probability. This presupposition can be validated by the well-known Kuhn-Tucker theorem (see, e.g., [12], p. 91) as follows. Let R^0 be the midpoint between points Q^0 and Q^{M-1}; that is, R^0 is the point that minimizes $D(Q^0 \| R^0)$ subject to the constraint

$$D(Q^0 \| R^0) = D(Q^{M-1} \| R^0). \tag{A10}$$

Thus, letting $R^0 = (r_0, r_1, \cdots, r_{M-1})$, it follows from the variational calculus that

$$\begin{cases} r_0 = 1/(1+\exp(-\theta_2)) \\ r_k = (1/(1-(1-\eta)^{M-1})) \binom{M-1}{k} \eta^k (1-\eta)^{M-1-k} \\ \qquad \times \exp(-\theta_2)/(1+\exp(-\theta_2)) \quad k=1, \cdots, M-1 \end{cases} \tag{A11}$$

where θ_2 is defined in (15). Combining the last equations with (A10) leads to the equality

$$D(Q^0 \| R^0) = \log(1+\exp(-\theta_2)). \tag{A12}$$

On the other hand, it is not difficult to show that the set of inequalities

$$D(Q^0 \| R^0) \geqq D(Q^i \| R^0), \quad i=1, 2, \cdots, M-2 \tag{A13}$$

can be equivalently expressed as

$$u(\eta, i) \geqq t(\eta, i), \quad i=1, 2, \cdots, M-2 \tag{A14}$$

where

$$t(\eta, i) \triangleq \{i-(M-1)(1-(1-\eta)^i)/(1-(1-\eta)^{M-1})\} \times \log(1-\eta) \tag{A15}$$

$$u(\eta, i) \triangleq \sum_{j=1}^{i} \binom{i}{j} \eta^j (1-\eta)^{k-j} \log\left(\binom{M-1}{j} \Big/ \binom{i}{j}\right). \tag{A16}$$

Since it also can be easily verified after some calculation that each inequality in (A14) holds only for $0 \leqq \eta \leqq \eta_i$ with a positive real number η_i to be determined for each $i = 1, 2, \cdots, M-2$, then, letting the minimum among η_i's be denoted by $\eta_{\min}$, it must follow that the set of inequalities in (A13) jointly holds for $0 \leqq \eta \leqq \eta_{\min}$. Therefore, it is confirmed by the above argument that if $D(Q^0 \| R^0)$ is selected for the constant C, then C evidently satisfies the Kuhn-Tucker condition for the channel capacity. Consequently, channel capacity C in (13) can be derived directly from (A12).

Appendix B

Calculation of C in (21) and (22)

Let the ith row vector of the channel matrix in (19) be denoted as a probability distribution by Q^i and let V be the convex hull of these vectors, i.e.,

$$V \triangleq C(Q^0, Q^1, \cdots, Q^M). \tag{B1}$$

It is evident from the matrix form that if $\theta \doteqdot 1$, then every Q^i is an extreme point of V and $V \doteqdot \bar{\Delta}^{M+1}$. Therefore, in this case the equidistant point q^0 from all extreme points of V must exist in V and thus, according to the geometric characterization of the channel capacity [8], the channel capacity C must be given by the equidistant point as follows:

$$C = D(Q^0 \| q^0). \tag{B2}$$

The equidistant point $q^0 = (q_0, q_1, \cdots, q_M)$ can now be calculated easily from the equal distance condition

$$D(Q^0 \| q^0) = D(Q^1 \| q^0) = \cdots = D(Q^M \| q^0) \tag{B3}$$

associated with the normalization condition $\Sigma_{i=0}^{M} q_i = 1$ and is given by

$$\begin{cases} q_0 = 1/(1+M\exp(-\hat{H}/\theta)), & \text{(B4)} \\ q_k = \exp(-\hat{H}/\theta)/(1+M\exp(-\hat{H}/\theta)), & \\ \qquad k=1, \cdots, M & \text{(B5)} \end{cases}$$

where $\hat{H}$ is defined in (23). Consequently, when $\theta \doteqdot 1$, channel capacity C in (21) can be readily obtained by combining (B4) with (B2). As θ becomes less than one, it is no longer possible to postulate that $V \doteqdot \bar{\Delta}^{M+1}$. As a result, C in (21) is no longer the actual channel capacity and another channel capacity formula that is valid for small values of θ must be pursued. In this case, the geometric consideration of the channel capacity [8] and the Kuhn-Tucker theorem [12] can again be used profitably as follows.

Let $\hat{V}$ be the convex hull of the set of row vectors excluding Q^0, i.e.,

$$\hat{V} \triangleq C(Q^1, Q^2, \cdots, Q^M) \tag{B6}$$

and let $\hat{q}^0 \triangleq (\hat{q}_0, \hat{q}_1, \cdots, \hat{q}_M)$ be the equidistant point in $\hat{V}$ from all extreme points of $\hat{V}$. It is evident from a geometrical consideration that $\hat{q}^0$ can be calculated as the solution to the following minimization problem:

$$\min_{\hat{q}^0} D(Q^1 \| \hat{q}^0) \tag{B7}$$

subject to the constraint

$$\begin{cases} D(Q^1 \| \hat{q}^0) = D(Q^2 \| \hat{q}^0) = \cdots = D(Q^M \| \hat{q}^0), & \text{(B8)} \\ \sum_{k=0}^{M} \hat{q}_k = 1. & \text{(B9)} \end{cases}$$

Variational calculus leads to the solution

$$\begin{cases} \hat{q}_0 = 1-\theta, & \text{(B10)} \\ \hat{q}_k = \theta/M, \quad k=1, \cdots, M. & \text{(B11)} \end{cases}$$

It can be readily understood that

$$\hat{q}^0 = (0, 1/M, \cdots, 1/M)Q. \tag{B12}$$

This means that $\hat{q}^0$ is the probability distribution of the channel output symbol if all the input symbols but symbol ϕ are transmitted with equal probability $1/M$. Therefore, according to the Kuhn–Tucker Theorem, the channel capacity C is given by

$$C = D(Q^1 \| \hat{q}^0) \tag{B13}$$

if C satisfies the following:

$$C \geqq D(Q^0 \| \hat{q}^0). \tag{B14}$$

It follows from (B10) and (B11) that

$$C = \theta \log M \tag{B15}$$

and

$$D(Q^0 \| \hat{q}^0) = -\log (1-\theta). \tag{B16}$$

Therefore, inequality (B14) can be equivalently expressed by

$$\theta \log M + \log (1-\theta) \geqq 0. \tag{B17}$$

Since it can be easily shown that the last inequality has solution $0 \leqq \theta \leqq \theta_0$, with θ_0 being the unique positive solution to (24), channel capacity C in (22) is confirmed. On the other hand, it can be ascertained by some calculation that at $\theta = \theta_0$, values of C in (21) and (22) coincide with each other. Consequently, it is confirmed that C in (21) is actually valid for $\theta_0 \leqq \theta \leqq 1$.

References

[1] H. P. Yuen, "Two-photon coherent states of the radiation field," *Phys. Rev.*, vol. A13, no. 6, pp. 2226–2243, June 1976.

[2] R. E. Slusher, L. W. Hollberg, B. Yurke, J. C. Mertz, and J. F. Valley, "Observation of squeezed states generated by four-wave mixing in an optical cavity," *Phys. Rev. Lett.*, vol. 55, no. 22, pp. 2409–2412, Nov. 1985.

[3] H. P. Yuen and J. H. Shapiro, "Optical communication with two-photon coherent states—Part I: Quantum-noise reduction," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 657–668, Nov. 1978.

[4] J. H. Shapiro, H. P. Yuen, and J. A. Machado Mata, "Optical communication with two-photon coherent states—Part II: Photoemissive detection and structured receiver performance," *IEEE Trans. Inform. Theory*, vol. IT-25, pp. 179–192, Mar. 1979.

[5] Y. Yamamoto, N. Imoto, and S. Machida, "Amplitude squeezing in a semiconductor laser using quantum nondemolition measurement and negative feedback," *Phys. Rev.*, vol. A33, no. 5, pp. 3243–3261, May 1986.

[6] Y. Yamamoto and H. A. Haus, "Preparation, measurement and information capacity of optical quantum states," *Rev. Mod. Phys.*, vol. 58, no. 4, Oct. 1986.

[7] H. Takahashi, "Information theory of quantum-mechanical channels," in *Advances in Communication Systems, Vol. 1*, A. V. Balakrishnan, Ed. New York: Academic, 1965, pp. 227–310.

[8] K. Nakagawa and F. Kanaya, "A new geometric capacity characterization of a discrete memoryless channel," *IEEE Trans. Inform. Theory*, vol. 34, pp. 318–321, Mar. 1988.

[9] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 14–20, Jan. 1972.

[10] R. J. McEliece, "Practical codes for photon communication," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 393–398, July 1981.

[11] S. A. Butman, J. Katz, and J. R. Lesh, "Bandwidth limitations on noiseless optical channel capacity," *IEEE Trans. Commun.*, vol. COM-30, pp. 1262–1264, May 1982.

[12] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[13] M. H. A. Davis, "Capacity and cutoff rate for Poissonian-type channels," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 710–715, Nov. 1980.

★

Fumio Kanaya (M'85) received the B.S., M.S., and Dr.Eng. degrees in electrical engineering from The University of Tokyo, Tokyo, in 1963, 1965, and 1976, respectively.

In 1965, he joined the Electrical Communication Laboratories, NTT, and has engaged in research and development works on various communication systems including digital transmission and coding of images. Since July 1987, as a Research Fellow he has superintended Kanaya Research Laboratory at NTT Transmission Systems Laboratories, Kanagawa. His current research interests are in information theory and its application to communication theory and statistical decision theory.

Dr. Kanaya is a member of the Institute of Electronics, Information and Communication Engineers, and the Society of Information Theory and its Application.

★

Kenji Nakagawa (A'85) received the B.S., M.S., and D.S. degrees in mathematics from Tokyo Institute of Technology in 1980, 1982, and 1986, respectively.

In 1985, he joined the Basic Research Laboratories, NTT, and has engaged in research on information theory. Since July 1987, he has been a Research Engineer at Kanaya Research Laboratory, NTT Transmission Systems Laboratories. His research interests include information theory, especially an application of information geometry to information theory.

Dr. Nakagawa is a member of the Institute of Electronics, Information and Communication Engineers, and the Society of Information Theory and its Application.