Very Noisy Channels, Reliability Functions, and Exponentially Optimum Codes

Sangmin Lee and Kim A. Winick, *Member, IEEE*

***Abstract*—A very noisy channel (VNC), is a discrete-input memoryless channel whose capacity is close to zero. Very noisy channels are of interest, since they serve as models for some important physical channels. There are two distinct classes of VNC's: Reiffen's class I and Majani's class II. It is shown that the reliability function is known exactly for both classes of VNC's, by extending the results previously obtained only for class I. It is then shown that an exponentially optimum code can be constructed for a channel if it can be modeled as repeated uses of a binary-input class I VNC, a binary-input / binary-output class II VNC or a class II very noisy binary erasure channel. The theory developed is illustrated by considering the direct detection optical channel used with polarization modulation. The capacity, reliability function, and exponentially optimum code, for this channel, are derived.**

***Index Terms*—Very noisy channels, error exponents, reliability functions, exponentially optimum codes.**

I. Introduction

A VERY noisy channel is a discrete memoryless channel (DMC) whose capacity is close to zero. Very noisy channels (VNC's) were introduced by Reiffen [1] to model many physical channels operating at low signal-to-noise ratios. More importantly, a large class of physical channels, operating at arbitrary signal-to-noise ratios, can be modeled as repeated uses of a VNC. In particular, this is true for the infinite bandwidth additive white Gaussian noise (AWGN) channel [2] and the direct detection Poisson optical channel [3], [4].

Majani [5] undertook a systematic study of very noisy channels. He enlarged Reiffen's definition of VNC's to include two distinct classes. Class I VNC's are identical to those defined by Reiffen. The very noisy channel, as defined by Majani, is a DMC whose transition probabilities, $p(y|x)$, are given by

$$p(y|x) = \omega(y) + \epsilon\lambda(x, y) + O(\epsilon^2), \qquad (1.1)$$

where x and y are elements of the input and output alphabets $\mathscr{X}$ and $\mathscr{Y}$, $\omega(y)$ is a probability distribution on $\mathscr{Y}$, ϵ is a very small number, $\lambda(x, y)$'s are fixed numbers satisfying

$$\sum_{y\in\mathscr{Y}} \lambda(x, y) = 0, \qquad \forall\, x \in \mathscr{X}, \qquad (1.2)$$

and $O(\epsilon^n)$ is a quantity which contains terms of order n or larger in ϵ. If we consider the set $\mathscr{S}$ defined by $\mathscr{S} = \{y \in \mathscr{Y}: \omega(y) = 0\}$, then class I and class II VNC's are defined by the condition $\mathscr{S} = \phi$ and $\mathscr{S} \neq \phi$, respectively.

In most practical cases, a VNC serves as a model for a communication channel in which the information efficiency, measured in nats per resource (where the resource may be any abstract quantity such as energy, area, or time), is a primary measure of system performance. In such cases, the ϵ that appears in the definition (1.1) is a decreasing function of a parameter z, i.e., $\epsilon = \epsilon(z)$, where z denotes the resource expenditure per channel use. Consider, for example, binary on/off-keying on the direct detection Poisson optical channel with hard decision, maximum likelihood, demodulation [6]. The channel inputs are $x = 0$ or 1, where $x = 0$ corresponds to the transmission of no photons and $x = 1$ corresponds to the transmission of $\lambda_s\Delta$ photons. λ_s is the peak transmitted power in photons per second (photons/s), and Δ is the duration of a channel symbol in seconds. This channel can be modeled as the binary Z-channel shown in Fig. 1. The transition probabilities $\{p(y|x)\}$ for this channel are given by (1.3) below

$$\begin{pmatrix} p(0|0) & p(1|0) \\ p(0|1) & p(1|1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + \epsilon\begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix}, \qquad (1.3)$$

where

$$\epsilon = \epsilon(\Delta) = 1 - e^{-\lambda_s\Delta}. \qquad (1.4)$$

The capacity in nats per channel use (nats/c.u.), $C(\Delta)$, and the capacity in nats per second (nats/s), $C(\Delta)/\Delta$, of this Z-channel are

$$C(\Delta) = \ln\left[1 + \epsilon(1 - \epsilon)^{(1-\epsilon)/\epsilon}\right]$$

and

$$\frac{C(\Delta)}{\Delta} = \frac{\ln\left[1 + \epsilon(1 - \epsilon)^{(1-\epsilon)/\epsilon}\right]}{\Delta}, \qquad (1.5)$$

respectively. $C(\Delta)$ and $C(\Delta)/\Delta$ are plotted in Fig. 2 as a function of Δ for a fixed value of $\lambda_s = 1$ photon/s. Note that for a fixed peak power λ_s the channel capacity in nats/s is maximized as $\Delta \to 0$, and in this limit the

Manuscript received May 7, 1992; revised manuscript received September 7, 1993.

This work was presented in part at the Annual Conference on Information Sciences and Systems, Princeton, NJ, March, 1992.

Kim A. Winick is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109-2122.

Sangmin Lee is with the Department of Electronics Engineering, Kang Nung National University, Kang Nung, Kangwondo, Korea.

IEEE Log Number 9401972.

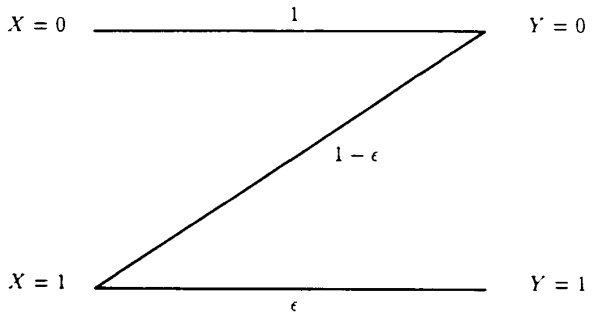

Fig. 1. Z-channel.

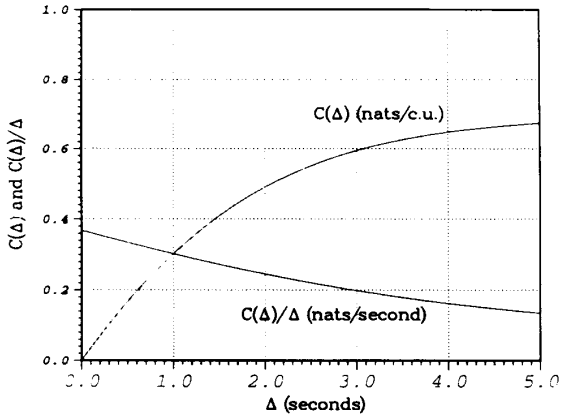

Fig. 2. Capacities for the direct detection optical channel.

Z-channel becomes a very noisy channel with the resource expenditure per channel use being $z = \Delta$.

The concept illustrated above can be generalized. Consider a code consisting of M code words, where each code word requires an expenditure, Z, of some resource each time it is transmitted. Transmission of a code word involves N consecutive channel uses, and each of these uses expends resource z, where

$$z = \frac{Z}{N}. \tag{1.6}$$

Define the code rate, $\boldsymbol{R}$, of this code, in nats per resource, as

$$\boldsymbol{R} = \frac{\ln M}{Z}. \tag{1.7}$$

Then the rate $\boldsymbol{R}$ in nats per resource is related to the rate R in nats per channel use by

$$\boldsymbol{R} = \frac{R}{z}. \tag{1.8}$$

Thus, the capacity per unit resource, $\boldsymbol{C}$ is defined by

$$\boldsymbol{C} = \sup_{z>0} \frac{C(z)}{z}, \tag{1.9}$$

where $C(z)$ is the capacity in nats per channel use for a given z. Define $\boldsymbol{C}^*$ by

$$\boldsymbol{C}^* = \lim_{z\to 0} \frac{C(z)}{z}. \tag{1.10}$$

Then, $\boldsymbol{C}^*$ is the capacity per unit resource in the very noise limit as $z \to 0$.

Abdel-Ghaffar and McEliece [7] have modeled several important physical channels as repeated uses of VNC's and computed $\boldsymbol{C}^*$. In many of these channels, $\boldsymbol{C}$ is equal to $\boldsymbol{C}^*$. They have also constructed codes for very noisy binary symmetric channel, which achieves $\boldsymbol{C}^*$. Chao [8] has extended this coding work to include all binary-input class I VNC's. Verdu [9] has studied the capacity per unit resource, $\boldsymbol{C}$, for memoryless channels, and has given conditions under which $\boldsymbol{C}$ is readily computed. In this paper, we will extend these coding works to compute the reliability function for the channels which can be modeled as repeated uses of a VNC. We will also construct exponentially optimum codes for some subclass of such channels.

The reliability function $E(R)$ for a DMC is the exponent with which the error probability of the best block code over the channel may be made to decrease exponentially as the block length, N, gets large. If $P_e(N, R)$ denotes the probability of code word error for a block code of length, N, and code rate, R, and $P_e^\dagger(N, R)$ denotes the minimum P_e for all such codes, then the reliability function, $E(R)$, is defined by

$$E(R) = \limsup_{N\to\infty} \frac{-\ln P_e^\dagger(N, R)}{N}. \tag{1.11}$$

It follows from (1.11) that

$$P_e^\dagger(N, R) = e^{-NE(R)+\Theta(N)} \quad \text{as } N \to \infty,\ R < C, \tag{1.12}$$

where $\Theta(\cdot)$ denotes a function such that $\Theta(x)/x \to 0$ as $x \to \infty$. When we are interested in the probability of error in terms of resource expenditure, (1.12) may be rewritten as

$$P_e^\dagger(Z, \boldsymbol{R}) = e^{-ZE(\boldsymbol{R})+\Theta(Z)} \quad \text{as } Z \to \infty,\ \boldsymbol{R} < \boldsymbol{C}. \tag{1.13}$$

where $\boldsymbol{E}(\boldsymbol{R})$, the reliability function in terms of resource expenditure, is given by

$$\boldsymbol{E}(\boldsymbol{R}) = \sup_{z>0} \frac{E(z\boldsymbol{R})}{z}. \tag{1.14}$$

Let $\boldsymbol{E}^*(\boldsymbol{R})$ be defined by

$$\boldsymbol{E}^*(\boldsymbol{R}) = \lim_{z\to 0} \frac{E(z\boldsymbol{R})}{z}. \tag{1.15}$$

Then, $\boldsymbol{E}^*(\boldsymbol{R})$ is the reliability function in the very noisy limit as $z \to 0$. Codes which achieve the probability of error bound (1.12) or (1.13) are said to be exponentially optimum. Thus, in the very noisy limit, exponentially optimum codes will have the probability of error given by $P_e^\dagger(Z, \boldsymbol{R}) = \exp[-Z\boldsymbol{E}^*(\boldsymbol{R}) + \Theta(Z)]$ as $Z \to \infty$.

The reliability function is, in general, not known exactly, but it has been bounded [10]–[12]. Consider a DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities $p(y|x)$. Let $\mathbf{q}$ denote the input probability distribution on $\mathcal{X}$. For this channel, a lower bound on $E(R)$ is given by [10]

$$E(R) \geq E_L(R) = \max\{E_r(R), E_{ex}(R)\} = \begin{cases} E_{ex}(R), & 0 \leq R \leq R_{ex} \\ E_r(R), & R_{ex} \leq R \leq C, \end{cases} \tag{1.16}$$

where

$$E_{ex}(R) = \max_{\mathbf{q}} \sup_{\rho \geq 1} [E_x(\rho, \mathbf{q}) - \rho R], \quad \text{(expurgated error exponent)}, \tag{1.17}$$

$$E_r(R) = \max_{\mathbf{q}} \max_{0 \leq \rho \leq 1} [E_0(\rho, \mathbf{q}) - \rho R], \quad \text{(random coding error exponent)}, \tag{1.18}$$

$$E_x(\rho, \mathbf{q}) = -\rho \ln \left\{ \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} q(x) q(x') \cdot \left[\sum_{y \in \mathcal{Y}} \sqrt{p(y|x)p(y|x')} \right]^{1/\rho} \right\}, \tag{1.19}$$

$$E_0(\rho, \mathbf{q}) = -\ln \sum_{y \in \mathcal{Y}} \left[\sum_{x \in \mathcal{X}} q(x) p(y|x)^{1/(1+\rho)} \right]^{1+\rho}, \tag{1.20}$$

and

$$R_{ex} = \max_{\mathbf{q}} \left. \frac{\partial E_x(\rho, \mathbf{q})}{\partial \rho} \right|_{\rho = 1}. \tag{1.21}$$

Furthermore, it can be shown that [10]

$$E_r(R) = E_r(0) - R, \qquad 0 \leq R \leq R_{cr} \tag{1.22}$$

$$\frac{\partial E_0(\rho, \mathbf{q})}{\partial \rho} \geq 0, \qquad \rho \geq -1 \tag{1.23}$$

$$\frac{\partial E_x(\rho, \mathbf{q})}{\partial \rho} \geq 0, \qquad \rho > 0 \tag{1.24}$$

where the critical rate R_{cr} is defined by

$$R_{cr} = \max_{\mathbf{q}} \left. \frac{\partial E_0(\rho, \mathbf{q})}{\partial \rho} \right|_{\rho = 1}. \tag{1.25}$$

The results of [11], [12] give the following upper bound on $E(R)$:

$$E(R) \leq E_U(R) = \begin{cases} E_{ex}(0), & R = 0, \\ E_{sl}(R), & 0 < R \leq R_{sl}, \\ E_{sp}(R), & R_{sl} \leq R \leq C, \end{cases} \tag{1.26}$$

where

$$E_{sp}(R) = \max_{\mathbf{q}} \sup_{\rho \geq 0} [E_0(\rho, \mathbf{q}) - \rho R], \quad \text{(sphere-packing error exponent)}, \tag{1.27}$$

$$E_{sl}(R) = E_{ex}(0) - \lambda_{sl} R, \quad \text{(straight-line error exponent)} \tag{1.28}$$

and $-\lambda_{sl}$ is the slope of the straight line that passes through the point $E_{ex}(0)$ and is tangent to the curve $E_{sp}(R)$. The point of tangency on the R axis is denoted R_{sl}. As an illustration, a plot of these upper and lower bounds is given in Fig. 3 for the case of binary symmetric channel with cross-over probability $p = 0.005$. It is known [11], [12] that these upper and lower bounds are identical, and therefore tight, for all rates greater than or equal to the critical rate R_{cr}, i.e.,

$$E_{sp}(R) = E_r(R), \qquad R_{cr} \leq R < C. \tag{1.29}$$

For $R < R_{cr}$, the bounds differ, and further tightening remains an unsolved problem.

There are only a few channels for which the reliability function is known exactly at all rates less than capacity. These are Reiffen's VNC [12], energy limited channels considered by Gallager [13], the infinite bandwidth AWGN channel [2], and the direct detection Poisson optical channel [3], [4]. Exponentially optimum codes have been constructed only for the later two channels. In this paper we will extend the number of channels for which the reliability function is known exactly and will construct exponentially optimum codes for a subclass of these channels.

The remainder of this paper is organized into five sections. In Section II, it is shown that the reliability function is known exactly for all class II VNC's. In Section III, exponentially optimum codes are constructed for channels that can be modeled as a binary-input class I VNC. Similar results are in Section IV for channels that can be modeled as a class II binary-input/binary-output VNC or a class II binary erasure VNC. The fundamental performance limits of polarization switching direct detection optical channel is examined in Section V. The channel is modeled as repeated uses of a very noisy binary channel. The capacity, C, and reliability function, $E(R)$, are determined for this optical channel, and an exponentially optimum code is constructed for all rates less than the capacity, C. Section VI summarizes our results.

II. Reliability Functions for Very Noisy Channels

In this section we will prove that the reliability function is known exactly for all VNC's defined by (1.1) and (1.2). The proof will rely on the following well known theorem.

Theorem 2.1: The reliability function, $E(R)$, is known exactly and is equal to $E_r(R)$ for all rates $R < C$, provided

$$E_{ex}(0) = E_r(0). \tag{2.1}$$

Proof: It is known that $E(R) = E_r(R)$ for $R_c \leq R < C$ [see (1.29)]. Thus it suffices to consider the region $0 \leq R$

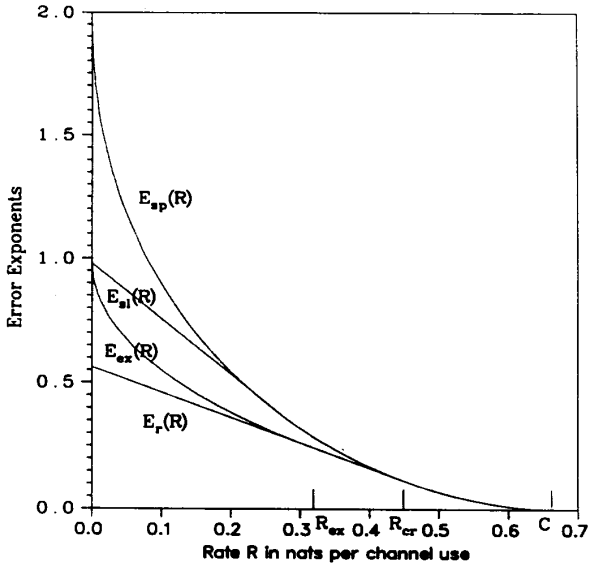

Fig. 3. Error exponents for BSC ($p = 0.005$).

$\leq R_c$. Now $E_r(R) = E_r(0) - R$ for $0 \leq R \leq R_c$ [see (1.22)], and this straight line is tangent to $E_{sp}(R)$ at $R = R_c$. Thus, if condition (2.1) is satisfied, then $R_{sl} = R_c$ and $E_{sl}(R) = E_{ex}(0) - R = E_r(R)$ for $0 \leq R \leq R_c$. □

We now proceed to show that the reliability function is known exactly for all VNC's. Only class II VNC's are considered, since the result is known to be true for class I VNC's [12]. We begin by computing $E_r(0)$. The random coding error exponent, $E_r(R)$, for class II VNC's has been derived by Majani [5]. For these channels, $E_o(\rho, \boldsymbol{q})$ is given by

$$E_0(\rho, \boldsymbol{q}) = \epsilon \sum_{y \in \mathcal{S}} \left[\sum_{x \in \mathcal{X}} q(x)\lambda(x,y) - \left(\sum_{x \in \mathcal{X}} q(x)\lambda(x,y)^{1/(1+\rho)} \right)^{1+\rho} \right] + O(\epsilon^2). \quad (2.2)$$

Therefore, it follows from (1.18), (1.23), and (2.2) that

$$E_r(0) = \max_{\boldsymbol{q}} \max_{0 \leq \rho \leq 1} E_o(\rho, \boldsymbol{q}) = \max_{\boldsymbol{q}} E_o(1, \boldsymbol{q}) \quad (2.3)$$

$$= \max_{\boldsymbol{q}} \epsilon \sum_{y \in \mathcal{S}} \left[\sum_{x \in \mathcal{X}} q(x)\lambda(x,y) - \left(\sum_{x \in \mathcal{X}} q(x)\sqrt{\lambda(x,y)} \right)^2 \right] + O(\epsilon^2). \quad (2.4)$$

We now compute the zero rate expurgated error exponent, $E_{ex}(0)$. It follows from (1.17), (1.19), and (1.24) that

$$E_{ex}(0) = \max_{\boldsymbol{q}} \sup_{\rho \geq 1} E_x(\rho, \boldsymbol{q}) = \max_{\boldsymbol{q}} \lim_{\rho \to \infty} E_x(\rho, \boldsymbol{q}) \quad (2.5)$$

$$= \max_{\boldsymbol{q}} \left\{ \sum_{x \in \mathcal{X}} \sum_{x' \in \mathcal{X}} q(x)q(x') \cdot \ln \left[\sum_{y \in \mathcal{Y}} \sqrt{p(y|x)p(y|x')} \right] \right\}, \quad (2.6)$$

where (2.6) is derived using L'Hopitals rule. For class II VNC's, we have

$$p(y|x)p(y|x') = \begin{cases} \omega^2(y) + \epsilon\omega(y)[\lambda(x,y) + \lambda(x',y)] + O(\epsilon^2), & y \notin \mathcal{S}, \\ \epsilon^2\lambda(x,y)\lambda(x',y) + O(\epsilon^3), & y \in \mathcal{S}. \end{cases} \quad (2.7)$$

Therefore,

$$\sqrt{p(y|x)p(y|x')} = \begin{cases} \omega(y) + \frac{1}{2}\epsilon[\lambda(x,y) + \lambda(x',y)] + O(\epsilon^2), & y \notin \mathcal{S}, \\ \epsilon\sqrt{\lambda(x,y)\lambda(x',y)} + O(\epsilon^2), & y \in \mathcal{S}. \end{cases} \quad (2.8)$$

It follows from (2.8) that

$$\sum_{y \in \mathcal{Y}} \sqrt{p(y|x)p(y|x')} = \sum_{y \notin \mathcal{S}} \sqrt{p(y|x)p(y|x')} + \sum_{y \in \mathcal{S}} \sqrt{p(y|x)p(y|x')}$$

$$= \sum_{y \notin \mathcal{S}} \omega(y) + \frac{\epsilon}{2} \sum_{y \notin \mathcal{S}} [\lambda(x,y) + \lambda(x',y)] + \epsilon \sum_{y \in \mathcal{S}} \sqrt{\lambda(x,y)\lambda(x',y)} + O(\epsilon^2)$$

$$= 1 - \frac{\epsilon}{2} \sum_{y \in \mathcal{S}} [\lambda(x,y) + \lambda(x',y)] + \epsilon \sum_{y \in \mathcal{S}} \sqrt{\lambda(x,y)\lambda(x',y)} + O(\epsilon^2), \quad (2.9)$$

where the last step in (2.9) follows from (1.2) and the fact that $\omega(y)$ is a probability distribution on $\mathcal{Y}$. Since $\ln(1 + z) = z + O(z^2)$, we can write

$$\ln \sum_{y \in \mathcal{Y}} \sqrt{p(y|x)p(y|x')} = -\frac{\epsilon}{2} \sum_{y \in \mathcal{S}} [\lambda(x,y) + \lambda(x',y)] + \epsilon \sum_{y \in \mathcal{S}} \sqrt{\lambda(x,y)\lambda(x',y)} + O(\epsilon^2). \quad (2.10)$$

Therefore, the expurgated error exponent at zero rate, as given by (2.6), is

$$E_{ex}(0) = \max_{q} \epsilon \left[\sum_{y \in \mathcal{S}} \left(\sum_{x \in \mathcal{X}} q(x)\lambda(x,y) \right) - \sum_{y \in \mathcal{S}} \left(\sum_{x \in \mathcal{X}} q(x)\sqrt{\lambda(x,y)} \right)^2 \right] + O(\epsilon^2). \quad (2.11)$$

It immediately follows from (2.4) and (2.11) that

$$E_r(0) = E_{ex}(0) \quad \text{as } \epsilon \to 0. \quad (2.12)$$

Thus, by Theorem 2.1 the reliability function for class II VNC's is known exactly at all rates less than capacity as $\epsilon \to 0$ and is given by

$$E(R) = E_r(R), \qquad R < C. \quad (2.13)$$

III. Exponentially Optimum Codes for Binary-Input Class I VNC's

In this section we will construct exponentially optimum codes for channels that can be modeled as repeated uses of a binary-input class I VNC with resource expenditure per channel use z. It is known [5], [10]–[12] that in the limit as $\epsilon(z) \to 0$, the reliability function for binary-input class I VNC's is given by

$$E(R) = E_r(R) = \begin{cases} \dfrac{C(z)}{2} - R, & \text{for } 0 \le \dfrac{R}{C(z)} \le \dfrac{1}{4}, \\ \left(\sqrt{C(z)} - \sqrt{R}\right)^2, & \text{for } \dfrac{1}{4} \le \dfrac{R}{C(z)} \le 1, \end{cases} \quad (3.1)$$

where $C(z)$ is the channel capacity in nats/c.u. and is given by

$$C(z) = \frac{\epsilon^2(z)}{8} \sum_{y \in \mathcal{Y}} \frac{1}{\omega(y)} [\lambda(0,y) - \lambda(1,y)]^2 + O(\epsilon^3(z)). \quad (3.2)$$

We will restrict ourselves to channels whose capacity per resource is nonzero in the very noisy limit, that is,

$$\boldsymbol{C}^* = \lim_{z \to 0} \frac{C(z)}{z} > 0 \quad \text{or} \quad \lim_{z \to 0} \frac{\epsilon^2(z)}{z} = c_1 > 0, \quad (3.3)$$

where c_1 is a strictly positive constant. Suppose that we are given a channel that is modeled as repeated uses of a binary-input class I VNC. Combining (1.10), (1.15), and (3.1)–(3.3) yields the following expression for the reliability function of the channel in the very noisy limit:

$$\boldsymbol{E}^*(\boldsymbol{R}) = \begin{cases} \dfrac{\boldsymbol{C}^*}{2} - \boldsymbol{R}, & \text{for } 0 \le \dfrac{\boldsymbol{R}}{\boldsymbol{C}^*} \le \dfrac{1}{4}, \\ \left(\sqrt{\boldsymbol{C}^*} - \sqrt{\boldsymbol{R}}\right)^2, & \text{for } \dfrac{1}{4} \le \dfrac{\boldsymbol{R}}{\boldsymbol{C}^*} \le 1, \end{cases} \quad (3.4)$$

where $\boldsymbol{C}^*$ is given by

$$\boldsymbol{C}^* = \lim_{z \to 0} \frac{C(z)}{z} = \frac{1}{8} c_1 \sum_{y \in \mathcal{Y}} \frac{1}{\omega(y)} [\lambda(0,y) - \lambda(1,y)]^2. \quad (3.5)$$

In what follows, we construct a binary code for this channel, and show that the code is exponentially optimum, in the sense that it achieves the probability of code word error given by $P_e = \exp\{-Z\mathbf{E}^*(\mathbf{R}) + \Theta(Z)\}$ as $Z \to \infty$, where $\boldsymbol{E}^*(\boldsymbol{R})$ is as given by (3.4).

Given M, k, let $\mathcal{A}$ be an $M \times \binom{M}{k}$ binary matrix, whose columns are those $\binom{M}{k}$ binary vectors containing exactly k ones. For example, for $M = 5$ and $k = 2$, $\binom{M}{k} = 10$ and the matrix $\mathcal{A}$ is given by

$$\mathcal{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}. \quad (3.6)$$

Let x_{mn} denotes the mnth entry of $\mathcal{A}$. We define a binary code, $\mathcal{C}(M,k)$, as the set of M binary code words, $\boldsymbol{x}_m = (x_{m1}, x_{m2}, \cdots, x_{mN})$, $m = 1, 2, \cdots, M$, where $N = \binom{M}{k}$. These $\mathcal{C}(M,k)$ codes have been previously examined by Wyner for use on the direct detection Poisson optical channel [3], [4]. The following theorem shows that this family of codes is exponentially optimum on the channel being considered.

Theorem 3.1: Given a channel that can be modeled as repeated uses of a binary-input class I VNC satisfying (3.3), then for any rate $\boldsymbol{R}$, $\boldsymbol{R} < \boldsymbol{C}^*$, the code $\mathcal{C}(M, M/2)$ with $M = \exp[\boldsymbol{R}Z]$ is exponentially optimum in terms of resource expenditure per nat as $Z \to \infty$.

Proof: The resource per channel use, z, for this code is given by

$$z = \frac{Z}{N} = \frac{Z}{\binom{M}{M/2}} = \frac{1}{\boldsymbol{R}} \frac{\ln M}{\binom{M}{M/2}}. \quad (3.7)$$

Thus for fixed $\boldsymbol{R}$, M approaches infinity and z approaches zero as $Z \to \infty$. The first part of the proof relies on a minor modification of ideas first developed by Abdel-

Ghaffar and McEliece [7] and later extended by Chao [8], therefore our discussion will be brief. Consider a binary-input, L-ary output class I VNC with $\mathscr{X} = \{0, 1\}$ and $\mathscr{Y} = \{0, 1, \cdots, L-1\}$. Suppose the code word $\boldsymbol{x}_0$ is transmitted and the received N-vector is $\boldsymbol{y} = (y_1, y_2, \cdots, y_N)$. Consider maximum likelihood decoding, then a correct decision is made if

$$\prod_{j=1}^{N} p(y_j|x_{0j}) > \prod_{j=1}^{N} p(y_j|x_{ij}) \quad \text{for all } i = 1, 2, \cdots, M-1. \tag{3.8}$$

Let $N_i^0(y)$, $y \in \mathscr{Y}$ be the number of components in $\boldsymbol{y}$ equal to y where $x_{0j} = 0$ and $x_{ij} = 1$. Similarly let $N_i^1(y)$, $y \in \mathscr{Y}$ be the number of components in $\boldsymbol{y}$ equal to y where $x_{0j} = 1$ and $x_{ij} = 0$. With these definition, (3.8) can be written as

$$S_i = \sum_{y \in \mathscr{Y}} \left[N_i^0(y) \ln \frac{p(y|0)}{p(y|1)} + N_i^1(y) \ln \frac{p(y|1)}{p(y|0)} \right] > 0 \quad \text{for all } i = 1, 2, \cdots, M-1. \tag{3.9}$$

Then the probability, P_c, of decoding a code word correctly satisfies

$$P_c \geq \Pr[S_i > 0, \forall\, i = 1, 2, \cdots, M-1]. \tag{3.10}$$

The mean, variance and covariance of the S_i may be computed [8].[1] The results are

$$\boldsymbol{E}(S_i) = \frac{N}{4}\epsilon^2(z)\left(\sum_{y \in \mathscr{Y}} \frac{1}{\omega(y)} (\lambda(0, y) - \lambda(1, y))^2 \right) + O(\epsilon^3(z)), \tag{3.11}$$

$$\operatorname{Var}(S_i) = \frac{N}{2}\epsilon^2(z)\left(\sum_{y \in \mathscr{Y}} \frac{1}{\omega(y)} (\lambda(0, y) - \lambda(1, y))^2 \right) + O(\epsilon^3(z)) \tag{3.12}$$

$$\operatorname{Cov}(S_i, S_j) = \frac{1}{2} \operatorname{Var}(S_i) \qquad i \neq j. \tag{3.13}$$

It follows from (3.2), (3.11), (1.6), and (1.10) that

$$\boldsymbol{E}(S_i) = 2NC(z) + O(\epsilon^3(z)) \to 2ZC^* \quad \text{as } Z \to \infty \text{ (i.e., } M \to \infty \text{ and } z \to 0.) \tag{3.14}$$

Similarly,

$$\operatorname{Var}(S_i) \to 4ZC^* \quad \text{or} \quad \operatorname{Var}(S_i)/\boldsymbol{E}(S_i) \to 2 \quad \text{as } Z \to \infty. \tag{3.15}$$

Invoking the central limit theorem, Chao [8] argues the S_i's become jointly Gaussian random variables as $N = \binom{M}{M/2} \to \infty$. Define a new set of random variable T_i by

$$T_i = \frac{S_i - \boldsymbol{E}(S_i)}{\sqrt{\operatorname{Var}(S_i)}} \sqrt{2}. \tag{3.16}$$

Then the T_i's are jointly Gaussian, and it follows from (3.12), (3.13), and (3.16) that

$$\boldsymbol{E}(T_i) = 0 \tag{3.17}$$

$$\operatorname{Var}(T_i) = 2 \tag{3.18}$$

$$\operatorname{Cov}(T_i, T_j) = 1 \qquad i \neq j. \tag{3.19}$$

Thus, the T_i's can be approximated by [8, Lemma 2.4]

$$T_i \approx W_i - W_0, \tag{3.20}$$

where $W_0, W_1, \cdots, W_{M-1}$ are independent, zero mean, unit variance Gaussian random variables. Combining (3.10) and (3.14)–(3.16) with this fact yields

$$P_c \geq \Pr\left[T_i > -\sqrt{\boldsymbol{E}(S_i)} \ \forall\, i = 1, 2, \cdots, M-1 \right] \approx \Pr\left[W_i - W_0 > -\sqrt{2ZC^*} \ \forall\, i = 1, 2, \cdots, M-1 \right] \quad \text{as } Z \to \infty. \tag{3.21}$$

In the second part of this proof, $P_e = 1 - P_c$ will be bounded using a technique described in [14, pp. 341–346]. It follows from (3.21) that

$$P_c \geq \int_{-\infty}^{\infty} \left\{ \Pr\left[W_i > -\sqrt{2ZC^*} + w_o \right] \right\}^{M-1} p(w_0)\, dw_0, \tag{3.22}$$

where

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2). \tag{3.23}$$

Therefore,

$$P_c \geq \int_{-\infty}^{\infty} \{\Pr[W_i > -\alpha]\}^{M-1} p(\alpha - \sqrt{2ZC^*})\, d\alpha = \int_{-\infty}^{\infty} \{1 - Q(a)\}^{M-1} p(\alpha - \sqrt{2ZC^*})\, d\alpha \tag{3.24}$$

$$P_e = 1 - P_c \leq \int_{-\infty}^{\infty} p(\alpha - \sqrt{2ZC^*}) \cdot \left\{ 1 - [1 - Q(\alpha)]^{M-1} \right\} d\alpha, \tag{3.25}$$

where the Q-function is given by

$$Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} \exp(-s^2/2)\, ds \tag{3.26}$$

and is upper bounded by [14, p. 84]

$$Q(\alpha) < \exp(-\alpha^2/2), \qquad \alpha > 0. \tag{3.27}$$

Note that

$$1 - [1 - Q(\alpha)]^{M-1} \leq (M-1)Q(\alpha) < M \exp(-\alpha^2/2), \qquad \alpha > 0 \tag{3.28}$$

[1] Although the codes in $\mathscr{C}$ are not necessarily triply orthogonal codes, the method of computation used by Chao for triply orthogonal codes can also be employed for the codes in $\mathscr{C}$ without any significant modification.

and

$$1 - [1 - Q(\alpha)]^{M-1} = 1 - \{\Pr[W_i > -\alpha]\}^{M-1} < 1. \tag{3.29}$$

Combining (3.25), (3.28), and (3.29) yields

$$P_e < \int_{-\infty}^{a} p(\alpha - \sqrt{2ZC^*})\, d\alpha + M\int_{a}^{\infty} p(\alpha - \sqrt{2ZC^*}) \cdot \exp(-\alpha^2/2)\, d\alpha, \qquad a > 0, \tag{3.30}$$

where a is an arbitrary constant. The constant a should be chosen to minimize the right-hand side of (3.30). Taking the derivative of the right-hand side of (3.30), and setting it to zero, yields

$$\exp(a^2/2) = M \quad \text{or} \quad a^2/2 = \ln M. \tag{3.31}$$

Now it follows from (3.23), (3.26), and (3.27) that

$$\int_{-\infty}^{a} p(\alpha - b)\, d\alpha = Q(b - a) \le \exp\left[-(a - b)^2/2\right] \quad \text{for } b \ge a. \tag{3.32}$$

It also follows from (3.23) and (3.26) that

$$\int_{a}^{\infty} p(\alpha - b) e^{-\alpha^2/2}\, d\alpha = \frac{1}{\sqrt{2}} \exp(-b^2/4) Q\left(\sqrt{2}\left[a - \frac{b}{2}\right]\right). \tag{3.33}$$

Combining (3.27) and (3.33) yields

$$\int_{a}^{\infty} p(\alpha - b) \exp(-\alpha^2/2)\, d\alpha < \begin{cases} \exp[-b^2/4], & a \le b/2, \\ \exp\left[-b^2/4 - (a - b/2)^2\right], & a \ge b/2. \end{cases} \tag{3.34}$$

Now let b be defined by

$$b = \sqrt{2ZC^*}. \tag{3.35}$$

Combining (3.30)–(3.35) yields

$$P_e < \begin{cases} \exp\left[-(a - b)^2/2\right] + \exp(a^2/2)\exp(-b^2/4), & 0 \le a \le b/2, \\ 2\exp\left[-(a - b)^2/2\right], & b/2 \le a \le b. \end{cases} \tag{3.36}$$

Observe that

$$\frac{(a - b)^2}{2} - \left(\frac{b^2}{4} - \frac{a^2}{2}\right) = \left(a - \frac{b}{2}\right)^2 \ge 0. \tag{3.37}$$

Thus,

$$P_e < \begin{cases} 2\exp(-b^2/4 + a^2/2), & 0 \le a \le b/2 \\ 2\exp\left[-(a - b)^2/2\right], & b/2 \le a \le b. \end{cases} \tag{3.38}$$

Finally, combining (3.31), (3.35), and (3.38), and using the fact that

$$R = \frac{\ln M}{Z} \tag{3.39}$$

yields

$$P_e < \begin{cases} 2\exp\left[-Z\left(\frac{C^*}{2} - R\right)\right], & 0 \le R \le \frac{C^*}{4}, \\ 2\exp\left[-Z(\sqrt{C^*} - \sqrt{R})^2\right], & \frac{C^*}{4} \le R \le C^* \end{cases} \tag{3.40}$$

Thus, it follows from (3.40) and (3.4) that the code $\mathscr{C}(M, M/2)$ is exponentially optimum as $Z \to \infty$. □

IV. Exponentially Optimum Codes for Class II Binary-Input / Binary-Output VNC's and Class II Binary Erasure VNC's

In this section we will construct exponentially optimum codes for channels that can be modeled as repeated uses of a class II binary-input/binary-output VNC or a class II binary erasure VNC. The result for binary-input/binary-output class II VNC's depends on a minor modification of the techniques developed by Wyner for the intensity modulated direct detection optical channel [3], [4]. The result of this paper extends Wyner's result to prove the exponential optimality of the family of code $\mathscr{C}$ on class II binary-input/binary-output VNC's. A key step in doing this is the inequality (4.13) which allows us to deal with Poisson statistics instead of the binomial statistics of VNC's.

A. Class II Binary-Input / Binary-Output VNC

Without loss of generality the probability transition matrix for a binary-input/binary-output class II VNC can be written as

$$\{p(y|x)\} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \epsilon(z)\begin{bmatrix} -\lambda & \lambda \\ -\mu & \mu \end{bmatrix}, \tag{4.1}$$

where λ and μ are arbitrary numbers with $0 < \lambda < \mu$. The capacity $C(z)$ of this channel is given by

$$C(z) = \max_{q}\left[\epsilon(z)\left(q_0\lambda \ln\frac{\lambda}{q_0\lambda + q_1\mu} + q_1\mu \ln\frac{\mu}{q_0\lambda + q_1\mu}\right)\right] + O(\epsilon^2(z)), \tag{4.2}$$

where $q = (q_0, q_1)$ is a channel input probability vector. As in Section III, we restrict ourselves to channels whose capacity per resource is nonzero, when they are operated

in the very noisy limit, i.e.,

$$C^* = \lim_{z\to 0} \frac{C(z)}{z} > 0 \quad \text{or} \quad \lim_{z\to 0} \frac{\epsilon(z)}{z} = c_2 > 0, \quad (4.3)$$

where c_2 is a strictly positive constant. Suppose that we are given a channel that is modeled as repeated uses of a binary-input/binary-output class II VNC satisfying (4.3). By direct substitution of (4.1) into (1.20), it is easy to verify that

$$E_0(\rho, \boldsymbol{q}) = \epsilon(z)\Big[q_0\lambda + q_1\mu - (q_0\lambda^{1/(1+\rho)} + q_1\mu^{1/(1+\rho)})^{1+\rho}\Big] + O(\epsilon^2(z)). \quad (4.4)$$

Then, it follows from (1.18), (2.13), (4.3), and (1.15) that the reliability function for this channel is given by

$$E^*(\boldsymbol{R}) = \max_{\boldsymbol{q}} \max_{0\le\rho\le 1} \Big[c_2\Big\{q_0\lambda + q_1\mu - (q_0\lambda^{1/(1+\rho)} + q_1\mu^{1/(1+\rho)})^{1+\rho}\Big\} - \rho\boldsymbol{R}\Big]. \quad (4.5)$$

In the following theorem, we show that the family $\mathscr{C}$ of binary codes constructed in Section III is exponentially optimum for channels which can be modeled as repeated uses of a binary-input/binary-output class II VNC.

Theorem 4.1: Given a channel that can be modeled as repeated uses of a binary-input/binary-output class II VNC satisfying (4.3), then for any rate $\boldsymbol{R}$, $\boldsymbol{R} < \boldsymbol{C}^*$, the code $\mathscr{C}(M, q_1^0 M)$, with $M = \exp[\boldsymbol{R}Z]$, is exponentially optimum as $Z \to \infty$, where $\boldsymbol{q}^0 = (q_0^0, q_1^0)$ denotes the optimal input probability vector that achieves the maximum in (4.5).

Proof: For any given input probability vector $\boldsymbol{q} = (q_0, q_1)$, let $k = q_1 M$. Consider now the code $\{\boldsymbol{x}_m = (x_{m1}, x_{m2}, \cdots, x_{mN})\colon m = 1, 2, \cdots, M\}$ in the family $\mathscr{C}$, with parameters (M, k) and $N = \binom{M}{k}$. Note that the resource per channel use z, for this code is given by

$$z = \frac{Z}{\binom{M}{q_1 M}} = \frac{1}{\boldsymbol{R}} \frac{\ln M}{\binom{M}{q_1 M}}. \quad (4.6)$$

Thus for fixed $\boldsymbol{R}$, M approaches infinity and z approaches zero as $Z \to \infty$. For each given m, let $S_m = \{n \in [1, \cdots, N]\colon x_{mn} = 1\}$. Then S_m is the set of positions in the mth binary code word which are equal to 1. By symmetry each row of the code matrix $\mathscr{A}$ has the same number of ones. Furthermore the total number of ones in the matrix is kN, thus each row (i.e., code word) must contain kN/M ones. If we let $|\cdot|$ denotes the cardinality of a set, then it follows that $|S_m| \to q_1 N$ and $|S_m^c| \to q_0 N$ as $Z \to \infty$, where S_m^c is the complement of set S_m. Also note that

$$\frac{|S_m \cap S_{m'}^c|}{\binom{M}{k}} = \frac{\binom{M-2}{k-1}}{\binom{M}{k}} = \frac{\left(1 - \frac{k}{M}\right)\frac{k}{M}}{\left(1 - \frac{1}{M}\right)} \to q_0 q_1 \quad \text{as } M \to \infty,\ m' \ne m,$$

$$\frac{|S_m \cap S_{m'}|}{\binom{M}{k}} = \frac{\binom{M-2}{k-2}}{\binom{M}{k}} = \frac{\frac{k}{M}\left(\frac{k}{M} - \frac{1}{M}\right)}{\left(1 - \frac{1}{M}\right)} \to q_1^2 \quad \text{as } M \to \infty,\ m' \ne m,$$

$$\frac{|S_m^c \cap S_{m'}^c|}{\binom{M}{k}} = \frac{\binom{M-2}{M-k-2}}{\binom{M}{k}} = \frac{\left(1 - \frac{k}{M}\right)\left(1 - \frac{k}{M} - \frac{1}{M}\right)}{\left(1 - \frac{1}{M}\right)} \to q_0^2 \quad \text{as } M \to \infty,\ m' \ne m. \quad (4.7)$$

Let $\boldsymbol{y} = (y_1, y_2, \cdots, y_N)$ be the received vector, and let Ψ_m denote the number of positions in which the code word $\boldsymbol{x}_m$ and the received vector $\boldsymbol{y}$ are both 1, i.e., $\Psi_m = |\{n \in [1, 2, \cdots, N] | y_n = 1, x_{mn} = 1\}|$. Similarly, let Φ_m denote the number of positions in which the code word $\boldsymbol{x}_m$ equals 0 and the received vector $\boldsymbol{y}$ equals 1, i.e., $\Phi_m = |\{n \in [1, 2, \cdots, N] | y_n = 1, x_{mn} = 0\}|$. Suppose the decoding rule is given by

$$\text{choose } \boldsymbol{x}_m \quad \text{iff} \quad \Psi_m > \Psi_{m'}, \qquad \forall\, m' \ne m.$$

Given m, define $E_{m'}$, $m' \ne m$, as the decoding error event $\{\Psi_{m'} \ge \Psi_m\}$. Then the probability of decoding error, $P_{e,m}$, given that code word $\boldsymbol{x}_m$ has been transmitted, is

$$P_{e,m} \le \Pr(\cup_{m' \ne m} E_{m'} | \boldsymbol{x}_m). \quad (4.8)$$

This probability of error will now be bounded by using a modified version of technique developed by Wyner for the direct detection Poisson channel [3]. Let $V_1 = \Psi_m$ and $V_0 = \Phi_m$. Then $V = V_1 + V_0$ is the total number of 1's in the received vector $\boldsymbol{y}$. Note that given $\boldsymbol{x}_m$, V_1 and $V_0 = V - V_1$ are mutually independent binomial random variables with distributions

$$\Pr(V_1 = l | \boldsymbol{x}_m) = b(l; Nq_1; \epsilon(z)\mu), \quad (4.9)$$

$$\Pr(V_0 = l | \boldsymbol{x}_m) = b(l; Nq_0; \epsilon(z)\lambda), \quad (4.10)$$

where $b(l; n; p)$ is the binomial distribution

$$b(l; n; p) = \binom{n}{l} p^l (1-p)^{n-l}. \quad (4.11)$$

Conditioning the probability in (4.8) on the two variables V and V_1, we have

$$\begin{aligned} P_{e,m} &\le \sum_{n=0}^{N} \sum_{n_1=0}^{n} \Pr[V = n, V_1 = n_1 | \boldsymbol{x}_m] \\ &\quad \cdot \Pr[\cup_{m' \ne m} E_m | \boldsymbol{x}_m, V = n, V_1 = n_1] \\ &= \sum_{n=0}^{N} \sum_{n_1=0}^{n} b(n - n_1; Nq_0; \epsilon(z)\lambda) \\ &\quad \cdot b(n_1; Nq_1; \epsilon(z)\mu) \\ &\quad \cdot \Pr\Big[\bigcup_{m' \ne m} E_m | \mathbf{X}_m, V = n, V = n_1\Big] \quad (4.12) \end{aligned}$$

Using the inequality [15, pp. 171–172]

$$b(l; n; p) < e^{-np}\frac{(np)^l}{l!}e^{(np+1)p}, \tag{4.13}$$

(4.12) becomes

$$P_{e,m} \le \exp\left[N\epsilon^2(z)(q_0\lambda^2 + q_1\mu^2) + \epsilon(z)(\lambda+\mu)\right]\cdot P', \tag{4.14}$$

where

$$P' = \sum_{n=1}^{N}\sum_{n_1=0}^{n}\frac{e^{-\Lambda}\Lambda^n}{n!}\binom{n}{n_1}\pi^{n_1}(1-\pi)^{n-n_1} \cdot \Pr\left[\bigcup_{m'\neq m} E_{m'}|\boldsymbol{x}_m, V=n, V_1=n_1\right], \tag{4.15}$$

$$\Lambda = N\epsilon(z)(q_0\lambda + q_1\mu), \tag{4.16}$$

$$\pi = \frac{q_1\mu}{q_0\lambda + q_1\mu}. \tag{4.17}$$

The set A and the function $Q(n_1, n)$ are now defined as follows

$$A = \{(n_1, n): 0 \le n_1 \le n, (n-n_1)q_1 - n_1q_0 < 0\}, \tag{4.18}$$

$$Q(n_1, n) = \frac{e^{-\Lambda}\Lambda^n}{n!}\binom{n}{n_1}\pi^{n_1}(1-\pi)^{n-n_1}. \tag{4.19}$$

Then, (4.15) can be written as

$$P' \le P_1' + P_2', \tag{4.20}$$

where

$$P_1' = \sum_{(n_1,n)\notin A} Q(n_1, n) \tag{4.21}$$

$$P_2' = \sum_{(n_1,n)\in A} Q(n_1,n)\Pr[\cup_{m'\neq m}E_{m'}|\boldsymbol{x}_m, n, n_1] \le \sum_{(n_1,n)\in A} Q(n_1,n)\left[\sum_{m'\neq m}\Pr \cdot[E_{m'}|\boldsymbol{x}_m, V=n, V_1=n_1]\right]^{\rho}, \quad 0\le\rho\le 1. \tag{4.22}$$

We now upper bound P_1' and P_2' using the Chernoff bound as in [3]. Applying the Chernoff bound on P_1' yields

$$P_1' = \Pr[V_0q_1 - V_1q_0 \ge 0] \le E[e^{(V_0q_1 - V_1q_0)\tau}], \tag{4.23}$$

where τ is any positive number, and E denotes the expectation operator. Since V_0 and V_1 are independent random variables given that code word $\boldsymbol{x}_m$ was transmitted, (4.23) becomes

$$P_1' \le E[e^{V_0q_1\tau}]E[e^{-V_1q_0\tau}] \tag{4.24}$$

$$= \sum_{l=0}^{Nq_0} e^{\tau q_1 l_b}(l; Nq_0; \epsilon(z)\lambda)\sum_{l=0}^{Nq_1} e^{-\tau q_0 l_b}(l; Nq_1; \epsilon(z)\mu). \tag{4.25}$$

Combining (4.13) and (4.25) yields

$$P_1' \le \Gamma\exp\left[N\epsilon^2(z)(q_0\lambda^2 + q_1\mu^2) + \epsilon(z)(\lambda+\mu)\right], \tag{4.26}$$

where

$$\Gamma \le \sum_{l=0}^{Nq_0} e^{-\Lambda_0}\frac{\Lambda_0^l}{l!}e^{q_1l\tau}\sum_{l=0}^{Nq_1} e^{-\Lambda_1}\frac{\Lambda_i^l}{l!}e^{-q_0l\tau}, \tag{4.27}$$

$$\Lambda_0 = Nq_0\epsilon(z)\lambda, \tag{4.28}$$

$$\Lambda_1 = Nq_1\epsilon(z)\mu. \tag{4.29}$$

It follows from (4.27) that

$$\Lambda < \sum_{l=0}^{\infty} e^{-\Lambda_0}\frac{\Lambda_0^l}{l!}e^{q_1l\tau}\sum_{l=0}^{\infty} e^{-\Lambda_1}\frac{\Lambda_1^l}{l!}e^{-q_0l\tau} = \exp\left[-(\Lambda_0+\Lambda_1) + \Lambda_0e^{\tau q_1} + \Lambda_1e^{-\tau q_0}\right]. \tag{4.30}$$

The right-hand side of (4.30) is minimized by differentiating it with respect to τ and setting the result to 0. This yields

$$e^{\tau} = \frac{\Lambda_1q_0}{\Lambda_0q_1}. \tag{4.31}$$

Combining (4.26)–(4.31), we have

$$P_1' \le \exp\left[-N\epsilon(z)(q_0\lambda + q_1\mu - \mu^{q_1}\lambda^{q_0})\right] \cdot \exp\left[N\epsilon^2(z)(q_0\lambda^2 + q_1\mu^2) + \epsilon(z)(\lambda+\mu)\right] = \exp\left[-Z\frac{\epsilon(z)}{z}(q_0\lambda + q_1\mu - \mu^{q_1}\lambda^{q_0}) + \delta_1\right], \tag{4.32}$$

where

$$\delta_1 = N\epsilon^2(z)(q_0\lambda^2 + q_1\mu^2) + \epsilon(z)(\lambda+\mu) = Z\frac{\epsilon^2(z)}{z}(q_0\lambda^2 + q_1\mu^2) + \epsilon(z)(\lambda+\mu). \tag{4.33}$$

Note that $\epsilon(z) \to 0$ and $\epsilon(z)/z \to c_2$ as $z \to 0$. Thus, it follows from (4.33) that

$$\frac{\delta_1}{Z} \to 0 \quad \text{as } Z \to \infty \text{ (i.e., as } z \to 0). \tag{4.34}$$

Thus, using (4.3), the bound in (4.32) can be written as

$$P_1' \le e^{-ZE_1^* + \Theta(Z)} \quad \text{as } Z \to \infty, \tag{4.35}$$

where

$$E_1^* = c_2(q_0\lambda + q_1\mu - \mu^{q_1}\lambda^{q_0}). \tag{4.36}$$

We now proceed to bound the term P_2' given by (4.22).

$$\Pr[E_{m'}|\boldsymbol{x}_m, V=n, V_1=n_1] = \Pr[\Psi_{m'} - \Psi_m \ge 0 \ge 0|\boldsymbol{x}_m, V=n, V_1=n_1]. \tag{4.37}$$

Let W_A be the number of positions in the set $S_m \cap S_{m'}^c$, where $y_n = 1$, i.e., $W_A = |\{n \in [1,2,\cdots,N] | n \in S_m \cap S_{m'}^c, y_n = 1\}|$, and W_c be the number of positions in the set $S_m^c \cap S_{m'}$, such that $y_n = 1$, i.e., $W_C = |\{n \in [1,2,\cdots,N] | n \in S_m^c \cap S_{m'}, y_n = 1\}|$. Then,

$$\Psi_{m'} - \Psi_m = W_C - W_A \tag{4.38}$$

Given x_m, n, and n_1, the random variables W_A and W_C are independent and have a binomial distribution. It fol-

lows from (4.7) that $|S_m \cap S_{m'}^c| = |S_m^c \cap S_{m'}| = Nq_1q_0$. Therefore,

$$\Pr[W_A = l|\boldsymbol{x}_m, V = n, V_1 = n_1] = b(l; n_1; p_a), \quad (4.39)$$

$$\Pr[W_C = l|\boldsymbol{x}_m, V = n, V_1 = n_1] = b(l; n - n_1; p_c), \quad (4.40)$$

where

$$p_a = \frac{|S_m \cap S_{m'}^c|}{|S_m|} = \frac{Nq_1q_0}{Nq_1} = q_0, \quad (4.41)$$

$$p_c = \frac{|S_m^c \cap S_{m'}|}{|S_m^c|} = \frac{Nq_0q_1}{Nq_0} = q_1. \quad (4.42)$$

Combining (4.22), (4.37), and (4.38) yields

$$P_2' \le \sum_{(n_1,n)\in A} Q(n_1,n)\Big[\sum_{m'\neq m} \Pr[W_C - W_A \ge 0|\boldsymbol{x}_m, V = n, V_1 = n_1]\Big]^{\rho}, \quad 0 \le \rho \le 1. \quad (4.43)$$

As was done with the P_1' term, the right-hand side of (4.43) is now upper bounded using the Chernoff bound. The result is [3, pp. 1458–1459]

$$P_2' \le M^{\rho}(\alpha + \beta)^{1+\rho} \exp\left[-\Lambda + \Lambda(\alpha+\beta)^{1+\rho}\right] \quad \text{for any } \rho, 0 \le \rho \le 1, \quad (4.44)$$

where

$$\alpha = (q_1^{\rho}\pi)^{1/(1+\rho)}, \quad (4.45)$$

$$\beta = [(1-\pi)q_0^{\rho}]^{1/(1+\rho)} \quad (4.46)$$

and Λ and π are given by (4.16) and (4.17), respectively. Combining (4.16), (4.17), and (4.44)–(4.46) yields

$$\begin{aligned} P_2' &\le M^{\rho}\exp\Big[-N\epsilon(z)\Big\{(q_0\lambda + q_1\mu) - (q_0\lambda^{1/(1+\rho)} + q_1\mu^{1/(1+\rho)})^{1+\rho}\Big\} + \delta_2\Big] \\ &= \exp\left[-Z\left\{\frac{\epsilon(z)}{z}\Big[(q_0\lambda + q_1\mu) - (q_1\mu^{1/(1+\rho)} + q_0\lambda^{1/(1+\rho)})^{1+\rho}\Big] - \rho\boldsymbol{R}\right\} + \delta_2\right], \end{aligned} \quad (4.47)$$

where

$$\delta_2 = \ln\left[N\epsilon(z)(q_0\lambda^{1/(1+\rho)} + q_1\mu^{1/(1+\rho)})^{1+\rho}\right] = \ln\left[Z\frac{\epsilon(z)}{z}(q_0\lambda^{1/(1+\rho)} + q_1\mu^{1/(1+\rho)})^{1+\rho}\right]. \quad (4.48)$$

Note that $\delta_2/Z \to 0$ as $Z \to \infty$. Thus, using (4.3), the bound in (4.47) can be written as

$$P_2' \le e^{-ZE_2^*(\boldsymbol{R};\rho,\boldsymbol{q})+\Theta(Z)} \quad \text{as } Z \to \infty, \quad (4.49)$$

where

$$E_2^*(\boldsymbol{R};\rho,\boldsymbol{q}) = c_2\Big[(q_0\lambda + q_1\mu) - (q_1\mu^{1/(1+\rho)} + q_0\lambda^{1/(1+\rho)})^{1+\rho}\Big] - \rho\boldsymbol{R}. \quad (4.50)$$

Combining (4.14), (4.20), (4.33)–(4.35), and (4.49) yields the following upper bound on the code word error probability $P_{e,m}$:

$$P_{e,m} \le e^{-ZE_1^*+\Theta(Z)} + e^{-ZE_2^*(\boldsymbol{R};\rho,\boldsymbol{q})+\Theta(Z)} \quad \text{as } Z \to \infty. \quad (4.51)$$

$E_2^*(\boldsymbol{R};\rho,\boldsymbol{q})$ can now be upper bounded as indicated below [3].

$$\begin{aligned} E_2^*(\boldsymbol{R};\rho,\boldsymbol{q}) &\le c_2\Big[(q_0\lambda + q_1\mu) - (q_1\mu^{1/(1+\rho)} + q_0\lambda^{1/(1+\rho)})^{1+\rho}\Big] \\ &\le c_2\Big[(q_0\lambda + q_1\mu) - \big(q_1\sqrt{\mu} + q_0\sqrt{\lambda}\big)^2\Big] \\ &= c_2q_1q_0\big(\sqrt{\lambda} - \sqrt{\mu}\big)^2. \end{aligned} \quad (4.52)$$

Thus, combining (4.36) and (4.52) yields

$$E_1^* - E_2^*(\boldsymbol{R};\rho,\boldsymbol{q}) \ge c_2\lambda\Big[(q_0 + q_1t)^2 - t^{2q_1}\Big] \ge c_2\lambda[(q_0 + q_1t) - t^{q_1}] \ge 0, \quad (4.53)$$

where

$$t = 4\sqrt{\frac{\mu}{\lambda}} \quad \text{(note that } t > 1, \text{ since } \mu > \lambda) \quad (4.54)$$

The last inequality in (4.52) follows from the that $t > 1$, t^{q_1} is concave downward in t, and t^{q_1} is tangent to the straight line $q_0 + q_1t$ at $t = 1$. Combining (4.53) and (4.51) yields

$$P_{e,m} \le e^{-ZE_2^*(\boldsymbol{R};\rho,\boldsymbol{q})+\Theta(Z)} \quad 0 \le \rho \le 1 \text{ as } Z \to \infty. \quad (4.55)$$

for the code $\mathscr{C}(M, q_1M)$. Since the bound (4.55) is valid for any $0 \le \rho \le 1$ and for any choice of $\boldsymbol{q}$, we have for the code $\mathscr{C}(M, q_1^oM)$

$$P_{e,m} \le \exp\left[-Z\Big\{\max_{0\le\rho\le1} E_2^*(\boldsymbol{R};\rho,\boldsymbol{q}^o)\Big\} + \Theta(Z)\right] \quad \text{as } Z \to \infty. \quad (4.56)$$

Note that the quantity $\max_{0\le\rho\le1} E_2^*(\boldsymbol{R};\rho,\boldsymbol{q}^o) = \max_{\boldsymbol{q}}\max_{0\le\rho\le1} E_2^*(\boldsymbol{R};\rho,\boldsymbol{q})$ in (4.56) is simply that reliability function $E^*(\boldsymbol{R})$ of the channel given by (4.5). Thus the code is exponentially optimum as $Z \to \infty$. □

B. Class II Binary Erasure VNC

The probability transition matrix for a class II binary erasure VNC is given by

$$\{p(y|x)\} = \begin{bmatrix}0 & 0 & 1\\ 0 & 0 & 1\end{bmatrix} + \epsilon(z)\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & -1\end{bmatrix} + O(\epsilon^2(z)). \quad (4.57)$$

The capacity $C(z)$ of this channel can be easily computed and is given by

$$C(z) = \epsilon(z)\ln 2 + O(\epsilon^2(z)). \tag{4.58}$$

Suppose that we are given a channel that is modeled as repeated uses of a class II binary erasure VNC satisfying (4.3). By direct substitution of (4.57) into (1.20), it is easy to verify that

$$E_0(\rho, \boldsymbol{q}) = \epsilon(z)\left[1 - (q_0^{1+\rho} + q_1^{1+\rho})\right] + O(\epsilon^2(z)). \tag{4.59}$$

Then, it follows from (1.18), (2.13), (1.15), and (4.3) that the reliability function for this channel is given by

$$\boldsymbol{E}^*(\boldsymbol{R}) = \max_{\boldsymbol{q}} \max_{0\le\rho\le 1} \left[c_2\{1 - (q_0^{1+\rho} + q_1^{1+\rho})\} - \rho\boldsymbol{R}\right]. \tag{4.60}$$

Performing the maximization over $\boldsymbol{q}$ in (4.60) yields

$$\boldsymbol{E}^*(\boldsymbol{R}) = \max_{0\le\rho\le 1} [c_2\{1 - 2^{-\rho}\} - \rho\boldsymbol{R}]. \tag{4.61}$$

In the following theorem, we show that the family of binary codes, $\mathscr{C}$, is exponentially optimum for channels which can be modeled as repeated uses of a class II binary erasure VNC given by (4.57).

Theorem 4.2: Given a channel that can be modeled as repeated uses of a class II binary erasure VNC satisfying (4.3), then for any rate $\boldsymbol{R}$, $\boldsymbol{R} < \boldsymbol{C}^*$, the code $\mathscr{C}(M, M/2)$, with $M = \exp[(\boldsymbol{RZ}]$, is exponentially optimum as $Z \to \infty$.

Proof: Consider the code $\mathscr{C}(M, M/2) = \{\boldsymbol{x}_m = (x_{m1},\cdots, x_{mN})\colon m = 1,2,\ldots,M\}$ in the family $\mathscr{C}$. The block length of this code is $N = \binom{M}{M/2}$, and the resource per channel use, z, for this code is given by

$$z = \frac{Z}{\binom{M}{M/2}} = \frac{1}{\boldsymbol{R}}\frac{\ln M}{\binom{M}{M/2}}. \tag{4.62}$$

Thus, for fixed $\boldsymbol{R}$, M approaches infinity and z approaches zero as $Z \to \infty$. Let $S_m = \{n \in [1,2,\cdots,N] | x_{mn} = 0\}$. Then, $|S_m| = |S_m^c| = N/2$ for the given $\mathscr{C}(M, M/2)$ code. Let $\boldsymbol{y} = \{y_1, y_2,\cdots, y_N\}$ be the received N-vector, and let Ψ_m denote the number of positions in which the code word $\boldsymbol{x}_m$ and the received vector y are both 0, i.e., $\Psi_m = |\{n \in [1,2,\cdots,N] | y_n = 0, x_{mn} = 0\}|$. Similarly, let Φ_m denote the number of positions in which the code word $\boldsymbol{x}_m$ and the received vector $\boldsymbol{y}$ are both 1, i.e., $\Phi_m = |\{n \in [1,2,\cdots,N] | y_n = 1,\ x_{mn} = 1\}|$. Then, given that the code word $\boldsymbol{x}_m$ has been transmitted, Ψ_m and Φ_m are mutually independent random variables with distributions

$$\Pr(\Psi_m = k|\boldsymbol{x}_m) = b(k, N/2; \epsilon(z)), \tag{4.63}$$

$$\Pr(\Phi_m = k|\boldsymbol{x}_m) = b(k, N/2; \epsilon(z)), \tag{4.64}$$

where $b(\ )$ denotes the binomial distribution given by (4.11). Suppose the decoder uses the following decoding rule:

choose code word m iff $\Psi_m + \Phi_m > \Psi_{m'} + \Phi_{m'}$, for all $m' \neq m$.

Given m, define $E_{m'}$, $m' \neq m$, to be the event $\{\Psi_{m'} + \Phi_{m'} \ge \Psi_m + \Phi_m\}$. Then the probability of a decoding error $P_{e,m}$ can be upper bounded as follows

$$\begin{aligned} P_{e,m} &\le \Pr\left\{\bigcup_{m'\neq m} E_{m'}|\boldsymbol{x}_m\right\} \\ &= \sum_{k=0}^{N/2}\sum_{l=0}^{N/2} \Pr\{\Psi_m = k, \Phi_m = l|\boldsymbol{x}_m\} \\ &\quad\cdot \Pr\left\{\bigcup_{m'\neq m} E_{m'}|\boldsymbol{x}_m, \Psi_m = k, \Phi_m = l\right\} \\ &\le \sum_{k=0}^{N/2}\sum_{l=0}^{N/2} b(k, N/2; \epsilon(z))b(l, N/2; \epsilon(z)) \\ &\quad\cdot\left[\sum_{m'\neq m} \Pr\{E_{m'}|\boldsymbol{x}_m, \Psi_m = k, \Phi_m = l\}\right]^{\rho} \end{aligned} \tag{4.65}$$

for any ρ, $0 \le \rho \le 1$. In the binary erasure channel, it is impossible that an input 0 is received as 1 or an input 1 is received as 0. Thus, given that $\boldsymbol{x}_m$ has been transmitted, we must have $\Psi_{m'} \le \Psi_m$ and $\Phi_{m'} \le \Phi_m$. Therefore, the event $E_{m'}$ occurs if and only if $\Psi_{m'} = \Psi_m$ and $\Phi_{m'} = \Phi_m$. That is, given the error event $E_{m'}$ and the conditioning event $\{\boldsymbol{x}_m, \Psi_m = k, \Phi_m = l\}$, each of the k 0's in the received vector $\boldsymbol{y}$ must occur in the common region $S_m \cap S_{m'}$, and each of the l 1's in the received vector $\boldsymbol{y}$ must occur in the common region $S_m^c \cap S_{m'}^c$. Since by (4.7) $|S_m \cap S_{m'}| = N/4$, and since, given that $\Psi_m = k$, each of the k 0's are equally likely to occur in any position in S_m, it follows that

$$\Pr\{k \text{ 0's in } S_m \cap S_{m'}|\Psi_m = k\} = 2^{-k}. \tag{4.66}$$

Similarly,

$$\Pr\{l \text{ 1's in } S_m^c \cap S_{m'}^c|\Phi_m = l\} = 2^{-l}. \tag{4.67}$$

Combining (4.65)–(4.67) and (4.13) yields

$$\begin{aligned} P_{e,m} &\le e^{N\epsilon^2(z)+2\epsilon(z)}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \\ &\quad\cdot e^{-N\epsilon(z)/2}\frac{\left(\frac{N\epsilon(z)}{2}\right)^k}{k!} \\ &\quad\cdot e^{-N\epsilon(z)/2}\frac{\left(\frac{N\epsilon(z)}{2}\right)^l}{l!}(M-1)^{\rho}2^{-k\rho}2^{-l\rho} \\ &\le \exp[N\epsilon^2(z) + 2\epsilon(z)]M^{\rho} \\ &\quad\cdot\exp[-N\epsilon(z)]\left(\sum_{k=0}^{\infty}\frac{\left(\frac{N\epsilon(z)}{2}2^{-\rho}\right)^k}{k!}\right)^2 \\ &= \exp[N\epsilon^2(z) + 2\epsilon(z)]\exp[\rho\boldsymbol{RZ}] \\ &\quad\cdot\exp[-N\epsilon(z)(1-2^{-\rho})] \\ &= \exp\left[-Z\left\{\frac{\epsilon(z)}{z}(1-2^{-\rho}) - \rho\boldsymbol{R}\right\} + \delta_3\right], \end{aligned} \tag{4.68}$$

where

$$\delta_3 = N\epsilon^2(z) + 2\epsilon(z) = Z\frac{\epsilon^2(z)}{z} + 2\epsilon(z). \quad (4.69)$$

Note that $\delta_3/Z \to 0$ as $z \to 0$. Thus, using (4.3), we may write (4.68) as

$$P_{e,m} \le \exp[-Z\{c_2(1-2^{-\rho}) - \rho R\} + \Theta(Z)], \quad \text{as } Z \to \infty. \quad (4.70)$$

Since the bound in (4.70) is valid for any $0 \le \rho \le 1$, we have

$$P_{e,m} \le \exp\left[-Z \max_{0\le\rho\le 1}\{c_2(1-2^{-\rho}) - \rho R\} + \Theta(Z)\right] = \exp[-ZE^*(R) + \Theta(Z)], \quad \text{as } Z \to \infty, \quad (4.71)$$

where $E^*(R)$ is the reliability function given in (4.61). Thus the code is exponentially optimum. □

V. Reliability Function and the Exponentially Optimum Code for the Polarization Modulated Direct Detection Optical Channel (PM DDOC)

The ultimate efficiency of an optical communication system can be specified as the minimum number of photons/bit required to achieve an arbitrary small probability of error. For the intensity modulated direct detection optical channel (IM DDOC), the message signal modulates the intensity of the laser output, and the receiver detects photon arrivals. In [3], [4], it is shown that the most efficient modulation for this channel is binary on/off keying. At very high data rates, however, it is often impractical to intensity modulate a laser directly. In these cases, the laser is run at constant output power, and the modulation is done externally. Consequently, the laser output is simply "thrown away" during an off pulse.

For the polarization modulated direct detection optical channel (PM DDOC), the message signal modulates the polarization angle, $\theta(t)$, of the optical field instead of its intensity. Thus, the average power transmitted remains constant and no laser power is discarded. The receiver consists of two photon detectors, one for each orthogonal polarization component (denoted ν and μ) of the transmitted field. Polarization modulation reduces to intensity modulation when the receiver consists of only a single photon detector for either the ν or μ polarization component. Thus, IM is a special case of PM and must have a lower efficiency. In this section the reliability function of the PM DDOC will be computed, and an exponentially optimum code constructed.

For PM, the intensity of the ν and μ polarization components can be written as

$$\lambda_\nu(t) = \lambda_s \cos^2\theta(t) \quad \text{and} \quad \lambda_\mu(t) = \lambda_s \sin^2\theta(t) \quad (5.1)$$

where λ_s equals the total number of photons transmitted per second, and $\theta(t)$ is the polarization angle of the transmitted field. The channel output consists of two independent Poisson counting processes, $\nu(t)$ and $\mu(t)$, with intensities equal to $\lambda_s \cos^2\theta(t)$ and $\lambda_s \sin^2\theta(t)$, respectively.

A (T, R) code is defined as a set of $M = \exp[RT]$ waveforms $\{\theta_m(t),\ m = 1,2,\cdots,M\}$ where each waveform is of duration T seconds, and satisfies the constraint $0 \le \theta_m(t) \le \pi/2$, $\forall\ t \in [0,T]$ and $\forall\ m$. A decoder mapping D is a function D: $\{(\nu(t), \mu(t)),\ 0 \le t \le T\} \to \{1,2,\cdots,M\}$. The probability of a decoding error, when code word $\theta_n(t)$ is transmitted and the decoder mapping is D, is denoted by $P_{e,n}(\{\theta_m(\cdot)\}; D)$. Intuitively, there should be little loss in performance if the waveforms $\{\theta_m(t),\ m = 1,2,\cdots,M\}$ are required to be piecewise constant over very short intervals. In addition, if the waveforms are piecewise constant then the number of photons detected at the receiver in each of these intervals is a sufficient decision statistic. Thus, we have the following theorem, which may be rigorously proved using the development given in [4].

Theorem 5.1: Given T and R, let $M = \exp[RT]$. Let $\{\theta_m(t),\ m = 1,2,\cdots,M\}$ be a (T, R) code with decoder mapping D. Then, for any $\delta > 0$, there exists another (T, R) code $\{\theta'_m(t),\ m = 1,2,\cdots,M\}$ with a decoder mapping D' for which $P_{e,n}(\{\theta'_m\}; D') < P_{e,n}(\{P_{e,n}(\{\theta_m\}; D) + \delta$, $\forall\ n$, and such that for some sufficiently large N, (a) the waveform $\theta'_m(t)$, for each m, is constant on the subinterval $[(k-1)\Delta, k\Delta)$, $1 \le k \le N$, and (b) the decoder mapping D' depends only on N uniformly spaced samples of the output process, $\{(\nu(k\Delta), \mu(k\Delta)),\ k = 1,2,\cdots,N\}$, where $\Delta = T/N$. □

It follows from Theorem 5.1 that we can model the PM DDOC as a discrete-time memoryless channel, where each channel use corresponds to Δ seconds. The input alphabet is $A_X = \{\theta:\ 0 \le \theta \le \pi/2\}$, and the output alphabet is the set of all nonnegative integer 2-tuples (n_ν, n_μ), where n_ν and n_μ denote the number of photons detected by the ν and μ detectors, respectively, during the interval of time Δ. Let $y = 0$, 1, and 2 denote the channel output events $\{n_\nu = 1,\ n_\mu = 0\}$, $\{n_\nu = 0,\ n_\mu = 1\}$, and $\{n_\nu = 0,\ n_\mu = 0\}$, respectively. The complement of the union of these three events will be denoted $y = 3$. The transition probabilities of the channel are then given by

$$p(0|\theta) = \lambda_s\Delta(\cos^2\theta)e^{-\lambda_s\Delta},$$
$$p(1|\theta) = \lambda_s\Delta(\sin^2\theta)e^{-\lambda_s\Delta},$$
$$p(2|\theta) = e^{-\lambda_s\Delta}, \qquad p(3|\theta) = 1 - e^{-\lambda_s\Delta} - \lambda_s\Delta e^{-\lambda_s\Delta}. \quad (5.2)$$

Since $\Delta \to 0$, (5.2) can be written as

$$p(0|\theta) = \epsilon(\Delta)\cos^2\theta, \qquad p(1|\theta) = \epsilon(\Delta)\sin^2\theta,$$
$$p(2|\theta) = 1 - \epsilon(\Delta) + O(\epsilon^2(\Delta)),$$
$$p(3|\theta) = O(\epsilon^2(\Delta)), \quad (5.3)$$

where

$$\epsilon(\Delta) = \lambda_s\Delta. \quad (5.4)$$

The terms of order $\epsilon^2(\Delta)$ may be neglected in (5.3), and the channel output reduces to only three possibilities $y = 0$, 1,or 2. Note that this is a class II VNC, with resource per channel use Δ. The following theorem [16, pp. 96, 145] may now be invoked.

Theorem 5.2: Let m be the smallest number of inputs that can be used with nonzero probability to achieve the capacity (or random coding error exponent) for a discrete-input memoryless channel whose output alphabet size is J. Let A be such a set of inputs. Then $m \leq J$, and the input probability assignment on A to achieve capacity (or random coding error exponent) using only inputs in A is unique.

It follows from Theorem 5.2 and (2.13) that the capacity (and the reliability function) of the PM DDOC is achieved using at most three inputs. Let these three inputs be denoted by θ_0, θ_1, and θ_2. Then (5.3) reduces to

$$p(0|\theta_i) = \epsilon(\Delta)z_i, \qquad p(1|\theta_i) = \epsilon(\Delta)(1 - z_i),$$
$$p(2|\theta_i) = 1 - \epsilon(\Delta), \qquad i = 0, 1, 2, \tag{5.5}$$

where

$$z_i = \cos^2 \theta_i, \qquad i = 0, 1, 2. \tag{5.6}$$

The mutual information of the channel (in nats) may be calculated and is equal to

$$I(X:Y) = \epsilon(\Delta)\left[h\left(\sum_{i=0}^{2} q_i z_i\right) - \sum_{i=0}^{2} q_i h(z_i)\right], \tag{5.7}$$

where $h(x) = -x \ln x - (1 - x)\ln(1 - x)$ is the binary entropy function, and the q_i is the probability of the input θ_i. By the Kuhn–Tucker conditions the input probabilities which maximize the mutual information, i.e., achieve capacity must satisfy [16, p. 87]

$$\frac{\partial I(X:Y)}{\partial q_i} = \epsilon(\Delta)\left[\left\{\ln\left(1 - \sum_{i=0}^{2} q_i z_i\right) - \ln\left(\sum_{i=0}^{2} q_i z_i\right)\right\} \cdot z_i - h(z_i)\right] \leq \text{const} \qquad \forall\, i, \tag{5.8}$$

with equality if $q_i \neq 0$. If $(q_0, q_1$, and q_2 are all nonzero, then the Kuhn–Tucker condition cannot be satisfied. To see this suppose that the Kuhn–Tucker condition is satisfied with q_0, q_1, and q_2 all nonzero. Then (5.8) may be written as

$$az_i - h(z_i) = c, \qquad i = 0, 1, 2, \tag{5.9}$$

where a and c are constants, and no two z_i's are identical. The z_i solutions to (5.9) are given by the intersection of a straight line, $y = az - c$, with the binary entropy function $y = h(z)$. The binary entropy function, however, is concave, and consequently it can intersect a straight line at no more than two points. Thus, the channel capacity is achieved using only two inputs, say θ_0 and θ_1. Equation (5.7) then becomes

$$I(X:Y) = \epsilon(\Delta)[h(q_0 z_0 + q_1 z_1) - \{q_0 h(z_0) + q_1 h(z_1)\}]. \tag{5.10}$$

The binary entropy function, h, is always positive and has maximum value ln 2, therefore $I(X:Y) \leq \epsilon(\Delta)\ln 2$. Thus, $I(X:Y)$ is maximized by choosing $q_0 = q_1 = 1/2$, $z_0 = 0$, and $z_1 = 1$, which implies $\theta_0 = 0$ and $\theta_1 = \pi/2$. Substitution of these values in (5.10) yields the capacity $C(\Delta)$, in nats/c.u., for the equivalent VNC as

$$C(\Delta) = \epsilon(\Delta)\ln 2 \text{ nats/c.u.} \tag{5.11}$$

Thus, the capacity in nats/s in the very noisy limit is

$$C^* = \lim_{\Delta \to 0} \frac{\delta(\Delta)\ln 2}{\Delta} = \lambda_x \ln 2 \text{ nats/s}. \tag{5.12}$$

Note that since $\epsilon(\Delta)/\Delta$ does not depend on Δ, we have $C = C^*$ for this channel. That is, the channel capacity per unit time, C of the PM DDOC is achieved in the very noisy limit as $\Delta \to 0$.

Since this channel is a class II VNC, $E(R) = E_r(R)$ for $R < C$ by the result of Section II. It follows from Theorem 5.1 that the error exponent, $E(R)$, is achieved using no more than three channel inputs. According to (1.18), $E(R)$ is achieved if and only if the inputs and their probability of use are chosen to maximize $E_o(\rho, \boldsymbol{q})$, or equivalently to minimize the function $F_o(\rho, \boldsymbol{q}) \triangleq \exp[-E_o(\rho, \boldsymbol{q})]$. The function F is convex cap in $\boldsymbol{q}$ [16, p. 144]. Thus it follows from the Kuhn–Tucker conditions that a necessary and sufficient condition on the probability vector $\boldsymbol{q}$ for $E_o(\rho, \boldsymbol{q})$ to be maximized is

$$\frac{\partial F}{\partial q_i} \leq \text{const}, \qquad i = 0, 1, 2 \tag{5.13}$$

with equality if and only if $q_i \neq 0$. $F_o(\rho, \boldsymbol{q})$ can be computed using (1.20) and (5.5), and the result is

$$F_o(\rho, \boldsymbol{q}) = (1 - \epsilon(\Delta)) + \epsilon(\Delta)\left[\sum_{i=0}^{2} q_i z_i^{1/(1+\rho)}\right]^{1+\rho} + \epsilon(\Delta)\left[\sum_{i=0}^{2} q_i (1 - z_i)^{1/(1+\rho)}\right]^{1+\rho}. \tag{5.14}$$

Combining (5.13) and (5.14) yields

$$az_i^{1/(1+\rho)} + b(1 - z_i)^{1/(1+\rho)} = c \qquad \forall\, i \text{ such that } q_i \neq 0, \tag{5.15}$$

where

$$a = \left(\sum_{i=0}^{2} q_i z_i^{1/(1+\rho)}\right)^{\rho}, \qquad b = \left(\sum_{i=0}^{2} q_i (1 - z_i)^{1/(1+\rho)}\right)^{\rho}, \tag{5.16}$$

and c is a constant. Given $q_i \neq 0$, $i = 0, 1, 2$, however, the three simultaneous equations (5.15) cannot be satisfied with three distinct z_i's. To see this let $x_i = z_i^{1/(1+\rho)}$ and $y_i = (1 - z_i)^{1/(1+\rho)}$. Then, the simultaneous equations

(5.15) imply that the line

$$ax_i + by_i = c \tag{5.17}$$

and the convex downward curve

$$y_i = (1 - x_i^{1+\rho})^{1/(1+\rho)} \tag{5.18}$$

must intersect at three distinct points, (x_0, y_0), (x_1, y_1), and (x_2, y_2), which is impossible. Thus one of the q_i's, say q_2, must be zero. Then, from (5.14), we have

$$\begin{aligned} F_0(\rho, \boldsymbol{q}) &= (1 - \epsilon(\Delta)) + \epsilon(\Delta)\left[\sum_{i=0}^{1} q_i x_i\right]^{1+\rho} \\ &\quad + \epsilon(\Delta)\left[\sum_{i=0}^{1} q_i y_i\right]^{1+\rho} \\ &\geq (1 - \epsilon(\Delta)) + 2\epsilon(\Delta) \\ &\quad \cdot \sqrt{\left[\sum_{i=0}^{1} q_i x_i\right]^{1+\rho}\left[\sum_{i=0}^{1} q_i y_i\right]^{1+\rho}}, \end{aligned} \tag{5.19}$$

where the equality holds if and only if

$$\sum_{i=0}^{1} q_i x_i = \sum_{i=0}^{1} q_i y_i. \tag{5.20}$$

The inequality in (5.19) follows from the fact that the average of two positive numbers is never smaller than their geometric mean. For any given $0 \leq \rho \leq 1$ we need to choose the q_i's and z_i's to minimize $F_o(\rho, \boldsymbol{q})$. Note that condition (5.20) is satisfied for any given ρ by $q_0 = q_1 = 1/2$, $x_0 = 0$, and $x_1 = 1$. Thus the optimum choice is $\theta_0 = 0$ and $\theta_1 = \pi/2$. With this choice $E_0(\rho, \boldsymbol{q})$ becomes

$$\begin{aligned} E_0(\rho, \boldsymbol{q}) &= -\ln[1 - \epsilon(\Delta)(1 - 2^{-\rho})] + O(\epsilon^2(\Delta)) \\ &= \epsilon(\Delta)(1 - 2^{-\rho}) + O(\epsilon^2(\Delta)). \end{aligned} \tag{5.21}$$

Therefore,

$$E(R) = \max_{0 \leq \rho \leq 1} [\epsilon(\Delta)(1 - 2^{-\rho}) - \rho R + O(\epsilon^2(\Delta))], \quad 0 \leq R \leq C(\Delta) \tag{5.22}$$

Since $\epsilon(\Delta)/\Delta$ does not depend on Δ, it follows from (5.22), (5.4), (1.14), and (1.15) that

$$\boldsymbol{E}(\boldsymbol{R}) = \boldsymbol{E}^*(\boldsymbol{R}) = \max_{0 \leq \rho \leq 1} \left[\lambda_s\left(1 - \left(\frac{1}{2}\right)^{\rho}\right) - \rho \boldsymbol{R}\right], \quad 0 \leq \boldsymbol{R} \leq \boldsymbol{C}^* \tag{5.23}$$

Performing the maximization over ρ in (5.23) yields

$$\boldsymbol{E}(\boldsymbol{R}) = \boldsymbol{E}^*(\boldsymbol{R}) = \begin{cases} \dfrac{\lambda_s}{2} - \boldsymbol{R}, & 0 \leq \boldsymbol{R} < \lambda_s \dfrac{\ln 2}{2} \\ \lambda_s - \left[1 + \ln\left(\dfrac{\lambda_s \ln 2}{\boldsymbol{R}}\right)\right]\dfrac{\boldsymbol{R}}{\ln 2}, & \lambda_s \dfrac{\ln 2}{2} \leq \boldsymbol{R} < \lambda_s \ln 2 \end{cases} \tag{5.24}$$

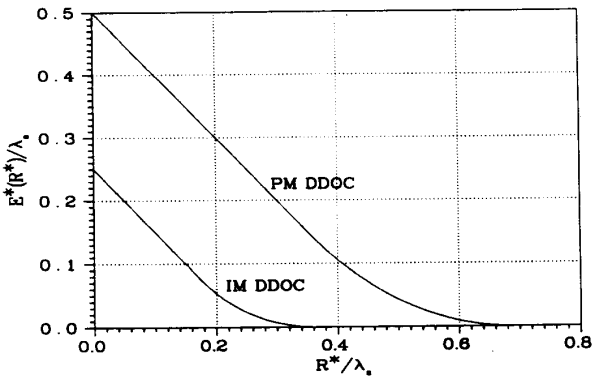

Fig. 4. Reliability function for PM DDOC.

which is plotted in Fig. 4. For comparison Fig. 4 also shows $\boldsymbol{E}(\boldsymbol{R})$ for the intensity modulation direct detection optical channel with peak transmission power λ_s photons/s [3].

It follows from the above discussion that the PM direct detection optical channel can be modeled as repeated uses of a class II binary erasure VNC with the resource per channel use being Δ and $\epsilon(\Delta) = \lambda_s \Delta$. Thus, using Theorem 4.2, we can easily construct an exponentially optimum code for this channel. For any given rate $\boldsymbol{R} < \boldsymbol{C}$, in nats/s, let $M = \exp[\boldsymbol{R}T]$ where T is sufficiently large. We now choose a binary code $\mathscr{C}(M, M/2)$ from the family of codes, $\mathscr{C}$. Let $\{\boldsymbol{x}_m: m = 1, 2, \cdots, M\}$ denotes the code words of the code $\mathscr{C}(M, M/2)$. Then construct a $(T, \boldsymbol{R})$ waveform code $\{\theta_m(t): m = 1, 2, \cdots, M\}$ by letting $\theta_m(t) = (\pi/2)x_{mn}$, $t \in [(n-1)\Delta, n\Delta)$, $1 \leq n \leq N$, $1 \leq m \leq M$, where $\Delta = T/N$ and $N = \binom{M}{M/2}$. Using this waveform code over the PM DDOC is equivalent to using the binary code $\mathscr{C}(M, M/2)$ over the modeling channel, i.e., the class II binary erasure VNC. Thus, by Theorem 4.2, this code is exponentially optimum.

As a final point, we note that the Bhattacharyya distance is a useful quantity for designing coded waveforms and bounding the probability or error in communication systems. For a channel with input alphabet $\mathscr{X}$, and output $\mathscr{Y}$, and transition probabilities $p(y/x)$, the Bhattacharyya distance between two channel inputs $x_i, x_j \in \mathscr{X}$ is given by

$$d_{ij} = -\ln \sum_{y \in \mathscr{Y}} \sqrt{p(y|x_i)p(y|x_j)} \tag{5.25}$$

We will derive the Bhattacharyya distance for the PM

DDOC. Let the two channel inputs be the polarization angles θ_i and θ_j, respectively. Then

$$d_{ij} = -\ln \sum_{n_\nu=0}^{\infty} \sum_{n_\mu=0}^{\infty} \sqrt{p(n_\nu, n_\mu|\theta_i)p(n_\nu, n_\mu|\theta_j)}, \quad (5.26)$$

where n_ν and n_μ denote the numbers of photons received during some observation interval by the ν and μ detectors, respectively. For a given transmitted polarization angle, the quantities n_ν and n_μ are mutually independent Poisson random variables. Thus,

$$p(n_\nu, n_\mu|\theta) = \frac{(\lambda_s T \cos^2\theta)^{n_\nu} e^{-(\lambda_s T \cos^2\theta)}}{n_\nu!} \cdot \frac{(\lambda_s T \sin^2\theta)^{n_\mu} e^{-(\lambda_s T \sin^2\theta)}}{n_\mu!}, \quad (5.27)$$

where λ_s is the total number of photons transmitted per second and T is the observation interval. Combining (5.26) and (5.27) yields

$$\begin{aligned} d_{ij} &= -\ln\left[e^{-\lambda_s T} \sum_{n_\nu=0}^{\infty} \frac{(\lambda_s T \cos\theta_i \cos\theta_j)^{n_\nu}}{n_\nu!} \cdot \sum_{n_\mu=0}^{\infty} \frac{(\lambda_s T \sin\theta_i \sin\theta_j)^{n_\mu}}{n_\mu!}\right] \\ &= -\ln\left[e^{-\lambda_s T} e^{\lambda_s T \cos\theta_i \cos\theta_j} e^{\lambda_s T \sin\theta_i \sin\theta_j}\right] \\ &= \lambda_s T\left[1 - \cos(\theta_i - \theta_j)\right] \\ &= \frac{\lambda_s T}{2} \sin^2\left(\frac{\theta_i - \theta_j}{2}\right). \end{aligned} \quad (5.28)$$

VI. Conclusions

The reliability function is a fundamental quantity, which specifies the minimum probability of error obtainable on a channel as a function of the transmission rate and code block length. Although the reliability function can always be bounded, it is known exactly for only a few channels. In this paper, we have shown that the reliability function is known exactly for an extended class of very noisy channels as defined by Majani. Exponentially optimum codes have been constructed, at all rates less than capacity, for channels that can be modeled as repeated uses of a binary-input class I VNC, a binary-input/binary-output class II VNC, or a class II very noisy binary erasure channel. In addition, the concept of VNC's has been used to derive the capacity, error exponent, and exponentially optimum code for the direct detection polarization modulated optical channel.

References

[1] B. Reiffen, "*A note on 'very noisy' channels*," *Inform, Contr.*, vol. 6, pp. 126–130, June 1963.

[2] A Wyner, "On the probability of error for communication in white gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 86–90, Jan. 1967.

[3] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Part I" *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1449–1461, Nov. 1988.

[4] A. D. Wyner, "Capacity and error exponent for the direct detection photon channel—Part II," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 1462–1471, Nov. 1988.

[5] E. Majani, A model for the study of very noisy channels, and application. Ph.D. thesis, California Institute of Technology, Pasadena, 1987.

[6] J. Pierce, E. Posner, and E. Rodemich, "The capacity of the photon counting channel," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 61–77, Jan. 1981.

[7] K. Abdel-Ghaffer and R. McEliece, "The ultimate limits of information density," in *Proceedings of the NATO Advanced Study Institute on Performance Limits in Communication Theory and Practice*, III Ciocco, Italy), July 1986.

[8] C. Chao, Error-correction coding for reliable communication in the presence of extreme noise. Ph.D. thesis, California Institute of Technology, Pasadena, 1989.

[9] S. Verdu, "On capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 1019–1030, Sept. 1990.

[10] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 3–18, Jan. 1965.

[11] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels I," *Inform. Contr.*, vol. 10, pp. 65–103, Jan. 1967.

[12] C. Shannon, R. Gallager, and E. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels II," *Inform. Contr.*, vol. 10, pp. 522–552, May 1967.

[13] R. G. Gallager, "Power limited channels: coding, multiaccess, and spread spectrum," in *Proc. 1988 Conf. Information Sciences and Systems* (Princeton, NJ), p. 372, Mar. 1988.

[14] Wozencraft and I. Jacobs, *Principles of Communication Engineering*. New York: Wiley, 1965.

[15] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 1, New York: Wiley, 1971.

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