October 1983 LIDS-P-1337

ERROR EXPONENTS FOR A DIRECT DETECTION OPTICAL CHANNEL*

by

Pierre A. Humblet**

Abstract

Error exponent bounds for a noise free optical channel using a direct detection receiver are computed for the cases of Pulse Position Modulation (PPM) and for the basic channel with binary output and average intensity constraint. Some consequences are derived regarding the behavior of the interleaving scheme suggested by Massey, the optimality of the PPM format and the significance of the R_o parameter.

---

*Support for the National Science Foundation under grant NSF/ECS 79-20834 is gratefully acknowledged.

**Room No. 35-203, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Cambridge, Massachusetts 02139

I. Introduction

There has been much interest recently [1] in direct detection optical communications, particularly for space applications. Under some conditions [2] a noiseless model is appropriate. In this model time is divided into slots. During the ith slot the transmitter sends an optical pulse and the receiver counts the number of photons received. That number is assumed to be a Poisson random variable with mean λ_i, where λ_i is proportional to the energy of the transmitted signal. Statistically, the input to this channel is thus a real number (λ_i), the output is an integer (the number of photons) and the transition probability is Poisson, with parameter λ_i. Under an average energy constraint ($E\ \lambda_i \leq \lambda$) the capacity, in nats/photon units, increases without bound as λ decreases at the expense of increased bandwidth per nat and of increased peakedness of the intensity. Throughout the paper we will be interested in the small λ case.

Pierce [1] has suggested that Pulse Position Modulation is effective in that case. In his system, slots are grouped in blocks of M, and a pulse is transmitted during exactly one of the M slots, depending on the information to be transmitted. The receiver can only make an error if no photon is received, which occurs with probability $\varepsilon = \exp(-M\lambda)$. Thus at this "higher level" the channel behaves as a M-ary symmetric erasure channel (SEC) the capacity of which is easily found to be $\frac{\ln\ (M)}{M\ \lambda}\ (1-\varepsilon)$ nats/photon. It can be made as large as desired by increasing M while keeping $M\lambda$ constant.

McEliece [2] has shown how Reed-Solomon codes can be implemented together with PMM modulation and has analyzed the performance of the joint system. Massey [3] has made the clever observation that if $M = \prod_{i=1}^{m} M_i$

then a M-ary SEC can be decomposed into m M_i-ary correlated SEC's. He showed that effective coding schemes can be realized by interleaving m codes, specially when the interleaved codes are short constraint length binary convolutional codes.

Most of the theoretical studies of the channels described above have considered the capacity and R_o parameters of the channel. It is striking that R_o is less than one [6] (in nats/photon units) whereas the capacity is unbounded, this has renewed the debate on which parameter best characterizes the practical limitation of a channel.

In this paper we analyze the complete error exponents curves as functions of rate. We start with a brief review of the theory in section II, then we treat the PPM channel (section III) and the basic binary output channel (section IV). This analysis adds insight to Massey's observation, characterizes what parameters of PPM modulation are appropriate for a given λ, and settles the question of R_o versus capacity.

II. Review of Error Exponents for Discrete Memoryless Channels

It is well known [4] that for a given discrete memoryless channel characterized by transition probabilities $P(y|x)$ there exists block codes using the channel N times, transmitting information at the rate of R nats/channel use and meeting the following bounds on the block error probability.

$$P(\text{error}) \leq \exp(-N(E_o(\rho, Q) - \rho R)) \qquad 0 \leq \rho \leq 1$$

where $$E_o(\rho,Q) = -\ln \sum_y [\sum_x Q(x) P(y|x)^{\frac{1}{1+\rho}}]^{1+\rho}$$

and $Q(x)$ is a probability distribution on the channel input alphabet. In particular

$$P(\text{error}) \leq \exp(-N E_r(R))$$

where $$E_r(R) = \max_{0\leq\rho\leq 1} \max_Q [E_o(\rho,Q) - \rho R)$$

For small ρ, $E_o(\rho,Q) \sim \rho I(x;y) + 0(\rho^2)$, where $I(x;y)$ is the mutual information [4] between channel input and output when the input is distributed according to $Q(x)$. Thus the E_r bound gives non trivial result for R up to capacity, where the capacity is defined as $\max_Q I(x;y)$.

Taking $\rho = 1$ yields the weaker "R_o bound":

$$P(\text{error}) \leq \exp(-N(R_o - R))$$

where $$R_o = \max_Q E_o(1,Q).$$

The R_o bound is trivial when $R > R_o$, which can be much smaller than capacity. However when the optimizing ρ in $E_r(R)$ is equal to 1, which occurs for small R, $E_r(R) = R_o - R$.

These bounds are obtained by analyzing the expected behavior of codes generated randomly, with a per letter distribution Q(x). For small R the E_r bound is not tight as the probability of selecting bad codes dominates the probability of error of good codes. Expurgating bad codes gives the bound [5, problem 3.21].

$$P(\text{error}) \leq \exp(-N\, E_x(R))$$

where $$E_x(R) = -\max_{\rho \geq 1} \max_{Q} \ln\Big[\sum_x \sum_{x'} Q(x)Q(x') \Big[\sum_y \sqrt{P(y|x)P(y|x')}\Big]^{1/\rho} - R$$

Moreover, asymptotically over N, the following sphere packing bounds [4] applies to all codes

$$P(e) \gtrsim \exp(-N\, E_{sp}(R))$$

where $$E_{sp}(R) = \max_{0 \leq \rho} \max_{Q} \; (E_o(\rho,Q) - \rho R)$$

Note that if $E_{sp}(R)$ is achieved by $\rho \leq 1$, then $E_{sp}(R) = E_r(R)$. Thus the E_r and E_{sp} bounds are asymptotically tight when the optimal ρ is small, which can be shown to correspond to large R. When R is small it has been conjectured that the maximum of E_r and E_x is also asymptotically tight.

III. Application to the M-ary Symmetric Channel

We specialize the error exponents to the M-ary symmetric erasure channel. In all cases the optimizing input probability assignment is uniform so that we change our notation, removing the argument Q but adding M, thus defining in a straightforward fashion $E_o(\rho,M)$, $R_o(M)$, $E_r(R,M)$, $E_x(R,M)$, $E_{sp}(R,M)$.

On obtains the following values, some of which appear in figure 1 for $\varepsilon = .125$

$$E_o(\rho,M) = -\ln(\varepsilon + (1-\varepsilon)M^{-\rho}) \tag{1}$$

$$E_{sp}(R,M) = -R' \ln\left(\frac{1-\varepsilon}{R'}\right) - (1-R')\ln\frac{\varepsilon}{1-R'}, \quad 0 \le R' \le 1-\varepsilon$$

where $R' = R/\ln M$.

This indirectly shows that the channel capacity is

$$C = (1-\varepsilon)\ln M \tag{2}$$

Note that E_{sp} is Kullback's information divergence between R' and $1-\varepsilon$.

One also finds

$$R_o(M) = -\ln(\varepsilon + (1-\varepsilon)M^{-1})$$

and

$$E_r(R,M) = \begin{cases} E_{sp}(R,M) & R' > \dfrac{1-\varepsilon}{M\varepsilon + 1-\varepsilon} \\ R_o(M) - R & \text{else} \end{cases}$$

Finally

$$E_x(R,M) = \max_{\rho \ge 1} \left(-\rho \ln\left(M^{-1} + (1-M^{-1})\,\varepsilon^{1/\rho}\right) - \rho^R\right)$$

We now turn to Massey's observation that the M-ary SEC can be decomposed into $\frac{1}{p}$ $(0 < p \le 1)$ M^p-ary completely correlated SEC's. Instead of using a (N,R) code for the large channel one can equivalently use (N,pR) codes for the subchannels and still achieve the same total information rate.

Exponential bounds derived previously still hold for the subchannels, using the proper R and M. We make two observations:

a) $E_{sp}(R,M) = E_{sp}(pR, M^p)$

thus the channel and subchannels have the same sphere packing exponent. However the more subchannels one uses (the smaller p), the smaller is the rate region where E_{sp} is equal to E_r. (see Figure 1).

b) From (1) $E_o(\rho,M) = E_o(1,M^\rho)$

thus if $E_{sp}(R,M)$ is achieved by $\rho = p$, then

$$E_{sp}(R,M) = E_o(1,M^p) - pR.$$

In other words the curve $E_{sp}(R,M)$ could be generated as the upper envelope of the R_o bound for subchannels, if it were not for the diophantine constraints on 1/p and M^p. This intepretation gives a "physical meaning" to ρ

Let us now examine how the error bounds relate to the performances of actual codes. For example Massey [3] noted that a 16-ary N = 15 K = 8 Reed-Solomon code performs as well on a M = 16 channel as 4 interleaved N = 24 K = 12 binary Golay codes with maximum likelihood decoding (K denotes the number of information digits in a codeword.).

$M\lambda$, in photons per PPM channel use, is given by $M\lambda = \frac{K}{N} \frac{\ln M}{R}$ where R is the total rate in nats/photon. For R = 2/3 one finds for the Reed Solomon system

$$M\lambda = 2.22 \quad \varepsilon = .109 \quad E_r = .183 \text{ nats/photon}$$

upper bound on P(error) = 2.3 10^{-3} actual P(error) = 5.3 10^{-5}

For the Golay system

$$M\lambda = 2.08 \quad \varepsilon = .125 \quad E_r = .110 \text{ nats/photon}$$

$$\text{bound on } P(\text{error}) = 4.2\ 10^{-3} \quad \text{actual } P(\text{error}) = 4.5\ 10^{-5}$$

For the Reed Solomon system $E_r = E_{sp}$, but for the Golay system $E_r = R_o - R$, thus requiring a larger N for same probability of error. The actual P(error) were computed according to the methods outlined in [2] and [3]. In this example the error exponent E_r predicts very accurately the equivalence of the two systems. However the codes perform much better than what the bounds indicate.

Repeating the computation for $R = .5$ nats/photon yields for the Reed Solomon system

$$M\lambda = 2.96 \quad \varepsilon = .052 \quad E_r = .243 \text{ nats/photon}$$

$$\text{upper bound on } P(\text{error}) = 2.1\ 10^{-5} \quad \text{actual } P(\text{error}) = 2.3\ 10^{-7}$$

and for the Golay system

$$M\lambda = 2.77 \quad \varepsilon = .062 \quad E_r = .110 \quad E_r = .103 \text{ nats/photon}$$

$$\text{upper bound on } P(\text{error}) = 6.7\ 10^{-4} \quad \text{actual } P(\text{error}) = 1.8\ 10^{-7}$$

The agreement is not as good in this case, reflecting the weakness of the upperbounds at low rates for small M.

IV. The Binary Output Poisson Channel

Instead of restricting ourselves to PPM modulation we now allow each basic pulse making up the PPM signal to be modulated separately. The receiver is now binary, i.e., all decisions are based on whether any photon is received in each basic interval.

The average intensity per interval must be constrained for the problem to make sense and this complicates the computation of the error exponents. An upperbound on the probability of error can still be obtained by maximizing $E_o(\rho,Q)$ over all input distributions satisfying the constraint. The resulting bound will be non trivial for rates up to capacity, but it will not be sharp. Gallager [4 , section 7.3] explains how a better bound (the true $E_r(R)$) can be obtained by introducing and optimizing an extra parameter. We limit ourselves to the simple bound. This is usually also how R_o is defined in presence of constraints.

We first show that the distribution maximizing $E_o(\rho,Q)$ has mass only at 0 and at another point. As this holds for all $\rho > 0$ it is also true for the distribution achieving capacity.

We wish to minimize

$$\exp\left(-E_o(\rho,Q)\right) = \left(\int_0^\infty dQ(x)\, e^{-x/1+\rho}\right)^{1+\rho} + \left(\int_0^\infty dQ(x)\,(1-e^{-x})^{\frac{1}{1+\rho}}\right)^{1+\rho}$$

subject to the constraints $\int_0^\infty x\, dQ(x) \leq \lambda$

$$Q(\infty) = 1$$

$Q(x)$ non decreasing.

As the right hand side above is convex in Q, a necessary and sufficient condition for optimality is that

$$\left(\int dQ(y)\, e^{-y/1+\rho}\right)^{\rho} e^{-x/1+\rho} + \left(\int dQ(y)\,(1-e^{-y})^{\frac{1}{1+\rho}}\right)^{\rho}(1-e^{-x})^{\frac{1}{1+\rho}} \qquad (3)$$

$$\geq -ax + b \qquad \forall\, x \geq 0$$

for some $a > 0$ and b, with equality with probability one. For arbitrary non negative q_o, q_1 and ρ the function $q_o\, e^{-\frac{x}{1+\rho}} + q_1(1-e^{-x})^{\frac{1}{1+\rho}}$ has a second derivative with respect to x equal to

$$q_o\left[\left(\frac{1}{1+\rho}\right)^2 e^{-\frac{x}{1+\rho}}\right] - q_1\left[\frac{\rho}{(1+\rho)^2}\,(1-e^{-x})^{-\left(\frac{1+2\rho}{1+\rho}\right)} e^{-2x} + \frac{1}{1+\rho}\,(1-e^{-x})^{-\frac{\rho}{1+\rho}}\, e^{-x}\right]$$

This last function has a single zero in $(0,\infty)$, as the term in the first bracket decreases more slowly than the ones in the second. Thus if condition (3) is true it can be satisfied with equality by only one x in $(0,\infty)$. This shows that the optimizing Q has a mass, say $1-p$, at $x = 0$, and mass p at λ/p. Optimization over Q is thus reduced to a simple search over p. The resulting lowerbound on the true $E_r(R)/\lambda$ is displayed in Figure 2. Note that the value of $E_r(R)/\lambda$ depends on R and λ virtually only through the ratio of R to capacity.

The capacity (per channel use) can be written in turn as

$$C = \max_{0\leq p\leq 1} \left(H(y) - H(y|x)\right)$$

$$= \max_{0\leq p\leq 1} \mathcal{H}\left(p(1-e^{-\lambda/p})\right) - p_1\, \mathcal{H}(1-e^{-\lambda/p}) \qquad (4)$$

where $\mathcal{H}$ denotes the binary entropy function
Choosing $p = \lambda$ yields

$$C \geq \mathcal{H}(\lambda(1-e^{-1})) - \lambda\, \mathcal{H}(1-e^{-1})$$

The non negativity of H yields the upperbound

$$C \leq \max_{0 \leq p \leq 1} H(p(1-e^{-\lambda/p}))$$

$$= \begin{cases} \ln 2 & 1-e^{-\lambda} > .5 \\ H(1-e^{-\lambda}) & \text{else} \end{cases}$$

$$\leq \begin{cases} \ln 2 & \lambda > .5 \\ H(\lambda) & \text{else} \end{cases}$$

Keeping the low order terms in the power series expansion of the right hand side of (4) yields the following approximation

$$C = \max_{0 \leq p \leq 1} \lambda \left[\frac{\lambda}{2p} - \ln p + \frac{\lambda}{2} \frac{\ln p}{p} + 0(\lambda)\right] \text{ nats}$$

The optimizing p satisfies the relation

$$p^* = \frac{\lambda}{2} \ln \frac{1}{p^*}$$

thus $$p^* = \frac{\lambda}{2} \left[\ln \frac{1}{\lambda} - \ln \ln \frac{1}{\lambda} + 0\left(\frac{\ln \ln \frac{1}{\lambda}}{\ln \frac{1}{\lambda}}\right)\right]$$

and $$C = \lambda \left[\ln \frac{1}{\lambda} - \ln \ln \frac{1}{\lambda} + 0\left(\frac{\ln \ln \frac{1}{\lambda}}{\ln \frac{1}{\lambda}}\right)\right] \text{ nats}$$

Comparing this with McEliece's bound [2 ,(4.5)]

$$C \leq \lambda \left[\ln \left(\frac{1}{\lambda} + 1\right) + 1\right]$$

for multilevel channel output shows that, as expected, restriction to binary output entails no asymptotic loss in capacity.

Finally, when one is given an optical channel with a given λ per interval and one wishes to use Pierce's PPM scheme, the question arises on how large to make M. The optimal values of p found above suggests that M in the range $\frac{2}{\lambda \ln(\frac{1}{\lambda})} \leq M \leq \frac{1}{\lambda}$ would be suitable. Figure 3 shows the error exponent E_r of various PPM channels, when $\lambda = 10^{-3}$ per interval, and compares then with the bound on E_r derived in this section. The true E_r would always dominate, but its bound does not, specially at low rates. $M = 361$ is the value of M maximizing the capacity per photon when $\lambda = 10^{-3}$ as can be found by going back to (2). One concludes that a well chosen PPM format is close to optimality not only in terms of capacity, but also as far as error exponents are concerned.

In the intermediate rate region $(R \sim \frac{C}{2})$, $M \simeq 1/\lambda$ does about as well as the M maximizing the capacity of the channel. The advantage of $M \simeq 1/\lambda$ is that $\varepsilon \simeq e^{-1}$, thus one expects usual codes, such as rate 1/2 Reed-Solomon [3], to perform well. Being closer to capacity requires a smaller M, thus a larger ε and low rate codes tolerating channels that erase most symbols.

For example, when $M = \frac{1}{\lambda}$, $E_{sp} = E_r = .205$ nats/photon (or per PPM channel use) for $R = .5(1-e^{-\lambda})\ln \frac{1}{\lambda}$ nats/photon (or per PPM channel use); this rate is half the capacity of the PPM channel, or about $.5(1-e^{-1})$ of the capacity of the binary output photon channel.

A probability of error of $2\ 10^{-6}$ can be achieved by using $N = 64$. The theory of section III predicts that 4-ary codewords of that length (with $K \simeq 20$) can be interleaved on the $M = \frac{1}{\lambda}$ PPM channel while still achieving the same probability of error.

V. The Significance of R_o

As indicated previously R_o is the maximum value of $E_r(R)$, $R_o = E_r(0)$ or possibly less when there are constraints on the input signal. The notation is unfortunate as "R_o" connotes the idea of "rate", not of value of error exponent. This is probably due to the fact that the "R_o bound" is non trivial only for $R < R_o$, but this has little significance as the R_o bound is not tight.

For the Poisson channel any lower bound on error exponents must be at most 1/photon (when base e is used; I hesitate writing "nats/photon"): as any code with P equally likely codewords and an average of λ photons/codeword has an average error probability of at least $(1-\frac{1}{P})\exp(-\lambda(\frac{P}{P-1}))$ due to the possiblity of confusion when no photon is detected.

Thus it is not surprising that R_o be less than 1 nat/photon. R_o approaches 1 in the limit of few nats/dimension, when the erasure probability dominates the probability that randomly selected codewords are "close".

The boundedness of error exponents has no implications on the width of the capacity region, which has been shown to be arbitrarily large on a bit per photon basis. It has no implication on decoding complexity either: the analysis reported above shows that there is essentially no loss in implementing Massey's interleaving scheme. Increasing M while keeping ε fixed and interleaving log M fixed (N,R) codes allows arbitrarily high rates per photon while keeping the probability of error and the amount of decoding computation per bit constant.

Finally it has been suggested [3] that R_o be considered as a rough measure of the necessary code length in channel symbols required to achieve a given probability of error. The example reported at the end of section III tends to show that this role belongs more properly to $E_r(R)$ and $E_x(R)$ whose values bear little relationship to that of R_o.

We thus conclude that in general there is no evidence that R_o has any particular significance for communication systems, except as it relates to sequential decoding of convolutional codes [5]. Also, the practical limit on the achievable rate per photon on the optical channel lies not in the necessary complexity of the decoders, or in other information theoretic concepts, but on purely physical reasons: the difficulty to generate the vanishingly narrow but energyful pulses required [1], [7].

Figure Captions

Figure 1: $E_r(R,M)$, $E_x(R,M)$ and $E_{sp}(R,M)$ (nats per photon) for M = 2, 4, 8, 16 PPM channel with $\varepsilon = .125$

Figure 2: Lower bound on $E_r(R)$, $\lambda = 1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}$ Binary output Poisson channel

Figure 3: $E_r(R,M)$ for PPM channel with M = 10, 30, 100, 361, 1000, 3000
Lower bound on $E_r(R)$ for binary output Poisson channel
$\lambda = 10^{-3}$ in all cases.

REFERENCES

[1] J. R. Pierce, "Optical Channels: Practical Limits With Photon Counting", IEEE Trans. Commun., Vol. COM-26, pp. 1819-1821, Dec. 1978.

[2] R. J. McEliece, "Practical Codes for Photon Communication", IEEE Trans. Inform. Theory, IEEE Trans. I.T., Vol. IT-27, No. 4, July 1981.

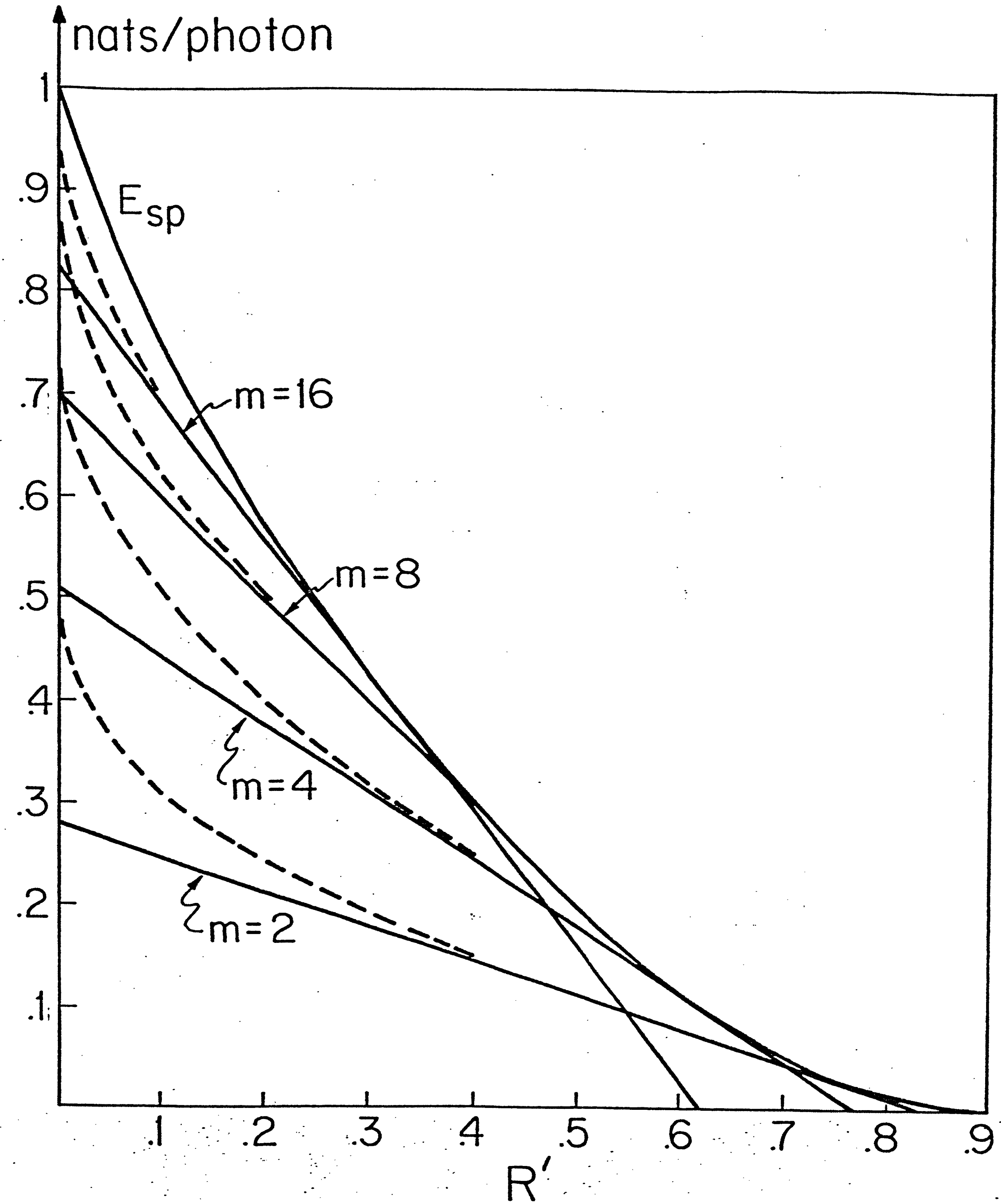
[3] J. L. Massey, "Capacity, Cutoff Rate, and Coding for a Direct-Detection Optical Channel", IEEE Trans. Commun., Vol. COM-29, No. 11, pp. 1615-1621, Nov. 1981.

[4] R. G. Gallager, Information Theory and Reliable Communication, Wiley and Sons, New York, 1968.

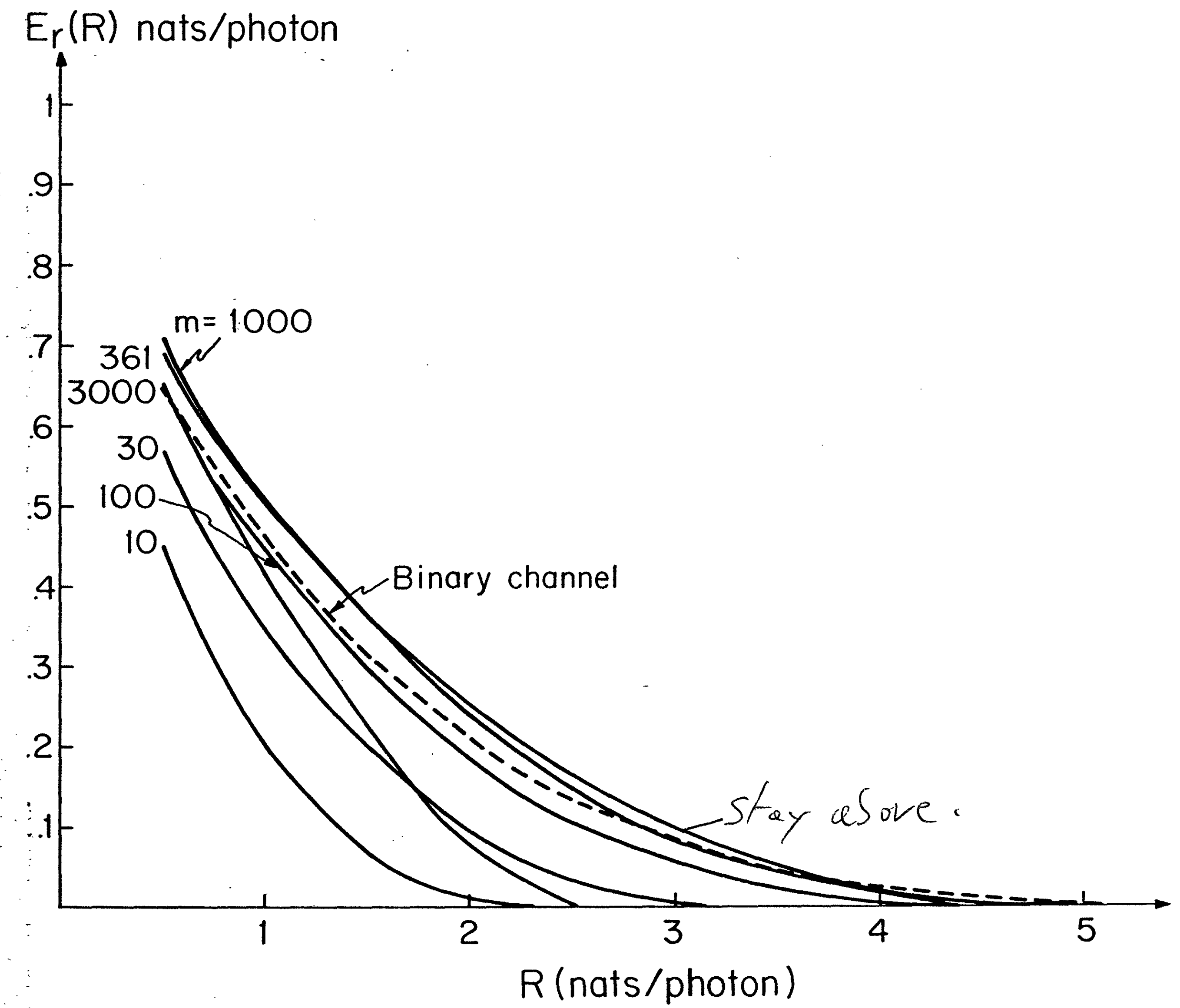
[5] A. J. Viterbi and J. K. Omura, Principles of Digital Communication and Coding, New York, McGraw-Hill, 1978.

[6] D. L. Snyder and I. B. Rhodes, "Some Implications of the Cutoff-Rate Criterion for Coded Direct-Detection Optical Communication Systems", IEEE Trans. Inforation Theory, Vol. IT-26, No. 3, May 1980.

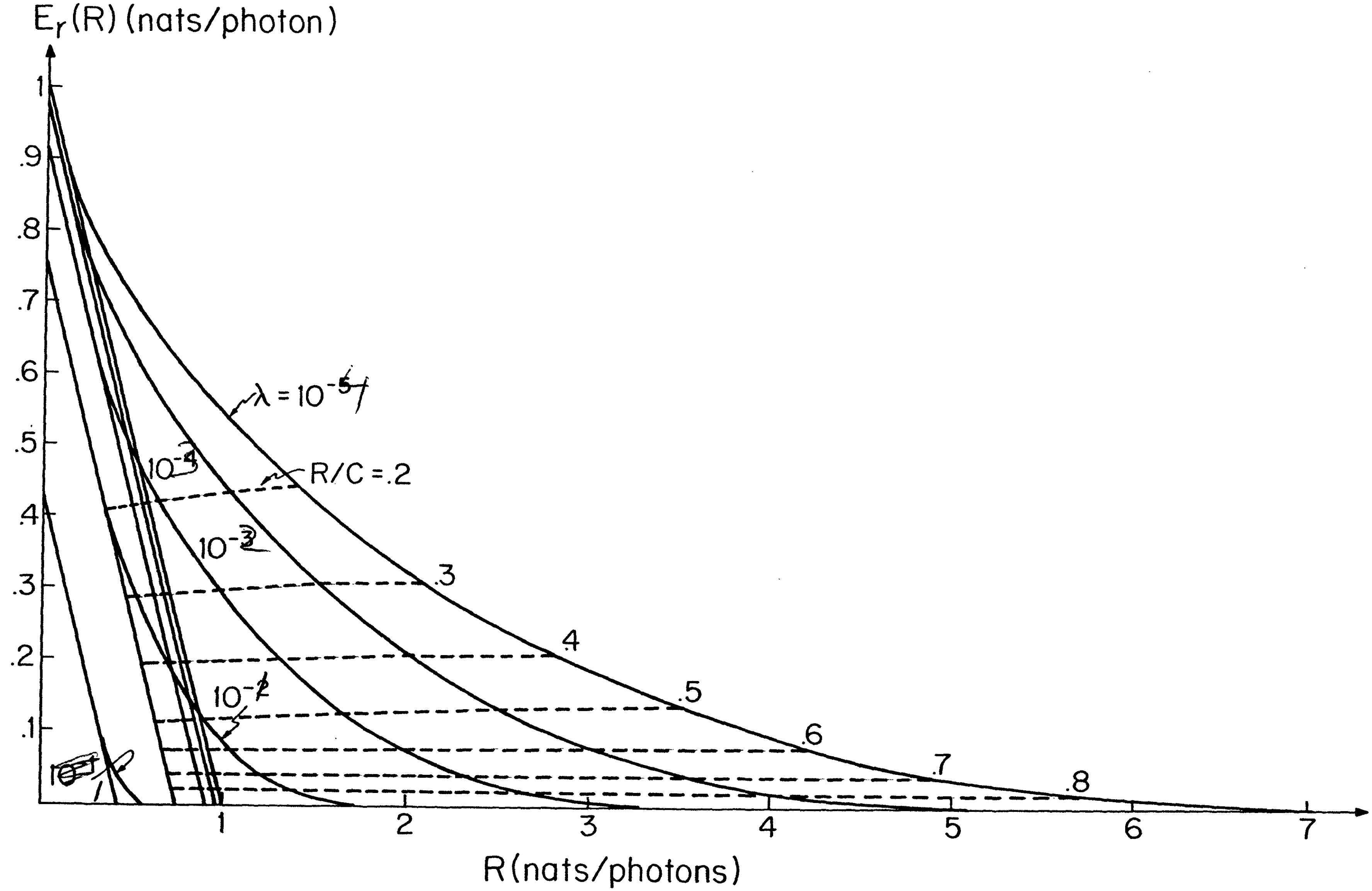
[7] S. A. Butman, J. Katz, and J. R. Lesh, "Practical Limitations on Noiseless Optical Channel Capacity", Jet Propulsion Lab., Pasadena, CA, DSN Progress Rep., Vo. 42-55, pp. 12-14, Nov.-Dec., 1979.



nats/photon
1
.9
.8
.7
.6
.5
.4
.3
.2
.1
E_{sp}
m=16
m=8
m=4
m=2
.1
.2
.3
.4
.5
.6
.7
.8
.9
R′



Er(R) nats/photon
1
.9
.8
.7
.6
.5
.4
.3
.2
.1
m= 1000
361
3000
30
100
10
Binary channel
stay above.
1
2
3
4
5
R (nats/photon)



Er(R) (nats/photon)
R(nats/photons)
λ = 10
R/C = .2
.3
.4
.5
.6
.7
.8