

# Exact Evaluation of the Capacity for Intensity-Modulated Direct-Detection Channels With Optical Amplifier Noises

Keang-Po Ho, *Senior Member, IEEE*

**Abstract**—Based on a rigorous model without approximation, the channel capacity is evaluated numerically for intensity-modulated direct-detection systems dominated by optical amplifier noises. With a discrete probability at zero intensity, the optimal input distribution is both discrete and continuously distributed. For small signal-to-noise ratio less than about 5 dB, the optimal signal is binary on–off keying.

**Index Terms**—Channel capacity, direct-detection, on–off keying.

## I. INTRODUCTION

WHEN AN optical signal is contaminated by amplifier noises, the channel capacity or the ultimate spectral efficiency is determined by the information-theoretic capacity per unit bandwidth, in the unit of bits/second/hertz [1]. With optical amplifiers to periodically compensate for fiber loss, most long-haul lightwave transmission systems are limited by amplifier noises. An intensity-modulated direct-detection (IMDD) signal is used in almost all commercial lightwave transmission systems.

The Shannon limit of [1]

$$C = \log_2(1 + \rho_s) \quad (1)$$

is well known for a power-constrained Gaussian channel where  $\rho_s$  is the signal-to-noise ratio (SNR) as defined in [2]. Using coherent detection and with the allowance of high complexity and long delay, the Shannon limit can be closely approached using Turbo codes [3]–[5]. When those codes are used for IMDD systems limited by amplifier noise [6], the channel capacity must be derived rigorously to understand the limitation. For single-channel optical signal limited by fiber nonlinearities, the ultimate spectral efficiency is also limited by that for IMDD signal [2], [7].

The channel capacity of optically amplified IMDD systems was studied by [8], [9]. A lower bound is derived in [9] based on half-Gaussian distribution; the asymptotic spectral efficiency is equal to 0.5 b/s/Hz less than half of (1), or  $(1/2)\log_2 \rho_s - 0.5$  using the definition of  $\rho_s$  of (1). Here, the optimal input distribution is found numerically. Binary modulation is found to be optimal at low SNR.

Manuscript received September 27, 2004; revised December 12, 2004. This work was supported in part by the National Science Council of Taiwan under Grant NSC-93-2213-E-002-061 and Grant NSC92-2218-E-002-055.

The author is with the Institute of Communication Engineering and Department of Electrical Engineering, National Taiwan University, Taipei 106, Taiwan, R.O.C. (e-mail: kpho@cc.ee.ntu.edu.tw).

Digital Object Identifier 10.1109/LPT.2005.843947

## II. CHANNEL CAPACITY OF IMDD CHANNELS

For IMDD systems limited by amplifier noise, the discrete-time model of the channel with input electric field  $X$  and output intensity  $Y$  is [9]

$$Y = |X + N|^2 \quad (2)$$

where the amplifier noise  $N$  is equivalently either a complex zero-mean Gaussian random variable or a two-dimensional zero-mean Gaussian random vector with a variance of  $\sigma_n^2$  per dimension. The mean output intensity is  $m_y = \sigma_x^2 + 2\sigma_n^2$ , where  $\sigma_x^2$  is the variance or power of the input  $X$ . The SNR is  $\rho_s = \sigma_x^2/2\sigma_n^2$ .

For the IMDD channel of (2), the conditional probability density function (pdf) of  $p(y|x)$  is a noncentral  $\chi^2$ -distribution with two degrees of freedom [10, pp. 41–44]

$$p(y|x) = \frac{1}{2\sigma_n^2} \exp\left[-\frac{y + x_1^2 + x_2^2}{2\sigma_n^2}\right] I_0\left(\frac{\sqrt{y(x_1^2 + x_2^2)}}{\sigma_n^2}\right), \quad y \geq 0 \quad (3)$$

with noncentrality parameter of  $x_1^2 + x_2^2$ , and  $I_0(\cdot)$  is the zeroth-order modified Bessel function of the first kind. The pdf (3) conditions solely on the input amplitude of  $S = |X|$ . The channel output is the intensity  $Y \geq 0$  but monotonic one-to-one transfer to the amplitude of  $R = \sqrt{Y} \geq 0$  does not change the channel capacity. Using the input and output amplitude random variables of  $S$  and  $R$ , respectively, the conditional pdf becomes a Rice distribution of

$$p(r|s) = \frac{r}{\sigma_n^2} \exp\left[-\frac{r^2 + s^2}{2\sigma_n^2}\right] I_0\left(\frac{rs}{\sigma_n^2}\right), \quad r, s \geq 0. \quad (4)$$

The channel capacity, or the maximum spectral efficiency limit, is equal to the maximum mutual information of

$$C = \max_{p(s)} E \left\{ \log_2 \frac{p(r|s)}{p(r)} \right\} \quad (5)$$

where  $E\{\cdot\}$  denotes expectation,  $p(s)$  and  $p(r) = \int_0^\infty p(r|s)p(s) ds$  are the pdf of the input and output amplitudes, respectively. The channel capacity is also equal to  $C = \max_{p(s)} \{H(R) - H(R|S)\}$ , where the output entropy of  $H(R)$  and the conditional entropy of  $H(R|S)$  are [1]

$$H(R) = - \int p(r) \log_2 p(r) dr \quad (6)$$

$$H(R|S) = - \int \int p(s)p(r|s) \log_2 p(r|s) dr ds \quad (7)$$

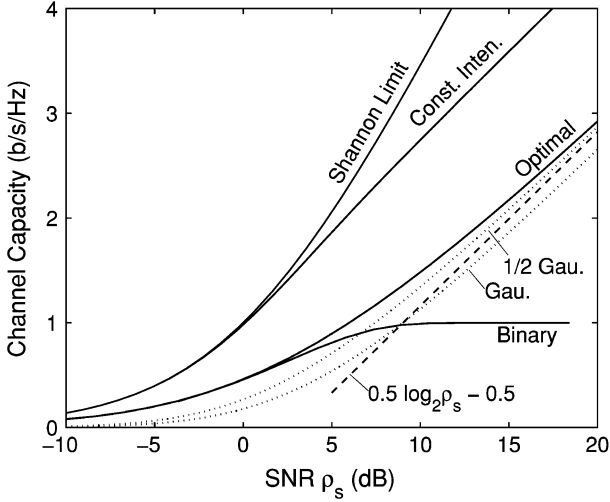


Fig. 1. Channel capacity as a function of SNR  $\rho_s$ . For IMDD channel, the input has either the optimal, half-Gaussian, or Gaussian distribution.

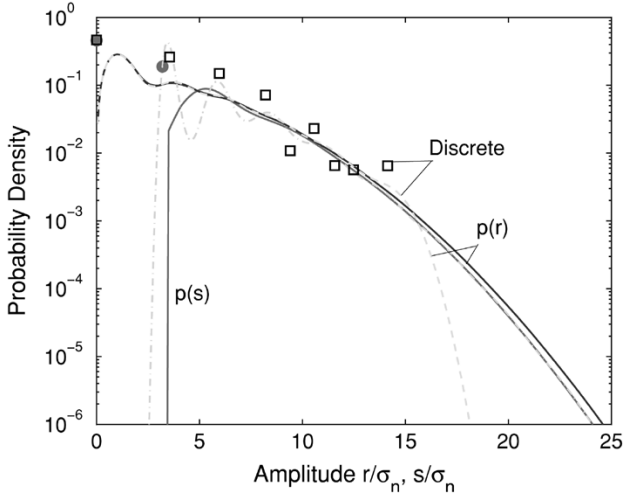


Fig. 2. Input and output probability density as a function of normalized input and output amplitude of  $r/\sigma_n, s/\sigma_n$  for  $\rho_s = 10$  dB.

where all integrations are from 0 to  $+\infty$ . The capacity of (5) should be evaluated together with the average power and probability constraints of

$$\int s^2 p(s) ds = \sigma_x^2, \quad \int p(s) ds = 1. \quad (8)$$

### III. NUMERICAL RESULTS

Based on different assumptions, three algorithms are used to find the optimal input distribution to maximize the channel capacity of (5). Fig. 1 shows the channel capacity as a function of SNR for IMDD channel of (2). Different algorithms give different input distributions as shown in Fig. 2 for  $\rho_s = 10$  dB but the same channel capacity in Fig. 1. The Shannon limit of (1) is also shown in Fig. 1 for comparison. Fig. 1 also shows the capacity of constant-intensity modulation [11]–[13] that has a spectral efficiency 1.6 b/s/Hz better than IMDD channels at large SNR.

Other than the optimal input distribution arrived from different algorithms, Fig. 1 also calculates the channel capacity when the input amplitude of  $p(s)$  is half-Gaussian distributed [9] ( $p(x)$  is confined to one dimension and is Gaussian) or the input electric field of  $p(x)$  is two-dimensionally Gaussian distributed [2], [14]. Fig. 1 also shows the approximated channel capacity of  $(1/2) \log_2 \rho_s - (1/2)$  [9].

When an artificial peak-power constraint is imposed to the IMDD channel, the optimal input distribution is discrete [15]–[17]. Similar to the algorithm of [15] and [16] based on numerical optimization, Fig. 2 shows the optimal discrete input distribution as square for a peak-power constraint ten times the average power. Beginning with binary signal, the number of discrete points is increased one by one until the channel capacity converges. Fig. 2 shows the optimal nine discrete points with the corresponding probability.

Without the artificial peak-power constraint, the optimal input distribution has continuous component or infinite number of points very close to each other in its tail. Using the asymptotic expression of  $I_0(x) \sim e^x/\sqrt{2\pi x}$  for large SNR,  $p(r|s)$  becomes asymptotically a Gaussian distribution with variance of  $\sigma_n^2$  and a constant conditional entropy of  $H(R|S) = (1/2) \log_2(2\pi e\sigma_n^2)$ . With both large amplitudes of  $r$  and  $s$ , the optimal distribution is asymptotically having a Gaussian profile of  $\sim e^{-\kappa_1 s^2}$  and  $\sim e^{-\kappa_2 r^2}$  for input and output, respectively, where  $\kappa_1$  and  $\kappa_2$  are two related factors.

To find the optimal continuous distribution, the continuous IMDD channel is transformed to an  $N$  to  $M$  discrete channel with input and output probabilities of  $p(s_i)\Delta s$  and  $p(r_j)\Delta r$ , where  $0 \leq i < N$  and  $0 \leq j < M$ ,  $\Delta s$  and  $\Delta r$  are the step size. The traditional Arimoto algorithm [18] with  $N = M = 1000$  is modified to include the average-power constraint and the known tail profile. The Arimoto algorithm gives only a single discrete point at zero intensity. Instead of using continuous distribution as the initial assumption, the algorithm is modified with discrete probability at zero intensity. Fig. 2 shows the optimal input distribution with discrete probability at zero intensity (overlapped with the square there) and continuous-distribution as a dashed-dotted line. Note that unlike the Rayleigh channel in [16], the Arimoto algorithm converges very fast for IMDD channel.

Although the first two algorithms give the same channel capacity shown in Fig. 1 and we may prove the existence of a discrete probability exactly at zero intensity, more discrete points may exist between the continuous distribution at large amplitude and the zero-intensity discrete point. The third algorithm is the combination of the Arimoto algorithm [18] and numerical optimization [15], [16] with the prior assumption that there are several discrete points at small input amplitude. The Arimoto algorithm is used to find the optimal continuous distribution and numerical optimization is used to find the optimal discrete points. The two procedures are used alternatively with increase channel capacity in each iteration. Fig. 2 also shows the optimal input distribution with two discrete points at low intensity (solid circles, the circle at zero intensity overlapped with the square) and the continuous distribution at large amplitude (solid line). Not shown in Fig. 2, more than two discrete points are used for low intensity in further calculations that do not give a channel capacity with observable difference with that in Fig. 1.

All three algorithms converge to the same channel capacity without observable difference. However, Fig. 2 shows that the input distribution of  $p(s)$  has significant difference from one algorithm to others. At  $\rho_s = 10$  dB, the three algorithms give a channel capacity within  $\pm 0.05\%$ . The output distributions of  $p(r)$  in Fig. 2 from the three input distributions also have no significant difference at small amplitude. Only the tail distribution of  $p(r)$  has major difference when the input distribution is totally or partially discrete.

Fig. 1 also shows the channel capacity for binary signal (two discrete points in the input distribution) calculated by numerical optimization. Binary signal achieves the channel capacity for SNR less than about 5 dB. The channel capacity of binary signal was also calculated in [9] with the assumption that the two levels are equally probable. Except for large SNR ( $\rho_s > 12$  dB), the optimal binary signal has larger probability at zero-intensity and smaller probability at nonzero-intensity. For example, if only 10% probability at nonzero-intensity, the nonzero-intensity is ten times the averaged intensity as compared with twice the averaged intensity for equal-probable case. Compared to a similar curve in [9], the binary signal can achieve better channel capacity at low SNR. By sending occasional pulses with large intensity above the noise, the system is similar to the essence of return-to-zero signaling. Systems with powerful forward error correction [6] can operate around  $\rho_s = 5 \sim 7$  dB. Binary instead of multilevel signals may be sufficient for those systems.

Based on the same channel model of (2), the half-Gaussian distribution of [9] provides a lower bound and is 0.07–0.21 b/s/Hz worse than the optimal distribution. As discuss earlier, the optimal distribution has a tail profile of  $e^{-\kappa_1 r^2}$ , similar to that of half-Gaussian distribution. For very large SNR ( $\rho_s > 30$  dB), the discrete region of Fig. 2 at low intensity becomes insignificant and the half-Gaussian distribution should be very close to the optimal distribution.

In [19], the channel capacity is derived for nonlinear channel. In the linear case, the channel capacity is equal to the one-dimensional Shannon limit of  $(1/2) \log_2(1 + \rho_s)$  by assuming that the output of  $Y$  is Gaussian distributed. In practice, as  $Y$  has only positive value but the Gaussian distribution implicitly assumes that the random variable may have negative value. The approximation of [19] overestimates the channel capacity by 0.5 b/s/Hz. As shown in [9], input distribution is not discussed in [8]. The model of [8] gave a capacity between  $\pm 1$  b/s/Hz of  $(1/2) \log_2 \rho_s$  depending on the definition of SNR. The SNR in [8] is approximately the ratio of the power of signal–noise and noise–noise beating, twice larger than that in this letter.

Fig. 1 also includes the channel capacity for Gaussian input-distribution of [14] that is 0.19–0.35 b/s/Hz worse than the optimal distribution. The asymptotic spectral limit of [14] is 0.19 b/s/Hz less than that of [9] using half-Gaussian distribution.

The optimization problem to find the channel capacity is a convex problem with an unique global maxima. The variations of Fig. 2 show that a large variety of input distributions is very close to the global maxima. With a power limit, the discrete point at zero intensity increases the input entropy but not the power. The spectral efficiency of Fig. 1 shows that the input

distributions with the profile of Fig. 2 provide larger channel capacity than the previous proposal of [9] and [14].

#### IV. CONCLUSION

The capacity of IMDD channel, to our knowledge, has been calculated rigorously the first time based on numerical optimization. Three algorithms find significantly different input distributions but the channel capacity converges to the same value. The optimal input distribution has at least a single discrete point at zero intensity. At low SNR, the optimal distribution is found to be binary on–off keying signal. At very large SNR, when the single discrete point at zero intensity becomes insignificant, the optimal distribution is very close to half-Gaussian distribution.

#### REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–646, 1948.
- [2] J. M. Kahn and K.-P. Ho, "Spectral efficiency limits and modulation/detection techniques for DWDM systems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, no. 2, pp. 259–272, Mar./Apr. 2004.
- [3] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes (1)," in *Proc. IEEE Int. Conf. Commun. (ICC '93)*, vol. 2, pp. 1064–1070.
- [4] S.-Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, no. 2, pp. 58–60, Feb. 2001.
- [5] C. Berrou, "The ten-year-old Turbo codes are entering into service," *IEEE Commun. Mag.*, vol. 41, no. 8, pp. 110–116, Aug. 2003.
- [6] T. Mizuochi, Y. Miyata, T. Kobayashi, K. Ouchi, K. Kuno, K. Kubo, K. Shimizu, H. Tagami, H. Yoshida, H. Fujita, M. Akita, and K. Motohima, "Forward error correction based on block Turbo code with 3-bit soft decision for 10-Gb/s optical communication systems," *IEEE J. Sel. Topics Quantum Electron.*, vol. 10, no. 2, pp. 376–386, Mar./Apr. 2004.
- [7] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Phys. Rev. Lett.*, vol. 91, no. 203 901, 2003.
- [8] E. Desurvire, "Fundamental information-density limits in optically amplified transmission: An entropy analysis," *Opt. Lett.*, vol. 25, pp. 701–703, 2000.
- [9] A. Mecozzi and M. Shtaif, "On the capacity of intensity modulated systems using optical amplifiers," *IEEE Photon. Technol. Lett.*, vol. 13, no. 9, pp. 1029–1031, Sep. 2001.
- [10] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.
- [11] J. M. Geist, "Capacity and cutoff rate for dense M-ary PSK constellations," in *Proc. IEEE Mil. Commun. Conf. (MilCom '90)*, pp. 168–170.
- [12] J. P. Aldis and A. G. Burr, "The channel capacity of discrete time phase modulation in AWGN," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 184–185, Jan. 1993.
- [13] K.-P. Ho and J. M. Kahn, "Channel capacity of WDM systems using constant-intensity modulation formats," in *Optical Fiber Commun. Conf. (OFC 2002)*, Anaheim, CA, Paper ThGG85.
- [14] M. J. W. Hall, "Gaussian-noise and quantum-optical communication," *Phys. Rev. A*, vol. 50, pp. 3295–3303, 1994.
- [15] J. G. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Inf. Control*, vol. 18, pp. 203–219, 1971.
- [16] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.
- [17] J. Huang and S. Meyn, "Characterization and computation of optimal distributions for channel code," *IEEE Trans. Info. Theory*, submitted for publication.
- [18] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 14–20, Jan. 1972.
- [19] L. G. L. Wegener, M. L. Povinelli, A. G. Green, P. P. Mitra, J. B. Stark, and P. B. Littlewood, "The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: Cross-phase modulation and four-wave mixing," *Physica D*, vol. 189, pp. 81–99, 2004.