Coded Modulation Techniques for Direct-Detection Optical Fiber Communication Systems with Optical Amplifiers

Roy You* and Joseph M. Kahn†

Abstract: In this paper, we investigate spectrally efficient coded multilevel modulation for direct-detection optical fiber communication systems with optical amplifiers. The dominant noise in the received photocurrent is the beat noise of amplified spontaneous emission with signal and itself. Although the decision variables are higher-order Chi-square distributed, we approximate them by first-order Chi-square random variables. This allows us to work with the electric field magnitudes of the transmitted optical signals and to approximate the amplifier noise as additive white Gaussian noise. Using this approximation, we develop a pairwise error probability bound, which shows that the code design criterion is to maximize the minimum Euclidean distance between the electric field amplitudes of transmitted signal sequences. Hence, the coded modulation schemes developed for additive white Gaussian noise channels, such as trellis-coded modulation and multi-level coding, can also be applied in our case. We evaluate the performance of various coded multilevel modulation schemes and compare their performance to coded and uncoded on-off keying. Moreover, the error performance can be improved by using asymmetrical coded modulation techniques, which jointly optimize the coded modulation and the signal constellation.

Keywords: optical fiber communication, trellis-coded modulation, multi-level coding, asymmetrical modulation, spectral efficiency

*Roy You is with Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 (e-mail: ryou@eecs.berkeley.edu).

†Joseph M. Kahn was with Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720. He is now with Department of Electrical Engineering, Stanford University, Stanford, CA 94305 (e-mail: jmk@ee.stanford.edu).

1 Introduction

Over the last two decades, optical fiber communication systems have developed rapidly to meet the increasing transmission capacity demand. As the physical device technologies mature, more advanced communication techniques are needed to further improve the performance of optical fiber systems [1]. In this paper, we focus on coded multilevel modulation techniques, which can improve system spectral efficiency and resistance to chromatic dispersion and polarization-mode dispersion. While most current optical fiber systems use on-off keying (OOK) with symbol-by-symbol hard-decision detection that offers simplicity and good power efficiency, their spectral efficiencies are far below the predicted limits [2]. When error-correcting codes are used, the spectral efficiency is further reduced to improve the error performance. With the deployment of long-haul optical fiber systems with optical amplifiers, spectral efficiency has become an important issue. By improving spectral efficiency, we can achieve higher data throughput in the bands (e.g. C and L) supported by existing fiber amplifiers, dispersion compensators, and other components. By lowering the symbol rate, we can improve the resistance to uncompensated chromatic dispersion and polarization-mode dispersion, and also utilize lower-speed electronics. One common way of improving spectral efficiency is to use multilevel amplitude shift keying (ASK) instead of on-off keying. However, uncoded ASK modulation can lead to significant reduction in power efficiency. In this paper, we explore the use of coded ASK modulation techniques, which trades off spectral efficiency and power efficiency. If coded ASK modulation is used as an inner code and a convolutional binary code is used as an outer code, then the performance improvement achieved by an improved coded modulation scheme can be combined with the gain of the outer binary code.

This paper is organized as follows. In section II, we introduce the channel model for amplified direct-detection systems. The dominant noises in the received photocurrent are the beat noise between the transmitted signal and the amplified spontaneous emission (ASE) and the beat noise of ASE with itself [3]. Although the received photocurrent is a higher-order Chi-square distributed random variable conditioning on input, an approximate model using a first-order Chi-square distribution is introduced, which simplifies the subsequent development of sequence detection. In section III, we analyze the symbol-by-symbol detection of OOK and ASK using the simplified model, which serves as a basis of comparison with sequence detection. In section IV, we examine both hard-decision and soft-decision sequence detection techniques, which defines the error performance of coded modulation schemes. It is shown that the maximum likelihood sequence detection (MLSD) for optical system dominated by ASE noise has similar structure to MLSD for additive white Gaussian noise (AWGN) channels. Therefore, traditional coded modulation schemes, such as trellis-coded modulation (TCM) [4] and multi-level coding (MLC) [5], can be applied to our case. The results obtained using these coded modulation techniques are discussed in section V. As for AWGN channels, further improvement can be obtained by optimizing the signal constellation jointly with coded modulation. Hence, asymmetrical TCM [6] and asymmetrical MLC techniques are discussed in section VI. In section VII, we compare various coded and uncoded modulation schemes and show that coded modulations combined with multilevel signaling can improve system spectral efficiency and improve error performance. The conclusions are given in section VIII.

2 System Model

The optical fiber communication system under study is shown in Figure 1. Although the performance of such a system is usually impacted not only by optical amplifier noise but also by fiber dispersion and nonlinearity, in this paper, we ignore the effects of dispersion and nonlinearity. Under this assumption, the system is linear and memoryless in electric field and can be modeled by considering its channel response during one symbol interval. Hence, without loss of generality, we consider transmitted and received signals during a particular symbol interval in modeling the channel.

Assuming a rectangular non-return-to-zero pulse shape, with symbol interval T and optical carrier frequency f_c, the transmitted optical signal can be expressed in the symbol interval $[0, T)$ as

$$x(t) = \sqrt{2P}\cos(2\pi f_c t), t \in [0, T). \quad (1)$$

Information is encoded in the intensity P of the optical signal, which is drawn from a predefined set of intensity levels $\{P_k, k = 0, \ldots, L-1\}$ and the average power of transmitted signal is defined as the mean of input intensity

$$P_{av} = E[P]. \quad (2)$$

For simplicity, we assume an infinite extinction ratio, which means that the lowest optical intensity is $P_0 = 0$. For L-level ASK modulation with infinite extinction ratio, it has been shown [7] that the quadratic level set $\{P_k = (k\Delta)^2, k = 0, \ldots, L-1\}$ is optimal in symbol-by-symbol detection when the dominant noise is the signal-ASE beat noise. Since the signal-ASE beat noise also dominates in our model, we use the quadratic level set for uncoded signal constellations. However, as we shall show later, the quadratic level set is not optimal when coding is considered, and power performance can be improved by jointly optimizing the intensity level set with the code. The parameter Δ denotes the minimum separation between constellation points in electric field amplitude and is related to the error performance of uncoded detection. We define the constellation shaping factor as $\gamma_s = P_{av}/\Delta^2$, which is unitless and is indicative of system average power efficiency. The system spectral efficiency (bit/s/Hz) is defined as $\mu = (\log(L)R_c)/(TB_s)$, where L is the number of intensity levels, R_c is the code rate for coded modulation (equal to 1 for an uncoded system). We will take $B_s = 1/T$ as the measure of the minimum optical bandwidth required for L-ASK modulation with proper pulse shaping and/or optical filtering, which is approximately the minimum channel spacing that can be achieved in a wavelength-division multiplexing system [8]. The energy per bit (J/bit) is then $E_b = P_{av}T/\mu$. These definitions will be used later for comparison between different modulation schemes.

In long-haul systems, optical amplifiers are inserted periodically along the fiber to compensate for transmission losses. When the optical signal passes through each optical amplifier, ASE noise is accumulated and becomes the dominant noise for the system where other sources of noise (shot noise, thermal noise, etc.) can be ignored. The effect of the cascaded optical amplifiers can be modeled as a single amplifier, which has a wide-band white Gaussian noise process $n(t)$ with power spectral density $\frac{N_0}{2}$ and a frequency-flat intensity gain of G taking account of the overall effect of fiber attenuation and amplification [3]. The amplified optical signal is denoted as

$$x_A(t) = \sqrt{G}x(t) + n(t). \tag{3}$$

At the receiver, the optical signal is filtered by a polarizer, which passes the transmitted signal and blocks out the noise in the polarization orthogonal to the signal.[1] After the polarizer, an optical band-pass filter is used to select the desired optical signal and limit ASE noise from outside of the signal band. The optical signal can be implemented by either Fabry-Perot filter or Bragg grating filter. However, to simplify our analysis, we use the idealized filter model provided in [9]. The optical filter bandwidth is assumed to be an integral multiple M of the signal bandwidth and is denoted as $B_f = M/T$. The overall effect of the optical filter is represented by the impulse response in electric field:

$$h_o(t) = \begin{cases} \sqrt{2}\frac{M}{T}\cos 2\pi f_c t, & t \in [0, \frac{T}{M}) \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

[1]Although for direct-detection systems a polarizer is not required, the presence of this polarizer allows us to eliminate the noise in orthogonal polarization, thus simplifying the mathematical analysis without changing the nature of the problem. Therefore, we use the polarizer here for mathematical convenience.

The output of the optical filter is

$$x_o(t) = x_A(t) \otimes h_o(t), \tag{5}$$

where $\otimes$ denotes convolution. A photodetector with responsivity R is used to convert the filtered optical signal into photocurrent. The photodetector is a square-law device converting optical intensity to electrical current, where the photocurrent is

$$i(t) = R \cdot x_o^2(t). \tag{6}$$

Because the optical band-pass filter's bandwidth is greater than the signal's bandwidth, an electrical low-pass filter is used to filter out excessive noise. The electrical filter bandwidth is typically assumed to be the same as the input signal bandwidth. Following [9], a discrete-time low-pass filter model is used, where the current is sampled and averaged over one symbol duration. The impulse response of the electrical filter is

$$h_e(t) = \frac{1}{M} \sum_{l=1}^{M} \delta\big(t - l\frac{T}{M}\big), \tag{7}$$

and the output to decision device is

$$I = i(t) \otimes h_e(t)\Big|_{t=T}. \tag{8}$$

After simplification, the optical fiber communication system with optical amplifiers can be represented by the abstract model shown in Figure 2, where the decision variable I is an M-th order Chi-square random variable conditioning on input intensity P:

$$I = R \sum_{l=1}^{M} (\sqrt{\frac{GP}{M}} + n_l)^2. \tag{9}$$

The n_l's are zero-mean IID Gaussian random variables with variance $\sigma^2 = \frac{N_0}{2T}$. The conditional mean and variance of the decision variable are

$$E[I|P] = RGP + \frac{RN_0M}{2T}, \tag{10}$$

$$VAR[I|P] = \frac{2R^2GPN_0}{T} + \frac{R^2N_0^2M}{2T^2}. \tag{11}$$

This agrees with [10]. The first term of the mean is from signal. The second term of the mean is from the average power of ASE noise. It introduces a small shift proportional to the optical filter bandwidth and ASE noise power spectral density. The first term of the variance is from signal-ASE beat noise. The second term of the variance is from ASE-ASE beat noise in the optical band-pass filter.

As shown in [11], the received decision variable I can be expressed in an alternative form:

$$I = R \cdot \left[(\sqrt{GP} + N)^2 + \sum_{i=1}^{M-1} N_i^2 \right], \tag{12}$$

where N and N_i are IID random variables with Gaussian distribution $\mathcal{N}(0, \sigma^2)$. From equation (12), we can interpret N as the ASE noise within the signal bandwidth and the N_i's as ASE noises within the optical filter but outside the signal band. It is obvious that signal only beats with ASE noise in the signal band and that in other bands, there is only ASE-ASE beat noise. Therefore, it is desirable to limit the bandwidth of optical filter in order to reduce the amount of ASE-ASE beat noise. However, since the optical filter bandwidth is usually greater than the signal bandwidth in practice, the additional ASE-ASE noise cannot be totally eliminated.

Although it is possible to work with the M-th order Chi-square distributed decision variable directly in symbol-by-symbol detection, the solution is usually numerical, and this approach cannot provide much insight for the case of sequence detection. Thus, we want to approximate the decision variable I with a simple distribution. One possibility is to apply the central limit theorem and to approximate I by a Gaussian random variable. However, since the central limit theorem converges slowly in M, it has been shown that the Gaussian approximation is not very accurate [12]. In this paper, we use a first-order Chi-square distribution approximation for the decision variable, which is introduced in [13]:

$$I \approx R \cdot \Big(\big(\sqrt{GP + \frac{1}{2}\sigma_{M-1}^2} + N \big)^2 + \frac{1}{2}\sigma_{M-1}^2 \Big), \tag{13}$$

where $\sigma_{M-1}^2 = (M-1)\sigma^2$, and $N \sim \mathcal{N}(0, \sigma^2)$. This approximation has the same mean and variance as the original decision variable I and is also non-negative. Figure 3 shows that the approximation is very accurate for $GP \gg \sigma_{M-1}^2$ and provides an upper-bound of tail probability for $P = 0$. As we can see, the second term of equation (12) is the ASE-ASE beat noise from outside of the signal band and it has two effects on the received signal, introducing a constant shift and increasing the variance of the received noise. To take account of these effects in the approximate model, the transmitted signal is shifted up by a constant amount ($\frac{1}{2}\sigma_{M-1}^2$) to introduce an artificial beat noise accounting for the increased variance. The received signal is further adjusted by another constant amount ($\frac{1}{2}\sigma_{M-1}^2$) to account fully for the constant shift introduced by ASE-ASE beat noise. This approximation is easier to work with than the original M-th order Chi-squared distributed decision variable since it only involves one random variable, and is much more accurate than the simple Gaussian approximation. Using this model, we can get approximations to the pairwise error probabilities that give engineering intuition and lead to simple code design criteria. Therefore, we use this approximate model for the rest of the paper.

3 Symbol-by-Symbol Detection

With the approximate channel model, we first look at the case of symbol-by-symbol detection, which serves as the basis of comparison with other schemes. We use the quadratic level set for signal constellation and the average power for L-ASK is

$$P_{av} = \frac{(L-1)(2L-1)}{6}\Delta^2. \tag{14}$$

Assuming all levels are equi-probable, the optimal detection technique is the maximum-likelihood detection

$$\hat{P} = \max_k p(I|P_k), \tag{15}$$

where the hypothesis is chosen to maximize the conditional pdf $p(I|P_k)$. The conditional pdf $p(I|P_k)$ can be expressed as

$$p(I|P_k) \approx \frac{1}{\sqrt{4RI - 2R^2\sigma_{M-1}^2}} p_N\left(\sqrt{\frac{I}{R} - \frac{\sigma_{M-1}^2}{2}} - \sqrt{GP_k + \frac{\sigma_{M-1}^2}{2}}\right), \tag{16}$$

where $p_N(x)$ denotes the pdf of $N \sim \mathcal{N}(0, \sigma^2)$.

In the case of OOK, using maximum-likelihood detection, the pairwise error probability between P_0 and P_1 is

$$\begin{aligned} Pr(P_0 \to P_1) &= Pr(P_1 \to P_0) = Q\left(\frac{|\sqrt{GP_0 + \frac{\sigma_{M-1}^2}{2}} - \sqrt{GP_1 + \frac{\sigma_{M-1}^2}{2}}|}{2\sigma}\right) \\ &\approx Q(\frac{\sqrt{G}\Delta}{2\sigma}), \quad \text{if } G\Delta^2 \gg \frac{\sigma_{M-1}^2}{2}. \end{aligned} \tag{17}$$

Since $\frac{\sigma_{M-1}^2}{2}$ is usually much smaller than $G\Delta^2$, it is often ignored in performance analysis. However, it is worth noting that when $\frac{\sigma_{M-1}^2}{2}$ increases with M, the argument of the Q function decreases and a loss of performance is incurred. This observation agrees with our earlier assertion that increasing the optical filter bandwidth degrades error performance. The overall bit-error probability for OOK is

$$P_b = \frac{1}{2}Pr(P_0 \to P_1) + \frac{1}{2}Pr(P_1 \to P_0) \approx Q(\sqrt{\frac{G\Delta^2}{4\sigma^2}}) = Q(\sqrt{\frac{\mu G E_b}{2\gamma_s N_0}}), \tag{18}$$

where the spectral efficiency and the constellation shaping factor are respectively $\mu = 1$, $\gamma_s = 1/2$. The received signal-to-noise ratio can be defined as SNR $= GE_b/N_0$. The factor in front of SNR in the Q function argument indicates the power efficiency of the modulation schemes and we define it as the asymptotic power gain for the modulation scheme, which is denoted by Γ and is unitless. For uncoded OOK,

$$\Gamma(\text{Uncoded, OOK}) = \frac{\mu}{2\gamma_s} = 1. \tag{19}$$

For L-level ASK, a similar procedure can be carried out to calculate the error performance. Using maximum-likelihood detection, the pairwise error probability between between any two hypotheses P_i and P_j is

$$Pr(P_i \to P_j) = Pr(P_j \to P_i) \approx Q(\frac{|i-j|\sqrt{G}\Delta}{2\sigma}), \quad \text{if } G\Delta^2 \gg \frac{\sigma_{M-1}^2}{2}. \tag{20}$$

Since the dominant pairwise error happens between adjacent signal levels, the bit-error probability is approximately

$$P_b \approx \sum_{i=0}^{L-2} \frac{1}{L} Pr(P_i \to P_{i+1}) + \sum_{i=1}^{L-1} \frac{1}{L} Pr(P_i \to P_{i-1}) \approx Q(\sqrt{\frac{G\Delta^2}{4\sigma^2}}) = Q(\sqrt{\frac{\mu G E_b}{2\gamma_s N_0}}), \tag{21}$$

where the spectral efficiency and constellation shaping factor are $\mu = 2$, $\gamma_s = 7/2$ for uncoded 4-ASK. The asymptotic power gain is

$$\Gamma(\text{Uncoded, 4-ASK}) = \frac{\mu}{2\gamma_s} = \frac{2}{7}. \tag{22}$$

Comparing symbol-by-symbol detection for 4-ASK and OOK, we have a 2-to-1 gain in spectral efficiency, but there is a $10\log_{10}(\frac{7}{2}) = 5.44$ dB power loss in error performance. This power loss is inherently due to multilevel signaling and will be larger if higher-level modulation is used. Therefore, we limit our attention to 4-ASK for the rest of the paper and try to find a trade-off between spectral and power efficiencies using coded modulation techniques.

4 Sequence Detection

When coded modulation is used, information bits are mapped onto sequences of symbols, introducing redundancy to improve error performance. Sequence detection is then used to detect the transmitted information bits. In this section, we study the performances of hard-decision and soft-decision sequence detection, which can be used to characterize the performance of coded modulation schemes.

Let $\vec{P}_i = [P_i[1], P_i[2], \ldots]$ denote a particular transmitted sequence and let $\vec{I} = [I[1], I[2], \ldots]$ denote the received sequence. Since fiber dispersion and nonlinearity are not considered, the channel is memoryless and the relationship between $I[k]$ and $P_i[k]$ is governed by the channel model as described above. An error event happens when $\vec{P}_i$ is transmitted but another sequence $\vec{P}_j$ is favored based on the received sequence $\vec{I}$. We define this probability as the pairwise error probability $Pr(\vec{P}_i \rightarrow \vec{P}_j)$. The system error performance P_b is characterized by the maximum pairwise error probability, which depends on the code and sequence detection technique used.

A simple extension of symbol-by-symbol detection is to make a decision of the symbol transmitted during each interval and to use the ensemble symbol sequence to detect the bit sequence transmitted. This technique is called hard-decision sequence detection. In this case, a pairwise error occurs when based on the received symbol sequence, the wrong bit sequence is favored over the original information bit sequence. If we let p denote the uncoded bit-error probability, this pairwise error probability between the code sequences can be approximated by Bhattacharyya bound [14]

$$Pr(\vec{P}_i \rightarrow \vec{P}_j) \approx [2\sqrt{p(1-p)}]^{d_h}, \tag{23}$$

where d_h denotes the Hamming distance between bit sequences corresponding to $\vec{P}_i$ and $\vec{P}_j$.

For OOK, the uncoded bit-error probability is $p = Q(\sqrt{\frac{\mu G E_b}{2\gamma_s N_0}})$, where the spectral efficiency and the constellation shaping factor are $\mu = R_c$, $\gamma_s = 1/2$, and R_c is the coding rate. The worst-case pairwise error probability is approximately

$$\max_{i \neq j} Pr(\vec{P}_i \to \vec{P}_j) \approx [2\sqrt{p(1-p)}]^{d_{min}} \approx Q(\sqrt{\frac{d_{min}\mu G E_b}{4\gamma_s N_0}}) \tag{24}$$

where d_{min} is the minimum Hamming distance between any two underlying bit sequences. Comparing this with the error probability of symbol-by-symbol detection, we can see that the error probability is reduced when $d_{min} > 2$ and we denote this error performance gain as the coding factor for hard-decision sequence detection, $\gamma_c = d_{min}/2$. Because of its simplicity, hard-decision sequence detection is often used with binary convolutional codes. With some of the best convolutional codes [15], we can get large coding gain. However, the spectral efficiency is reduced from unity to the code rate R_c. The asymptotic power gain is

$$\Gamma(\text{Hard, OOK}) = \frac{\mu\gamma_c}{2\gamma_s} = \frac{d_{min}R_c}{2}. \tag{25}$$

The results for using the convolutional codes with OOK are tabulated in Table 1.

For 4-ASK, similar procedure can be carried out, where the uncoded bit-error probability is $p = Q(\sqrt{\frac{\mu G E_b}{2\gamma_s N_0}})$ and the maximum pairwise error probability is

$$\max_{i \neq j} Pr(\vec{P}_i \to \vec{P}_j) \approx Q(\sqrt{\frac{d_{min}\mu G E_b}{4\gamma_s N_0}}). \tag{26}$$

The expression is very similar to OOK hard-decision sequence detection, but the spectral efficiency and shaping factor for 4-ASK are $\mu = 2R_c$, $\gamma_s = 7/2$. The coding gain is still defined as $\gamma_c = d_{min}/2$. The asymptotic power gain is

$$\Gamma(\text{Hard, 4-ASK}) = \frac{\mu\gamma_c}{2\gamma_s} = \frac{d_{min}R_c}{7}. \tag{27}$$

Comparing hard-decision sequence detection with symbol-by-symbol detection for the same modulation, the error performance is improved due to the coding gain, but the spectral efficiency is reduced. Just as in symbol-by-symbol detection, if the same code is used from OOK and 4-ASK, the 4-ASK has twice the spectral efficiency but incurs a 5.44 dB power loss.

The system error performance can be improved by using soft-decision sequence detection, which preserves received decision variables during each interval and uses the decision variable sequence to detect the transmitted bit sequence. Assuming all transmitted sequences are equally probable, the optimal detection is the maximum-likelihood sequence detection, where the joint conditional sequence probability of $\vec{I}$ is used to select the best hypothesis:

$$\tilde{P}_i = \arg\max_{\vec{P}_i}[p(\vec{I}|\vec{P}_i)]. \tag{28}$$

$p(\vec{I}|\vec{P}_i)$ denotes the conditional probability density function and can be expressed as the product of per symbol conditional pdf since the channel is memoryless:

$$p(\vec{I}|\vec{P}_i) = \prod_k p(I[k]|P_i[k]) \tag{29}$$

If we define the path metric as

$$PM(\vec{I},\vec{P}_i) = \sum_{P_i[k]} \left\| \sqrt{\frac{I[k]}{R} - \frac{\sigma_{M-1}^2}{2}} - \sqrt{GP_i[k] + \frac{\sigma_{M-1}^2}{2}} \right\|^2, \tag{30}$$

the maximum-likelihood sequence detection rule can be expressed as

$$PM(\vec{I},\vec{P}_i) \underset{\hat{P}\neq\vec{P}_i}{\overset{\hat{P}\neq\vec{P}_j}{\lessgtr}} PM(\vec{I},\vec{P}_j). \tag{31}$$

The pairwise error probability is

$$Pr(\vec{P}_i \to \vec{P}_j) = Pr(\ PM(\vec{I}, \vec{P}_i) - PM(\vec{I}, \vec{P}_j) > 0 \mid \vec{P}_i\ ). \tag{32}$$

Defining the distance function between two code sequences as

$$d_{i,j}^2 = \sum_k \| \sqrt{P_i[k] + \frac{\sigma_{M-1}^2}{2G}} - \sqrt{P_j[k] + \frac{\sigma_{M-1}^2}{2G}} \|^2 \approx \sum_k \| \sqrt{P_i[k]} - \sqrt{P_j[k]} \|^2, \tag{33}$$

then, the pairwise error probability is

$$Pr(\vec{P}_i \to \vec{P}_j) \approx Q(\sqrt{\frac{G\ d_{i,j}^2}{4\sigma^2}}). \tag{34}$$

Defining the free Euclidean distance as $d_f^2 = \min_{i \neq j}\ (d_{i,j}^2)$ and defining the coding factor for soft-decision detection as $\gamma_c = d_f^2/\Delta^2$, then, the maximum pairwise error probability is

$$\max_{i \neq j} Pr(\vec{P}_i \to \vec{P}_j) = Q(\sqrt{\frac{Gd_f^2}{4\sigma^2}}) = Q(\sqrt{\frac{\gamma_c \mu G E_b}{2\gamma_s N_0}}). \tag{35}$$

When OOK modulation is used, the spectral efficiency and constellation shaping factor are respectively $\mu = R_c$, $\gamma_s = 1/2$. The overall asymptotic power gain is

$$\Gamma(\text{Soft, OOK}) = \frac{\mu \gamma_c}{2\gamma_s} = \frac{d_f^2 R_c}{\Delta^2}. \tag{36}$$

When 4-ASK is used, the spectral efficiency and constellation shaping factor are respectively $\mu = 2R_c$, $\gamma_s = 7/2$, and the overall asymptotic power gain is

$$\Gamma(\text{Soft, 4-ASK}) = \frac{\mu \gamma_c}{2\gamma_s} = \frac{2d_f^2 R_c}{7\Delta^2}. \tag{37}$$

We notice that the relationship between d_{min} and d_f is $d_f^2 \geq d_{min}\Delta^2$. The equality holds when OOK is used. Therefore, for the same code, the coding factor of hard-decision detection is half of soft-decision detection. Hence, there is a 3 dB power gain using soft-decision detection.

We see that the MLSD can be done by using Viterbi algorithm with the defined path metric as in equation (30). The pairwise error probability between two coded sequences depends approximately on the Euclidean distance between the electric field amplitude $\sqrt{P_i[k]}$ of these sequences. This gives us a criterion for designing coded modulation: the goal is to maximize the minimum Euclidean distance between the electric field amplitude of the transmitted symbol sequences. This means that we can use coded modulation schemes for this channel, just as for the AWGN channel.

5 Coded Modulation Schemes

In previous section, we showed that, under MLSD, the pairwise error probability is related to the Euclidean distance between the electric field amplitudes of transmitted signal sequences, where the average power of the transmitted sequences is measured by the mean of optical intensities, which is the mean square of the electric field amplitudes. Therefore, if we treat the electric field amplitude of the optical signal as the transmitted symbol, the same coded modulation schemes designed for AWGN channel can be used for optical amplified fiber communication systems. In this section, we evaluate the improvement of performance with coded modulation schemes such as trellis-coded modulation and multi-level coding.

Trellis-coded modulation for AWGN channels was first introduced in [16] and has been summarized in [4]. The essence of TCM is to expand the signal constellation and to use a convolutional encoder to select the signal points forming coded sequences. The mapping of constellation points is done to maximize the minimum Euclidean distance between code sequences. The decoding is done by MLSD, usually using a Viterbi algorithm with Euclidean distance metric. Because coding is done on the expanded signal constellation, the overall data rate loss is less than using simple convolutional coding.

For 4-ASK, the signal set is partitioned into two subset as shown in Figure 4. Since MLSD is used for decoding, the error performance is

$$P_b \approx Q(\sqrt{\frac{Gd_f^2}{4\sigma^2}}) = Q(\sqrt{\frac{\gamma_c \mu G E_b}{2\gamma_s N_0}}). \quad (38)$$

For 4-ASK, only rate-$\frac{1}{2}$ trellis codes can be used, since the denominator of the code rate must be $\log(L)$. Therefore, the spectral efficiency and the shaping factor are respectively $\mu = 1$, $\gamma_s = 7/2$. Since the electric field amplitude of the signal constellation has the shape of a $\mathbb{Z}_1$ lattice, the free Euclidean distance d_f^2 is the same as shown in [4]. The coding factor is still defined as $\gamma_c = d_f^2/\Delta^2$. The asymptotic power gain is

$$\Gamma(\text{Soft, 4-ASK TCM}) = \frac{\mu\gamma_c}{2\gamma_s} = \frac{d_f^2}{7\Delta^2}. \quad (39)$$

The results for TCM are shown in Table 1.

Using some simple trellis codes, we can achieve the same spectral efficiency as uncoded OOK, while achieving better error performance. However, the drawback of TCM is that it uses Euclidean distance, which means it has to use soft-decision detection output for Viterbi decoder. Such a design requires very high speed A/D converters, which are just becoming available for 10 Gb/s systems and are not currently available for 40 Gb/s systems. Also, for 4-ASK, only rate-$\frac{1}{2}$ trellis codes can be used, which limits the flexibility of system design. Therefore, we consider other coded modulation schemes.

Another branch of coded modulation is the multi-level coding technique. It was first proposed in [17] and has been summarized in [5]. Similar to the set partition of TCM shown in Figure 4, MLC also partition the constellation points into $\log_2(L)$ levels with two cosets at each level. One address bit is associated with each level and binary component codes are used to protect the address bits. The component codes are decoded in multiple stages, where the lower-level component codes are decoded first and then the higher-level component codes are decoded using the information of the lower-level address bits. Since each level partition in MLC can be considered as an equivalent channel with a binary input, component codes with different rates can be used as long as their rates are within the capacity of the equivalent channels [5].

Because the quadratic level set constellation is used, the shaping factor is still $\gamma_s = 7/2$. Since binary codes are used for each level, the overall spectral efficiency is the sum of the code rates of the individual codes, $\mu = \sum_{k=0}^{\log_2(L)-1} R_k$, where R_k is the individual code rate. We denote the Hamming distance of the binary code at level k as h_k. Then the performance of hard-decision decoding at level k is characterized as $P_b^{(k)} \approx [2\sqrt{p^{(k)}(1-p^{(k)})}]^{h_k}$, where $p^{(k)}$ is the uncoded bit-error probability at that level. Under the set partition, the minimum intra-set distances are $\{\Delta_0 = \Delta,\ \Delta_1 = 2\Delta\}$. Therefore, the uncoded bit-error probability for level k is $p^{(k)} = Q(\sqrt{G\Delta_k^2/4\sigma^2})$. Defining the coding gain at level k as $\gamma_k = (h_k\Delta_k^2)/(2\Delta^2)$, the overall bit-error rate is

$$P_b = \sum_{k=0}^{\log_2(L)-1} R_k \cdot P_b^{(k)} \approx \max_k Q(\sqrt{\frac{\gamma_k \mu G E_b}{2\gamma_s N_0}}), \tag{40}$$

which is dominated by the worst level k. The overall asymptotic power gain is

$$\Gamma(\text{Hard, 4-ASK MLC}) = \min_k \frac{\mu\gamma_k}{2\gamma_s} = \Big(\min_k \frac{h_k\Delta_k^2}{14\Delta^2}\Big)\cdot\Big(\sum_{k=0}^{\log_2(L)-1} R_k\Big). \tag{41}$$

Besides hard-decision sequence detection, soft-decision sequence detection can be performed at each level as well. The error performance at level k is characterized as $P_b^{(k)} = Q(\sqrt{Gd_{f,k}^2/4\sigma^2})$, where $d_{f,k}^2$ is the free distance at level k. As in section IV-B, $d_{f,k}^2 \geq h_k\Delta_k^2$. The coding gain of each level is $\gamma_k = d_{f,k}^2/\Delta^2$ and the overall bit-error rate is

$$P_b \approx \sum_k R_k \cdot Q(\sqrt{\frac{\gamma_k \mu G E_b}{2\gamma_s N_0}}) \approx \max_k Q(\sqrt{\frac{\gamma_k \mu G E_b}{2\gamma_s N_0}}). \tag{42}$$

The overall coding gain is defined as

$$\Gamma(\text{Soft, 4-ASK MLC}) = \min_k \frac{\mu\gamma_k}{2\gamma_s} = \Big(\min_k \frac{d_{f,k}^2}{7\Delta^2}\Big) \cdot \Big(\sum_{k=0}^{\log_2(L)-1} R_k\Big). \tag{43}$$

Since the overall gain of MLC is limited by the worst case among all component codes, it is possible to find optimal combinations of code rate and code complexity to get the highest coding gain. Since Δ_k is larger at higher levels, a higher-rate code with smaller Hamming distance can be used. In Table 1, we show the results of combining rate-$\frac{1}{4}$ or rate-$\frac{1}{2}$ rate codes with rate-$\frac{3}{4}$ codes. Moreover, since soft-decision detection has better performance, we can decide whether to use soft- or hard-decision decoding for each component code individually to optimize performance [18].

6 Asymmetric Modulation

Up to this point, we have assumed that the signal constellation is the quadratic level set. Although this constellation is approximately optimal for symbol-by-symbol detection, it can be modified to obtain power gain in coded modulation. Since each signal point is protected differently by the coding scheme, it is possible to shift these signal points to balance the protection achieved by coding and by signal separation. By optimizing the shifts of signal points, the overall asymptotic power gain can be increased. In this section, we explore the performance gain by using a non-quadratic level set with TCM and MLC.

Asymmetric TCM, where signal constellation and trellis-coded modulation are jointly optimized, has been proposed for AWGN channel in [6]. The constellation and set partition are shown in Figure 5. We allow $\sqrt{P_1}$ and $\sqrt{P_3}$ to be shifted by $\delta\Delta$, where δ is the shift ratio. The range of the shift ratio is $-1 \leq \delta \leq 1$, which allows the signal points to be shifted either way. The average power is

$$P_{av} = \frac{1}{4}(0^2 + (1-\delta)^2 + 2^2 + (3-\delta)^2)\Delta^2 = \frac{7 - 4\delta + \delta^2}{2}\Delta^2, \quad (44)$$

and the shaping factor is $\gamma_s = (7 - 4\delta + \delta^2)/2$. As in the case of regular TCM, the pairwise error probability is

$$P(\vec{P}_i \to \vec{P}_j) = Q(\sqrt{\frac{Gd_{i,j}^2}{4\sigma^2}}) \approx \exp(-\frac{\mu G E_b d_{i,j}^2}{4N_0 P_{av}}), \quad (45)$$

where the spectral efficiency is $\mu = 2R_c$. As in [6], we define $\gamma_c = d_f^2/\Delta^2 = \min_{i \neq j}\ (d_{i,j}^2/\Delta^2)$, $T_{i,j}^2 = d_{i,j}^2\ \mu/P_{av}$, $D = \exp(-GE_b/4N_0)$. Then, the asymptotic gain is

$$\frac{\gamma_c \mu}{2\gamma_s} = \frac{1}{2}\min_{i \neq j} T_{i,j}^2 = \frac{1}{2}\lim_{D \to 0} \log_2 \frac{T(2D,1)}{T(D,1)}, \quad (46)$$

where $T(D,I)$ is the super-state transfer function.

The super-state function can be derived from the super-state transition diagram. For example, for rate-$\frac{1}{2}$ trellis code with two states, the state trellis diagram and the super-state transition diagram are shown in Figure 6. The super-state transfer function is

$$\begin{aligned} T(D,I) &= \frac{2a(c+d)}{1-2b}, \\ a &= \frac{1}{2}ID^{\frac{4\mu}{\gamma_s}}, \quad b = \frac{1}{2}ID^{\frac{(1-\delta)^2\mu}{\gamma_s}}, \\ c &= \frac{1}{2}D^{\frac{(3-\delta)^2\mu}{\gamma_s}}, \quad d = \frac{1}{2}D^{\frac{(1+\delta)^2\mu}{\gamma_s}}. \end{aligned} \quad (47)$$

Thus,

$$\Gamma(\text{Soft, Asym. 4-ASK TCM}) = \max_{-1<\delta<1} \frac{\mu\gamma_c}{2\gamma_s} = \max_{-1<\delta<1} \frac{4 + (1+\delta)^2}{7 - 4\delta + \delta^2}. \quad (48)$$

The maximum gain $\Gamma = 2$ is achieved when δ approaches 1. However, when $\delta = 1$, the two partitioned subsets merge, which leads to a catastrophic code. Thus, in practice, δ need to be chosen empirically and cannot be 1. Using similar analysis, we found that for 4-state and

8-state asymmetrical trellis codes, the maximum coding gain $\Gamma = 2$ is also approached when δ approaches 1. Compared with OOK, a maximum of 3 dB can be gained in each case.

Similar principles can be applied to multi-level coding. When different codes are used for the equivalent binary channels, the coding gains for the binary channels are different. Therefore, we can shift the signal constellation to change the uncoded bit-error probabilities for the equivalent channels to improve the worst-case binary channel which limits the system performance. The overall asymptotic power gain is maximized when the equivalent binary channels provide equal error protection.

The constellation and set partition are shown in Figure 5. The minimum intra-set distances are $\{\Delta_0 = (1-\delta)\Delta,\ \Delta_1 = 2\Delta\}$. As we have shown in previous section, the spectral efficiency and the shaping factor are respectively $\mu = \sum_{k=0}^{\log_2(L)-1} R_k$, $\gamma_s = (7 - 4\delta + \delta^2)/2$. If hard-decision sequence detection is used for the component codes, the coding gain is $\gamma_k = (h_k \Delta_k^2)/(2\Delta^2)$ and

$$\begin{aligned}\Gamma(\text{Hard, Asym. 4-ASK TCM}) &= \max_{-1<\delta<1} \min_k \frac{\mu\gamma_k}{2\gamma_s} \qquad (49)\\ &= \max_{-1<\delta<1} \Big(\min_k \frac{h_k\Delta_k^2}{2(7-4\delta+\delta^2)\Delta^2}\Big)\cdot\Big(\sum_{k=0}^{\log_2(L)-1} R_k\Big).\end{aligned}$$

If soft-decision sequence detection is used, the coding gain at level-k is $\gamma_k = d_{f,k}^2/\Delta^2$ and

$$\begin{aligned}\Gamma(\text{Soft, Asym. 4-ASK TCM}) &= \max_{-1<\delta<1} \min_k \frac{\mu\gamma_k}{2\gamma_s} \qquad (50)\\ &= \max_{-1<\delta<1} \Big(\min_k \frac{d_{f,k}^2}{(7-4\delta+\delta^2)\Delta^2}\Big)\cdot\Big(\sum_{k=0}^{\log_2(L)-1} R_k\Big).\end{aligned}$$

Thus, it is possible to choose h_i and δ such that Γ is maximized. For example, with the rate-$\frac{1}{2}$ and rate-$\frac{3}{4}$ component codes, the rate-$\frac{1}{2}$ code limits the overall power gain. By varying δ to increase the overall gain with the rate-$\frac{1}{2}$ code while decreasing the gain with the rate-$\frac{3}{4}$ code, the maximal power gain can be found when the equivalent channel gains are equal.

7 Comparison of Schemes and Conclusions

The results of different coded and uncoded modulation schemes with OOK and 4-ASK are show in Table 1. The SNR requirements to achieve $P_b = 10^{-12}$ versus the achieved spectral efficiency are plotted for the coded and uncoded modulation schemes in Figure 7. Only soft-decision decoding gains are presented in the table and figures. When hard-decision decoding is used, 3 dB should be subtracted from the soft-decision coding gain. In Figure 7, the channel capacity of the binary symmetric channel corresponding to OOK, the channel capacity bound corresponding to 4-ASK and the channel capacity bound for intensity modulated systems using optical amplifiers [2] are also plotted for comparison.

In Figure 7, the channel capacity bounds clearly indicates that although OOK is appropriate for low SNR, employing 4-ASK is beneficial when SNR is moderately high. Moreover, in the SNR range of 15 dB to 20 dB, it is sufficient to use 4-ASK. Since uncoded 4-ASK modulation will introduce 5.44 dB loss of performance, it is necessary to use coded modulation schemes such as TCM or MLC. TCM provides 1 to 3 dB of performance gain with moderate complexity, but can only be used with soft-decision decoding. When MLC with soft-decision decoding is used, coding gain comparable to TCM can be achieved while allowing hard-decision decoding. With longer constraint length, hard-decision MLC can achieve a few dB of performance gain over OOK while achieving the same, or higher, spectral efficiency. Further performance gain can be achieved by using asymmetrical coded modulation schemes, which jointly optimize the code and signal constellation.

We have shown here that coded modulations are effective in improving spectral efficiency while maintaining power efficiency. In our opinion, it is particularly attractive to use asymmetrical MLC for optical fiber systems with optical amplifiers, since it can improve both power and spectral efficiency while allowing hard-decision detection. We expect that other techniques, such as multi-dimensional TCM and shaping codes, can also be effective.

References

[1] Jack H. Winters, Richard D. Gitlin, and Sanjay Kasturia. Reducing the effects of transmission impairments in digital fiber optic sytems. *IEEE Communications Magazine*, June 1993.

[2] Antonio Mecozzi and Mark Shtaif. On the capacity of intensity modulated systems using optical amplifiers. *IEEE Photonics Technology Letters*, September 2001.

[3] Govind P. Agrawal. *Fiber-Optic Communication Systems*. John Wiley & Sons, Inc., 1997.

[4] Gottfried Ungerboeck. Trellis-coded modulation with redundant signal sets, part I and II. *IEEE Communications Magazine*, February 1987.

[5] Udo Wachsmann, Robert F. H. Fischer, and Johannes B. Huber. Multilevel codes: Theoratical concepts and practical design rules. *IEEE Transactions on Information Theory*, July 1999.

[6] Dariush Divsalar, Marvin K. Simon, and Joseph H. Yuen. Trellis coding with asymmetric modulations. *IEEE Transactions on Communications*, February 1987.

[7] J. Rebola and A. Cartaxo. Optimization of level spacing in quaternary optical communication systems. In *Proceedings of the SPIE Applications of Photonic Technology*, Quebec, Canada, June 2000.

[8] S. Banerjee, D. F. Grosz, A. P. Kung, D. N. Maywar, A. Gurevich, T. H. Wood, and A. Agarwal. Ultra-high-capacity long-haul 40-Gb/s WDM transmission with 0.8-b/s/Hz spectral efficiency by means of strong optical filtering. *IEEE Photonics Technology Letters*, 2003.

[9] Ozan K. Tonguz and Leonid G. Kazovsky. Theory of direct-detection lightwave receivers using optical ampifiers. *Journal of Lightwave Technology*, Februry 1991.

[10] N. A. Olsson. Lightwave systems with optical ampifiers. *Journal of Lightwave Technology*, July 1989.

[11] H. O. Lancaster. *The Chi-squared Distribution*. John Wiley & Sons, Inc., 1969.

[12] Dietrich Marcuse. Derivation of analytical expressions for the bit-error probability in lightwave systems with optical amplifiers. *Journal of Lightwave Technology*, December 1990.

[13] Munuswamy Sankaran. Approximations to the non-central chi-square distribution. *Biometrika*, 50:199–204, June 1963.

[14] Stephen B. Wicker. *Error Control Systems for Digital Communication and Storage*. Prentice Hall, Inc., 1995.

[15] S. Lin and D.J. Costello Jr. *Error Control Coding: Fundamentals and Applications*. Prentice Hall, Inc., 1973.

[16] Gottfried Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Transactions on Information Theory*, January 1982.

[17] H. Imai and S. Hirakawa. A new multilevel coding method using error correcting codes. *IEEE Transactions on Information Theory*, May 1977.

[18] Jay N. Livingston and Hadi G. Jardak. The performance of multilevel coding employing hard decision and hybrid decoders on fading and gaussian channels. *IEEE Transactions on Communications*, February/March/April 1995.

Modulation Type	Constr. Length	Asymmetric Shift Ratio δ	Shaping Factor γ_s	Coding Factor γ_c	Spectr. Efficiency μ (Bit/Sec./Hz)	Asymp. Power Gain Γ (dB)
OOK	0	0	1/2	1	1	0
Conv. Code (Soft)	1	0	1/2	3	1/2	0.38
	2	0	1/2	5	1/2	2.61
	3	0	1/2	6	1/2	3.39
	4	0	1/2	7	1/2	4.06
	5	0	1/2	8	1/2	4.64
	6	0	1/2	10	1/2	5.61
	7	0	1/2	10	1/2	5.61
4-ASK	0	0	7/2	1	2	-5.44
Trellis Coded Modulation	1	0	7/2	5	1	-1.46
	2	0	7/2	9	1	1.09
	3	0	7/2	10	1	1.55
	4	0	7/2	11	1	1.96
	5	0	7/2	13	1	2.69
	6	0	7/2	14	1	3.01
	7	0	7/2	16	1	3.59
	8	0	7/2	17	1	3.85
Asymm. TCM	1	1	2	8	1	3.01
	2	1	2	8	1	3.01
	3	1	2	8	1	3.01
Multi-level Coding (Soft)	6	0	7/2	16	1/4+3/4	3.59
	9	0	7/2	24	1/4+3/4	5.35
	12	0	7/2	32	1/4+3/4	6.60
	6	0	7/2	8	1/2+3/4	1.55
	9	0	7/2	12	1/2+3/4	3.31
	12	0	7/2	15	1/2+3/4	4.28
Asymm. MLC (Soft)	6	0.057	3.387	16	1/4+3/4	3.73
	6	-0.414	4.414	16	1/2+3/4	3.55
	9	-0.414	4.414	24	1/2+3/4	5.31
	12	-0.461	4.527	32	1/2+3/4	6.45

Table 1: Comparison table of various uncoded and coded modulation schemes. Only results for soft-decision detection are listed. When hard-decision detection is possible, 3 dB of power loss should be subtracted. δ, γ_s, γ_c, μ and Γ are defined in the text. Convolutional codes assume OOK modulation while the coded modulations assume 4-ASK modulation.

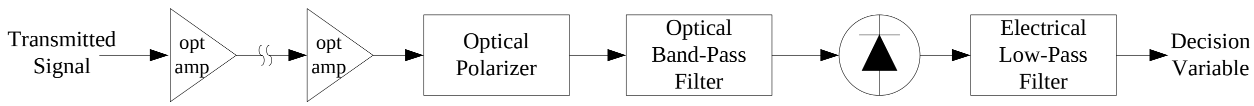


Figure 1: System model.

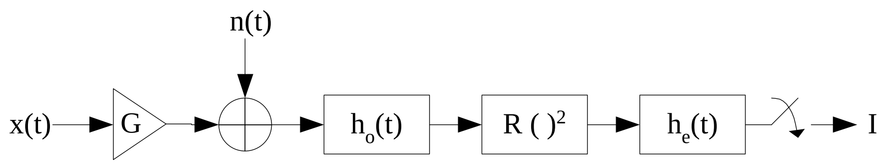


Figure 2: Channel model.

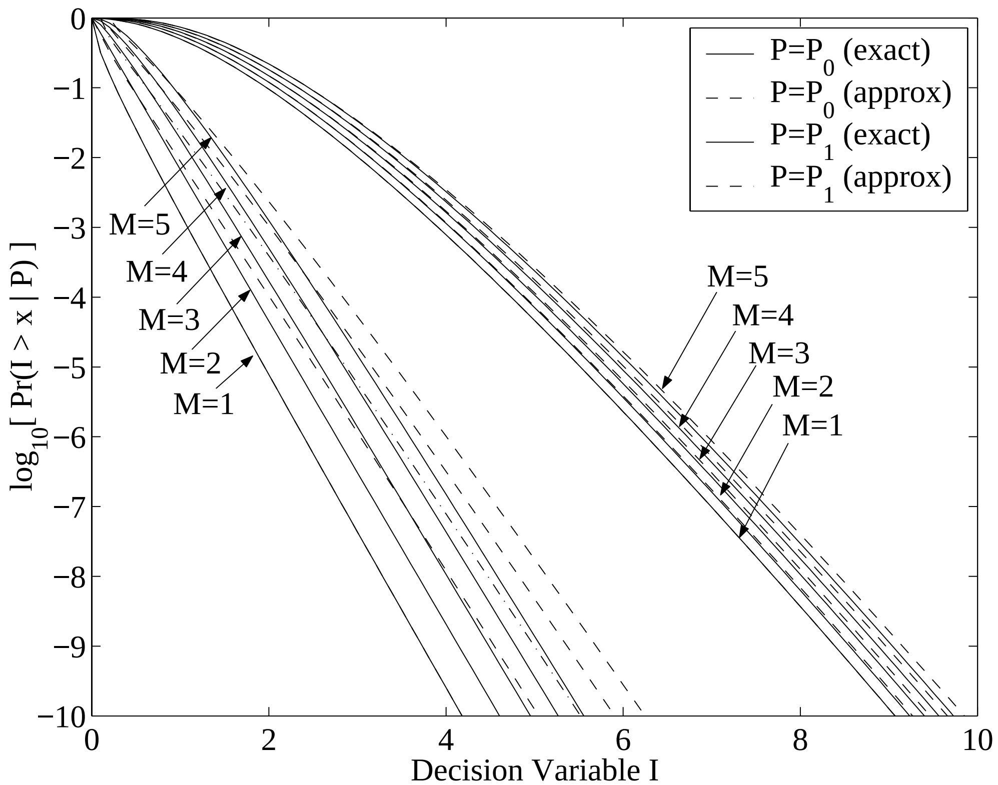


Figure 3: Comparison of exact and approximate complementary cdf of the decision variable I conditioning on $P = P_0$ or $P = P_1$ in logarithmic scale. M is the ratio between optical filter bandwidth and signal bandwidth. This figure shows that the approximation is accurate when P is big or M is small.

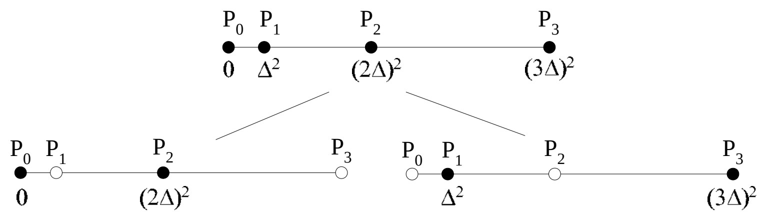


Figure 4: Quadratic 4-ASK constellation partition.

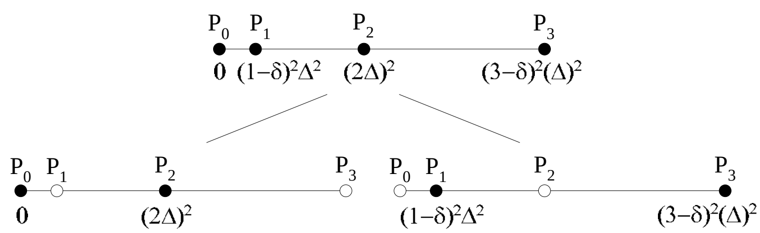
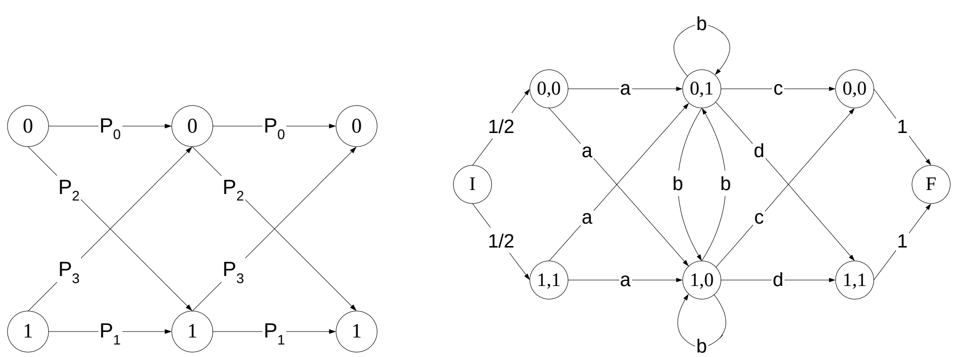


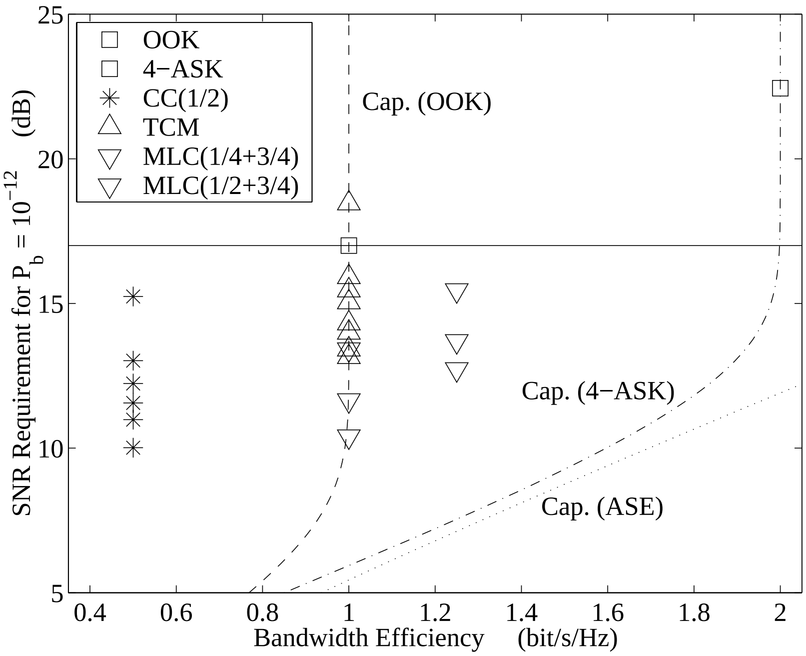
Figure 5: Non-quadratic 4-ASK constellation partition.



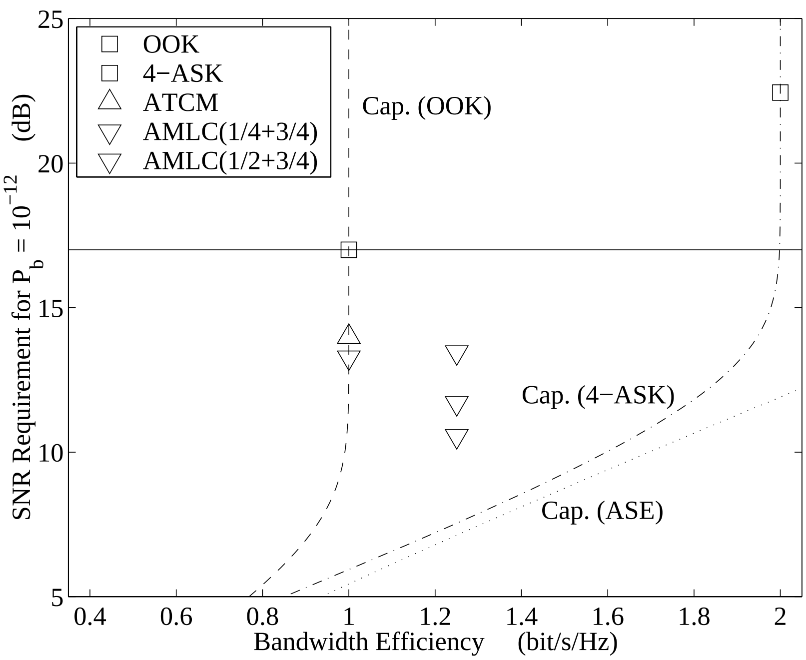
(a) Two state trellis.

(b) Transition diagram.

Figure 6: Two state trellis diagram and super-state transition diagram.



(a) Soft-decision detection with quadratic level set.



(b) Soft-decision detection with asymmetrical modulation.

Figure 7: Comparison of uncoded and coded modulation schemes. The spectral efficiencies of various modulation schemes are compared with the spectral efficiency limit bounds of OOK, 4-ASK and unconstrained direct-detection.