Maximum-Likelihood Classification for Digital Amplitude-Phase Modulations

Wen Wei and Jerry M. Mendel, *Fellow, IEEE*

Abstract—**In this paper, we apply the maximum-likelihood (ML) method to the classification of digital quadrature modulations. We show that under an ideal situation as stated later, the I-Q domain data are sufficient statistics for modulation classification and obtain a generic formula for the error probability of a ML classifier. Our study of asymptotic performance shows that the ML classifier is capable of classifying any finite set of distinctive constellations with zero error rate when the number of available data symbols goes to infinity.**

Index Terms—**Maximum-likelihood detection, pattern classification, quadrature amplitude modulation.**

I. Introduction

MODULATION classification (MC) is a technique to identify the modulation type of a modulated signal corrupted by noise. A formal description of the MC problem is as follows.

Given a measurement $r(t), 0 \leq t \leq \tau$, find the modulation type of $r(t)$ from c possible modulations, $\{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_c\}$.

In the past decade, various methods have been developed for this problem. In some of the earlier works [1], [2], one or more signal statistics, such as spectrum, moments, zero crossings, etc., are picked as discriminant functions. These methods are usually suboptimal because the discriminating statistics are empirical.

From Bayes decision theory, minimum error-rate classification can be achieved by finding the maximum among c *a posteriori* probabilities $P(\mathcal{I}_i \mid r(t))$, $i = 1, 2, \cdots, c$. If all modulation types are equally likely, then the optimal classifier is the maximum-likelihood (ML) classifier, which finds the maximum among c conditional probabilities $P(r(t) \mid \mathcal{I}_i)$. Polydoros and Kim [4] and Kuo and Lin [3] pioneered in the area of likelihood MC. These works lean more toward the practical side. Their performance analyzes are either simulation-based or are limited to a small set of modulations.

In this study, we consider the MC problem in an ideal situation where all signal parameters as well as the noise power are known, and, in addition, the data symbols are independent and the pulse shape is rectangular. Our goal is to develop a theoretical performance analysis of the generic ML classifier that is applicable to any digital amplitude-phase modulation, including phase-shift keying (PSK0, pulse amplitude modulation (PAM), quadrature amplitude modulation (QAM), etc., which are widely used in modern communications. For example, QAM's are often used in digital microwave communications, and V.29 and V.34 (which are QAM modulations) are used in telephone-line modems. The technologies for high-speed internet access, such as ISDN, HDSL, and ADSL, also belong to this category of modulations.

Paper approved by B. Vucetic, the Editor for Modulation of the IEEE Communications Society. Manuscript received September 29, 1997; revised September 22, 1998 and September 23, 1998. This work was supported by the Center for Research on Applied Signal Processing, University of Southern California. This paper was presented in part at the 29th Asilomar Conference on Signals, Systems, and Computers, Asilomar, CA, October 29–November 1, 1995.

W. Wei was with the Signal and Image Processing Institute, Department of Electrical Engineering Systems, University of Southern California, Los Angeles, CA 90089-2564 USA. He is now with Netscreen Technologies, Santa Clara, CA 95051 USA.

J. M. Mendel is with the Signal and Image Processing Institute, Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089–2564 (e-mail: mendel@sipi.usc.edu).

Publisher Item Identifier S 0090-6778(00)01585-3.

II. ML Classification under Ideal Conditions

A. Preliminaries

A digital amplitude-phase modulation uses the amplitude and phase information of a signal during time segments (symbols) to carry information and has the following generic form:

$$s(t) = \mathcal{R}\left\{ \sum_{n=-\infty}^{k} A\hat{S}_n p(t - nT) e^{j(2\pi f_c t + \theta_c)} \right\}, \quad kT < t \leq (k+1)T, \quad k = 0, 1, 2, \cdots \tag{1}$$

where f_c and θ_c are the carrier frequency and phase, respectively, T is the symbol period, $p(t)$ is a spectrum-shaping pulse function, A is the signal amplitude, and $\hat{S}_n$ assumes a value from a set of M complex numbers $\{S_1, S_2, \cdots, S_M\}$, which is called the *constellation* of the modulation. Our MC problem is to identify the transmitted constellation based on the the following noise-corrupted received signal:

$$r(t) = s(t) + n(t), \qquad 0 \leq t \leq NT \tag{2}$$

where $n(t)$ is a white Gaussian noise with power density N_0.

In the ideal situation stated earlier, $A, p(t), f_c$, and θ_c, as well as the noise power N_0, are known. Although this situation is normally not the case for noncooperative communications, we believe that this study is important for the following reasons. 1) The method to develop a classifier under ideal conditions can be used as a reference for developing variations when one or more parameters are unknown, e.g., we can use estimates of unknown parameters in the ideal classifier. 2) This ideal classifier provides an upper bound on the performance of any classifier and therefore provides a way to determine the minimum resources (signal-to-noise ratio (SNR), number of symbols) needed to achieve a desired performance.

B. Sufficient Statistics for MC

From detection theory [6], a discrete set of sufficient statistics can be extracted for the detection of continuous-waveforms. Here, we follow this idea and investigate the sufficient statistics for the MC problem. Assume that the pulse-shaping function $p(t)$ is causal and has a finite time span of L symbols, i.e.,

$p(t) = 0$ if $t < 0$ or $t > LT$. In order to utilize the existing results in detection theory, we consider the following detection problem:

$$\text{Hypothesis } H_m\text{: } r(t) = s(t \mid \mathbf{s}_m) + n(t), \quad 0 \le t \le NT, \quad m = 1, 2, \cdots, N_s$$

where $\mathbf{s}_m$ is a particular realization of complex random sequence $\{\hat{S}_n, n = -L+1, \cdots, 0, 1, \cdots, N-1\}$, i.e., $\mathbf{s}_m = \{\hat{S}_{mn}, n = -L+1, \cdots, 0, 1, \cdots, N-1\}, N_s$ is the total number of all possible realizations of $\{\hat{S}_n\}$, and $s(t \mid \mathbf{s}_m)$ is a realization of random signal $s(t)$, based on the above realization of $\hat{S}_n$. Note that the index n starts at a negative number $-L+1$ because the pulse function $p(t)$ has a time span of L symbols.

Suppose there are c possible constellations, and M_i is the number of points in constellation $\mathcal{I}_i$; then, the maximum number of different $\mathbf{s}_m$ for the entire c modulations is $N_\mathbf{s} = \sum_{i=1}^{c} M_i^{N+L}$. From detection theory [6], the following is a set of sufficient statistics for detecting $\mathbf{s}_m$:

$$\hat{L}_m = \int_0^{NT} r(t) s(t \mid \mathbf{s}_m)\, dt, \qquad m = 1, 2, \cdots, N_\mathbf{s}. \tag{3}$$

Substitute $s(t \mid \mathbf{s}_m)$ from (1) into (3); after some calculus operations, we obtain

$$\begin{aligned}\hat{L}_m &= \int_0^{NT} r(t) \sum_{n=-L+1+k}^{k} \mathcal{R}\left\{ A\hat{S}_{\rm mn} p(t-nT) \cdot e^{j(2\pi f_c t + \theta_c)} \right\} dt \\ &= \sum_{n=-L+1}^{N-1} \mathcal{R}\left\{ A\hat{S}_{\rm mn} \int_{\max(0,n)T}^{NT} r(t) p(t-nT) \cdot e^{j(2\pi f_c t + \theta_c)} \, dt \right\}.\end{aligned} \tag{4}$$

Define

$$r_n = A \int_{\max(0,n)T}^{NT} r(t) p(t-nT) e^{j(2\pi f_c t + \theta_c)}\, dt, \quad n = -L+1, \cdots, 0, \cdots, N-1 \tag{5}$$

then

$$\hat{L}_m = \sum_{n=-L+1}^{N-1} \mathcal{R}\left\{ r_n \hat{S}_{\rm mn} \right\} = \mathbf{s}_m^t \mathbf{r} \tag{6}$$

where $\mathbf{r} = \mathrm{col}\{\mathrm{r_n, n = -L+1, \cdots, 0, 1, \cdots, N-1}\}$. Let $\hat{\mathbf{L}} = \mathrm{col}\{\hat{\mathrm{L}}_\mathrm{m}, \mathrm{m} = 1, 2, \cdots, \mathrm{N_s}\}$, then from (6), we see that random vector $\mathbf{r}$ is the only random factor to determine random vector $\hat{\mathbf{L}}$ (note that $\mathbf{s}_m$ is deterministic); therefore

$$P(H_m \mid \hat{\mathbf{L}}) = P(H_m \mid \mathbf{r}). \tag{7}$$

Since $\hat{\mathbf{L}}$ is a set of sufficient statistics, we have $P(H_m \mid r(t)) = P(H_m \mid \hat{\mathbf{L}})$, and

$$\begin{aligned}P(\mathcal{I}_i \mid r(t)) &= \sum_m P(\mathcal{I}_i \mid H_m) P(H_m \mid r(t)) \\ &= \sum_m P(\mathcal{I}_i \mid H_m) P(H_m \mid \hat{\mathbf{L}}).\end{aligned} \tag{8}$$

Because $P(\mathcal{I}_i \mid H_m)$ does not depend on $r(t)$, from (8) and (7), we conclude that $\mathbf{r}$ is a set of sufficient statistics for the MC problem.

C. ML Modulation Classifier

In the rest of this paper, we will only consider the case when $p(t)$ is a rectangular function over interval $[0, T]$ (i.e., $L = 1$). In this special case, it is easy to see from (5) that $r_n = 0$ for $n < 0$, and (5) reduces to

$$r_n = A \int_{nT}^{(n+1)T} r(t) e^{j(2\pi f_c t + \theta_c)}\, dt, \quad n = 0, 1, \cdots, N-1. \tag{9}$$

From this point on, we change the index of r_n from n to k, the symbol index, to reflect the fact that each r_n is related to only one symbol of received data.

The in-phase and quadrature portion of r_k is

$$r_{I,k} = \frac{AT}{2} \mathcal{R}\{\hat{S}_k\} + n_{I,k} \tag{10}$$

$$r_{Q,k} = -\frac{AT}{2} \mathcal{I}\{\hat{S}_k\} + n_{Q,k} \tag{11}$$

where

$$n_{I,k} = \int_{(k-1)T}^{kT} n(t) \cos(2\pi f_c t + \theta_c)\, dt \tag{12}$$

and

$$n_{Q,k} = \int_{(k-1)T}^{kT} n(t) \sin(2\pi f_c t + \theta_c)\, dt. \tag{13}$$

Using the assumption that the noise is Gaussian and white, it can easily be shown that $n_{I,k}$ and $n_{Q,k}$ are zero-mean white Gaussian sequences each with variance equal to $N_0 T/2$. Hence, r_n assumes a Gaussian distribution centered at a point in a scaled constellation. $n_{I,k}$ and $n_{Q,k}$ are generally not independent unless certain conditions are met, e.g., when $f_c T$ is an interger. Let

$$r_n' = r_n e^{j(\theta_c - 2\pi f_c T)}. \tag{14}$$

Then, it can be easily shown that the real and imaginary parts of r_n' are independent. Due to the phase rotation in (14), the mean of r_n' sits at a point of the scaled constellation that is rotated accordingly. Because the constellation under consideration is an arbitrary set of points, the rotation can be absorbed into the original constellation to form a new constellation; therefore, without loss of generality, we consider only the situation where $n_{I,k}$ and $n_{Q,k}$ are independent without the need for rotation.

Denote a group of c possible constellations by

$$\mathcal{I}_j = \{S_{j1}, S_{j2}, \cdots, S_{jM_j}\}, \qquad j = 1, 2, \cdots, c \tag{15}$$

where M_j is the number of points in constellation $\mathcal{I}_j$. Given a set of received data $X_N = \{\mathbf{x}_k = (r_{I,k}, r_{Q,k})^t, k = 1, 2, \cdots, N\}$, classification within the group of constellations can be considered as a test on the following c hypotheses:

$$H_j\text{: the underlying constellation is } \mathcal{I}_j, \qquad j = 1, 2, \cdots, c.$$

The ML classification method chooses the hypothesis whose likelihood or log-likelihood function is maximized, i.e.,

$$\begin{aligned}H_j^* &= \arg\max_{H_j} \ln(L(H_j \mid X_N)) \\ &= \arg\max_{H_j} \ln(p(X_N \mid H_j)).\end{aligned} \tag{16}$$

The joint probability density of $r_{I,k}$ and $r_{Q,k}$ at $(r_{I,k}, r_{Q,k}) = \mathbf{x}_k$, conditioned on H_j, is

$$p(\mathbf{x}_k \mid H_j) = \sum_{i=1}^{M_j} P(S_{ji} \mid \mathcal{I}_j) \frac{1}{\pi N_0 T} \cdot \exp\left(-\frac{1}{N_0 T}|\mathbf{x}_k - ATS_{ji}/2|^2\right) \quad (17)$$

where $P(S_{ji} \mid \mathcal{I}_j)$ is the *a priori* probability of S_{ji} in $\mathcal{I}_j$. In the following, we assume that all points in a constellation have the same *a priori* probability, i.e., $P(S_{ji} \mid \mathcal{I}_j) = 1/M_j$.

Assuming that the data from different symbols are independent, then, the log-likelihood function, after a constant factor is omitted, is

$$\begin{aligned} l(H_j \mid X_N) &= \ln(L(H_j \mid X_N)) \\ &= \ln \prod_{k=1}^{N} p(\mathbf{x}_k \mid H_j) \\ &= \sum_{k=1}^{N} \ln\left\{\frac{1}{M_j}\sum_{i=1}^{M_j} \exp\left(-\frac{1}{N_0 T}|\mathbf{x}_k - ATS_{ji}/2|^2\right)\right\}. \end{aligned} \quad (18)$$

Define $\text{SNR} = (A^2 T/N_0)$, then it can be shown [7] that $p(\mathbf{x}_k \mid H_j)$ depends only on *SNR* instead of on the absolute value of all three parameters. Consequently, we can assume, without loss of generality, $T = 2$ and $N_0 = 1$ and only use A to represent the SNR. The log-likelihood function in (18) becomes

$$l(H_j \mid X_N) = \sum_{k=1}^{N} \ln\left\{\frac{1}{M_j}\sum_{i=1}^{M_j} \exp\left(-\frac{1}{2}\|\mathbf{x}_k - AS_{ji}\|^2\right)\right\}. \quad (19)$$

III. Performance Analysis

Suppose I_j, $1 \le j \le c$, is the true constellation, then the misclassification probability is

$$P(\text{error} \mid H_j) = 1 - P[l(H_j \mid X_N) > l(H_n \mid X_N), \forall n \neq j] \quad (20)$$

and the overall misclassification probability is

$$P(\text{error}) = \frac{1}{c}\sum_{j=1}^{c} P(\text{error} \mid H_j). \quad (21)$$

Define random variables

$$a_{jn}(X_N) = l(H_j \mid X_N) - l(H_n \mid X_N), \quad n = 1, 2, \cdots, c, \quad n \neq j. \quad (22)$$

Then, (20) can be rewritten as

$$P(\text{error} \mid H_j) = 1 - P[a_{jn}(X_N) > 0, \forall n \neq j]. \quad (23)$$

Let

$$g_j(\mathbf{x}_k) = \ln\left\{\frac{1}{M_j}\sum_{i=1}^{M_j} \exp\left(-\frac{1}{2}\|\mathbf{x}_k - AS_{ji}\|^2\right)\right\} \quad (24)$$

then

$$a_{jn}(X_N) = \sum_{k=1}^{N} [g_j(\mathbf{x}_k) - g_n(\mathbf{x}_k)]. \quad (25)$$

We introduce the following notation:

$$m_{jn} = E(g_j(\mathbf{x}_k) - g_n(\mathbf{x}_k) \mid H_j) \quad (26)$$

$$\mathbf{a}_j(X_N) = (a_{j1}(X_N), a_{j2}(X_N), \cdots, a_{j,j-1}(X_N)) \quad (27)$$

$$\mathbf{m}_j = (m_{j1}, m_{j2}, \cdots, m_{j,j-1}, m_{j,j+1}, \cdots, m_{jc})^t \quad (28)$$

and

$$\mathbf{b}_j(X_N) = (\mathbf{a}_j(X_N) - N\mathbf{m}_j)/\sqrt{N}. \quad (29)$$

In (24)–(29), the indices are $k = 1, 2, \cdots, N$ and $n, j = 1, 2, \cdots, c, n \neq j$.

Then, $(c-1) \times 1$ vector $\mathbf{b}_j(X_N)$ is a sum of N independently, identically distributed (i.i.d.) random vectors [from (27) and (25)]. It can be shown [7] that these random vectors satisfy the conditions for the multivariate central limit theorem; therefore, $\mathbf{b}_j(X_N)$ approaches a Gaussian distribution as $N \to \infty$. Let $\mathbf{K}_j$ denote the covariance matrix of $\mathbf{b}_j$, i.e., $\mathbf{K}_j = E(\mathbf{b}_j \mathbf{b}_j^t \mid H_j)$; then, for large N, the probability density of $\mathbf{b}_j$ is

$$p(\mathbf{b}_j \mid H_j) \approx \frac{1}{\sqrt{(2\pi)^{c-1}}\sqrt{|\mathbf{K}_j|}} \exp\left\{-\frac{1}{2}\mathbf{b}_j^t \mathbf{K}_j^{-1} \mathbf{b}_j\right\}. \quad (30)$$

Hence, from (23) and (29), we have

$$\begin{aligned} P(\text{error} \mid H_j) &= 1 - P[b_{jn} > -\sqrt{N} m_{jn}, \forall n \neq j] \\ &= 1 - \int_{-\sqrt{N} m_{j1}}^{\infty} \cdots \int_{-\sqrt{N} m_{j,j-1}}^{\infty} \cdot \int_{-\sqrt{N} m_{j,j+1}}^{\infty} \cdots \int_{-\sqrt{N} m_{jc}}^{\infty} p(\mathbf{b}_j \mid H_j)\, d\mathbf{b}_j. \end{aligned} \quad (31)$$

This gives the error probability under each hypothesis. Note that the dimension of the above integral is $c - 1$. Now, the problem is simplified to evaluating the first- and second-order statistics of $g_j(\mathbf{x}_k)$ in (24), which can be calculated with numerical integration methods. Refer to [7] for details.

IV. Asymptotic Behavior of ML Classifier

In this section, we study the following question: for a fixed SNR, will the correct classification rate saturate at a certain level, or can it always approach 100% as N becomes very large? Because the ML classifier has the lowest error rate of all classifiers that are based on complex-domain data, if we can show that there exists a classifier whose error rate goes to zero as N goes to infinity, then the error rate of the ML classifier must also go to zero.

First, we introduce the Kolmogorov–Smirnov (K–S) distance [5]

$$d_l(X_N) = \sup_{\mathbf{x}} |F(\mathbf{x} \mid H_l) - \hat{F}_N(\mathbf{x})|, \qquad l = 1, 2, \cdots, c$$

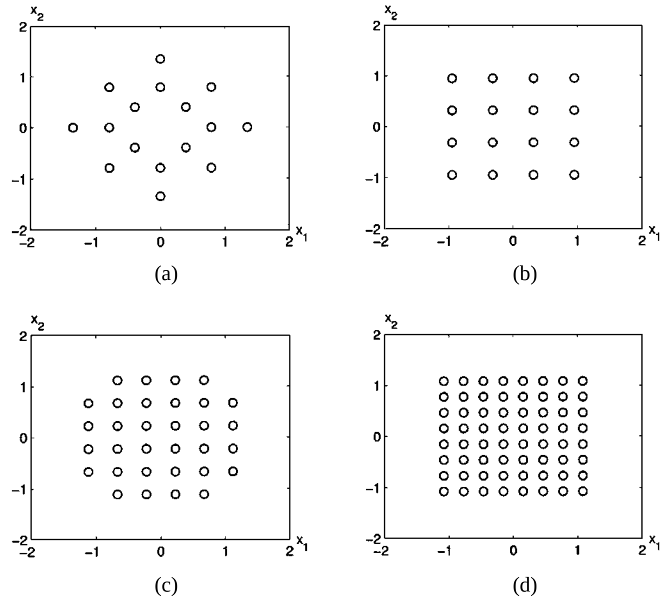

Fig. 1. Four constellations used in our simulations: (a) V.29, (b) 16-QAM, (c) 32-QAM, and (d) 64-QAM.

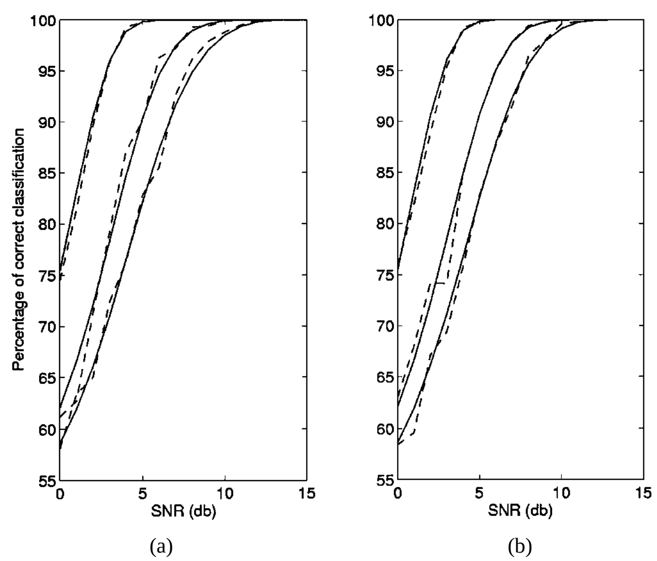

Fig. 2. Performance of ML classifier for 16-point V.29 and 16-QAM. The solid lines denote analysis, and the dashed lines denote simulation. (a) Signal is 16-point V.29. (b) Signal is 16-QAM. Lower curves: $N = 100$; middle curves: $N = 200$; upper curves: $N = 1000$.

(32)

where $F(\mathbf{x} \mid H_l)$ is the true probability distribution function under H_l, and $\hat{F}_N(\mathbf{x})$ is the sample distribution function of N symbols of incoming data X_N, which is defined as

$$\hat{F}_N(\mathbf{x}) = \frac{1}{N} \sum_{k=1}^{N} u(\mathbf{x}_k, \mathbf{x}) \tag{33}$$

where $u(\mathbf{x}_k, \mathbf{x}) = 1$ if $x_{k1} \le x_1$ and $x_{k2} \le x_2$, and $u(\mathbf{x}_k, \mathbf{x}) = 0$ otherwise $[\mathbf{x}_k = (x_{k1}, x_{k2})^t, \mathbf{x} = (x_1, x_2)]$.

Definition 1: A K–S classifier is one that uses the following decision rules:

IF $d_j(X_N) < d_l(X_N) \qquad \forall 1 \le l \le c, \quad l \ne j$

THEN Hypothesis H_j is selected.

Now we examine the asymptotic performance of the K–S classifier. Assume that H_j is true. It has been shown that [5]

$$\lim_{N\to\infty} P(d_j(X_N) = 0) = 1. \tag{34}$$

From (34) and (32), we see that

$$\lim_{N\to\infty} P(\hat{F}_N(\mathbf{x}) - F(\mathbf{x} \mid H_j) = 0) = 1 \qquad \forall \mathbf{x}. \tag{35}$$

From (32), we have the following inequality:

$$\begin{aligned} d_l(X_N) \ge\, & |F(\mathbf{x} \mid H_l) - F(\mathbf{x} \mid H_j)| \\ & - |\hat{F}_N(\mathbf{x}) - F(\mathbf{x} \mid H_j)|, \qquad l \ne j \quad \forall \mathbf{x}. \end{aligned} \tag{36}$$

From (34), (36), and (35), we see that

$$\lim_{N\to\infty} P[d_l(X_N) \ge |F(\mathbf{x} \mid H_l) - F(\mathbf{x} \mid H_j)|] = 1, \qquad l \ne j \quad \forall \mathbf{x}. \tag{37}$$

For every $l \ne j$ (we assume that $\mathcal{I}_l \ne \mathcal{I}_j$ for $l \ne j$), it can easily be shown that there exists at least one point $\mathbf{x}_l^*$ such that $\delta = |F(\mathbf{x}_l^* \mid H_l) - F(\mathbf{x}_l^* \mid H_j)| > 0$. Consequently, from (37), we have

$$\lim_{N\to\infty} P[d_l(X_N) \ge \delta] = 1. \tag{38}$$

By combining (34) and (38), we have $\lim_{N\to\infty} P[d_l(X_N) > d_j(X_N)] = 1$ for all $l \ne j$; therefore, the error probability (i.e., $1 - P[d_l(X_N) > d_j(X_N)]$) of the K–S classifier approaches zero as N goes to infinity, regardless of the SNR.

Using the fact that the ML classifier is optimal, we conclude the following. For any finite number of distinct constellations, the error rate of the ML classifier approaches zero as the number of samples goes to infinity, for any nonzero SNR's.

V. Examples and Discussions

In this section,we compare the results of our performance analyss with the results obtained from Monte Carlo simulations for randomly generated data. Four modulations are used: 16-QAM,16-point V.29, 32-QAM, and 64-QAM. Fig. 1 depicts the four constellations, normalized to the same power level. The comparisons were performed for the SNR ranging from 0 to 15 dB and the number of test symbols $N = 100, 200,$ and 1000. All six two-class combinations from the four modulations and all four three-class combinations were used. The simulations were performed for 1000 runs for each SNR and each value of N. Fig. 2 shows selected results for the two-class cases, i.e., for the following modulation pairs: V.29 and 16-QAM, V.29 and 32-QAM, and 32-QAM and 64-QAM. Fig. 3 shows selected results for the three-class cases, i.e., for 16-QAM, 32-QAM, and 64-QAM. All of our results [7] (not only the results shown in the figures) show that the analyses and the simulations match very well, except for some cases when SNR is small. Although we do not show all combinations of constellations due to space limitation, we note that all combinations were used in our simulations [7], and we observed matching experimental results and theoretical results for all cases.

VI. Conclusion

We have applied the ML method to the classification of digital amplitude-phase modulations and have obtained a generic for-

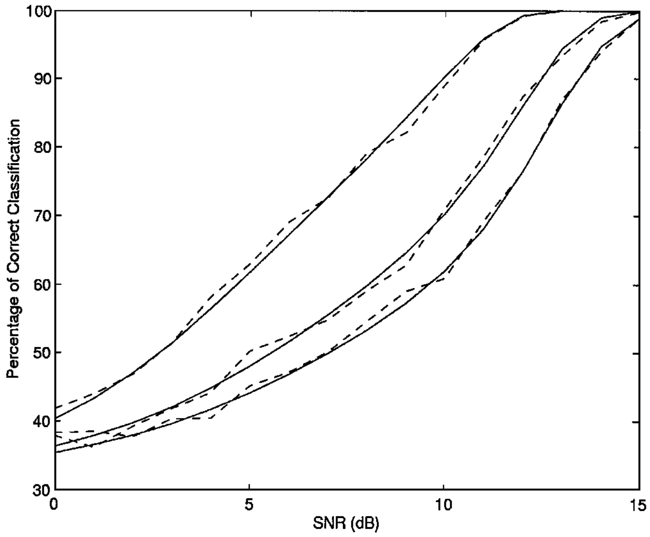

Fig. 3. Performance of ML classifier for three constellations. Signal is 16-QAM. The solid lines denote analysis, and the dashed lines denote simulation. Lower curves: $N = 100$, middle curves: $N = 200$, upper curves: $N = 1000$.

mula for the error probability of the ML classifier for any type of digital amplitude-phase modulation. The theorectical performance is derived under an ideal situation where all signal parameters as well as the noise power are known, and, in addition, the data symbols are independent and the pulse shape is rectangular. Because the ML classifier has optimal performance under ideal conditions, our analysis gives an upper bound of performance for any classifier that works under some nonideal condition; therefore, this study will be useful in evaluating other classifiers, as well as in determining the minimum resources (SNR and number of samples) needed for desired classification performance. Our study of asymptotic performance shows that the ML classifier is capable of classifying any finite set of distinctive constellations with zero error rate when the number of available symbols goes to infinity. We have also compared our theoretical performance results with simulation results that were obtained by means of actual classification, using randomly generated data, and they match well.

REFERENCES

[1] K. Assaleh, K. Farrell, and R. J. Mammone, “A new method of modulation classification for digitally modulated signals,” in *Proc. MILCOM*, 1992, pp. 712–716.

[2] S.-Z. Hsue and S. S. Soliman, “Automatic modulation classification using zero crossing,” *Proc. Inst. Elect. Eng.*, vol. 137, no. 6, pp. 459–464.

[3] Y.-C. Lin and C. C. Kuo, “Sequential modulation classification of dependent samples,” presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, Atlanta, GA, May 7–10, 1996.

[4] A. Polydoros and K. Kim, “On the detection and classification of quadrature digital modulations in broad-band noise,” *IEEE Trans. Commun.*, vol. 38, pp. 1199–1211, Aug. 1990.

[5] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980, pp. 28–37.

[6] H. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: Wiley, 1967.

[7] W. Wei, “Classification of Digital Modulations Using Constellation Analyzes,” Ph.D. dissertation, Univ. of Southern California, 1998.