Hierarchical Digital Modulation Classification Using Cumulants

Ananthram Swami, *Senior Member, IEEE*, and Brian M. Sadler, *Member, IEEE*

***Abstract*—A simple method, based on elementary fourth-order cumulants, is proposed for the classification of digital modulation schemes. These statistics are natural in this setting as they characterize the shape of the distribution of the noisy baseband I and Q samples. It is shown that cumulant-based classification is particularly effective when used in a hierarchical scheme, enabling separation into subclasses at low signal-to-noise ratio with small sample size. Thus, the method can be used as a preliminary classifier if desired. Computational complexity is order N, where N is the number of complex baseband data samples. This method is robust in the presence of carrier phase and frequency offsets and can be implemented recursively. Theoretical arguments are verified via extensive simulations and comparisons with existing approaches.**

***Index Terms*—Classification, higher order statistics, modulation.**

I. Introduction

AUTOMATIC recognition of digital modulation formats (i.e., identification of the underlying symbol constellation from observed noisy measurements) is increasingly important as the number and sophistication of digital signaling systems increase. There is an emerging need for intelligent modems capable of quickly discriminating signal types. Modulation classification may be used to identify interferences or to choose the appropriate demodulator in the cooperative scenario [27]. In this paper, we propose a simple, very low complexity, robust method based on fourth-order cumulants that can easily be applied in a hierarchical manner to classify various digital signaling formats. As we shall see, the use of higher order statistics (HOS) is natural in this setting because they characterize the shape of the distribution of the noisy baseband samples. The method is particularly effective for discriminating format subclasses, such as phase-shift keying (PSK) versus pulse amplitude modulation (PAM) versus quadrature amplitude modulation (QAM). Thus, it can be used as a simple preliminary classifier and may also be applied within each subclass to determine the exact modulation type if sufficient signal-to-noise ratio (SNR) and sample size are available. The robustness of this approach comes about not just from the resistance of HOS to additive colored Gaussian noise, but also from a natural robustness to constellation rotation and phase jitter; additive non-Gaussian noise can also be handled if its fourth-order cumulants are known, or via simple preprocessing. Further, the statistics used by the classifier can be recursively updated.

The problem of modulation classification within a set of known constellations has been fairly well studied in the recent literature. Under the typical independentl and identically distributed (i.i.d.) assumption, maximum-likelihood (ML) algorithms have been proposed. But the ML methods suffer from very high computational complexity (see, e.g., [28]); this has led to the development of the quasi-log-likelihood ratio (qLLR) classifiers [7], [12], [15], [19]; see also [13] for an interesting discussion of the ML and qLLR approaches. In addition, ML methods are not robust with respect to model mismatch such as phase and frequency offsets, residual channel effects, timing errors, and non-Gaussian noise distributions. Further, in a two-hypothesis problem, e.g., QAM(8, 8) versus QAM(16, 16), the ML solution tends to favor the denser constellation, i.e., QAM(16, 16). A number of feature-based classification schemes have been proposed. The features include the following attributes derived from: estimates of the instantaneous amplitude, frequency, and phase [1], [2]; outputs from simple zero-memory nonlinearities [4], [19], [20]; the empirical characteristic function and the Radon transform [29], [32]; moments of the extracted phase [19], [22]; and the characteristic function of the phase [21]. In the context of classifying M-ary frequency-shift keying (MFSK) signals, a method based on higher order correlations is studied in [3]; a method based on the moments of the "I-Q image" is proposed in [11].

When training data are available, neural network approaches [16], fuzzy classification and type-based classification [18] have been proposed. Some algorithms do not explicitly assume that the probability density function (pdf) is known, but they require enough training data so that the empirical noise pdf is a good estimate, e.g., [8], [18]. Various distance measures may be applied to determine the best match between the observed noisy data and templates based on the assumed noise pdf and symbol set; examples include entropy [18] and Hellinger distance [8]. A number of implementation issues arise with such methods, including choice of bin size and bin width when forming two-dimensional (2-D) histograms, and the typically large number of samples required. In addition, storage requirements are high, and estimating 2-D histograms (2-D pdf's) is not an easy task. In general, training-based approaches will not work well under conditions of carrier frequency offset (here, the received symbols trace arcs in the IQ plane) and phase jitter and are sensitive to the i.i.d. assumption. Similar to the model mismatch problem when using ML techniques, these methods are also sensitive to

Paper approved by W. E. Ryan, the Editor for Modulation, Coding, and Equalization of the IEEE Communications Society. Manuscript received August 10, 1998; revised April 27, 1999. This paper was presented in part at the Fedlab-ATIRP Annual Conference, College Park, MD, February 1998, and in part at the First Workshop on Signal Processing Advances in Wireless Communications, Paris, France, April 1997.

The authors are with the Army Research Laboratory, AMSRL–IS–TA, Adelphi, MD 20783 USA (e-mail: a.swami@ieee.org).

Publisher Item Identifier S 0090-6778(00)02276-5.

disparity between the training and test data, such as changes in the noise distribution or its variance, or the introduction or elimination of interferers.

II. Problem Statement

We assume that we are operating in a coherent, synchronous environment with single-tone signaling and that carrier, timing, and waveform recovery have been accomplished. We note that coarse estimates of the carrier frequency and signal bandwidth may be obtained from an estimate of the power spectral density. The baseband I and Q components can then be extracted, and a high resolution spectrum analysis of the fourth power of the signal can be used to refine the estimate of the carrier frequency. The baud rate can be estimated via a tracking loop, and symbol timing can be recovered with standard fractional sampling schemes. Constellation-independent algorithms (e.g., [10]) can be used to equalize the effects of the pulse-shaping filter and the (assumed) linear channel. We do not address these problems in this paper, although our model and simulations include residual baseband effects.

After preprocessing, we obtain a baseband sequence, composed of samples of the complex envelope, that can be written as

$$y(n) = A\, e^{j2\pi f_o T n + j\theta_n} \cdot \sum_{\ell=-\infty}^{\infty} x(\ell) h(nT - \ell T + \epsilon_T T) + g(n) \quad (1)$$

where $x(\ell)$ is the symbol sequence. Here, A is an unknown amplitude factor, $h(\cdot)$ represents residual channel effects (e.g., due to incomplete equalization), T is the symbol spacing, ϵ_T $(0 \le \epsilon_T < 1)$ represents timing errors, f_o is the residual carrier frequency or frequency offset, θ_n is the phase jitter, and $g(n)$ is the additive noise sequence. We regard f_0 as an unknown constant for each realization of $\{y(n)\}_{n=1}^{N}$, whereas θ_n is regarded as a random variable (r.v.) which varies from received symbol to symbol. The symbol sequence $\{x(\ell)\}$ may be drawn from one of K constellations with known symbols $\{\mu_{k,1}, \cdots, \mu_{k,s_k}\}$, where s_k is the number of symbols in the kth signal constellation. It is not necessary to assume that the symbols are equiprobable.

Given N samples, $\{y(n)\}_{n=1}^{N}$, we want to decide whether the signal is drawn from one of K equally likely known constellations. Note that the "equally likely" assumption is easily relaxed. We generally assume that the noise $g(n)$ is drawn from a (possibly colored) complex Gaussian process, although this assumption can be relaxed somewhat as we describe later. We assume that $y(n)$ has finite moments up to order eight, so that sample estimates of fourth-order cumulants have finite variance, a condition that is obviously met when $g(n)$ is Gaussian. Although we do not explicitly assume that the symbols are i.i.d., we do not exploit any intersymbol dependence either (such as that produced in trellis coding, for example).[1] For simplicity, we will assume symbol-spaced sampling. However, the algorithms are applicable to the fractionally sampled case so long as any transmitter pulse shaping is taken into account (e.g., via equalization).

[1]The variance analyses in Section IV and the Appendix assume i.i.d. symbols.

III. Classification Based on HOS

The optimal solution for this equiprobable multihypothesis problem is the ML solution. The high complexity of ML methods motivates the search for statistical features that will yield good results with low complexity. In addition, the residual baseband effects in (1) imply that some amount of robustness will be needed. In practice, we find that some classification problems are considerably easier than others. For example, it is much easier to discriminate PSK from the QAM subclasses than it is to discriminate QAM(8, 8) from QAM(16, 16). By easier, we mean that much less data are required (orders of magnitude less) at a lower SNR (tens of decibels lower). Thus, hierarchical methods may yield significant information even when reliable determination of the exact underlying class may be nearly impossible for a given data set. We propose to use normalized fourth-order cumulants of $y(n)$ as statistics for classification. These statistics characterize the shape of the distribution of the noisy signal constellation.

We will first define the fourth-order cumulants which we propose to use, discuss how they can be estimated from the data, and then give the theoretical values for various constellations of interest, assuming that the symbols are equiprobable. Expressions for the variances of the cumulant estimates are derived in the Appendix and discussed in the next section.

A. Definitions

For a complex-valued stationary random process $y(n)$, second-order moments can be defined in two different ways depending on placement of conjugation

$$C_{20} = E\left[y^2(n)\right] \qquad \text{and} \qquad C_{21} = E\left[|y(n)|^2\right]. \quad (2)$$

Similarly, fourth-order moments and cumulants can be written in three ways. Thus, fourth-order cumulants can be defined as

$$\begin{aligned} C_{40} &= \mathrm{cum}(y(n), y(n), y(n), y(n)) \\ C_{41} &= \mathrm{cum}(y(n), y(n), y(n), y^*(n)) \\ C_{42} &= \mathrm{cum}(y(n), y(n), y^*(n), y^*(n)). \end{aligned} \quad (3)$$

The statistics in (2) and (3) are the zeroth lags of the correlations and fourth-order cumulants of $y(n)$. For zero-mean r.v.'s w, x, y, and z, the fourth-order cumulant can be written as

$$\begin{aligned} \mathrm{cum}(w, x, y, z) = E(wxyz) - E(wx)E(yz) \\ - E(wy)E(xz) - E(wz)E(xy). \end{aligned} \quad (4)$$

We can use (4) to express C_{40}, C_{41}, or C_{42} in terms of the fourth- and second-order moments of $y(n)$, with the appropriate conjugations. See [6] for further details.

B. Sample Estimates

The cumulants in (2) and (3) can be estimated from the sample estimates of the corresponding moments. We assume

that $y(n)$ is zero-mean; in practice, the sample mean is removed before cumulant estimation. Sample estimates of the correlations are given by

$$\hat{C}_{21} = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^2$$
$$\hat{C}_{20} = \frac{1}{N}\sum_{n=1}^{N} y^2(n) \tag{5}$$

where the superscript $\hat{}$ denotes a sample average. This leads to the following estimates:

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N} y^4(n) - 3\hat{C}_{20}^2$$
$$\hat{C}_{41} = \frac{1}{N}\sum_{n=1}^{N} y^3(n)y^*(n) - 3\hat{C}_{20}\hat{C}_{21}$$
$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2. \tag{6}$$

We will assume, without loss of generality (wlog), that the constellations are normalized to have unit energy, implying that $C_{21} = 1$. In practice, we estimate the normalized cumulants

$$\tilde{C}_{4k} = \hat{C}_{4k} \Big/ \hat{C}_{21}^2, \qquad k = 0, 1, 2. \tag{7}$$

This self-normalizes the cumulant estimates and removes any scale problems in the data. The complexity of (6) and (7) is of order N, requiring only about $2N$ and $4N$ complex multiples for $\tilde{C}_{40}$ and $\tilde{C}_{42}$, respectively. In the case of noisy data, $\hat{C}_{21}$ in (7) must be replaced by $\hat{C}_{21} - \hat{C}_{21,g}$, where $\hat{C}_{21}$ is still given by (5) and $\hat{C}_{21,g}$ is an estimate of the variance of the additive noise $g(n)$; an estimate of $C_{21,g}$ is usually available in practice.

C. Theoretical Values

Here, we consider the theoretical values of the fourth-order cumulants in (3) for various signal constellations of interest; we assume that the symbols are equiprobable. The signal constellations considered are described in Table I. The theoretical values are obtained by computing the ensemble averages over the ideal noise-free constellation, under the constraint of unit energy. Table I shows the values of the statistics C_{40} and C_{42}. In Table I, PAM(∞) is the limiting value of PAM(n) as $n \to \infty$, and PSK(>4) holds for any PSK(n) constellation with $n > 4$. The last three columns of the table list the variances of the sample estimates of C_{40} and C_{42} and are explained in the next section. We note that for all the complex-valued constellations listed in Table I (excluding 8AMPM), $C_{20} = 0$ and $C_{41} = 0$. Note also that the statistics C_{42} and $|C_{40}|$ are unaffected by a phase rotation of the symbol set. In the case of C_{42}, this follows immediately from the symmetric conjugation in (3). In the case of C_{40}, if the phase rotation is ϕ, then C_{40} will have the complex multiplier $\exp(j4\phi)$; thus, $|C_{40}|$ will be unaffected.

The constellations in Table I naturally divide into the following four subclasses: binary PSK (BPSK) (binary real-valued), PAM (real-valued), PSK (constant-modulus), and QAM-V29-V32 (general complex-valued). We therefore propose a hierarchical classification structure which is shown in Fig. 1. We use C_{42} first to decide whether the constellation is real-valued (BPSK/PAM), circular (PSK), or rectangular (QAM). Then, if the unknown phase rotation can be assumed to be small, C_{40} may be used to help differentiate within each subclass. If the unknown phase rotation cannot be ignored, then $|C_{40}|$ must be used rather than C_{40}. As we see from Table I, when using $|C_{40}|$, there is a potential performance loss in some cases, e.g., V29 versus QAM(8, 8). If the initial decision is PSK, we use $|C_{40}|$ to decide whether it is PSK(4) [QAM(2, 2)] or PSK(>4). Similarly, if the initial decision is QAM, we use C_{40} or $|C_{40}|$ to decide whether it is V32, V29, QAM(8, 8), QAM(4, 4), or V29c. If the unknown phase rotation can be assumed to be small, we can distinguish between PSK(2), PSK(4), and their rotated versions. Note that the value of C_{40} for QAM(8, 8) and QAM(4, 4) are close to one another; $|C_{60}|$ may be a better statistic. The hierarchical approach attempts to first classify the data using "macro" characteristics; it then refines the membership using "micro" characteristics, in the spirit of [7] which considered the three-class BPSK/quaternary PSK (QPSK)/offset QPSK problem.

The 8AMPM constellation is distinct in that it does not have the fourfold symmetry of the other complex constellations; indeed, it has $C_{20} = j0.2$ and $C_{41} = -j0.08$; thus, it can be

TABLE I
THEORETICAL CUMULANT STATISTICS C_{40} AND C_{42} FOR VARIOUS CONSTELLATION TYPES, AND VARIANCES OF THEIR SAMPLE ESTIMATES

Constellation	C_{40}	C_{42}	N var ($\hat{C}_{40}$)	N var ($\hat{C}_{42}$)	N var$_1$ ($\hat{C}_{42}$)
BPSK	-2.0000	-2.0000	0.00	0.00	36.00
PAM(4)	-1.3600	-1.3600	2.56	2.56	34.72
PAM(8)	-1.2381	-1.2381	4.82	4.82	32.27
PAM(16)	-1.2094	-1.2094	5.52	5.52	31.67
PAM(32)	-1.2024	-1.2024	5.70	5.70	31.52
PAM(64)	-1.2006	-1.2006	5.74	5.74	31.49
PAM(∞)	-1.2000	-1.2000	5.76	5.76	31.47
PSK(4)	1.0000	-1.0000	0.00	0.00	12.00
PSK(>4)	0.0000	-1.0000	1.00	0.00	12.00
V32	0.1900	-0.6900	2.86	1.18	9.70
V29	0.5185	-0.5816	3.51	1.77	8.75
QAM(∞)	-0.6000	-0.6000	3.91	2.31	8.59
QAM(32,32)	-0.6012	-0.6012	3.89	2.29	8.61
QAM(16,16)	-0.6047	-0.6047	3.83	2.24	8.65
QAM(8,8)	-0.6191	-0.6191	3.58	2.06	8.82
QAM(4,4)	-0.6800	-0.6800	2.66	1.38	9.54
V29c	-1.2000	-0.6400	1.85	1.44	9.12
8AMPM	-0.5600	-0.7200	2.66	1.38	9.54

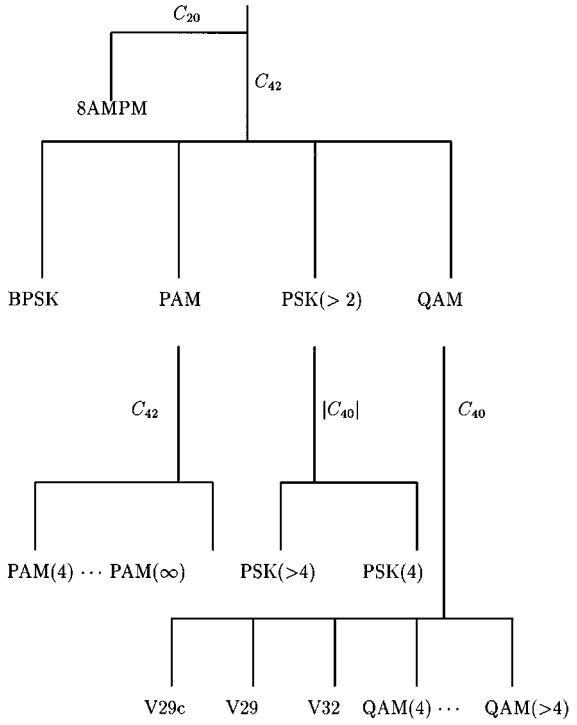

Fig. 1. Hierarchical classification scheme based on HOS.

easily distinguished from other constellations on the basis of C_{20} alone.

Within the QAM-type constellations, the V32, V29, and V29c symbol sets are also distinct since they do not satisfy $|C_{40}| = |C_{42}|$. For QAM signals, we have $C_{40} = C_{42}\exp(j4\phi)$, so that the phase rotation ϕ can be estimated easily.

The specific algorithm to be used depends upon the particular problem at hand, i.e., whether only some or all of the modulation types listed in Table I (or others) are active. Additional details are given in the next section.

IV. Asymptotic Threshold Analysis

In this section, we develop thresholds for the tests in the hierarchical classification scheme. In order to do this, we need to derive expressions for the variance of the sample estimates of the cumulants in (6). The variance expressions are given in (8)–(11); their derivations are given in the appendix, and assume that there is no additive noise. Equations (12)–(14) give the corresponding expressions when there is additive noise. In these expressions, $M_{km} := E[y^{k-m}(y^*)^m]$, the mixed moment of order k with m conjugations.

For the complex constellations, C_{20} is known (it is 0 for PSK and QAM); hence, with N samples

$$N\,\mathrm{var}[\hat{C}_{40}] = M_{84} - |M_{40}|^2. \tag{8}$$

If we assume that C_{21} is known, then

$$N\,\mathrm{var}[\hat{C}_{42}] = M_{84} - |M_{42}|^2. \tag{9}$$

If C_{21} is estimated from the data, then

$$N\,\mathrm{var}[\hat{C}_{42}] \approx \left[M_{84} - M_{42}^2\right] + 4M_{21}\left[3M_{42}M_{21} - 2M_{63} + 2M_{21}^3\right]. \tag{10}$$

For real constellations, $C_{20} = C_{21}$, $C_{40} = C_{42}$; if C_{21} is known, the variance expression is given by (8), otherwise, it is given by

$$N\,\mathrm{var}[\hat{C}_{42}] \approx \left[M_{84} - M_{42}^2\right] + 6M_{21}\left[5M_{42}M_{21} - 2M_{63} + 3M_{21}^3\right]. \tag{11}$$

Equations (8) and (9) were evaluated for various constellations, assuming unit energy $C_{21} = 1$, and are tabulated in columns 4 and 5 of Table I. Equations (10) and (11) were evaluated and are shown in the last column of Table I. Note that contrary to popular assumptions, if C_{21} is estimated from the data, then the variance of $\hat{C}_{42}$ can increase by a factor of 4–8 for the QAM constellations, even in the noise-free case.

In the appendix, we show that (8)–(10) can be expressed in terms of cumulants as

$$N\,\mathrm{var}[\hat{C}_{40}] \approx C_{84,y} + 16C_{63,y}C_{21,y} + 18C_{42,y}^2 + 72C_{42,y}C_{21,y}^2 + 24C_{21,y}^4 \tag{12}$$

$$N\,\mathrm{var}[\hat{C}_{42}] \approx C_{84,y} + 16C_{63,y}C_{21,y} + 17C_{42,y}^2 + 68C_{42,y}C_{21,y}^2 + 20C_{21,y}^4 + |C_{40,y}|^2 \tag{13}$$

$$N\,\mathrm{var}[\hat{C}_{42}] \approx C_{84,y} + 8C_{63,y}C_{21,y} + 17C_{42,y}^2 + 128C_{42,y}C_{21,y}^2 + 4C_{21,y}^4 + |C_{40,y}|^2. \tag{14}$$

In the above, the additional subscript y is introduced to reinforce the fact that these are cumulants of the noisy process $y(n) = x(n) + g(n)$. Since the symbol sequence $x(n)$ is assumed to be independent of the noise sequence $g(n)$, the cumulants of $y(n)$ can be expressed as the sum of the cumulants of $x(n)$ and $g(n)$, $C_{21,y} = C_{21,x} + C_{21,g}$, $C_{42,y} = C_{42,x} + C_{42,g}$, etc. If the additive noise is Gaussian, $C_{40,g} = C_{42,g} = C_{63,g} = C_{84,g} = 0$, $C_{21,g} = \sigma_g^2$. The moments and cumulants of the symbol sequences are easily evaluated via (21)–(23). This provides us with a vehicle for evaluating the variances of the sample estimates of C_{40} and C_{42}.

From Table I, we note that the value of C_{40} or C_{42} and their sample variances are different for different signal constellations. Consider a statistic S, which is Gaussian with mean μ_i and variance σ_i^2 under hypothesis H_i, $i = 0, 1$. Assume wlog that $\sigma_0^2 < \sigma_1^2$. Then, the likelihood ratio test (LRT) for achieving minimum probability of error, assuming equal priors, is an interval detector, which can be written as [24]

$$\text{decide } H_0 \text{ if } S \in [\mu - a,\, \mu + a] \tag{15}$$

where

$$\mu := \left(\frac{\mu_0}{\sigma_0^2} - \frac{\mu_1}{\sigma_1^2}\right)\frac{\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}$$

and

$$a^2 := \frac{\sigma_0^2\sigma_1^2}{\sigma_1^2 - \sigma_0^2}\left[\ln\frac{\sigma_1^2}{\sigma_0^2} + \frac{(\mu_1 - \mu_0)^2}{\sigma_1^2 - \sigma_0^2}\right].$$

If $\sigma_0^2 = \sigma_1^2$, we have a threshold detector; thus, if $\mu_0 < \mu_1$, we decide H_0 if $S < (\mu_0 + \mu_1)/2$. From Table I, we see that the variances of $\hat{C}_{40}$ are approximately the same for members of the QAM set and the PAM set, thus justifying the use of the

threshold detector. For any two-class problem, the probability of total error is given by $Q(d/2)$, where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2}\, du$$

is the tail probability function of the standard Gaussian, and $d = (\mu_1 - \mu_0)/\sigma$.

Under the equal variance hypothesis, we know that to obtain correct classification of 90%, 95%, and 99%, one would like to have the means under the two hypotheses separated by 3.3σ, 3.92σ, and 5.16σ, respectively. With σ obtained from Table I, this enables us to estimate the number of samples required to achieve a specified probability of correct classification, which we will denote by P_c.

Table II lists the number of samples required to obtain correct classification for various two-class problems in the noise-free case; for each class pair, the first row gives an estimate of the required N if C_{40} is used, and the second row if the absolute value $|C_{40}|$ is used in order to cope with an unknown phase rotation in the data. Note that the performance degradation may be dramatic in some cases but unchanged in others. The number of samples is overestimated, since we assume we are using a threshold detector rather than the interval detector. On the other hand, the number of samples is somewhat underestimated since we have neglected the effects of the additive noise. In fact, as the simulations demonstrate, the numbers in Table II are gross overestimates, and excellent classification results are obtained at low SNR's, with smaller number of samples, and with mismatch due to frequency offset, phase jitter, residual channel effects, cochannel interference, timing errors, and non-Gaussian noise. Indeed, for the four-class BPSK versus PAM(4) versus QAM(4, 4) versus PSK(8) problem, perfect classification results were obtained using 500 samples at an SNR of 10 dB.

Additive Gaussian noise does not affect $E\{\hat{C}_{40}\}$ or $E\{\hat{C}_{42}\}$, only their variances. Since we are using normalized statistics, it is important to have a good estimate of the noise variance; in practice, such an estimate is available. If the additive noise is non-Gaussian but circularly symmetric (a reasonable assumption), i.e., $C_{40,g} = 0$, then $\hat{C}_{40}$ will continue to be unbiased. Note that this is not true of $\hat{C}_{42}$, since the kurtosis of the non-Gaussian noise cannot be assumed to be zero. In this case, preprocessing the data by passing it through a zero-memory nonlinearity is important [23]; this idea is discussed further in the simulation section (see Example 6). Alternatively, if access to noise-only data are available, its kurtosis can be estimated.

Predistorting ("shaping") the constellations may help our classifier; for example, the difference between the $|C_{40}|$ values of QAM(4, 4) and V29 increases as the distortion parameter g in [5] increases; the difference in the $|C_{42}|$ values also increases, but not by much.

V. Simulation Results

In this section, a variety of simulation experiments are presented illustrating the performance of the proposed classification schemes. Some other simulation results are presented in [25] and [26]. For each Monte Carlo trial, the appropriate normalized statistics $|\tilde{C}_{40}|$ and $\tilde{C}_{42}$ were estimated via (5)–(7), based on N data samples, and the noise variance σ_g^2 was estimated from a noise-only data record of size N. Except where noted, the additive noise was complex white Gaussian. SNR in decibels is defined as

$$\text{SNR} = 10 \log_{10} \left(E_s/\sigma_g^2\right)$$

where E_s is the signal energy. Experiments will use two different sets of modulation types. We define a four-class problem based on the modulations given by

$$\Omega_4 = \{\text{BPSK, PAM}(4), \text{QAM}(4,4), \text{PSK}(8)\}. \quad (16)$$

We also define an eight-class problem based on

$$\Omega_8 = \{\text{BPSK, PAM, PSK}(4), \text{PSK}(8), \text{V32, V29, V29c, QAM}(4,4)\}. \quad (17)$$

All results are based on 1000 Monte Carlo trials, i.e., 4000 trials for the four-class problem and 8000 for the eight-class problem.

In Examples 1–11, we consider the four-class problem based on Ω_4 in (16). The eight-class problem of (17) is considered in Example 12. In Example 13, we compare our classifier with the qLLR classifier [19] for the binary problem: BPSK versus

TABLE II
NUMBER OF SAMPLES VERSUS P_c USING $\hat{C}_{40}$ IN THE FIRST ROW AND $|\hat{C}_{40}|$ IN THE SECOND ROW, FOR VARIOUS TWO-CLASS PROBLEMS

P_c	0.90	0.95	0.99
V29 vs V32	354	500	866
	354	500	866
V29 vs V29c	13	18	32
	82	116	201
V29 vs QAM(64)	31	44	76
	3,852	5,436	9,419
V29 vs 8AMPM	33	46	80
	22,194	31,317	54,264
QAM(16) vs QAM(64)	10,512	14,833	25,701
	10,512	14,833	25,701
BPSK vs PAM(4)	68	96	166
	68	96	166
PAM(4) vs QAM(16)	63	88	153
	63	88	153
PAM(4) vs PSK(8)	15	21	37
	15	21	37
QAM(16) vs PSK(8)	2,355	3,223	5,758
	2,355	3,223	5,758

QPSK. Finally, in Examples 14 and 15, we compare our classifier with the q_4 classifier of [12] and [15].

For the four-class problem, Ω_4, study of Table I reveals that it is advantageous to use $|C_{40}|$ for the test statistic (rather than C_{42}) because $|C_{40}| = 0$ for PSK(8); hence $|C_{40}|$ was used as the test statistic in Examples 1–11. For a given SNR, one can compute the optimal threshold under the assumption that $\hat{C}_{40}$ is Gaussian, using the expressions in Section IV. Let μ_k and σ_k^2 denote the mean and variance of the statistic, S, under the kth hypothesis; wlog, assume that the M hypotheses are ordered so that $\mu_1 < \mu_2 < \cdots < \mu_M$. A simplifying approximation is to consider that the variances are all equal: in this case, the decision rule is to choose H_k if $(\mu_{k-1} + \mu_k)/2 < S < (\mu_k + \mu_{k+1})/2$ with $\mu_0 = -\infty$ and $\mu_{M+1} = \infty$. We assumed that the variances were equal; thus, regardless of the actual noise variance, the decision rule used in Examples 1–11 was

$$\begin{aligned} |\tilde{C}_{40}| &< 0.34 \Rightarrow \text{PSK}(8) \\ 0.34 \le |\tilde{C}_{40}| &< 1.02 \Rightarrow \text{QAM}(4,4) \\ 1.02 \le |\tilde{C}_{40}| &< 1.68 \Rightarrow \text{PAM}(4) \\ 1.68 \le |\tilde{C}_{40}| &\Rightarrow \text{BPSK} \end{aligned} \tag{18}$$

with $\tilde{C}_{40}$ estimated as described in Section III-B. The theoretical probability of error is given by

$$P_c = \frac{1}{M} + \frac{1}{M} \sum_{k=1}^{M-1} Q\left(\frac{\mu_k - \mu_{k+1}}{2\sigma_{k+1}}\right) - Q\left(\frac{\mu_{k+1} - \mu_k}{2\sigma_k}\right).$$

Example 1: First, we study the performance of our C_{40}-based classifier of (18) under ideal conditions, i.e., in (1), $h(n) = \delta(n)$, $\theta_n = f_o = \epsilon_T = 0$; $g(n)$ was zero-mean complex Gaussian. The first two rows of Table III show the overall P_c for sample sizes of 100, 250, and 500 and SNR's of 5 and 10 dB. Note that increasing the sample size to $N = 500$ with SNR $= 10$ dB yields error-free results in this problem, demonstrating the ability of the proposed HOS-based test to attain very good discrimination among real, circular, and QAM-type modulations. This example also illustrates that the required number of samples listed in Table II is an overestimate. Because of lack of space, we present only representative confusion matrices in Table IV, corresponding to row 2 of Table III (see [26] for more examples).

Next, we introduce random phase jitter θ_n in (1). The phase was uniformly distributed over $\pm 7.5°$ and added to each baseband symbol independently from symbol to symbol. Results are shown in rows 3 and 4 of Table III.

We repeat the simulations, now with a (normalized) frequency offset f_0 in (1), modeled as uniformly distributed in the range $\pm 15/N$, where N is the number of samples. Here, f_0 is a fixed constant for each realization of $\{y(n)\}$, but varies randomly from realization to realization. Results are shown in rows 5 and 6 of Table III. The robustness of the classifier to phase errors and frequency offset is clearly apparent in the results which are essentially identical to those in rows 1 and 2.

Example 2: The performance of the detector of (18) is now studied as a function of SNR; Fig. 2 shows P_c versus SNR, with curves parameterized by sample size N. Note the essentially error-free performance with an SNR of about 8 dB for sample sizes of 250 or larger. At low SNR, the percentage of correct classification decreases to a limit of 25%, corresponding to guessing for a four-class problem. Also, with the small sample size of $N = 100$, the test becomes data-limited and does not appear to achieve completely error-free performance even at higher SNR's; the limiting value appears to be around 0.995.

TABLE III
(EXAMPLE 1) P_c FOR THE FOUR-CLASS PROBLEM

Case	SNR	$N = 100$	$N = 250$	$N = 500$
1. Regular	5 dB	0.7198	0.9180	0.9845
	10 dB	0.9685	0.9995	1.0000
2. Phase errors	5 dB	0.7288	0.9193	0.9845
	10 dB	0.9680	0.9992	1.0000
3. Frequency offset	5 dB	0.7330	0.9145	0.9870
	10 dB	0.9690	0.9995	1.0000

TABLE IV
CONFUSION MATRICES FOR EXAMPLE 1 WITH SNR = 10 dB, 1000 TRIALS, AND (TOP TO BOTTOM) $N = 100$, 250, AND 500 SAMPLES

	Classifier Output			
Class. Input	BPSK	PAM(4)	QAM(4,4)	PSK(8)
BPSK	992	8	0	0
PAM(4)	0	979	21	0
QAM(4,4)	0	9	952	39
PSK(8)	0	0	49	951
BPSK	1000	0	0	0
PAM(4)	0	1000	0	0
QAM(4,4)	0	1	998	1
PSK(8)	0	0	0	1000
BPSK	1000	0	0	0
PAM(4)	0	1000	0	0
QAM(4,4)	0	0	1000	0
PSK(8)	0	0	0	1000

Example 3—Non-Gaussian Noise: Next, we examine the robustness of the classification approach to the presence of outliers (impulsive noise) by utilizing a two-term Gaussian mixture for the pdf of $g(n)$ in (1), given by

$$f(g) = (1 - \epsilon) f_N(g) + \epsilon f_I(g)$$

where $0 < \epsilon < 1$ is the mixing parameter, and $f_N(g)$ and $f_I(g)$ are zero-mean Gaussian distributions with variances σ_N^2 and σ_I^2, respectively. We set $\epsilon = 0.01$ and $\sigma_I^2/\sigma_N^2 = 100$. The

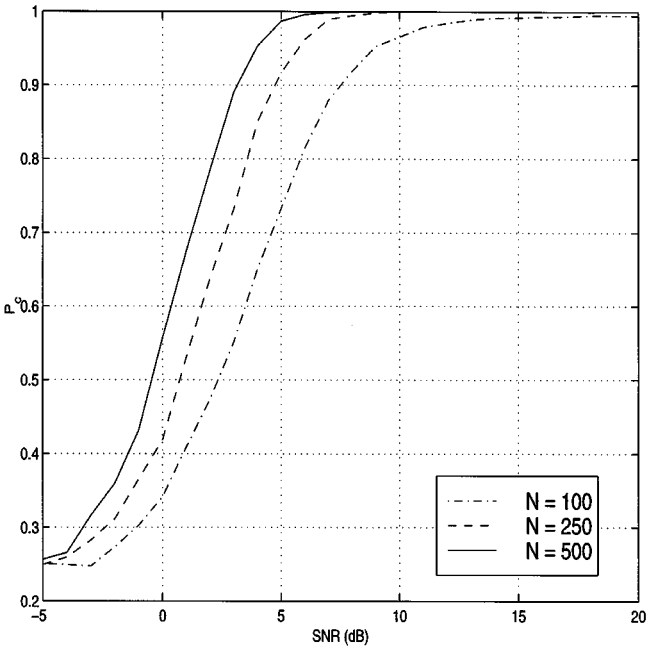

Fig. 2. (Example 2) P_c versus SNR for Ω_4 in AWGN, with curves parameterized by the number of symbols N.

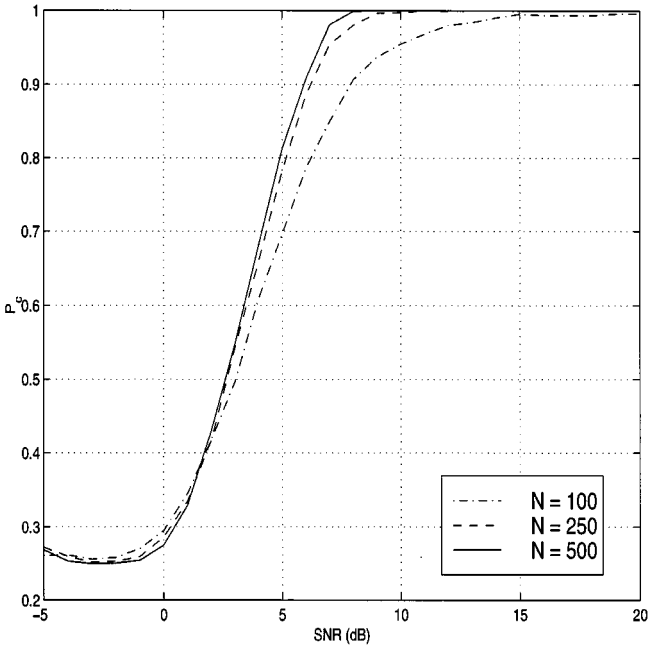

Fig. 3. (Example 3) P_c versus SNR for Ω_4 in impulsive non-Gaussian noise with random phase jitter. Curves are parameterized by the number of symbols N.

SNR for this example is set based on the total noise variance $\sigma_g^2 = (1-\epsilon)\sigma_N^2 + \epsilon\sigma_I^2$.

To ameliorate the effects of this heavy tailed non-Gaussian noise, a data-adaptive zero-memory nonlinearity (ZMNL) was employed. ZMNL-based approaches work well in a variety of unknown heavy-tailed noise environments, e.g., see [23] and references therein. The ZMNL was applied separately to the real and imaginary parts of the noisy complex samples and is given by

$$\phi(u) = \text{sign}(u)\ \min(|u|,\ \delta)$$

which is a nonreturn to zero clipper with a linear region determined by choice of δ. We used $\delta = 2\hat{\sigma}$, where $\hat{\sigma}$ is a robust estimate of the nominal standard deviation given by

$$\hat{\sigma} = \overline{\sigma}/0.7$$

and $\overline{\sigma}$ is the median absolute deviation (MAD) of the input sequence [14]. Clipping at $2\hat{\sigma}$ is motivated by the desire to avoid outliers that are too large in amplitude and thereby strongly affect the HOS estimates. The classifier used an estimate of the original noise variance σ^2 (before clipping), because we assume the noise power after clipping is not available.

In addition to the additive non-Gaussian noise, random phase jitter was introduced as in Example 1 above. Fig. 3 shows P_c versus SNR for three different sample sizes. These results may be compared with those of Fig. 2 from Example 2, in which the noise was Gaussian and there was no phase jitter. For $N = 100$, there is essentially no loss in performance due to the non-Gaussian noise and phase error. However, for this case, there is only one impulsive event per 100 samples on average ($\epsilon = 0.01$). The effect of the non-Gaussian noise is more noticeable as N increases, leading to a roughly 3-dB loss in performance for $N = 500$ versus that achieved in Gaussian noise. For the larger sample cases, there will be more clipping on average, and hence the estimate of the noise variance will be less accurate, leading to less accurate classification results.

Example 4—Recursive Classification: The relevant moments used by our classifier can be recursively updated as

$$\begin{aligned}
M_{20}(n) &= \lambda(n)M_{20}(n-1) + (1-\lambda(n))y^2(n)\\
M_{21}(n) &= \lambda(n)M_{21}(n-1) + (1-\lambda(n))|y(n)|^2 - a\sigma_g^2\\
M_{40}(n) &= \lambda(n)M_{40}(n-1) + (1-\lambda(n))y^4(n)\\
\tilde{C}_{40}(n) &= \big(M_{40}(n) - 3M_{20}^2(n)\big)\big/ M_{21}^2(n).
\end{aligned}$$

Here, a is a scale factor that might represent errors in estimating the noise variance σ_g^2. If we set $\lambda(n) = (n-1)/n$, we get unbiased estimates; in a nonstationary setting, one may want to use an exponential forgetting factor instead. In the context of modulation classification, recursive algorithms have been considered in [4]. Classification results are shown in Fig. 4 for three values of a: $a = 0,\ 1,\ 2$ as the number of samples was varied from 5 to 250; SNR was fixed at 15 dB. Notice that almost perfect classification is obtained with $N = 100$ samples, and that at this SNR, the estimate of the noise variance is not critical; setting $a = 0$ is equivalent to assuming that the noisy data are noise free, and setting $a = 2$ overestimates the noise variance by a factor of two.

Example 5—Phase Offsets: The statistic $|C_{40}|$ is theoretically insensitive to a fixed phase offset $\theta_n = \theta_c$, where θ_c is fixed over a realization but varies randomly, $U[-\pi,\ \pi)$ from realization to realization. With $N = 200$ and an SNR of 12 dB, we obtained P_c's of 0.9844, 0.9994, and 0.9895 corresponding to assumed noise variance of $a\sigma_g^2$, with a values of 0, 1, and 2. Note that there is no degradation in assuming that the noise variance is smaller ($a = 0$) or larger ($a = 2$). With $N = 200$, the dominant source of error is the finite-sample error in estimating $\hat{C}_{40}$.

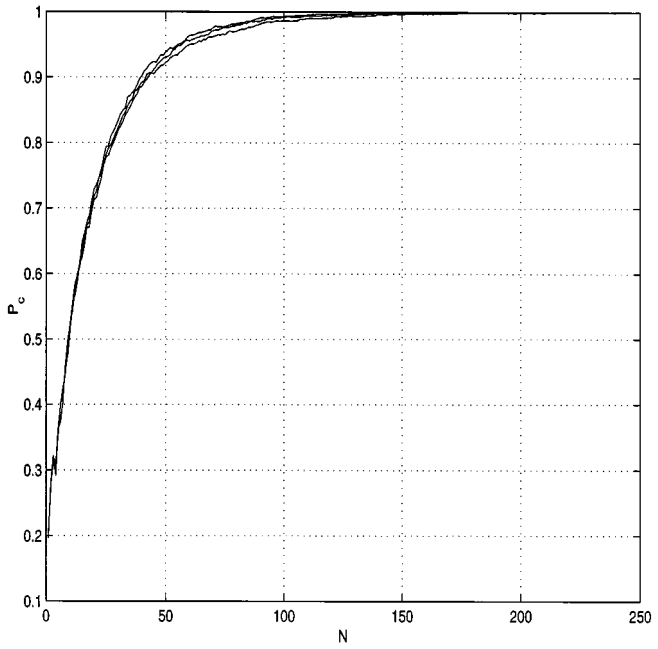

Fig. 4. (Example 4) P_c versus number of samples N for Ω_4. Statistics were recursively updated. The noise variance was assumed to be $a\sigma_g^2$, with $a = 0, 1, 2$.

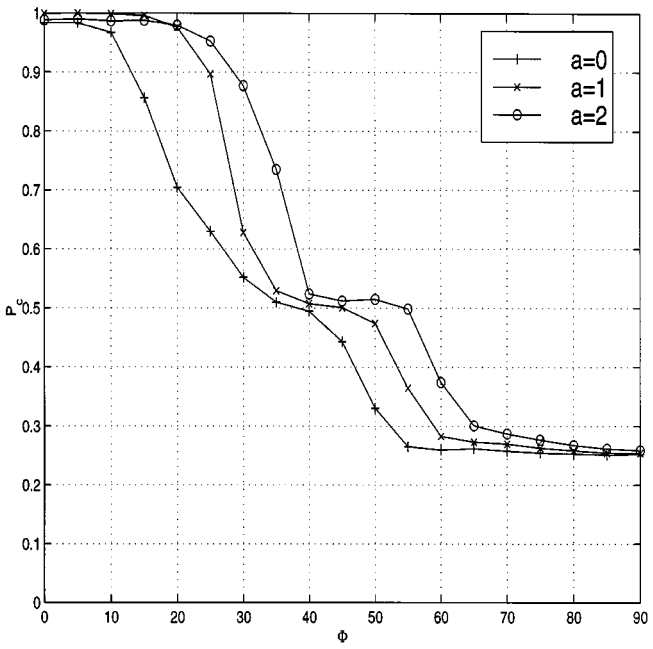

Fig. 5. (Example 6) P_c versus Φ for Ω_4; phase jitter θ_n is uniformly distributed over $[-\Phi, \Phi]$; $N = 200$, SNR = 12 dB. The noise variance was assumed to be $a\sigma_g^2$.

Example 6—Phase Jitter: Here, we see how performance is degraded by phase jitter; the phase ϕ_ℓ varies from symbol to symbol and is assumed to be uniformly distributed over $[-\Phi, \Phi]$. P_c is plotted against Φ in Fig. 5, for $N = 200$ samples and SNR=12 dB. Acceptable performance is obtained for $\Phi < 30°$, after which P_c drops rather sharply. Beyond $60°$, BPSK and PAM are classified as QAM and finally as PSK. QAM is classified as PSK for $\Phi > 30°$; as may be expected, there are no errors in classifying PSK(8) since the phase jitter does not alter the constant modulus property of the symbol set. Phase jitter increases the variance of the effective noise; thus, it is beneficial to assume a larger noise variance ($a = 2$).

Example 7—Frequency Offsets: Here, we see how performance is degraded by frequency offset, which causes the constellations to rotate; thus, symbol points are smeared along arcs. The number of samples was $N = 250$, SNR = 12 dB, and the normalized frequency f_o was varied from 0 to 0.001 (corresponding to a maximum rotation of $90°$). P_c is plotted against f_o in Fig. 6; acceptable performance is obtained for $f_o < 0.6 \times 10^{-3}$. One can think of frequency offset as being equivalent to a fixed phase offset plus phase jitter; thus, as in the previous example, it is beneficial to assume a larger noise variance ($a = 2$). Better performance might be obtained by estimating C_{40} of $z(n) = y^*(n-1)y(n)$, which serves to convert the frequency offset to a fixed phase offset;[2] we will now be computing a higher order cumulant or, equivalently, the cumulant of a correlated sequence; hence, more data would be required to achieve the same performance.

Example 8—Residual Channel Effects: If the symbol sequence $x(n)$ is passed through a finite-impulse response

[2]This idea is used to estimate the parameters of a chirp signal, e.g., see [31] and references therein. In the context of modulation classification, it was used in [15].

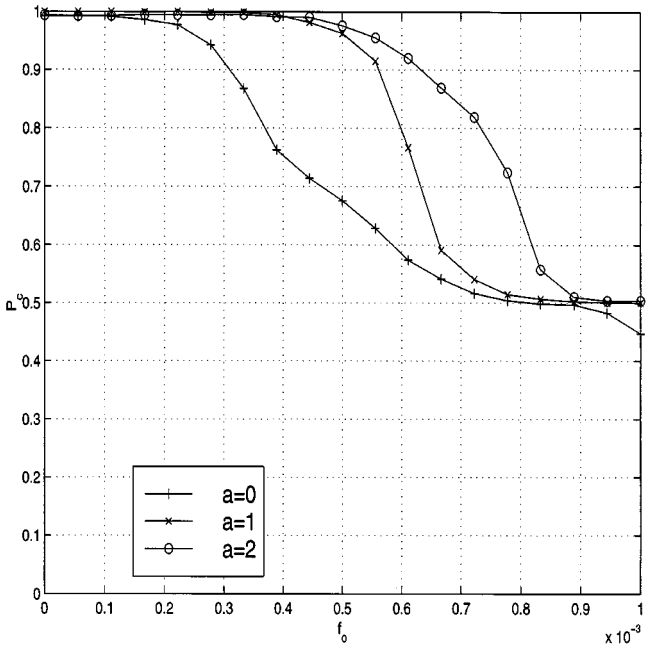

Fig. 6. (Example 7) P_c versus frequency offset f_c for Ω_4; $N = 250$, SNR = 12 dB. The noise variance was assumed to be $a\sigma_g^2$.

channel with coefficients $[h(0), h(1), \cdots, h(L-1)]$, the normalized statistics C_{40} and C_{42} will be multiplied by the factor

$$\beta = \sum_{n=0}^{L-1} |h(k)|^4 \Big/ \left(\sum_{n=0}^{L-1} |h(k)|^2 \right)^2 < 1 .$$

Since $\beta < 1$, the effect of the residual channel is to drive C_{40} and C_{42} toward 0 (filtering tends to make data more Gaussian),

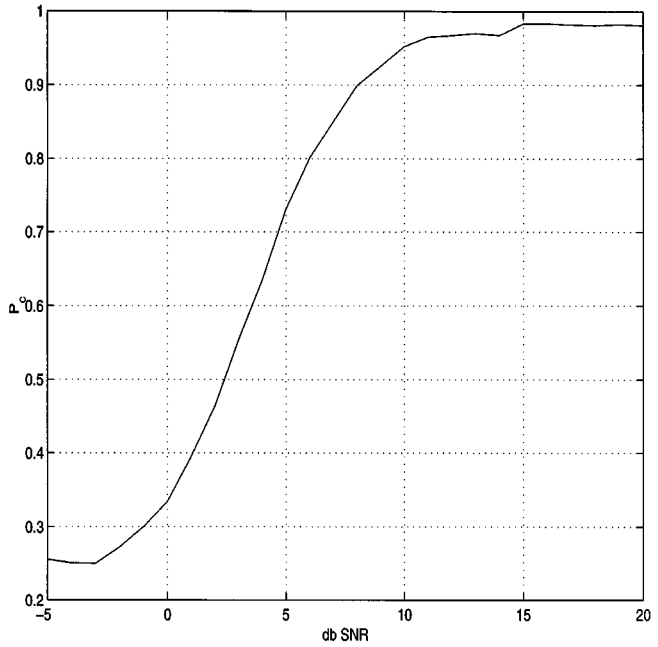

Fig. 7. (Example 8) P_c versus SNR for Ω_4 with a three-tap residual channel; $N = 100$.

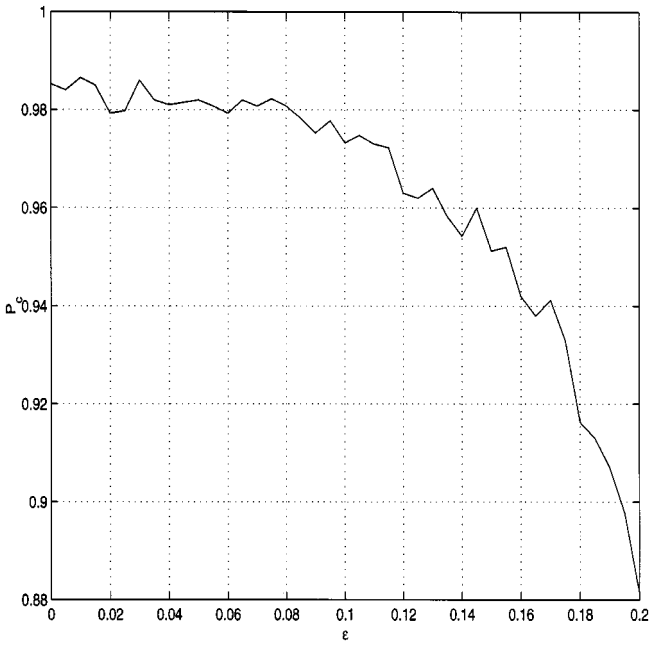

Fig. 8. (Example 9) P_c versus synchronization parameter ϵ_T for Ω_4; $N = 100$ and SNR $= 12$ dB.

thus decreasing the separation between the values of the statistics under the two hypotheses.

The residual channel was assumed to be a three-tap channel $[1, h(1), h(2)]$, where $h(1)$ and $h(2)$ were zero-mean independent Gaussian r.v.'s, each with variance $\sigma_h^2 = 0.01$. P_c is plotted against SNR for $N = 100$ samples in Fig. 7; we stress that the channel coefficients are not known to the classifier. As may be expected, higher SNR (or more samples) are required to achieve the same performance as in the ideal case (see Fig. 2).

Example 9—Synchronization Errors: For rectangular pulse shapes, a synchronization error of ϵ_T [see (1)] translates, after matched filtering, to an equivalent two-path channel $[1 - \epsilon_T, \epsilon_T]$. Other pulse shapes also lead to a two-path channel, but the channel coefficients will not be linear in ϵ_T. With $N = 100$ samples and an SNR of 12 dB, we varied ϵ_T from 0 (perfect synchronization) to 0.2 (20% error); P_c is plotted against ϵ_T in Fig. 8. Notice that excellent results are obtained even when $\epsilon_T = 0.15$ (the corresponding value of β is 0.94).

Example 10—Self-Interference: In duplex systems, where the channel is used for both forward and reverse communication links, the signal of interest is corrupted by a signal with the same symbol rate and modulation format, but at a much lower amplitude. Ignoring other effects in (1), the received signal can thus be written as $y(n) = x(n) + bx_1(n) + g(n)$, where $x_1(n)$ and $x(n)$ are i.i.d. sequences drawn from the same alphabet. We study the performance of our classifier for different values of b in Fig. 9. Excellent results are obtained for $b = 0.1$ and $b = 0.2$; performance degrades for $b \geq 0.25$ (i.e., when the signal-to-noise interference ratio (SIR) is below $20 \log_{10}(b) = 12$ dB). The effective noise in this case is non-Gaussian and highly structured.

Example 11—Cochannel Interference: Here, the interfering noise is from another symbol set assumed to have the same

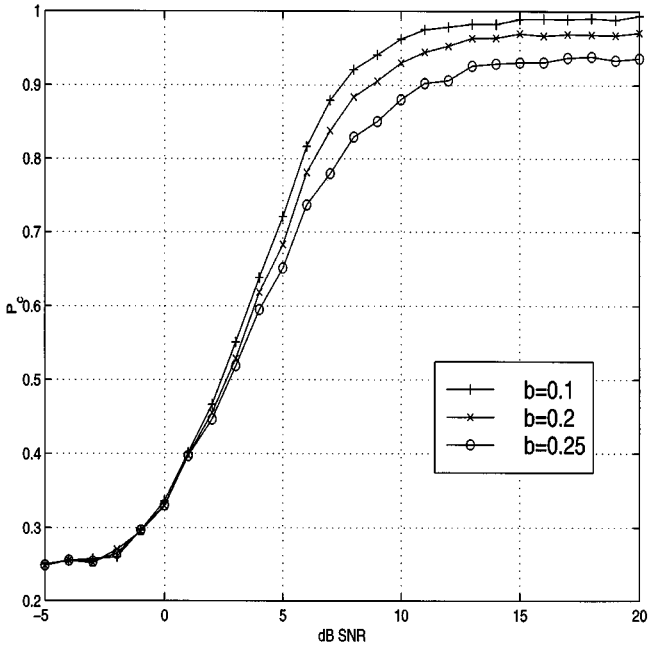

Fig. 9. (Example 10) P_c versus SNR, for Ω_4, with curves parameterized by mixing parameter b; $N = 100$. The interference was drawn from the same symbol set.

symbol timing. In the four-class problem, we used BPSK [PAM(4)] as the interferer for PAM(4) (BPSK), and QAM(4, 4) (PSK-8) as the interferer for PSK-8 [QAM(4, 4)]. The received signal is thus $y(n) = x(n) + cx_I(n) + g(n)$, where x_I denotes the interference. The performance of our classifier is depicted in Fig. 10 versus SNR for different values of c. Results are comparable to that in Example 10, indicating that our classifier is not sensitive to cochannel interference, at least when SIR >12 dB.

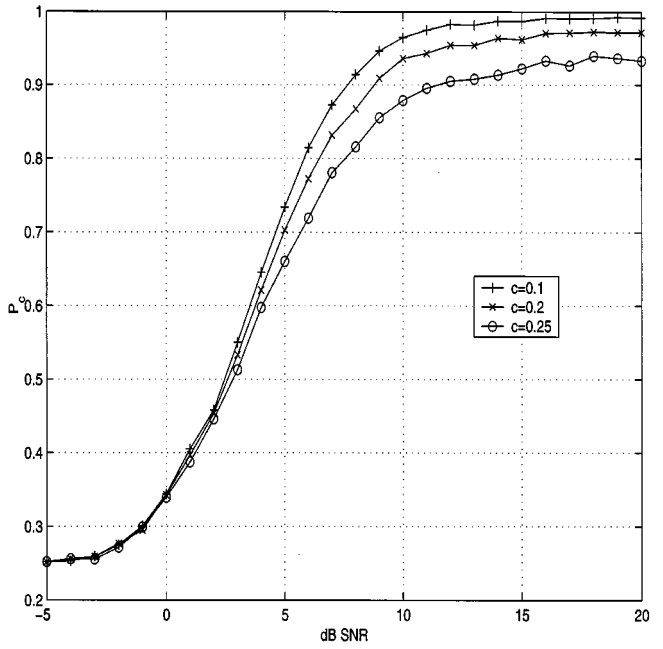

Fig. 10. (Example 11) P_c versus SNR, for Ω_4, with curves parameterized by mixing parameter c; $N = 100$. The interference was drawn from a competing symbol set.

Example 12—Eight-Class Problem: Next, we consider an eight-class problem based on Ω_8 in (17). Here we use both $\tilde{C}_{42}$ and $|\tilde{C}_{40}|$ in the tree-based classifier shown in Fig. 1. Results are shown in Tables V and VI for 1000 realizations of each class member. Table V shows results for the relatively small sample size of $N = 500$ with a relatively high SNR of 20 dB. Note the very good results for the PAM and PSK constellations. Misclassifications occur among the various QAM constellations, particularly between V29 and QAM(4, 4). Table VI depicts the case for a larger sample size of $N = 1000$ at a lower SNR of 10 dB. Results are very similar to those of Table V, showing that it is obviously possible to overcome lower SNR with more samples.

Example 13—Comparison with qLLR: The qLLR classifier [19] is a low-SNR large-sample approximation to the optimal classifier for testing BPSK versus QPSK. The carrier phase is assumed to be noncoherent, i.e., the received signal is modeled as $y(n) = x(n)\exp(j\theta_c) + g(n)$, where $g(n)$ is complex Gaussian and θ_c is $U[-\pi, \pi)$. A detailed analysis of the qLLR may be found in [7], [12], [15], and [19], where approximate expressions are derived for the pdf of the test statistic, enabling near-optimal choice of threshold; see also [3] and [13]. Here, we give some additional insight into the qLLR. Let $y(n) = y_I(n) + jy_Q(n) = x_I(n) + g_I(n) + jx_Q(n) + jg_Q(n)$. Also let σ_I^2, σ_Q^2, and ρ_{IQ} denote the variances and correlation coefficient of the in-phase (real) and quadrature (imaginary) components of $y(n)$. Let γ_{4I} and γ_{4Q} denote the kurtoses of x_I and x_Q. Denote the sample estimates by (where $A = I$ or Q)

$$\hat{\sigma_A^2} := \frac{1}{N}\sum_{n=1}^{N} y_A^2(n), \qquad \hat{\rho}_{IQ} := \frac{1}{N}\sum_{n=1}^{N} \frac{y_I(n)y_Q(n)}{\sqrt{\hat{\sigma_I^2}\hat{\sigma_Q^2}}}.$$

Then, the qLLR statistic can be written as

$$\mathrm{q_{LLR}} = \left(\hat{\sigma_I^2} - \hat{\sigma_Q^2}\right)^2 + 4\left(\hat{\rho}_{IQ}\hat{\sigma_I^2}\hat{\sigma_Q^2}\right)^2.$$

TABLE V
CONFUSION MATRIX FOR EXAMPLE 12, USING THE HIERARCHICAL CLASSIFIER OF Fig. 1. SNR = 20 dB, $N = 500$ SAMPLES, AND 1000 TRIALS. C1–C8 CORRESPOND TO BPSK, PAM(4), PSK(4), PSK(8), V32, V29, QAM(4, 4), AND V29c

	Classifier Output							
	C1	C2	C3	C4	C5	C6	C7	C8
C1	1000	0	0	0	0	0	0	0
C2	0	1000	0	0	0	0	0	0
C3	0	0	1000	0	0	0	0	0
C4	0	0	0	1000	0	0	0	0
C5	0	0	0	0	999	1	0	0
C6	0	0	0	0	12	843	145	0
C7	0	0	0	0	0	74	926	0
C8	0	0	0	0	0	0	0	1000

TABLE VI
CONFUSION MATRIX FOR EXAMPLE 12, USING THE HIERARCHICAL CLASSIFIER OF Fig. 1. SNR = 10 dB, $N = 1000$ SAMPLES, AND 1000 TRIALS. C1–C8 CORRESPOND TO BPSK, PAM(4), PSK(4), PSK(8), V32, V29, QAM(4, 4), AND V29c

	Classifier Output							
	C1	C2	C3	C4	C5	C6	C7	C8
C1	1000	0	0	0	0	0	0	0
C2	0	1000	0	0	0	0	0	0
C3	0	0	1000	0	0	0	0	0
C4	0	0	0	1000	0	0	0	0
C5	0	0	0	0	998	2	0	0
C6	0	0	0	0	5	879	116	0
C7	0	0	0	0	0	71	929	0
C8	0	0	0	0	0	0	0	1000

We can show that asymptotically

$$E\{\mathrm{q_{LLR}}\} \approx \left(\sigma_I^2 - \sigma_Q^2\right)^2 + \frac{1}{N}\left(\gamma_{4I} + \gamma_{4Q}\right) + \frac{2}{N}\left(\sigma_I^2 + \sigma_Q^2\right)^2 \approx \left(\sigma_I^2 - \sigma_Q^2\right)^2.$$

For the QPSK signal, $\sigma_I^2 = \sigma_Q^2$; for the real-valued BPSK, the pdf of y_I is bimodal and has larger variance (and kurtosis) than the unimodal quadrature component y_Q. Indeed, $\sigma_I^2 = \sigma_g^2/2 + 1$ and $\sigma_Q^2 = \sigma_g^2/2$. Hence, $E\{\mathrm{q_{LLR}}\} = 1$ for BPSK and $E\{\mathrm{q_{LLR}}\} = 0$ for QPSK, so that the asymptotic threshold is 0.5. Thus, the qLLR classifier exploits the *asymmetry* between the real and imaginary parts of the received signal.

We compare the performance of our classifier based on C_{40} with that of the qLLR classifier. In general, both have comparable performance for $N > 100$ samples and SNR > 0 dB; qLLR is optimized for the low-SNR case and performs better than C_{40} in the low-SNR regime $-5 \leq \text{SNR} \leq 0$, under ideal conditions. However, the C_{40}-based classifier is more robust to residual channel effects and synchronization errors. A two-path channel $h = [1-d, d]$ was assumed (recall, from Example 9, that this models synchronization errors). Performance of the two classifiers is plotted versus SNR for different values of d in Fig. 11. In the ideal case, with $d = 0$, qLLR has at least a 5-dB edge over C_{40} at SNR's below 0 dB. However, performance of qLLR degrades dramatically for $d \geq 0.15$; in this case, qLLR performance is limited by the channel effects and not by the number of samples or SNR. The performance of the qLLR classifier can be improved by dividing the observed data by $(1-d)$, so that the signal amplitude is normalized to unity; see Fig. 12 where the qLLR-based (but not the C_{40}) classifier knows the channel coefficient $h(0)$. In practice, we do not know the value of d; indeed, if d were known (i.e., known channel), one would opt to use a Viterbi equalizer (or synchronize correctly) before using the classifier.

The qLLR suggests that in some asymmetric cases, it may be advantageous to process the real and imaginary parts separately rather than deal directly with the complex signal. Since $C_{40} \equiv 0$ and $C_{42} \equiv -1$ for M-PSK, $M \geq 4$, several authors [21], [30] have looked at the polar decomposition instead. Recall that in the eight-class problem, most of the errors were due to the difficulty in distinguishing V29 from QAM(4, 4) (see Tables V and VI). Interestingly, it turns out that for V29, $C_{40} = 0.5185$, but $\gamma_{4R} = \gamma_{4I} = -0.1533$, indicating that it may be better to use γ_{4I} rather than C_{40}; this will involve modifications of our hierarchical approach, but is in keeping with the spirit of "macro"/"micro" classification.

Example 14—Comparison with the q_M Classifier: The qLLR classifier cannot distinguish M-PSK from M'-PSK (wlog, $M < M'$) if $M > 4$. By retaining terms in x^2 through x^M in the qLLR approximation, one can distinguish M-PSK from M'-PSK; however, the corresponding classifier appears intractable. The use of the q_m and p_m statistics is studied in [12]

$$q_m := \left| \sum_{n=1}^{N} y^m(n) \right|, \qquad p_m := \sum_{n=1}^{N} |y^m(n)|.$$

For $m = 4$, these correspond to $N|\hat{M}_{40}|$ and $N\hat{M}_{42}$. In general, setting the optimal threshold requires knowledge of SNR, and larger m requires higher SNR (or more samples) to achieve the same performance.

In [15], these classifiers were used in binary hypothesis tests involving QAM signals. The classifier was tested for V29 versus QAM(4, 4) with $N = 100$ samples: using q_4, P_c levels off at 0.67 at a 30-dB SNR; using p_4, it levels off at 0.7; another classifier based on $t_4 := 0.0135q_4 - 0.0246p_4$ achieves $P_c \approx 0.89$ at a 30-dB SNR (see [15, Fig. 5]). The performance of our $|C_{40}|$ and C_{42}-based classifiers are shown in Fig. 13. The performance of our classifier is better than that of q_4, p_4, and t_4 at SNR's

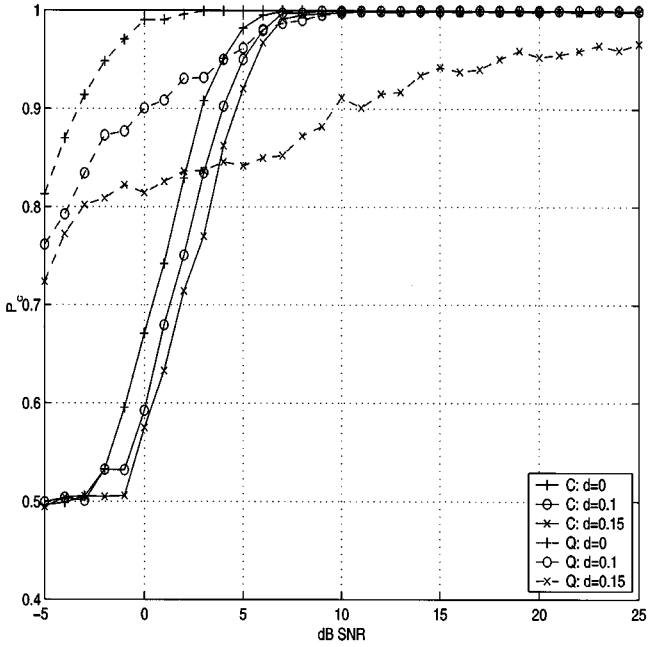

Fig. 11. (Example 13) P_c versus SNR for BPSK versus QPSK, for a two-path channel, $[1-d, d]$, $N = 100$. Performance of qLLR- and C_{40}-based classifiers is shown by dashed and solid lines, for different channels, $d = 0$ $(+)$, $d = 0.1$ (o), and $d = 0.15$ $(\times)$.

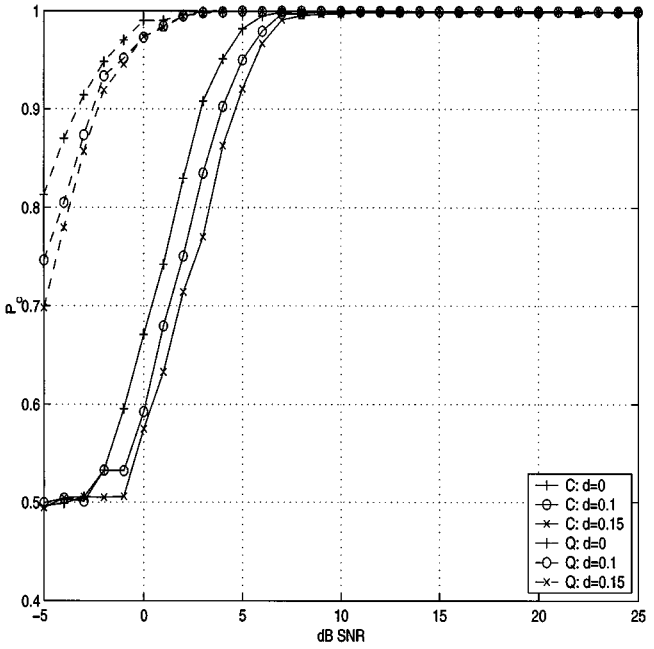

Fig. 12. (Example 13) P_c versus SNR for BPSK versus QPSK, for a two-path channel, $[1-d, d]$, $N = 100$. Performance of qLL- and C_{40}-based classifiers is shown by dashed and solid lines, for different channels, $d = 0$ $(+)$, $d = 0.1$ (o), and $d = 0.15$ $(\times)$. The qLLR classifier knows the value of d.

<10 dB. All these classifiers become data-limited (i.e., performance tends to flatten out with increasing SNR, for a fixed N). In the carrier coherent case, one can use C_{40} and obtain almost perfect classification at less than 10-dB SNR; performance is shown by $+$'s in Fig. 13

Example 15—PSK(16) versus QAM(4, 4): In [15], the q_4 statistic is used to discriminate between PSK(16) and QAM(4,

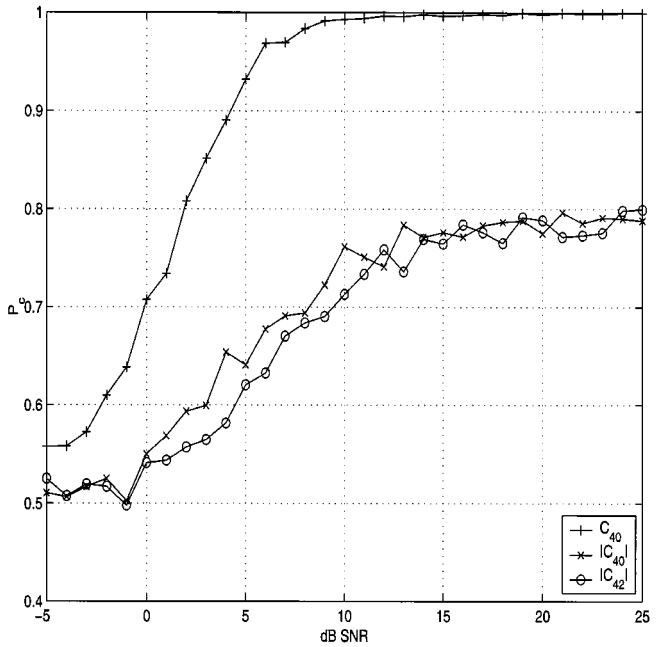

Fig. 13. (Example 14) P_c versus SNR, for V29 versus QAM(4, 4), with $N = 100$: based on C_{40} (+), $|C_{40}|$ (x), and C_{42} (o).

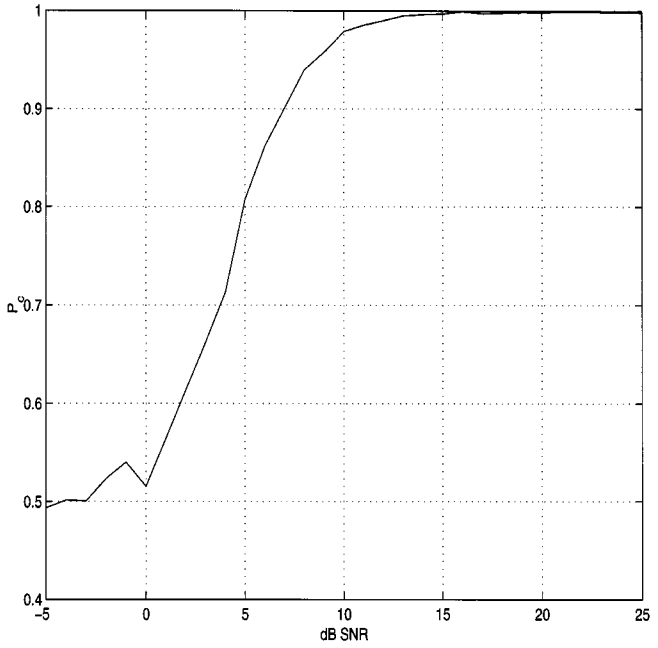

Fig. 14. (Example 15) P_c versus SNR, for PSK(16) versus QAM(4, 4), with $N = 100$.

4); using the optimal threshold, $P_c \approx 0.95$ was obtained at a 9–11-dB SNR with $N = 100$ samples (see [15, Fig. 2]), and $P_c \approx 1$ was obtained at a 20-dB SNR. The performance of our classifier based on C_{40} is shown in Fig. 14; we obtain $P_c \approx 0.95$ at about 8 dB, and $P_c \approx 0.99$ at 12 dB, which represents about 1–3-dB gain compared with the q_4 classifier. We note that the optimal and approximate thresholds for q_4/N in [15, eqs. (16), (23), (25), (26)] increase (unexpectedly) as $\sqrt{N}$; our threshold for $|C_{40}|$ (derived from Table I) was fixed at 0.34.

VI. Conclusions

We have shown that simple HOS are useful for classification of digitally modulated signals. They are particularly effective when used in a hierarchical scheme, allowing broad classification at low SNR. The approach is robust in the presence of carrier phase and frequency offsets, as well as impulsive non-Gaussian noise. Theoretical arguments were verified via extensive simulations and comparisons with existing approaches.

The decision thresholds were on the conservative side because they were obtained by assuming that the sample estimates of the test statistics C_{40} and C_{42} have equal variances under different hypotheses, and ignored the effects of additive noise. The performance could be improved by taking these issues into account.

The classification rule we have used is based on knowledge of the cumulants of the symbol set. In principle, our approach can be extended to the case where the constellation points are non-equiprobable (e.g., in unrandomized systems or in systems with shaped constellations) or when classifying other modulation formats such as V.34, V.33 trellis-coded, V.32 trellis-coded, etc.

Also of interest are classification problems that admit a "none of the above" hypothesis, where the observed data may be drawn from an unknown symbol set. The methods cited here generally do not address this question, nor do they address the question of determining the constellation size. These issues are addressed in [25] and [26], where we propose the use of hierarchical agglomerative clustering algorithms. Clustering provides a natural tool for offline analysis of unknown constellation types and can be used in conjunction with the HOS-based techniques proposed in this paper.

Appendix
Variance of Sample Estimates of Cumulants

We derive the variance of the estimates of the cumulants in (6). Usually, it is assumed that the variance of the estimate $\hat{C}_{21}$ is small enough to be ignored. We demonstrate that this is not a good assumption in general. Let $M_{km} := E[y^{k-m}(y^*)^m]$, the mixed moment of order k. Note that under the zero-mean assumption, $M_{20} = C_{20}$ and $M_{21} = C_{21}$. We will make use of the fact that sample estimates of $M_{k,m}$ are unbiased and asymptotically Gaussian with variance $(M_{2k,k} - |M_{k,m}|^2)/N$. Similar results hold asymptotically for sample estimates of the cumulants, e.g., see Brillinger [6]; however, we have not been able to find explicit expressions for the variances.

A. Known Second-Order Statistics

Assume that C_{20} and C_{21} are known exactly; recall that for the QAM and PSK symbol sets $C_{20} = 0$, and that C_{21} is the signal energy. In this case, the sample estimates of the cumulants are unbiased, and the variances are equal to those of the corresponding moments. We have

$$E[\hat{C}_{40}] = C_{40}$$
$$E[\hat{C}_{42}] = C_{42}$$
$$N \operatorname{var}[\hat{C}_{40}] = N \operatorname{var}[\hat{M}_{40}] = M_{84} - |M_{40}|^2 \quad (19)$$
$$N \operatorname{var}[\hat{C}_{42}] = N \operatorname{var}[\hat{M}_{42}] = M_{84} - |M_{42}|^2 \quad (20)$$

which establish (8) and (9).

B. Estimated Second-Order Statistics

For the QAM and PSK signals, we know that $C_{20} = 0$; hence, we set $\hat{C}_{20} = 0$ in (6). Since the signal energy may be unknown, we need to estimate C_{21}. Note that the variance expression for $\hat{C}_{40}$ continues to be given by (19), since $C_{40} = M_{40}$ under the assumption $C_{20} = 0$. The evaluation of the variance expression for $\hat{C}_{42}$ is complicated by the presence of the C_{21}^2 term. We have

$$\begin{aligned}\hat{C}_{42} &= \hat{M}_{42} - 2(\hat{C}_{21})^2 \\ E[\hat{M}_{42}] &= \frac{1}{N}\sum_{m=1}^{N} E|y(m)|^4 = M_{42} \\ E[(\hat{M}_{21})^2] &= \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N} E[|y(m)y(n)|^2] \\ &= M_{21}^2 + \frac{1}{N}\left(M_{42} - M_{21}^2\right) \\ E[\hat{C}_{42}] &= C_{42} - \frac{2}{N}\left(M_{42} - M_{21}^2\right).\end{aligned}$$

Hence, $\hat{C}_{42}$ is only *asymptotically* unbiased. It is usually argued that if N is large enough, the bias [which is in fact $\mathrm{var}(\hat{C}_{21})$] will be negligible, and that we can use the results in the previous subsection. For convenience, let $\alpha := M_{42} - M_{21}^2$. The exact error variance can be expressed as

$$\begin{aligned}\mathrm{var}[\hat{C}_{42}] &= \mathrm{var}[\hat{M}_{42}] + 4\mathrm{var}\left[\hat{M}_{21}^2\right] - 4\,\mathrm{cov}\left[\hat{M}_{42},\, \hat{M}_{21}^2\right] \\ \mathrm{var}[\hat{M}_{42}] &= \frac{1}{N}\left(M_{84} - M_{42}^2\right) \\ E\left(\hat{M}_{21}^2\right)^2 &= \frac{1}{N^4}\sum_{m=1}^{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} E|y(m)y(i)y(j)y(k)|^2.\end{aligned}$$

In the fourfold summation, the sum over $m \neq i \neq j \neq k$ yields an $O(1)$ term, and sums of the form $m = n \neq i \neq j$ yield $O(1/N)$ terms; the remaining terms are $O(1/N^2)$ and $O(1/N^3)$ and can be omitted in an asymptotic analysis. Hence, we obtain

$$\begin{aligned}\mathrm{var}\left[\hat{M}_{21}^2\right] &= \frac{(N-1)(N-2)(N-3)}{N^3} M_{21}^4 \\ &\quad + \frac{6(N-1)(N-2)}{N^3} M_{21}^2 M_{42} \\ &\quad - \left(M_{21}^2 + \frac{\alpha}{N}\right)^2 + O(1/N^2) \\ &\approx \frac{6}{N} M_{21}^2 M_{42} - \frac{2}{N} M_{21}^2 \alpha \\ &= \frac{2}{N} M_{21}^2 \left[2M_{42} + M_{21}^2\right].\end{aligned}$$

The covariance term is obtained as

$$\begin{aligned}\mathrm{cov}\left[\hat{M}_{42}\,\hat{M}_{21}^2\right] &= \frac{1}{N^2}\sum_{k=1}^{N}\sum_{l=1}^{M}\sum_{m=1}^{N} E[|y^2(k)y(l)y(m)|^2] \\ &\quad - E[\hat{M}_{42}]E\left[\hat{M}_{21}^2\right] \\ &\approx M_{42}M_{21}^2 + \frac{1}{N}\left[2M_{63}M_{21} + M_{42}^2\right] \\ &\quad - M_{42}\left[M_{21}^2 + \frac{1}{N}\left(M_{42} - M_{21}^2\right)\right] \\ &= \frac{M_{21}}{N}\left[2M_{63} + M_{42}M_{21}\right].\end{aligned}$$

Combining the various subexpressions, we obtain (10). Note that if C_{21} is known, the variance expression is given by the first term of (10) [see (9)]. In the case of real signals, $C_{42} = C_{40} = M_{42} - 3C_{21}^2$, which leads to (11).

C. Estimation in AWGN

Since the fourth-order cumulants of AWGN are zero, the sample estimates of the fourth-order cumulants are (asymptotically) unbiased estimates of the signal cumulants; the variances of the sample estimates will of course be affected by the AWGN, which is assumed to be complex and circularly symmetric. The various expressions in (8)–(11) continue to hold; however, the various moments are now expected values of powers of $g(n) + x(n)$ rather than $x(n)$ alone. We will first rewrite these equations in terms of cumulants rather than moments.

Both the symbol sequence and the additive noise are zero-mean. The complex QAM and PSK constellations have fourfold symmetry; hence, the symbol sequence $x(n)$ satisfies $C_{20} = 0$, $M_{41} = C_{41} = 0$. The additive noise is assumed to be circularly symmetric; hence, both the noise $g(n)$ as well as the noisy signal $y(n) = x(n) + g(n)$ satisfy $C_{20} = 0$, $M_{41} = C_{41} = 0$. We will make repeated use of this in the following derivation.

Moments and cumulants can be expressed in terms of one another [6]. For a collection of r.v.'s, $\{u_i\}_{i=1}^{8}$

$$\begin{aligned}M_8 &= C_8 + [M_2M_6]_{28} + [M_4M_4]_{35} - 2[M_2M_2M_4]_{210} \\ &\quad + 6[M_2M_2M_2M_2]_{105} \\ M_6 &= C_6 + [M_2M_4]_{15} - 2[M_2M_2M_2]_{15}\end{aligned}$$

where we have used the bracket notation of [17]. Here, $[M_2M_6]_{28}$ denotes 28 terms obtained by partitioning the eight r.v.'s into two disjoint sets of two and six elements; one of these terms is $E\{u_1u_2\}E\{u_3u_4u_5u_6u_7u_8\}$, and there are a total of 28 such partitions or terms. For evaluating M_{84}, we note that $u_i = y$ for $i = 1, \cdots, 4$, and $u_j = y^*$ for $j = 5, \cdots, 8$. Since $M_{20,\,y} = 0$, several terms vanish. We obtain

$$\begin{aligned}[M_2M_6]_{28} &= 16M_{21}M_{63} = 16C_{21}M_{63} \\ [M_2M_4]_{15} &= 9M_{21}M_{42} = 9C_{21}\left[C_{42} + 2C_{21}^2\right] \\ [M_2M_2M_2]_{15} &= 6M_{21}^3 = 6C_{21}^3 \\ [M_4M_4]_{35} &= |M_{40}|^2 + 18M_{42}^2 \\ &= |C_{40}|^2 + 18\left[C_{42} + 2C_{21}^2\right]^2 \\ [M_4M_2M_2]_{210} &= 72C_{21}^2M_{42} = 72C_{21}^2\left[C_{42} + 2C_{21}^2\right] \\ [M_2M_2M_2M_2]_{105} &= 24C_{21}^4.\end{aligned}$$

From the above, we obtain

$$M_{63} = C_{63} + 9C_{42}C_{21} + 6C_{21}^3 \tag{21}$$

$$\begin{aligned}M_{84} &= C_{84} + 16C_{63}C_{21} + |C_{40}|^2 + 18C_{42}^2 \\ &\quad + 72C_{42}C_{21}^2 + 24C_{21}^4.\end{aligned} \tag{22}$$

We also have

$$M_{42} = C_{42} + 2C_{21}^2, \qquad M_{40} = C_{40}, \qquad M_{21} = C_{21}. \tag{23}$$

We use (21)–(23) to reexpress the variance expressions in (8)–(11) in terms of the cumulants of the observed process $y(n) = x(n) + g(n)$, leading to (12)–(14).

REFERENCES

[1] K. Assaleh, K. Farrell, and R. J. Mammone, "A new method of modulation classification for digitally modulated signals," in *Proc. MILCOM*, vol. 2, San Diego, CA, 1992, pp. 712–716.

[2] E. E. Azzouz and A. K. Nandi, *Automatic Modulation Recognition of Communication Signals*. Norwell, MA: Kluwer, 1996.

[3] B. F. Beidas and C. L. Weber, "Higher-order correlation-based approach to modulation classification of digitally frequency-modulated signals," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 89–101, Jan. 1995.

[4] N. Benvenuto and T. W. Goeddel, "Classification of voiceband data signals using the constellation magnitude," *IEEE Trans. Commun.*, vol. 43, pp. 2759–2770, Nov. 1995.

[5] W. Betts, A. R. Calderbank, and R. Laroia, "Performance of nonuniform constellation on Gaussian channel," *IEEE Trans. Inform. Theory (*, vol. 40, pp. 1633–1638, Sept. 1994.

[6] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York: McGraw-Hill, 1981.

[7] K. M. Chugg, C.-S. Long, and A. Polydoros, "Combined likelihood power estimation and multiple hypothesis modulation classification," *Proc. 29th Asilomar Conf.*, pp. 1137–1141, 1995.

[8] D. L. Donoho and X. Huo, "Large-sample modulation classification using Hellinger representation," in *Proc. IEEE SPAWC*, Paris, France, Apr. 1997, pp. 133–136.

[9] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973, ch. 6.

[10] D. N. Godard, "Self-recovering equalization and carrier-tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.

[11] A. O. Hero and H. Hadinejad-Mehram, "Digital modulation classification using power moment matrices," in *Proc. ICASSP*, Seattle, WA, 1998, pp. 3285–3288.

[12] C.-Y. Huang and A. Polydoros, "Likelihood methods for MPSK modulation classification," *IEEE Trans. Commun.*, vol. 43, pp. 1493–1504, Feb./Mar./Apr. 1995.

[13] C. Le Martret and D. Boiteau, "A general maximum-likelihood classifier for modulation classification," in *Proc. EUSIPCO*, Rhodes, Greece, 1998, pp. 985–988.

[14] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[15] C. Long, K. Chugg, and A. Polydoros, "Further results in maximum likelihood classification of QAM signals," in *Proc. MILCOM*, Long Branch, NJ, 1994, pp. 57–61.

[16] C. Louis and P. Sehier, "Automatic modulation recognition with a hierarchical neural network," in *Proc. MILCOM*, vol. 3, Fort Monmouth, NJ, 1994, pp. 713–717.

[17] P. McCullagh, *Tensor Methods in Statistics*: Chapman & Hall, 1987.

[18] B. Paris, G. Orsak, H. Chen, and N. Warke, "Modulation classification in unknown dispersive environments," in *Proc. ICASSP*, vol. 5, Munich, Germany, Apr. 1997, pp. 3853–3856.

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

[20] J. Reichert, "Automatic classification of communication signals using higher order statistics," in *Proc. ICASSP*, vol. 5, San Francisco, CA, 1992, pp. 221–224.

[21] P. C. Sapiano, J. D. Martin, and R. J. Holbeche, "Classification of PSK signals using the DFT of phase histogram," in *Proc. ICASSP*, vol. 3, Detroit, MI, 1995, pp. 1868–1871.

[22] S. Soliman and S.-Z. Hsue, "Signal classification using statistical moments," *IEEE Trans. Commun.*, vol. 40, pp. 908–916, May 1992.

[23] B.M. Sadler, "Detection in colored impulsive noise using fourth-order cumulants," *IEEE Trans. Signal Processing*, vol. 44, pp. 2793–2800, Nov. 1996.

[24] M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan, *Introduction to Statistical Signal Processing With Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

[25] A. Swami and B. M. Sadler, "Modulation classification via hierarchical agglomerative cluster analysis," in *Proc. IEEE SPWC*, Paris, France, Apr. 1997, pp. 141–144.

[26] A. Swami and B. M. Sadler, "Hierarchical modulation classification," in *Proc. 2nd Annu. Fedlab Symp.*, College Park, MD, Feb. 1998, pp. 158–162.

[27] A. Swami and B. M. Sadler, "Issues in military communications," *IEEE Signal Processing Mag.*, vol. 16, pp. 31–33, Mar. 1999.

[28] W. Wei and J. M. Mendel, "A new maximum-likelihood method for modulation classification," in *Proc. Asilomar*, 1995, pp. 1132–1135.

[29] S. L. Wood, M. G. Larimore, and J. R. Treichler, "Modem constellation identification: A performance comparison of two methods," in *Proc. ICASSP*, vol. 3, Albuquerque, NM, 1990, pp. 1651–1654.

[30] Y. Yang and S. S. Soliman, "A suboptimal algorithm for modulation classification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, pp. 38–47, MONTH 1997.

[31] G. Zhou, G. B. Giannakis, and A. Swami, "On polynomial phase signals with time-varying amplitudes," *IEEE Trans. Signal Processing*, vol. 44, pp. 846–861, Apr. 1996.

[32] Q. Zhu, M. Kam, and R. Yeager, "Non-parametric identification of QAM constellations in noise," in *Proc. ICASSP*, vol. 4, Minneapolis, MN, 1993, pp. 184–187.

Ananthram Swami (S'79–M'84–SM'96) recieved the B.Tech. degree from the Indian Institute of Technology, Bombay, India, the M.S. degree from Rice University, Houston, TX, and the Ph.D. degree from the University of Southern California, Los Angeles, all in electrical engineering.

He has held positions with Unocal, USC, CS-3, and Malgudi Systems. He is currently a Senior Research Scientist at the Army Research Lab, Adelphi, MD, where his work is in the broad area of statistical signal processing with emphasis on applications to communications.

Dr. Swami is a a member of the IEEE Signal Processing Society's Technical Committee on Signal Processing for Communications. He was an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and coorganizer and cochair of the 1993 IEEE SPS Workshop on Higher Order Statistics, the 1996 IEEE SPS Workshop on Statistical Signal and Array Processing, and the ASA-IMA 1999 Workshop on Applications of Heavy-Tailed Distributions.

Brian M. Sadler (M'90) received the B.S. and M.S. degrees from the University of Maryland, College Park, in 1981 and 1984, respectively, and the Ph.D. degree from the University of Virginia, Charlottesville, in 1993, all in electrical engineering.

He has been a member of the Technical Staff of the Army Research Laboratory (ARL) and the former Harry Diamond Laboratories in Adelphi, MD, since 1982. He was a lecturer at the University of Maryland from 1985 to 1987, and has been lecturing at the Johns Hopkins University Whiting Institute since 1994 on statistical signal processing and communications. His research interests generally include statistical signal processing with applications in communications, radar, and aeroacoustics.

Dr. Sadler is a member of the IEEE Technical Committee on Signal Processing for Communications. He is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He cochaired the Second IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC'99).