Asymptotic Results in PSK Modulation Classification

Ethan Davis

E. Davis is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: ethan@mast.queensu.ca).

Abstract

Two asymptotic results in modulation classification are presented. First, a consideration of the noiseless case motivates a lower bound on misclassification probability. Contrary to central limit theorem approximations, the error probability does not go to zero as the signal-to-noise ratio goes to infinity for a fixed number of samples, but is bounded from below by a nonzero error floor if the constellations under different hypotheses have signal points in common. Second, Chernoff information, Bhattacharyya distance and Kullback-Liebler distances are calculated for phase-shift keying modulation classification problems. Simulation results show the utility of Chernoff and Bhattacharyya bounds for phase-shift keying modulation classification. It is shown that the Chernoff information is a valuable performance measure for modulation classification, both asymptotically and for a finite number of observations.

Index Terms

Chernoff bound, modulation classification.

I. Introduction

Modulation classification is the problem of identifying the modulation scheme employed by the transmitter of a communications signal. Although originally a problem in noncooperative communications theory, modulation classification is increasingly being applied in cooperative scenarios. One such application is in communication systems employing adaptive modulation techniques where blind modulation classification alleviates the need for overhead symbols carrying information about modulation format, thereby increasing the information throughput.

Previous work has focussed on investigating the performance of particular classifiers, by simulation and/or analysis[1]–[6]. This includes investigation of feature-based methods[3]–[6] as well as likelihood methods for modulation classification[1],[2] which are motivated by optimal decision theory rather than an ad-hoc or heuristic approach. For phase-shift keying modulation classification, feature-based detectors typically extract only phase information and then construct algorithms motivated by differences in moments[3], phase histograms (analyzed in [1],[2]), or the discrete Fourier transform of phase histograms[4] under each of the possible hypotheses. In [5] and [6], the received phase is processed through a likelihood test approximation using a Fourier expansion of the phase probability density functions. Likelihood methods are analyzed in [1] and [2], which develop so-called quasi-log-likelihood (qLLR) classifiers, based on an approximation to the log-likelihood ratio. Analytical probability of error expressions for the qLLR classifiers are obtained under a central limit theorem argument, and their validity for small values of signal-to-noise ratio(SNR) and small-to-moderate numbers of samples is confirmed using simulation results. Some work has been done on performance bounds for modulation classification. Reference [7] showed that, for a finite number of distinct constellations, the error rate of a maximum likelihood modulation classifier goes to zero as the number of samples approaches infinity.

Distance measures for probability distributions have been proposed which are easier to calculate than the analytical probability of error expression, but capture an essential feature of the performance of statistical tests or communications systems[8]–[13]. These measures include Bhattacharyya distance(BD)[8],[9], Kullback-Liebler distance(KLD)[14], Chernoff information(CI)[11], and more general information divergences[12],[13],[15]. Although they do not define true distance metrics on the space of probability distributions or densities, these quantities exhibit important properties about the probability of error that the usual distance metrics (e.g. the $\mathcal{L}_2$ norm) do not. Specifically, these coefficients relate to the probability of error asymptotically as "error exponents," and through bounds when the number of samples is finite. Results from large deviation theory[10],[14],[15] can be used to show that these quantities govern or bound the achievable logarithmic rate of decay to zero of probability of misclassification, in the limit as the number of samples goes to infinity. The CI is the optimal asymptotic error exponent for the Bayes error probability if the a priori probabilities are strictly positive[16]. Furthermore, in an M-ary hypothesis test, the minimum CI among all distinct pairs of the hypotheses determines the asymptotic error exponent of the test[16]. The BD has been proposed as a design criterion for communication systems[8],[9]. It provides both upper and lower bounds to the Bayes probability of error for a finite number of observations[8],[17], and upper and lower bounds on the CI. Bounds for the M-ary hypothesis testing problem have also been formulated which involve pairwise BDs[18]–[20], which further motivates its use as a performance measure. If a Neyman-Pearson criterion is used, the KLDs between the distributions are the relevant error exponents. Stein's lemma[14] shows that, in the limit as the number of samples goes to infinity, the best achievable logarithmic rate to zero of the type II error probability is given by a KLD, in the limit as the fixed type I error probability constraint goes to zero.

In this paper, we show that for a fixed number of samples, the probability of misclassification of an optimal classifier does not necessarily go to zero as the signal-to-noise ratio (SNR) goes to infinity. Instead, it is bounded from below by a nonzero error floor, if the constellations overlap under different hypotheses. Further, we calculate the CI, BD and KLDs for the classification of BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK. Simulation results are used to show the performances of the classifiers, and the practical implications of these distances.

This paper is organized as follows. In Section II, we establish the model for the binary modulation classification problem considering two-dimensional constellations in additive white Gaussian noise (AWGN). In Section III, we consider the problem in a noiseless environment, which motivates the derivation of a bound for the noisy case. It is shown that the error probability for this problem can exhibit a lower bound for a fixed number of samples. In Section IV, we calculate the CI, BD and KLDs for three phase-shift keying modulation classification problems. In Section V, we simulate classifiers for these problems using a Bayes minimum probability of error criterion, and a Neyman-Pearson criterion, to show their performances and the practical implications of the error exponent bounds.

II. System Model

Consider the binary modulation classification problem for two-dimensional modulation schemes transmitted in AWGN[21]. The received signal is given by

$$r(t) = s(t) + n(t),\ t \,\epsilon\, [-mT, (n+m-1)T], \tag{1}$$

where $\frac{1}{T}$ is the symbol rate, $2mT$ is the pulse duration[1] and $n(t)$ is AWGN with two-sided power spectral density $\frac{N_0}{2}$ W/Hz. Assuming perfect symbol synchronization, the signal component is given by

$$s(t) = \mathcal{R}e\Big\{\sqrt{\frac{2}{T}}\sum_{k=1}^{n} s_k\, p\big(t-(k-1)T\big)e^{j(\omega_c t+\theta_c)}\Big\}, \tag{2}$$

where $p(t)$ is a unit-energy pulse that is zero outside of $[-mT, mT]^1$ for an integer $m \geq 1$, ω_c and θ_c are the carrier frequency and phase, respectively, and $\{s_k\}$ is an iid sequence of complex values drawn according to a uniform distribution on one of two sets A or B. We assume that $p(t)$ is such that the Nyquist pulse-shaping criterion is met[22], so that there is no intersymbol interference at the output of the matched filter in the receiver. Under hypothesis H_0, which occurs with a priori probability π_0, the constellation is the set of complex points $A = \{\underline{a_1}, \underline{a_2}, \ldots \underline{a_{M_0}}\}$, which satisfies $\frac{1}{M_0}\sum_{i=1}^{M_0}|a_i|^2 = E_s$, where $|a_i|$ is the complex norm of $\underline{a_i}$, and E_s is the average energy per symbol. Under hypothesis H_1, which occurs with a priori probability π_1, the constellation is the set $B = \{\underline{b_1}, \underline{b_2}, \ldots \underline{b_{M_1}}\}$, which satisfies $\frac{1}{M_1}\sum_{i=1}^{M_1}|b_i|^2 = E_s$. The binary modulation classification problem is a hypothesis testing problem in which the receiver must decide which constellation is in use at the transmitter. We assume throughout that ω_c, θ_c, T, $p(t)$, E_s and N_0 are known to the receiver, so that effectively the only unknowns are the data sequence and the hypothesis. We further assume, without loss of generality, $0 < \pi_0, \pi_1 < 1$ to avoid trivial cases.

The optimal classifier structure for this model is known to consist of a matched filter followed by a threshold device[21]. Denote the received sequence as $\underline{r} \,\epsilon\, \mathcal{C}^n$, where $\underline{r}$ is a vector of the signal samples at the output of the receiver matched filter and $\mathcal{C}^n$ denotes an n-tuple of complex values, and the decision function as $D_{N_0} : \mathcal{C}^n \rightarrow \{0, 1\}$, which indicates a decision for hypothesis H_0 or H_1. Using a maximum a posteriori (MAP) criterion, an optimum decision function is[21]

$$D_{N_0}(\underline{r}) = \begin{cases} 0 & \text{if } L_{N_0}(\underline{r}) > 0 \\ 1 & \text{if } L_{N_0}(\underline{r}) \leq 0 \end{cases}, \tag{3}$$

[1]This formulation admits the use of windowed bandwidth efficient pulse shaping such as Nyquist pulse-shaping

where

$$L_{N_0}(\underline{r}) = \frac{1}{n}\log(\frac{\pi_0}{\pi_1}) + \log(\frac{M_1}{M_0}) + \frac{1}{n}\sum_{i=1}^{n}\log\Big(\frac{\sum_{j=1}^{M_0}\exp(\frac{-1}{N_0}\|\underline{r_i}-\underline{a_j}\|^2)}{\sum_{k=1}^{M_1}\exp(\frac{-1}{N_0}\|\underline{r_i}-\underline{b_k}\|^2)}\Big) \tag{4}$$

is the log-likelihood ratio.

In the case of phase-shift keying modulation, the constellation points are given by $a_m = \sqrt{E_s}e^{j\theta_m}$, where θ_m is an element of the set

$$\{k\frac{2\pi}{M}, k = 0\ldots M-1\}\,, \tag{5}$$

where M is the number of points in the constellation. Note that, under this definition, the constellations under different hypotheses have a nonempty intersection. For classification between BPSK and QPSK with a priori probabilities π_B and π_Q, an optimal test statistic using a maximum a posteriori (MAP) criterion is given by

$$L_{N_0}^{QB} = \frac{1}{n}\log\frac{\pi_Q}{\pi_B} + \log\frac{1}{2} + \frac{1}{n}\sum_{i=1}^{n}\log\Big(\frac{\cosh(\frac{2\sqrt{E_s}}{N_0}x_i) + \cosh(\frac{2\sqrt{E_s}}{N_0}y_i)}{\cosh(\frac{2\sqrt{E_s}}{N_0}x_i)}\Big)\,, \tag{6}$$

where x_i and y_i are the real and imaginary components, respectively, of the received signal sample r_i. For a test between BPSK and 8PSK, with a priori probabilities π_B and π_8, respectively, an optimal MAP test statistic is given by

$$\begin{aligned} L_{N_0}^{8B} &= \frac{1}{n}\log\frac{\pi_8}{\pi_B} + \log\frac{1}{4} + \\ &\frac{1}{n}\sum_{i=1}^{n}\log\Big(\frac{\cosh(\frac{2\sqrt{E_s}}{N_0}x_i)+\cosh(\frac{2\sqrt{E_s}}{N_0}y_i)+\cosh(\frac{2\sqrt{E_s}}{N_0}(x_i+y_i))+\cosh(\frac{2\sqrt{E_s}}{N_0}(x_i-y_i))}{\cosh(\frac{2\sqrt{E_s}}{N_0}x_i)}\Big)\,. \end{aligned} \tag{7}$$

For the test between QPSK and 8PSK with a priori probabilities π_Q and π_8, an optimal test statistic is given by

$$\begin{aligned} L_{N_0}^{8Q} = &\frac{1}{n}\log\frac{\pi_8}{\pi_Q} + \log\frac{1}{2} + \\ &\frac{1}{n}\sum_{i=1}^{n}\log\Big(\frac{\cosh(\frac{2\sqrt{E_s}}{N_0}x_i)+\cosh(\frac{2\sqrt{E_s}}{N_0}y_i)+\cosh(\frac{2\sqrt{E_s}}{N_0}(x_i+y_i))+\cosh(\frac{2\sqrt{E_s}}{N_0}(x_i-y_i))}{\cosh(\frac{2\sqrt{E_s}}{N_0}x_i)+\cosh(\frac{2\sqrt{E_s}}{N_0}y_i)}\Big)\,. \end{aligned} \tag{8}$$

III. Noiseless Bound

Consider the binary modulation classification problem in the noiseless case (i.e., $N_0 = 0$), which is a hypothesis test on a discrete probability space, since the received symbols are exactly the transmitted symbols. Under hypothesis H_0, which occurs with a priori probability π_0, the constellation is the set of complex points $A = \{\underline{a_1}, \underline{a_2}, \ldots \underline{a_{M_0}}\}$, while under hypothesis H_1, with a priori probability π_1, the constellation is the set $B = \{\underline{b_1}, \underline{b_2}, \ldots \underline{b_{M_1}}\}$. Let $M_a = |\{A \cap B\}|$ be the size of the intersection between the two constellations. Let $S_a = \{A \cap B\}^n$ denote the set of n-tuples from the intersection. These are called "ambiguous sequences" because they are possible under both hypotheses, and are the sequences that can lead to errors in the noiseless case. Define S_0 as the set of n-tuples containing at least one point from $A \setminus B$, i.e. a point in A that is not in B. If the observation is in this set, a maximum a posteriori (MAP) classifier will correctly choose H_0 because it is impossible that such a sequence could have been observed under H_1. Let S_1 be the set of n-tuples containing at least one point from $B \setminus A$. If the observation is in this set, the noiseless case classifier will correctly choose H_1.

An optimal test is again given by (3), using the likelihood function given by

$$L_0(\underline{r}) = \frac{1}{n}\log(\frac{\pi_0}{\pi_1}) + \log(\frac{M_1}{M_0}) + \frac{1}{n}\sum_{i=1}^{n}\log\Big(\frac{\sum_{j=1}^{M_0}\delta(\underline{r_j}-\underline{a_j})}{\sum_{k=1}^{M_1}\delta(\underline{r_k}-\underline{b_k})}\Big) \ , \tag{9}$$

where $\delta(\underline{r}) = 1$ if $\underline{r} = \underline{0}$ and zero otherwise. Then, the probability of error is given by

$$P_e(n) = \min\Big\{\pi_0(\frac{M_a}{M_0})^n,\ \pi_1(\frac{M_a}{M_1})^n\Big\} \ . \tag{10}$$

Thus, for a fixed number of samples, the noiseless case error probability is bounded from below by a nonzero error floor, if the intersection between the constellations is nonempty. For large enough n, the error probability is decreasing exponentially with n, except in the trivial case $M_a = M_0 = M_1$. The noiseless error probability expression in (10) is a bound in the sense that it is the limiting error probability of the noisy case test given by (3) and (4) as the SNR goes to infinity. That is, for modulation classification in noise,

$$\lim_{N_0 \to 0} P_e(n) = \min\Big\{\pi_0(\frac{M_a}{M_0})^n,\ \pi_1(\frac{M_a}{M_1})^n\Big\} \ . \tag{11}$$

The rigorous mathematical proof of this fact is nontrivial but too long to be included in this paper, owing to length restrictions. A rigorous proof of this is given in reference [23].

Consider the binary modulation classification problems between BPSK and QPSK, BPSK and 8PSK, and QPSK and 8PSK, in the noiseless case. Assume that the phases are given by (5), so that the signals possible under the hypothesis associated with the smaller constellation are also possible under the other hypothesis. For the binary test between H_0 (BPSK) and H_1 (QPSK) with a priori probabilities π_0 and $\pi_1 = 1 - \pi_0$, respectively, using (10), the error probability in the limit as the SNR goes to infinity is given by

$$P_e(n) = \min\{\pi_0, \pi_1\, 2^{-n}\}\,. \tag{12}$$

For the test between H_1 (QPSK) with a priori probability π_1 and H_2 (8PSK) with a priori probability $\pi_2 = 1 - \pi_1$, the error probability in the limit is given by

$$P_e(n) = \min\{\pi_1, \pi_2\, 2^{-n}\}\,. \tag{13}$$

Finally, for the test between H_0 (BPSK) with a priori probability π_0 and H_2 (8PSK) with a priori probability $\pi_2 = 1 - \pi_0$, the error probability in the limit is given by

$$P_e(n) = \min\{\pi_0, \pi_2\, 4^{-n}\}\,. \tag{14}$$

We note that the error probability expressions in [2] are derived through a central limit theorem argument. Simulation results are used to show that these expressions are accurate for small values of SNR and small-to-moderate numbers of samples. While [2] did not propose using the analytical expressions for large values of SNR, we note that the expressions indicate that the error probability tends to zero for large SNR where the noiseless bound is significant and the probability of misclassification does not go to zero for large SNR.

IV. Error Exponent Bounds

In this section, we calculate the CI, BD and KLDs for the modulation classification problems of BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK. We first state the definitions and bounds related to these quantities.

A. *Chernoff Bound*

Assume that under both hypotheses H_0 and H_1, a density for the received signal exists, and the single sample densities are given by $p_0(\underline{r})$ and $p_1(\underline{r})$, respectively. The alpha entropy between $p_0(\underline{r})$ and $p_1(\underline{r})$ is defined as

$$E_\alpha(p_0, p_1) = \int p_0^\alpha(\underline{r})\, p_1^{1-\alpha}(\underline{r})\, d\underline{r}\ . \tag{15}$$

For iid observations, an upper bound on the Bayesian probability of error for any $\alpha \in [0, 1]$ is given by[10]

$$P_n(e) \leq \pi_0^\alpha \pi_1^{1-\alpha} (E_\alpha(p_0, p_1))^n. \tag{16}$$

The Chernoff entropy is defined as

$$E^*(p_0, p_1) = \inf_{0 \leq \alpha \leq 1} E_\alpha(p_0, p_1)\ , \tag{17}$$

and α^* is defined as the minimizing value of α. The CI is defined as

$$C(p_0, p_1) = -\log_2 E^*(p_0, p_1)\ . \tag{18}$$

The best achievable asymptotic error exponent among all tests is given by[10]

$$\lim_{n \to \infty} \frac{1}{n} \log_2 P_n(e) = \log_2 E^*(p_0, p_1)\ . \tag{19}$$

The CI also determines a bound on the probability of error for a finite number of samples. From (17) and (18), the least upper bound of the form given in (16) is given by

$$P_n(e) \leq \pi_0^{\alpha^*} \pi_1^{1-\alpha^*} 2^{-nC(p_0, p_1)}. \tag{20}$$

B. Bhattacharyya Bounds

The Bhattacharyya bound is of the form (16) with $\alpha = \frac{1}{2}$. For iid observations, the Bhattacharyya bound on the Bayesian probability of error is given by[8]

$$P_n(e) \leq \sqrt{\pi_0 \pi_1} 2^{-nB(p_0,p_1)} \,, \tag{21}$$

where $B(p_0, p_1)$ is the BD, defined as

$$B(p_0, p_1) = -\log_2 \left\{ \int \sqrt{p_0(\underline{r}) p_1(\underline{r})} \, d\underline{r} \right\} . \tag{22}$$

An asymptotic version of (21) is given by

$$\lim_{n \to \infty} -\frac{1}{n} \log_2 P_n(e) \geq B(p_0, p_1) \,. \tag{23}$$

The expression in (23) is a lower bound on the CI.

For finite n and iid observations, a lower bound on the Bayesian probability of error is given by[8]

$$P_n(e) \geq \pi_0 \pi_1 2^{-n 2 B(p_0,p_1)} \,, \tag{24}$$

which admits an asymptotic version as

$$\lim_{n \to \infty} -\frac{1}{n} \log_2 P_n(e) \leq 2B(p_0, p_1) \,, \tag{25}$$

which defines an upper bound on the CI.

While the Bhattacharyya bounds are not as tight as the Chernoff bounds, the Bhattacharyya bounds are easier to calculate because the CI calculation involves the optimization problem implied by (17). For both small and large values of the SNR, the BD can be used to construct a good estimate of the CI. For small values of the SNR, the distributions under both hypotheses approach the same noise-only distribution, the alpha entropy in (15) tends to 1 for any α, and both the CI and BD approach zero. For large values of the SNR, the CI for the PSK modulation classification problems will approach twice the BD. Consider again the noiseless case. Under H_0, the probability distribution is uniform on the set of the M_0 possible symbols under H_0, while under H_1 the probability distribution is uniform on the set of the M_1 possible symbols under H_1. Let M_a be the cardinality of the intersection of the signal constellations. From (22), the BD in this case is given by

$$B_0 = -\log_2\left(M_a\sqrt{\frac{1}{M_0 M_1}}\right), \tag{26}$$

and the alpha entropy, using (15), is given by

$$E_\alpha = \frac{M_a}{M_1}\left(\frac{M_1}{M_0}\right)^\alpha . \tag{27}$$

Minimizing this over α gives a CI of

$$C_0 = -\log_2\left(\frac{M_a}{M_1}\right) . \tag{28}$$

Suppose $M_a = M_0$, so that all of the constellation points under H_0 are also possible under H_1. This is the case for the BPSK vs. QPSK, QPSK vs. 8PSK and BPSK vs. 8PSK modulation classification problems. Then,

$$\left(\frac{M_a}{\sqrt{M_0 M_1}}\right)^2 = \frac{M_a}{M_1} \tag{29}$$

$$\Rightarrow -\log_2\left(\frac{M_a}{M_1}\right) = -2\log_2\left(\frac{M_a}{\sqrt{M_0 M_1}}\right) \tag{30}$$

$$\Rightarrow C_0 = 2B_0 . \tag{31}$$

Thus, for large values of the SNR, the approximation of the CI by twice the BD will be a reasonable approximation for the PSK modulation classification problems under consideration.

Note that the Chernoff bound (20), Bhattacharyya upper bound (21) and Bhattacharyya lower bound (24) are all exponential bounds on the probability of error. Thus the bounds will appear linear on a log scale for a plot of the probability of error against the number of observations, with slopes given by the negatives of the error exponents $C(p_0, p_1)$, $B(p_0, p_1)$ and $2B(p_0, p_1)$. The CI has the additional property that it is the negative of the limiting slope on the tail of the probability of error as a function of the number of observations, which makes it a more useful performance measure than the BD.

C. Kullback-Liebler Bounds

The performance of a binary hypothesis test is characterized by the type I and type II error probabilities, given by $\alpha_n = P_n(e|H_0)$ and $\beta_n = P_n(e|H_1)$. Under the Neyman-Pearson criterion, β_n is minimized under the constraint that $\alpha_n < \epsilon$ for some fixed probability ϵ. For iid observations, Stein's lemma[14] states that

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log_2 \Big\{ \min_{D:\alpha_n < \epsilon} \beta_n \Big\} = -KL(p_0||p_1)\,, \tag{32}$$

where the minimization is over all decision functions D with type I error probability less than ϵ, and $KL(p_0||p_1)$ is the KLD, given by

$$KL(p_0||p_1) = \int p_0(\underline{r}) \log_2 (\frac{p_0(\underline{r})}{p_1(\underline{r})}) d\underline{r}\,. \tag{33}$$

The distance $KL(p_0||p_1)$ is an upper bound on the (base 2) asymptotic error exponent for the type II error probability if the type I error probability is allowed to remain constant as the number of samples goes to infinity. The type I error exponent is similarly bounded by $KL(p_1||p_0)$, which is not equal to $KL(p_0||p_1)$ in general.

D. Calculations for PSK Modulation Classification

Numerical integration was used to calculate the CI, BD and KLDs associated with classifiers for the following phase-shift keying modulations: BPSK, QPSK and 8PSK.

The CI values for the three binary tests (BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK) are plotted as functions of the SNR in Fig. 1. It is seen that the CI for the BPSK vs. 8PSK test is uniformly higher than the CI for the BPSK vs. QPSK and QPSK vs. 8PSK tests, while the CI for the QPSK vs. 8PSK test is uniformly lower than the CI for the BPSK vs. QPSK and BPSK vs. 8PSK tests. The CI for the BPSK vs. QPSK test is very close to the CI for the BPSK vs. 8PSK test for small values of the SNR, while it is very close to the CI of the QPSK vs. 8PSK test for large values of the SNR. In Section III, the error probability expressions in the noiseless case for the three pairs of PSK modulation classification problems were given in (12), (13) and (14). Assuming the a priori probabilities are nonzero, for large enough n, the probability of error for for both the BPSK vs. QPSK and QPSK vs. 8PSK tests are proportional to 2^{-n}, corresponding to an error exponent of one (using base 2). This is in agreement with Fig. 1, where it can be seen that the CI curves for BPSK vs. QPSK and QPSK vs. 8PSK are both asymptotic to one bit, i.e. for large values of the SNR we indeed approach the noiseless bound. For the BPSK vs. 8PSK test with nonzero a priori probabilities and sufficiently large n, the noiseless case error probability is proportional to 4^{-n}, corresponding to an error exponent of 2 (using base 2). This is verified by Fig. 1, in which the CI for this test approaches 2 bits as the SNR gets large. For an M-ary hypothesis test, the pair of hypotheses with the lowest CI will determine the asymptotic performance[16]. From Fig. 1, it is seen that the CI curve for the QPSK vs. 8PSK test is uniformly lower than the curves for the other two pairs. Thus, the logarithmic rate to zero of the misclassification probability for the 3-ary test among BPSK, QPSK and 8PSK is determined asymptotically by the misclassification probability in deciding between QPSK and 8PSK.

The CI, BD and KLDs for the modulation classification problems BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK are shown in Figs. 2, 3, and 4, respectively. The CI is shown in all figures to be bounded from below by the BD and bounded from above by twice the BD. The CI is close to the BD for small values of the SNR (indicating that $\alpha^* \approx \frac{1}{2}$ in this range), and close to twice the BD for large values of the SNR. Interestingly, the CI is very close to the BD for SNRs of 5 dB or less for the BPSK vs. QPSK and BPSK vs. 8PSK tests, and for SNRs of 10 dB or less for the QPSK vs. 8PSK test. It is shown that the KLDs $KL(BPSK||QPSK)$ and $KL(QPSK||8PSK)$ approach the value of one bit for large values of the SNR, while the KLD $KL(BPSK||8PSK)$ approaches the value of 2 bits for large values of the SNR.

In the following discussion, let H_0 be the hypothesis associated with the smaller constellation, and H_1 be the hypothesis associated with the larger constellation. From (32), these quantities are associated with the logarithmic rate to zero of the error probability given that hypothesis H_1 is true. This reflects the fact that, for large SNR, under hypothesis H_1, the probability of a transmitted sequence which is possible under hypothesis H_0 is on the order of 2^{-n} for the BPSK vs. QPSK and QPSK vs. 8PSK tests, and 4^{-n} for the BPSK vs. 8PSK test. These "ambiguous sequences" cause errors for sufficiently large SNR, when the probability of error given hypothesis H_0 is constrained to be less than ϵ, for ϵ sufficiently small. The KLDs $KL(QPSK||BPSK)$, $KL(8PSK||QPSK)$ and $KL(8PSK||BPSK)$ increase without bound as the SNR increases. These quantities are associated with the logarithmic rate to zero of the error probability given H_0. This result is reflecting the fact that, in the noiseless case, for large enough n, no error is made under hypothesis H_0.

V. Simulation Study

Simulations were run to demonstrate the performance of PSK classifiers as a function of the number of observations. In all cases, a sufficient number of simulation trials was employed to yield graphical accuracy. The required number of trials is over one hundred times the inverse of the error probability. Classifiers were simulated for the BPSK vs. QPSK, QPSK vs. 8PSK and BPSK vs. 8PSK problems. The results are then interpreted to ascertain the practical implications of the CI, BD and KLDs for these problems. Bayes tests, with assumed equal a priori probabilities, are shown for all pairs of distributions from the set BPSK, QPSK and 8PSK. Neyman-Pearson tests are shown for the BPSK vs. QPSK classifier.

The noiseless bound is evident in Fig. 5, which shows the probability of misclassification as a function of $\frac{E_s}{N_0}$ after one sample has been observed and after ten samples have been observed, for PSK classifiers using a Bayes criterion. The figure shows that the error probability for each classifier is subject to a nonzero error floor which is approached for large values of the SNR. For the same value of the SNR, the QPSK vs. 8PSK classifier has a higher misclassification probability than the BPSK vs. QPSK classifier, due to the fact that the constellation points are much closer together in the former test than in the latter, although the error probabilities approach the same error floor for large values of the SNR.

The performances of the Bayes classifiers and the associated Chernoff bound (20), Bhattacharyya upper bound (21) and Bhattacharyya lower bound (24) are shown in Fig. 6, as functions of the number of observations, for an SNR of 0 dB. From Fig. 1, at 0 dB, the CI values for the BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK tests are $.075$, $.074$ and $.0023$ bits, respectively. The absolute values of the slopes of least squares fit lines to the tails of the data (the last 50 samples shown) for the BPSK vs. QPSK and BPSK vs. 8PSK classifiers are $.079$ and $.076$ bits respectively, while that for the QPSK vs. 8PSK test is $.0027$ bits, using data for the 50 samples prior to the $1000th$ observation. These experimental values are close to the CI values, demonstrating the practicality of the CI for these classification problems at 0 dB. The values are larger than the CI because the slope of the line approaches the negative of the CI asymptotically in the number of samples, and these calculations are for relatively low numbers of samples. From Figs. 2, 3 and 4 it can be seen that, for each classifier at 0 dB, the CI is very close to the BD, and indeed, from Fig. 6, the Bhattacharyya upper bounds (21) are close to the Chernoff bounds (20) for all three classifiers. This verifies our earlier statement that, for small values of the SNR, the Bhattacharyya bound is a good approximation to the Chernoff bound. The Bhattacharyya lower bounds are seen to be not very tight at 0 dB, particularly for the BPSK vs. QPSK and BPSK vs. 8PSK classifiers. It can be seen from Figs. 2, 3, and 4 that twice the BD (which determines the lower bound (24)) is higher than the CI at small values of the SNR, where the Bhattacharyya lower bound is not a particularly useful bound.

In Fig. 7, the performances of the Bayes classifiers and the associated Chernoff and Bhattacharyya upper bounds are shown, as functions of the number of observations, for an SNR of 15 dB. At this SNR, the difference in the performance between the BPSK vs. QPSK and BPSK vs. 8PSK classifiers is more evident than at 0 dB. From Fig. 1, at 15 dB, the CI values are $.90$, 1.5, and $.68$ bits for the BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK tests, respectively. A least squares line was fit to the last 10 data points for each of the classifier performance curves in Fig. 7. The absolute values of the slopes of the lines were 1.0, 1.6 and $.69$ bits, for the BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK classifiers, respectively. As for the 0 dB simulation, we note that for 15 dB, these values are higher than, but close to the CI values, verifying our claim that the CI is a practical performance metric for these tests, although only 10 samples were used for the line. The Bhattacharyya distances at an SNR of 15 dB, from Figs. 2, 3 and 4, are $.50$, $.99$ and $.49$ bits for the BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK classifiers respectively. Neither the set of Bhattacharyya distances (which determine the upper bounds (21)) nor the set of twice the Bhattacharyya distances (which determine the lower bounds) are as good a match with the simulated least squares slopes as the set of CIs for all three classifiers, although twice the Bhattacharyya distance is close for the BPSK vs. QPSK test. This is because the curve has not yet become linear in the simulated range; with increasing numbers of samples the value will approach the CI. This is reflected in Figs. 7and 8, where the slopes of the Chernoff bounds are much closer to the simulated data that those of the Bhattacharyya bounds, considered across all three tests. The Bhattacharyya lower bounds for the modulation classification problems at 15 dB are shown in Fig. 8. From Figs. 6 and 7, it is seen that the Bhattacharyya lower bounds are significantly tighter for 15 dB than for 0 dB. This is expected as, from Figs. 2, 3 and 4, twice the BD is close to the CI at 15 dB for all three tests. At high SNR the Bhattacharyya lower bound approaches the Chernoff bound. The small discrepancy between the CI and twice the BD means that the Chernoff bound and the Bhattacharyya lower bound will have similar slopes when plotted on a log scale.

The performance of a BPSK vs. QPSK modulation classifier designed under a Neyman-Pearson criterion is shown in Figs. 9 and 10, at an SNR of 0 dB. Defining hypothesis H_0 as the BPSK hypothesis and hypothesis H_1 as the QPSK hypothesis, the criterion in Fig. 9 was to minimize the type II error probability $\beta_n = P(e|H_0)$ subject to the constraint that the type I error probability $\alpha_n = P(e|H_1) \leq \epsilon$, for fixed error probabilities $\epsilon = .05, .01$ and $.001$. In Fig. 10, the type I error probability α_n is minimized subject to the type II error probability constraint $\beta_n \leq \epsilon$, for the same values of ϵ. For a sequence of decreasing error probability constraints, we expect to see that the associated sequence of limiting slopes for the type II error probability curves is decreasing monotonically in magnitude to the KLD KL(BPSK|QPSK), from (32). For Figs. 9 and 10, The magnitudes of the slopes at the tails of the plot of the non-constant error probability curves were calculated. For Fig. 9, For the sequence of constraints $\alpha \leq .05$, $\alpha \leq .01$ and $\alpha \leq .001$, the absolute values of the slopes at the tails shown are given by $.20$, $.19$ and $.16$ bits, respectively. While, as expected, the magnitude of the slope decreases with decreasing ϵ, the sequence of simulated slopes does not seem to be approaching the KL distance KL(BPSK|QPSK) of $.25$ bits. This is because the absolute values of the slopes given above are clearly not the limiting values for the curves shown in Fig. 9. For the probability of error range shown, the curves are not particularly linear, and it is expected that the magnitude of the slope at the tail of these curves will be higher than those calculated for the small number of observations shown. Insufficient numbers of samples have been observed for a measurement of the limiting slope on the probability of error curves. In Fig. 10, as ϵ decreases, the magnitude of the limiting slope should tend to the KLD KL(QPSK|BPSK), which, from Fig. 2, is $.37$ bits at 0 dB. The magnitudes of the slopes at the tails of the plot of the type I error probability curves were calculated as $.29$, $.23$ and $.20$ bits, for the constraints $\beta \leq .05$, $\beta_n \leq .01$ and $\beta_n \leq .001$, respectively. As before, for the probability of error range shown, the curves are not linear and the limiting slopes will be higher in magnitude. For SNR of 0 dB, it does not appear that the Kullback-Liebler bounds are effective in performance determination for these problems.

VI. Conclusions

Consideration of modulation classification in a noiseless setting has provided a bound for the error probability in classifying two-dimensional constellations under noisy conditions. This bound shows that for a finite number of samples, the error probability is bounded from below by a nonzero error floor as the SNR goes to infinity, assuming that the constellations under different hypotheses have a nonempty intersection. The CI, BD and KLDs were calculated for phase-shift keying modulation classification problems. Modulation classifiers derived under both Bayes and Neyman-Pearson criteria were simulated. The results showed that, for Bayes tests, the CI was a more suitable performance measure than the BD. The CI provides a tighter upper bound on the error probability than the BD, and it yields estimates of the logarithmic rate to zero of the error probability curve, which were shown to be close even for relatively small numbers of observations, at both small and large values of the SNR. The BD provides useful upper and lower bounds on the error probability, is easier to calculate than the CI, and can be used to construct a good estimate of the CI at both small and large values of the SNR. The KLD provides a lower bound to the achievable logarithmic rate to zero of the type II error probability given a fixed constraint on the type I error probability, in the limit as the constraint goes to zero. For the classification problems studied, the error probability curves did not become linear fast enough to enable a determination of the limiting slope and meaningful comparison with the KLD.

REFERENCES

[1] A. Polydoros and K. Kim, "On the detection and classification of quadrature digital modulations in broadband noise," *IEEE Trans. Comm.*, vol. 38, pp. 1199–1211, Aug. 1990.

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

[3] S. Soliman and S.-Z. Hsue, "Signal classification using statistical moments," *IEEE Trans. Comm.*, vol. 40, no. 5, pp. 908–916, May 1992.

[4] P. C. Sapiano and J. D. Martin, "Maximum likelihood PSK classification using the DFT of phase histogram," *Proc. IEEE GLOBECOM '95*, vol. 2, pp. 1029–1033, Nov. 1995.

[5] Y. Yang and S. Soliman, "Optimum classifier for M-ary PSK signals," *Proc. IEEE Int. Conf. Commun. ICC'91*, vol. 3, pp. 1693–1697, June 1991.

[6] Y. Yang and C.-H. Liu, "An asymptotic optimal algorithm for modulation classification," *IEEE Communications Letters*, vol. 2, no. 5, pp. 117–119, May 1998.

[7] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Comm.*, vol. 48, pp. 189–193, Feb. 2000.

[8] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Comm.*, vol. 15, no. 1, pp. 52–60, Feb. 1967.

[9] G. T. Toussaint, "An upper bound on the probability of misclassification in terms of the affinity," *Proc. IEEE*, vol. 65, pp. 275–276, Feb. 1977.

[10] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Ann. Math. Statist.*, vol. 23, pp. 493–507, 1952.

[11] D. Kazakos, "Asymptotic error probability expressions for multihypothesis testing using multisensor data," *IEEE Trans. Systems, Man and Cybernetics*, vol. 21, no. 5, pp. 1101–1114, Sept.-Oct. 1991.

[12] M. E. Hellman and J. Raviv, "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. Info. Theory*, vol. 16, no. 4, pp. 368–372, July 1970.

[13] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Royal Stat. Soc.*, vol. 28, pp. 131–142, 1966.

[14] J. A. Bucklew, *Large Deviations Techniques in Decision, Simulation and Estimation*. New York: John Wiley and Sons, 1990.

[15] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston: Jones and Bartlett, 1993.

[16] C. C. Leang and D. H. Johnson, "On the asymptotics of M-hypothesis Bayesian detection," *IEEE Trans. Info. Theory*, vol. 43, no. 1, pp. 280–282, Jan. 1997.

[17] G. T. Toussaint, "Comments on "The divergence and Bhattacharyya distance measures in signal selection"," *IEEE Trans. Comm.*, vol. 20, no. 3, p. 485, June 1972.

[18] D. G. Lainiotis, "A class of upper bounds on probability of error for multihypothesis pattern recognition," *IEEE Trans. Info. Theory*, vol. 15, pp. 730–731, Nov. 1969.

[19] F. D. Garber and A. Djouadi, "Bounds on the Bayes classification error based on pairwise risk functions," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 281–288, Mar. 1988.

[20] A. Rényi, "On the amount of missing information and the Neyman-Pearson lemma," *Magyar Tud. Akad. Mat. Kutató Int. Köyl*, vol. 9, pp. 617–625, 1965.

[21] H. L. Van Trees, *Detection, Estimation and Modulation Theory, Part I*. New York: John Wiley and Sons, 1968.

[22] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.

[23] E. Davis, "Large deviations rate functions for decision processes with applications to modulation classification," Ph.D. dissertation, Queen's University, in progress.

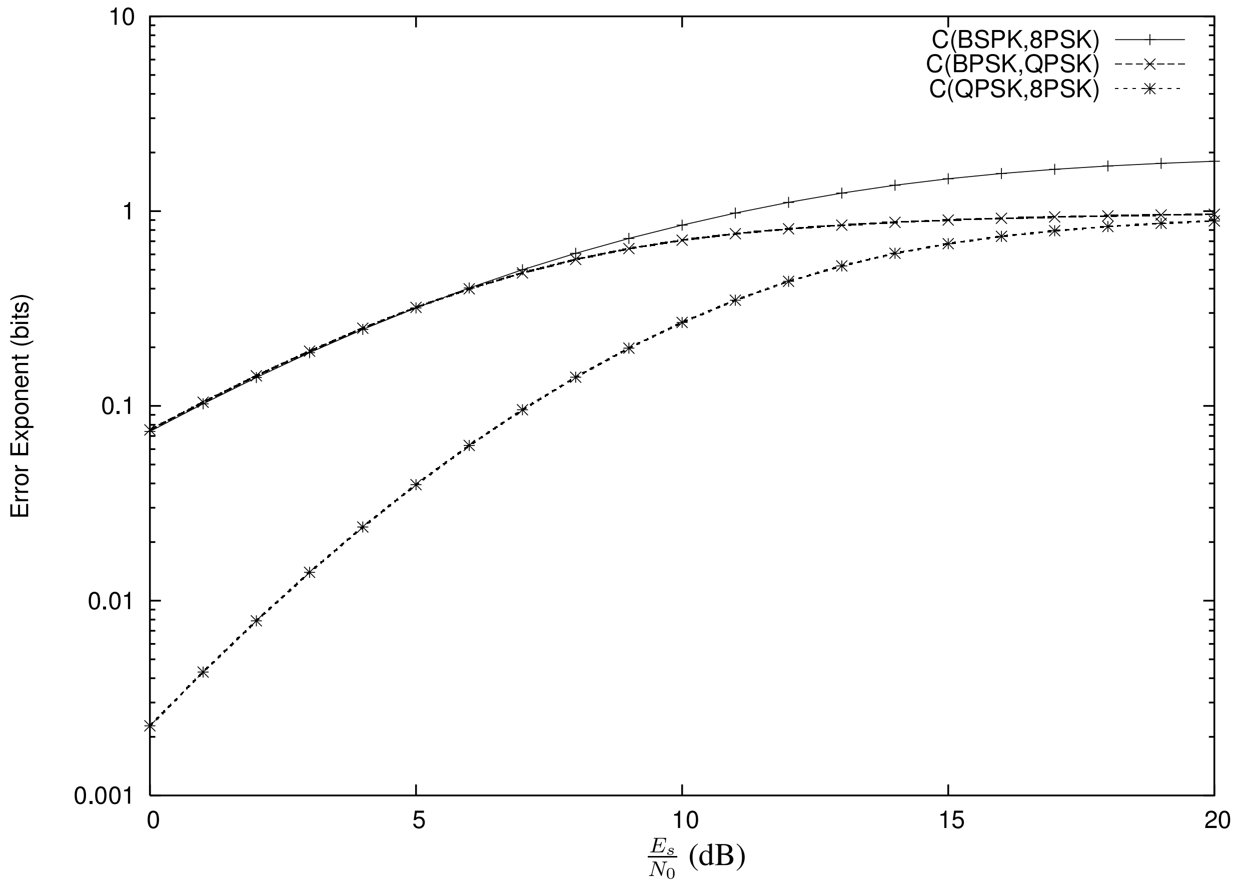

Fig. 1. CI vs. $\frac{E_s}{N_0}$ for BPSK vs. QPSK, BPSK vs. 8PSK and QPSK vs. 8PSK modulation classification problems.

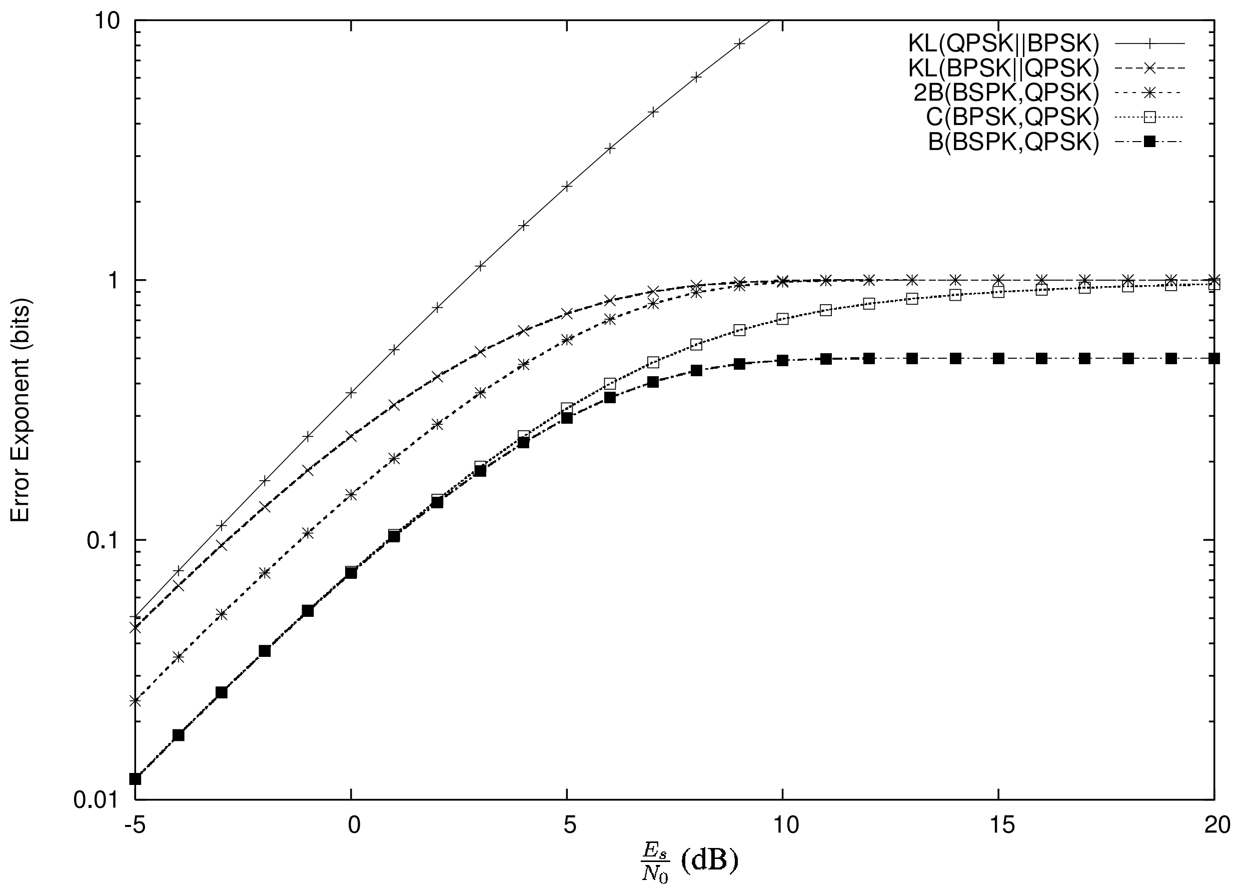

Fig. 2. KLDs in both directions, Bhattacharyya upper and lower bounds, and Chernoff information vs. $\frac{E_s}{N_0}$ for BPSK vs. QPSK modulation classification problems.

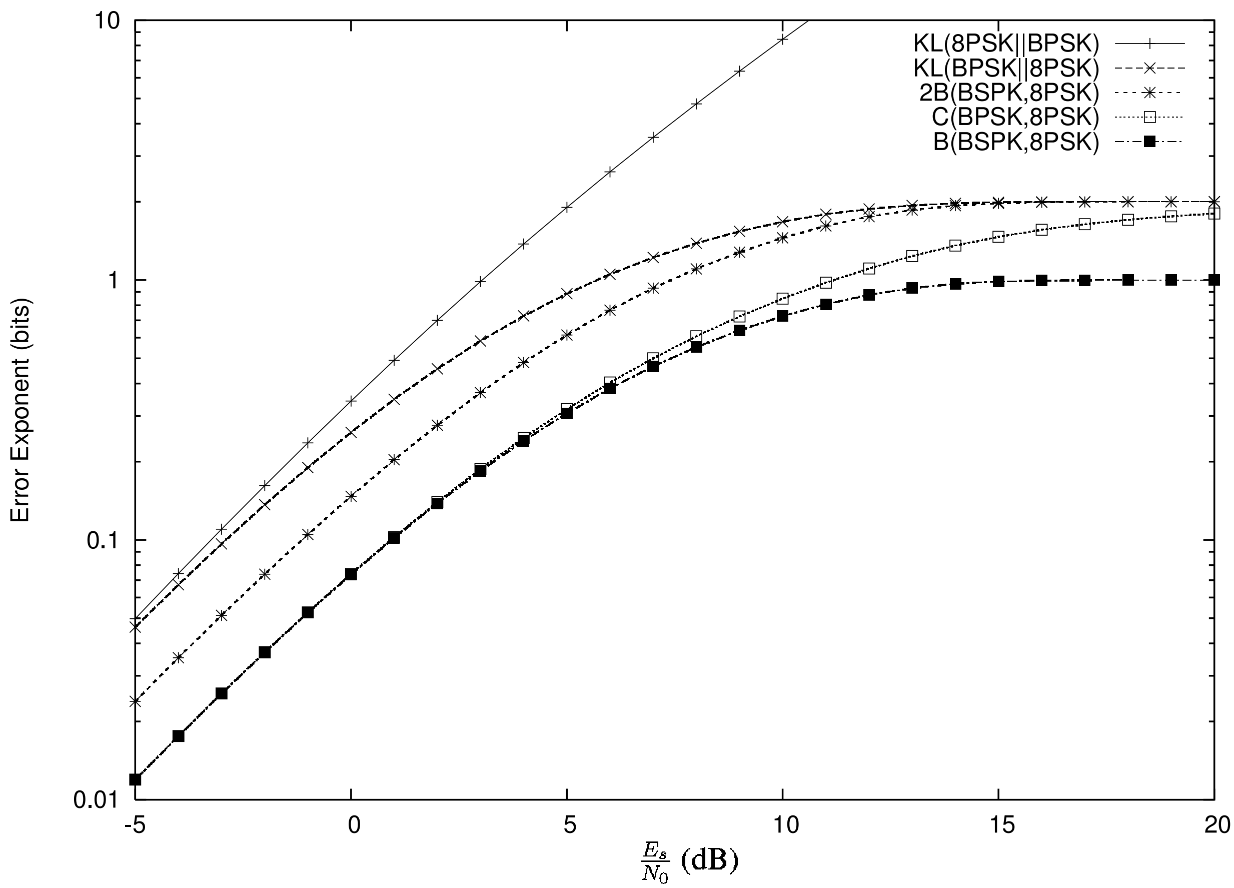

Fig. 3. KLDs in both directions, Bhattacharyya upper and lower bounds, and Chernoff information vs. $\frac{E_s}{N_0}$ for BPSK vs. 8PSK modulation classification problems.

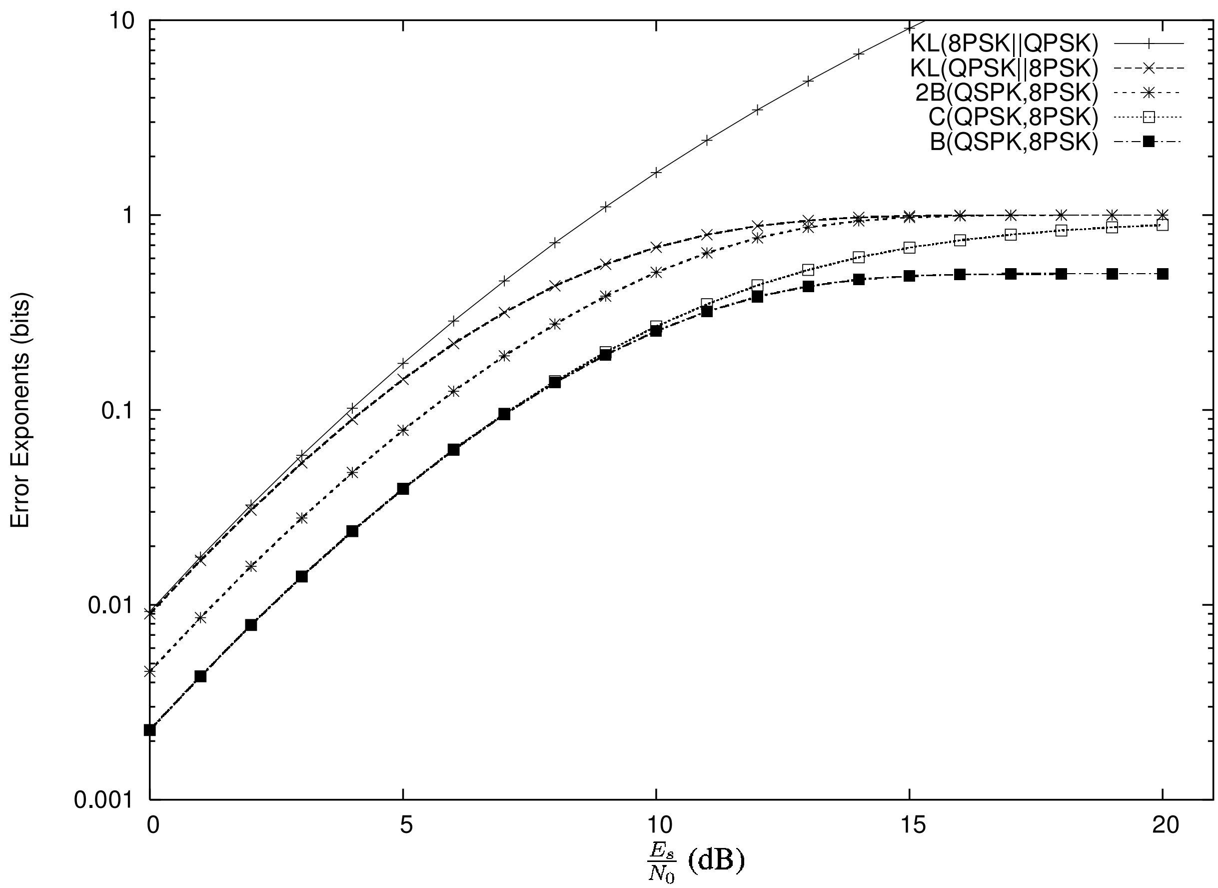

Fig. 4. KLDs in both directions, Bhattacharyya upper and lower bounds, and Chernoff information vs. $\frac{E_s}{N_0}$ for QPSK vs. 8PSK modulation classification problems.

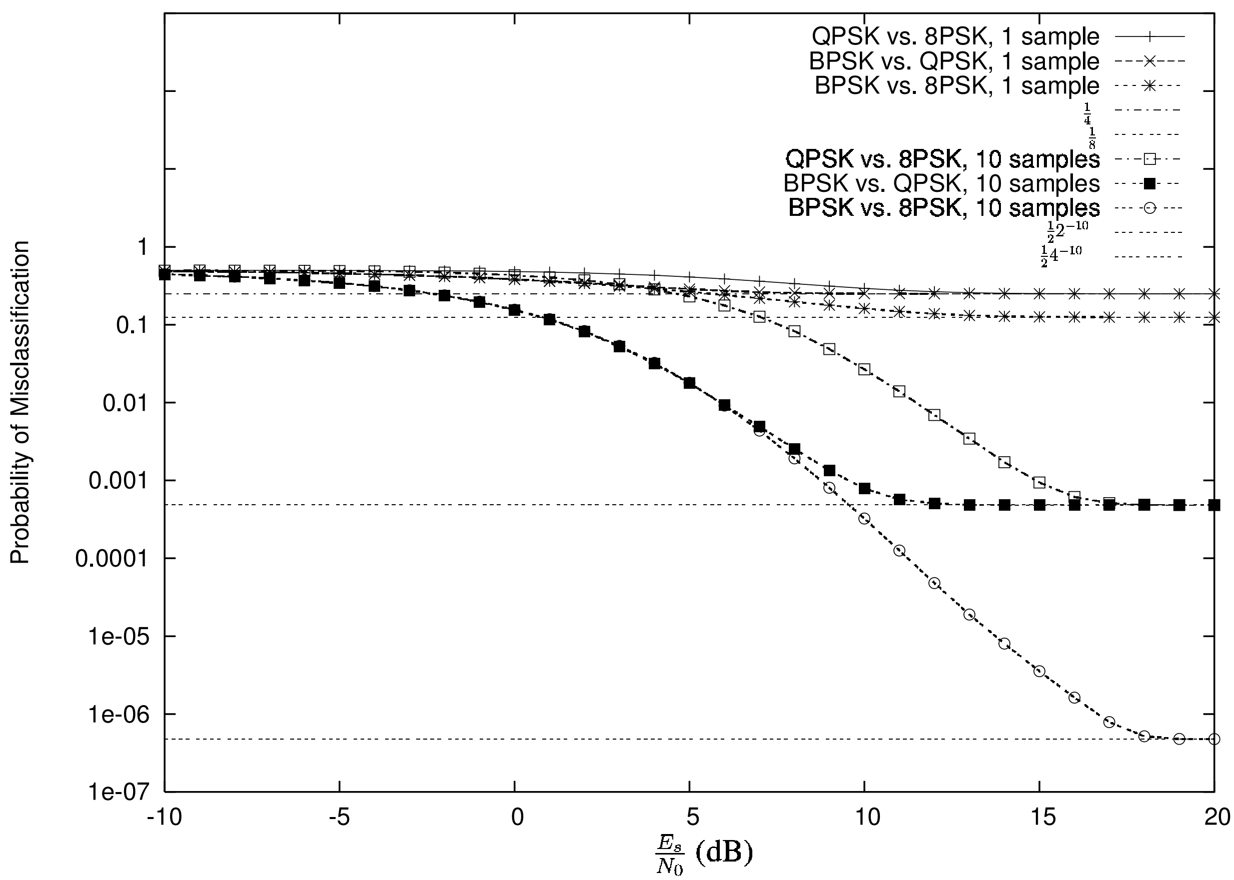

Fig. 5. The probability of misclassification vs. $\frac{E_s}{N_0}$ for Bayes tests for PSK modulation classification problems, after one observation and after 10 observations.

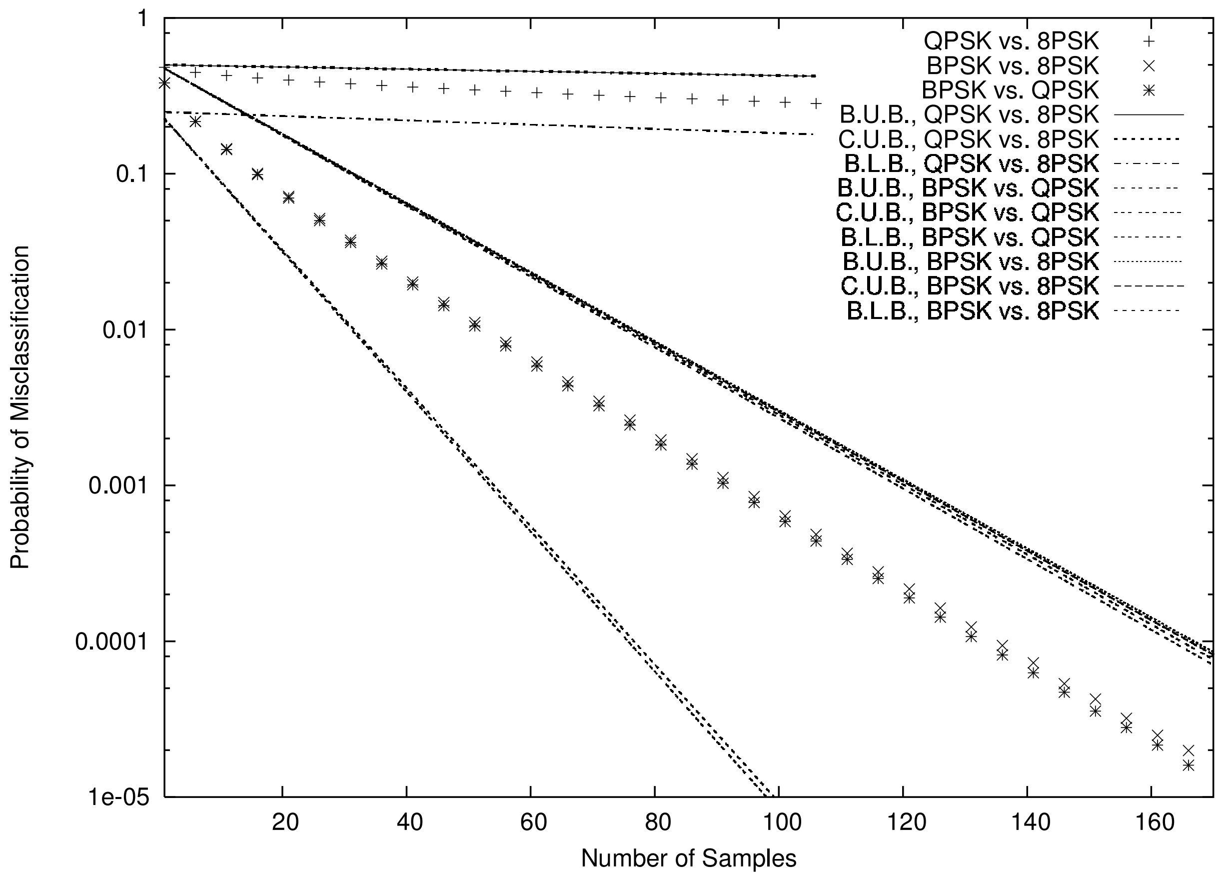

Fig. 6. The probability of misclassification, Chernoff upper bounds (C.U.B.), Bhattacharyya upper bounds (B.U.B.) and Bhattacharyya lower bounds (B.L.B.) vs. number of samples for PSK modulation classification problems using a Bayes criterion, at $\frac{E_s}{N_0} = 0$ dB.

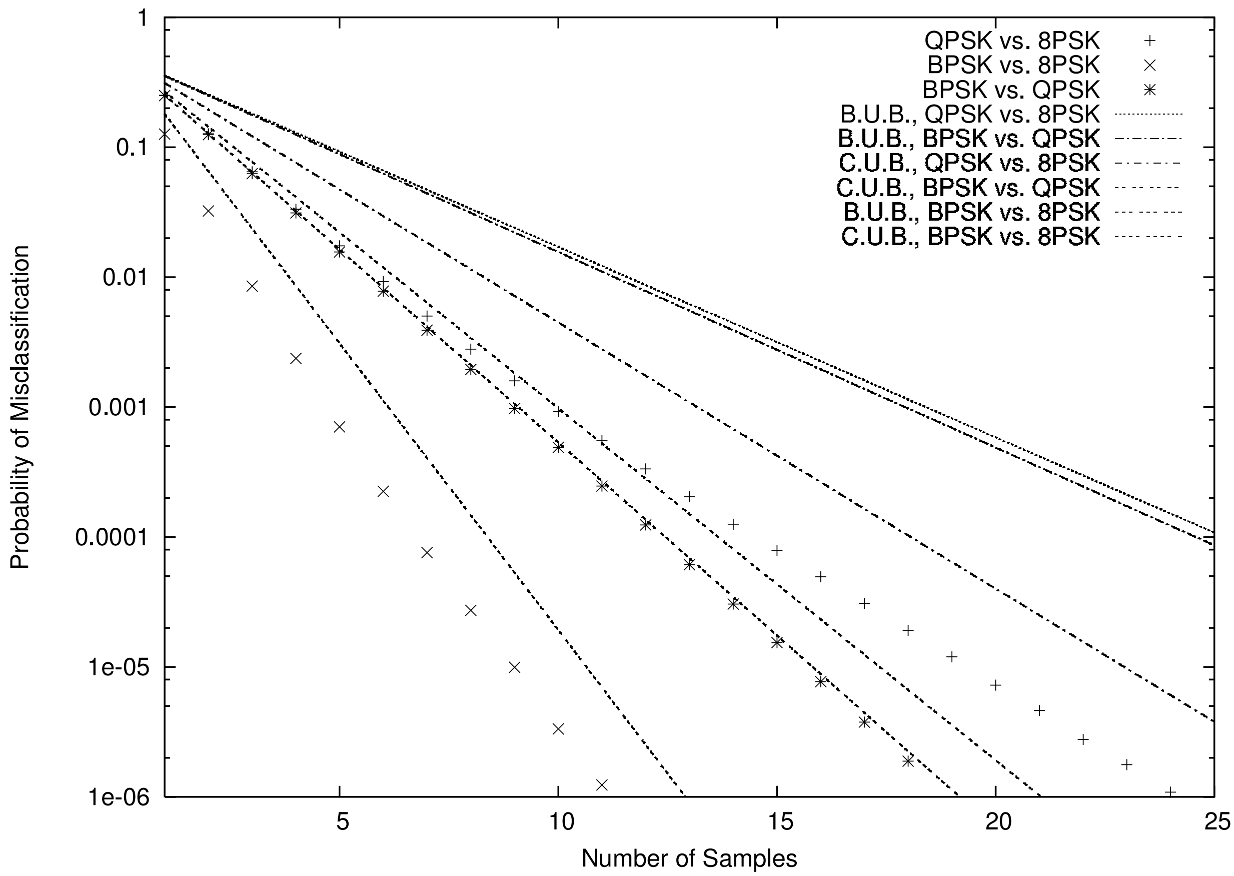

Fig. 7. The probability of misclassification, Chernoff upper bounds (C.U.B.) and Bhattacharyya upper bounds (B.U.B.) vs. number of samples for PSK modulation classification problems using a Bayes criterion, at $\frac{E_s}{N_0} = 15$ dB.

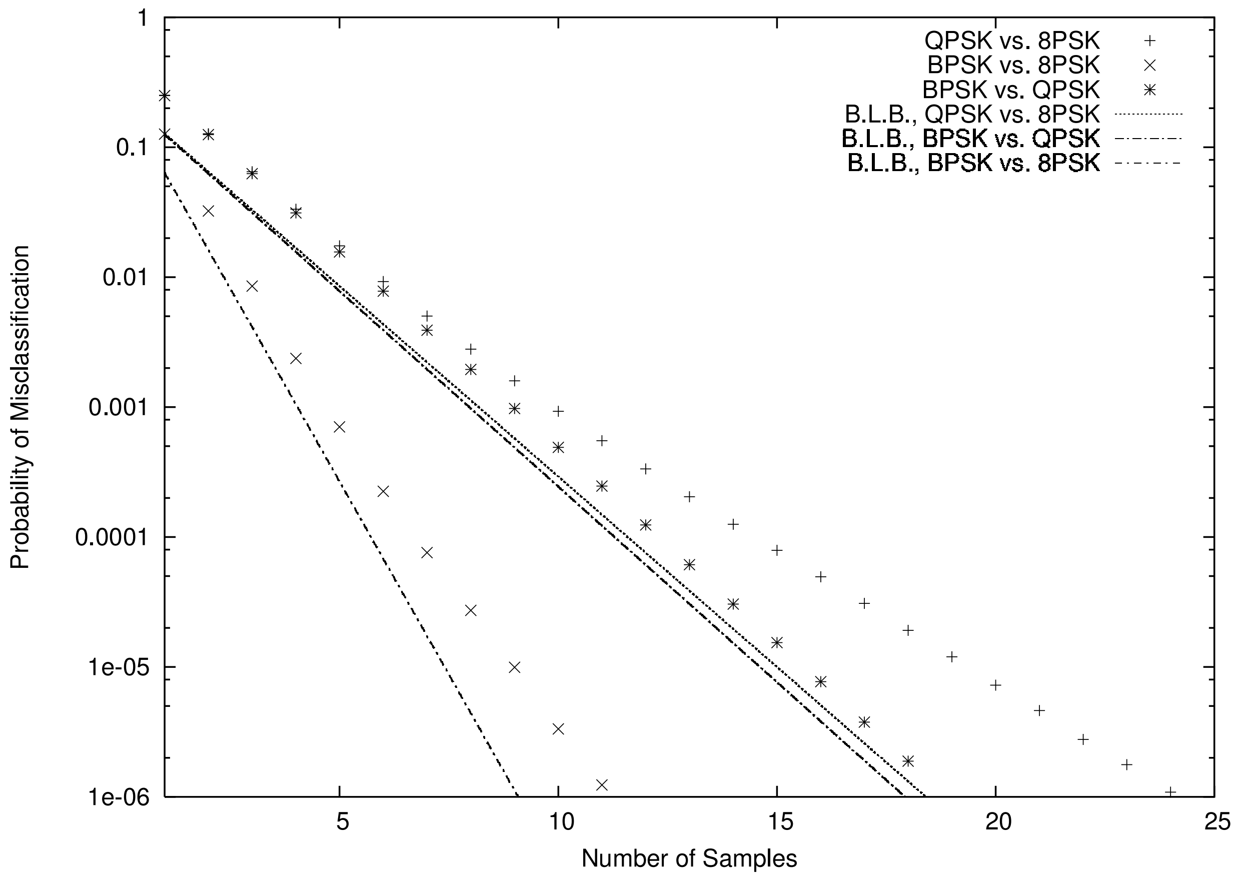

Fig. 8. The probability of misclassification and Bhattacharyya lower bounds (B.L.B.) vs. number of samples for PSK modulation classification problems using a Bayes criterion, at $\frac{E_s}{N_0} = 15$ dB.

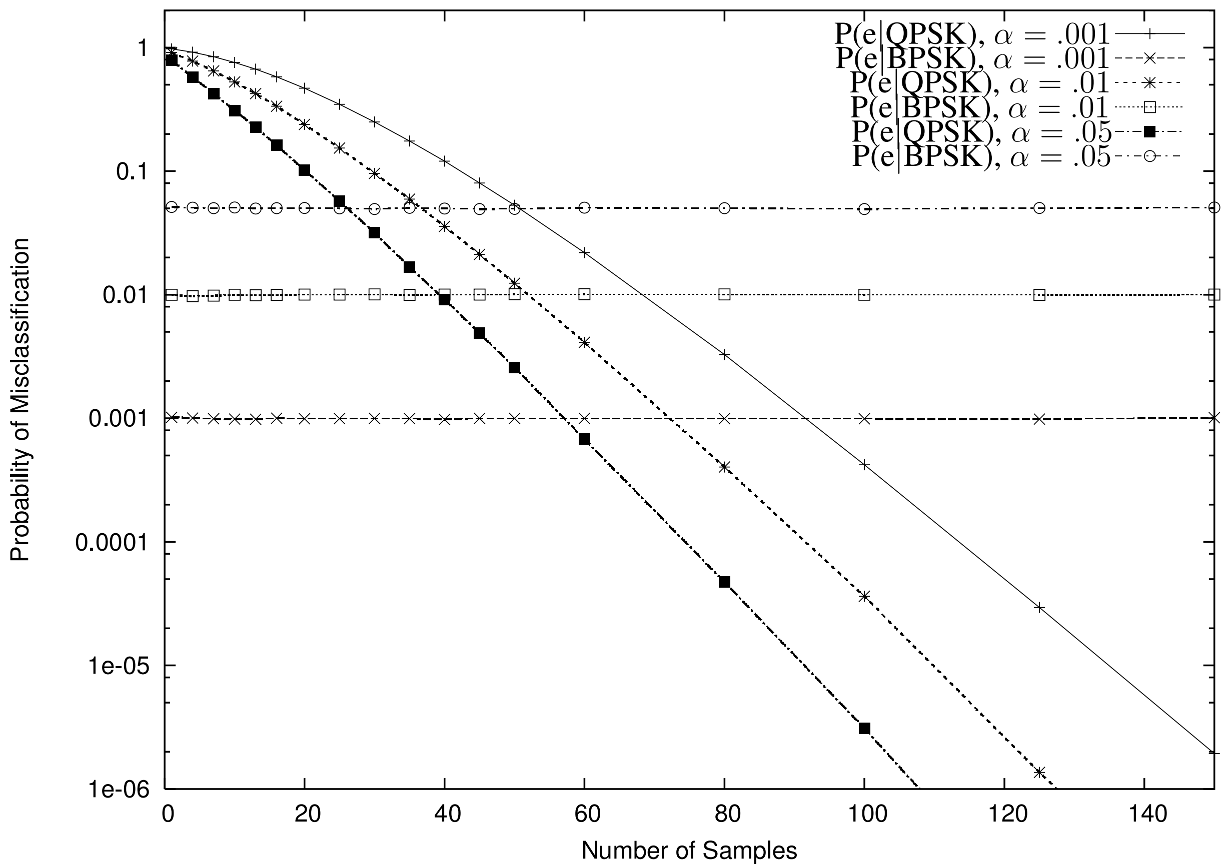

Fig. 9. The probability of misclassification vs. number of samples for BPSK vs. QPSK Neyman-Pearson tests, with fixed error probability constraints on the type I error probability, P(e|BPSK).

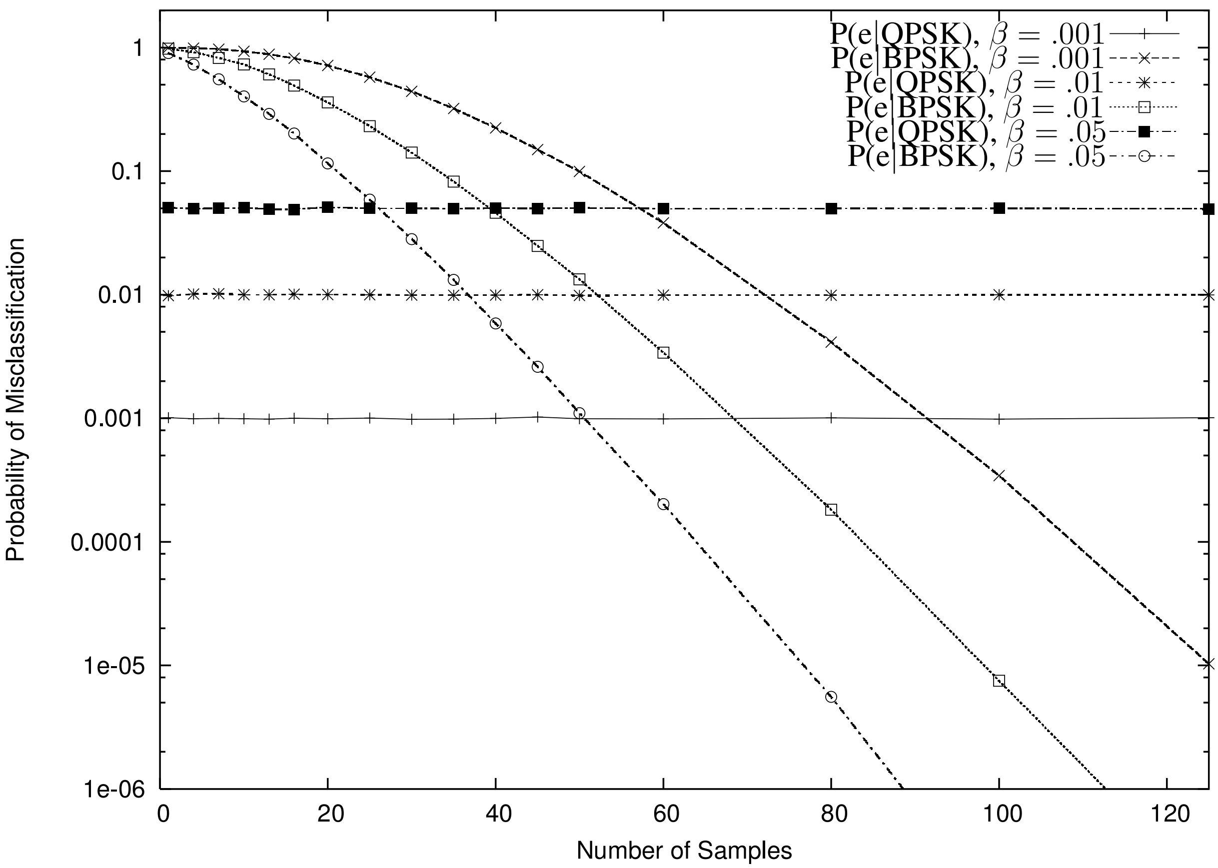

Fig. 10. The probability of misclassification vs. number of samples for BPSK vs. QPSK Neyman-Pearson tests, with fixed error probability constraints on the type II error probability, P(e|QPSK).