ELSEVIER

Signal Processing 60 (1997) 263–280

SIGNAL PROCESSING

Classification of quadrature amplitude modulated (QAM) signals via sequential probability ratio test (SPRT)[1]

Yu-Chuan Lin[a], C.-C. Jay Kuo[b,*]

[a] *Qualcomm Inc., San Diego, CA 92121, USA*

[b] *Integrated Media Systems Center and the Department of Electrical Engineering Systems, University of Southern California, MC 2564 Los Angeles, CA 90089-2564, USA*

Received 21 August 1996; revised 24 March 1997

Abstract

A novel approach to automatic classification of quadrature amplitude modulated (QAM) signals is presented in this research. Modulation classification has been traditionally treated as a hypothesis test problem with input signals of a fixed sample size. By formulating it as a variable sample size test problem, we propose a new classification algorithm based on the sequential probability ratio test (SPRT). It is demonstrated that the new approach has several important merits, including ease of error rate control, lower computational complexity and lower decision delay. © 1997 Published by Elsevier Science B.V.

Zusammenfassung

In diesem Artikel wird ein neuer Ansatz zur automatischen Klassifizierung eines Quadratur-Amplituden-Modulierten (QAM) Signals vorgestellt. Bisher wurde das Klassifizierungsproblem modulierter Signale als Hypothesentest mit vorgegebener Datenlänge behandelt. Wir formulieren das Problem nun mit variabler Datenlänge und schlagen einen neuen Algorithmus zur Klassifizierung vor, der auf einem sequentiellen Wahrscheinlichkeitsquotiententest (SPRT) basiert. Es wird gezeigt, daß dieser neue Ansatz verschiedene wichtige Vorteile besitzt. Das sind z.B. eine einfache Fehlerratenkontrolle, ein geringerer Rechenaufwand und eine geringere Entscheidungsverzögerung.© 1997 Published by Elsevier Science B.V.

Résumé

Ce travail présente une nouvelle approche pour la classification automatique de signaux à modulation d'amplitude en quadrature (QAM). La classification de modulation est traitée traditionnellement comme un problème de test d'hypothèse avec des signaux d'entrée de longueur fixe. En le formulant comme un problème de test de signaux de longueur variable, on propose un nouvel algorithme de classification basé sur le test de rapport de probabilité séquentiel (SPRT). On montre que cette nouvelle approche a plusieurs avantages importants, dont la facilité de contrôle du taux d'erreur, une complexité algorithmique plus faible ainsi qu'un délai de décision plus faible. © 1997 Published by Elsevier Science B.V.

Keywords: Modulation classification; Sequential probability ratio test; Multihypothesis test

* Corresponding author. Tel.: 213 740 4658; fax: 213 740 4651; e-mail: cckuo@sipi.usc.edu.

[1] This work was supported in part by the Office of Research and Development (ORD) CRASP program and in part by the National Science Foundation Presidential Faculty Fellow Award ASC-9350309.

PII S0165-1684(97)00077-7

1. Introduction

Automatic modulation classification is an important research problem in the receiver design for non-cooperative communication systems. The modulation classifier usually serves as a preprocessing unit for monitoring or interception systems. When the modulation scheme of a received signal is recognized, an appropriate demodulator can be selected to recover the information. The automatic modulation classification technique can also be applied to the control of radio frequency bands. Such a control task has so far been done by human operators. However, it becomes more difficult due to higher signal densities over a fixed bandwidth in recent years.

Two approaches to obtain the statistics of extracted features have been reported in previous works. The first approach derives the exact statistical descriptions of received signals under certain given assumptions [1, 7, 8, 15, 19, 24, 28]. This approach is often motivated by the receiver design of cooperative communication systems. It gives more insights into the cause–effect relationship among the performance and various uncertainty factors. An exact analysis of feature statistics is possible by imposing the synchronization assumptions, and the optimality can be theoretically guaranteed. The challenging part of this approach is the derivation of test statistics based on a set of reasonable assumptions. The second approach obtains the test statistics empirically rather than analytically, i.e. to design the classifier by using training data. Classifiers using features such as the periodogram, the bispectrum and the histograms of local frequency or phase estimates, have been reported in [5, 10, 11, 18, 22]. This approach is more realistic in the sense that it requires less a priori knowledge of received signals. However, the corresponding theoretical performance bound is very difficult to analyze.

Modulation classification has been traditionally formulated as a hypothesis test problem with a fixed size of samples, i.e. making decisions based on a fixed amount of received data. Classifiers using the fixed-sample-size test (FSST), which often operate in a batch mode, are suitable for bursty or packet type of data. For other applications, such as the transmission of continuum-type data, it is more natural to formulate the problem as a sequential process since received data are collected in a temporal order and their amount can be quite large. In this context, a more appropriate optimality is the least amount of data required for decision making while the classification performance such as the error probability meets a certain requirement.

This work focuses on the classification of signals with the quadrature amplitude modulation (QAM), which forms an important family of digital modulation schemes. We formulate this modulation classification problem as a variable-sample-size test problem, and propose a new classification algorithm based on the sequential probability ratio test (SPRT). Compared with FSST, SPRT is known as the optimal test in the sense that it can provide the same decision error probability with the least amount of samples [27]. Since the decision delay and the computational complexity increase as the number of samples used to make decisions increases, it is more favorable to use classifiers that use less samples to make decisions with the same error probability. Furthermore, SPRT can control the individual error probability given that one of the hypotheses is true. For some applications, it is an attractive feature to have the same decision quality regardless of the true hypothesis. It is possible to design a SPRT-based classifier that gives the same individual error probability. In contrast, FSST usually leads to very low error probabilities with 'distinguishable' hypotheses and high error probabilities with 'similar' hypotheses.

It will be demonstrated by experiments that the new approach has the above mentioned merits, including ease of error rate control, lower computational complexity and lower decision delay. The derivation of a reference-phase-free FSST for QAM classification has been an important problem [15, 19]. In this paper we incorporate SPRT with three different techniques, i.e. the transform, the generalized likelihood and the averaging likelihood methods, to solve the problem of reference phase uncertainty.

This paper is organized as follows. In Sections 2 and 3, we examine modulation classification algorithms using FSST and SPRT, respectively. In Section 4, three techniques are discussed to derive the reference-phase-invariant classifiers. Experimental results are shown in Section 5 and the concluding remarks are presented in Section 6.

2. Fixed-sample-size test (FSST)

2.1. Problem formulation

We assume that the received waveform $r(t)$ is a QAM modulated signal $x(t; \boldsymbol{s}_N, k, \boldsymbol{p})$ buried in AWGN $n(t)$ with zero mean and two-sided power spectral density of N_0 W/Hz, where $\boldsymbol{s}_N$ represents the symbol sequence $(s_0, s_1, \ldots, s_{N-1})$, N is the total number of received symbols, k indicates a different modulation format and $\boldsymbol{p}$ denotes a vector of communication parameters. Therefore, the received N symbol-periods waveform can be written as

$$r(t) = x(t; \boldsymbol{s}_N, k, \boldsymbol{p}) + n(t), \quad 0 \leqslant t \leqslant NT, \tag{1}$$

where T is the symbol duration. Here we assume that symbol timing is synchronized so that the signal is received at the beginning of each symbol period.

If the transmitted signal waveform $x(t; \boldsymbol{s}_N, k, \boldsymbol{p})$ is known, we can write down the a posteriori probability of the transmitted signal waveform based on the received signal $r(t)$ by using Bayes' rule as

$$p(x(t; \boldsymbol{s}_N, k, \boldsymbol{p})|r(t)) = \frac{p(x(t; \boldsymbol{s}_N, k, \boldsymbol{p}))\, p(r(t)|x(t; \boldsymbol{s}_N, k, \boldsymbol{p}))}{p(r(t))}.$$

By assuming that $\boldsymbol{p}$ is known or can be measured accurately from preprocessors, we can drop it for discussion in Sections 2 and 3. (Three methods will be described to handle the case of unknown $\boldsymbol{p}$, especially in resolving reference phase uncertainty, in Section 4.) We further assume that there is no preference for any particular symbol sequence or modulation format so that $p(x(t; \boldsymbol{s}, k, \boldsymbol{p}))$ and $p(r(t))$ are simply normalization terms which are independent of $\boldsymbol{s}$ or k. Then, the a posteriori probability of $x(t; \boldsymbol{s}_N, k)$ can be expressed as the following well-known form [26]:

$$p(x(t; \boldsymbol{s}_N, k)|r(t)) = C \exp\left\{ \frac{-1}{2N_0} \int_0^{NT} [r(t) - x(t; \boldsymbol{s}_N, k)]^2 \, \mathrm{d}t \right\}. \tag{2}$$

A QAM signal $x(t; \boldsymbol{s}_N, k)$ can be written as, for $0 \leqslant t \leqslant NT$,

$$x(t; \boldsymbol{s}_N, k) = X_{\mathrm{I}}(t; \boldsymbol{s}_N, k)\cos(2\pi f_{\mathrm{c}} t - \theta_{\mathrm{c}}) + X_{\mathrm{Q}}(t; \boldsymbol{s}_N, k)\sin(2\pi f_{\mathrm{c}} t - \theta_{\mathrm{c}}), \tag{3}$$

where f_{c} and θ_{c} denote the carrier frequency and the carrier reference phase, respectively, and

$$X_{\mathrm{I}}(t; \boldsymbol{s}_N, k) = \sum_{i=0}^{N} x_{\mathrm{I}}(i; k) u(t - iT),$$

$$X_{\mathrm{Q}}(t; \boldsymbol{s}_N, k) = \sum_{i=0}^{N} x_{\mathrm{Q}}(i; k) u(t - iT),$$

are the information bearing waveforms for I- and Q-channels, and where $u(t)$ is the spectrum shaping pulse function which is assumed to be known and absorbed into the communication parameter vector $\boldsymbol{p}$. In particular, we let $u(t)$ be a rectangular unit pulse function,

$$u(t) = \begin{cases} 1 & \text{if } 0 \leqslant t \leqslant T, \\ 0 & \text{otherwise.} \end{cases}$$

The I-, Q-channel sample sequences $\{x_{\mathrm{I}}(i; k);\ i = 1, 2, \ldots\}$ and $\{x_{\mathrm{Q}}(i; k);\ i = 1, 2, \ldots\}$ can be represented by $\{(x_{\mathrm{I}}(i; k), x_{\mathrm{Q}}(i; k)) \in \mathscr{S}_k;\ i = 1, 2, \ldots\}$, where $\mathscr{S}_k \triangleq \{(s_{\mathrm{I}}(m; k), s_{\mathrm{Q}}(m; k));\ m = 1, \ldots, M\}$ denotes the constellation or the symbol set of the modulation type k. The resulting two-dimensional plot of these M vectors $(s_{\mathrm{I}}(m; k), s_{\mathrm{Q}}(m; k))$, $m = 1, \ldots, M$, is called the constellation or signal space diagram, where each point in the diagram is called a constellation point. Examples are given in Fig. 1 to illustrate different constellation diagrams for QAM signals.

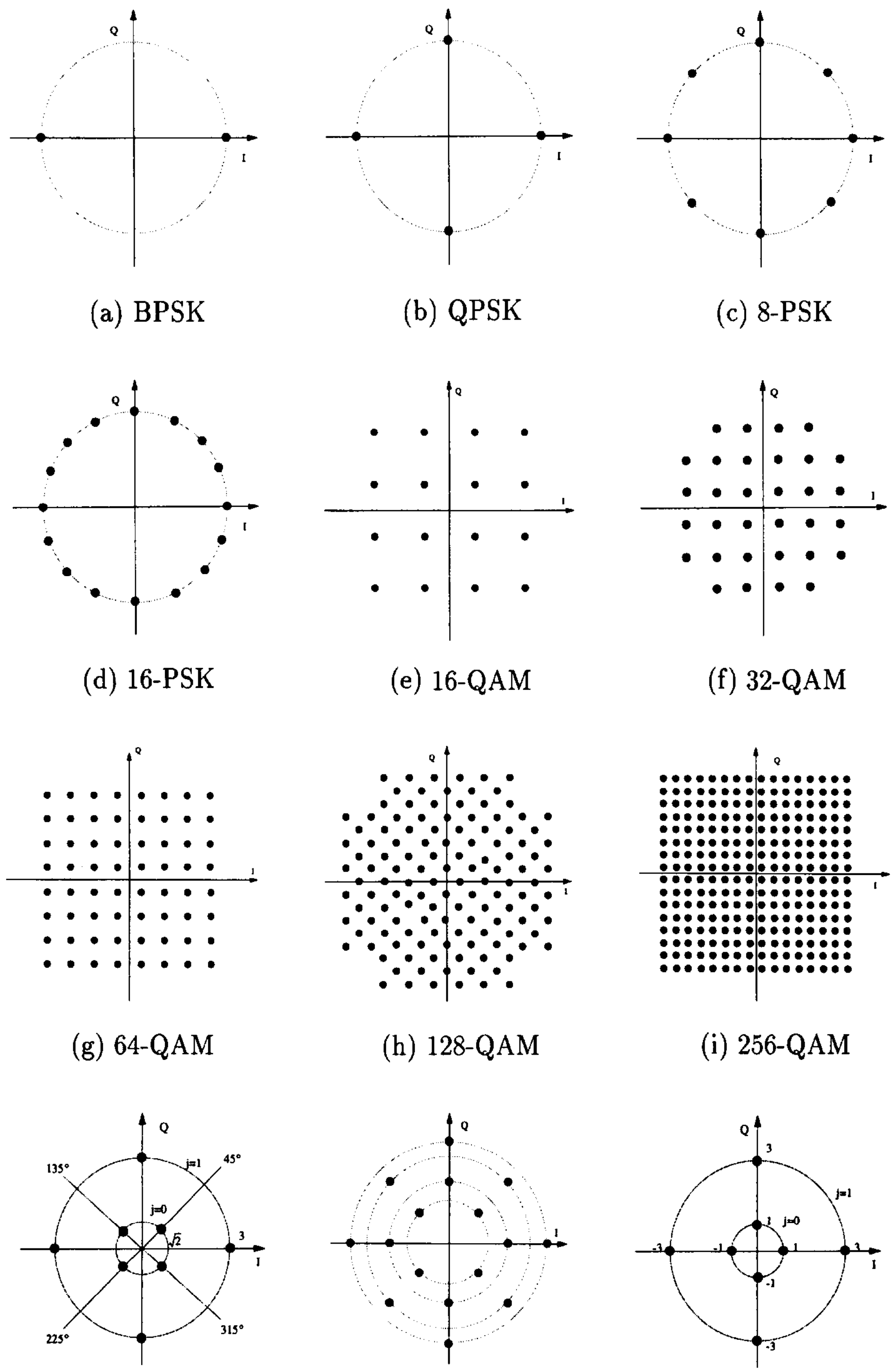
By imposing the assumptions of independent and identically distributed (i.i.d.) symbol sequences and an equal probable symbol set, we can derive the following equations from (2):

$$p(x(t; k)|r(t)) = C' \prod_{i=1}^{N} p(I_i, Q_i|k). \tag{4}$$

Note that I_i and Q_i are the ith one-symbol integration of I-, Q-channel and

$$p(I_i, Q_i|k) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{2\pi\sigma_T^2} \exp\left\{ \frac{-1}{2\sigma_T^2} [(I_i - s_{\mathrm{I}}(m; k))^2 + (Q_i - s_{\mathrm{Q}}(m; k))^2] \right\}, \tag{5}$$

where $(s_{\mathrm{I}}(m; k), s_{\mathrm{Q}}(m; k))$ is the mth constellation point of the kth modulation scheme and $\sigma_T^2 = N_0 T/2$.



(a) BPSK (b) QPSK (c) 8-PSK

(d) 16-PSK (e) 16-QAM (f) 32-QAM

(g) 64-QAM (h) 128-QAM (i) 256-QAM

(j) V29, 7200 bps (k) V29, 9600 bps (l) Star 8-QAM

Fig. 1. Examples of QAM constellations.

2.2. MAP classifier

Suppose that there are K possible modulation schemes in association with an N-symbol received waveform. Let us construct a multihypothesis test by associating each of the K hypotheses $H_1, H_2, \ldots, H_K$ with one modulation scheme. The a posteriori probability $p(x(t;k)|r(t))$ under H_k is specified by Eq. (4). Based on the maximum a posteriori (MAP) principle, one can minimize the total (or average) error probability

$$P_e = \sum_{i=1}^{K} P(H_i)\varepsilon_i,$$

where the individual decision error ε_i is defined by

$$\varepsilon_i = \sum_{j \neq i} \text{Prob}(\text{decided on } H_j | H_i \text{ is true}),$$

by deciding on $H_{\hat{k}}$ whenever its a posteriori probability is the largest. In other words, the modulation type $\hat{k}$ is chosen by using the following criterion:

$$\hat{k} = \arg\max_{k=1}^{K} p(x(t;k)|r(t)).$$

Note that the MAP classifier makes decision based on the probability calculation using a fixed number N of received symbols.

3. Variable-sample-size test with SPRT

A statistical test that uses a random number of samples is called the *sequential* test [12, 13, 20]. One important advantage of the sequential test is the flexibility in controlling the individual error probability ε_i for a multihypothesis test problem. The MAP test discussed above minimizes the total error probability, but has no control over the individual error probability conditioned on a given hypothesis. In a binary hypothesis test, the Neyman–Pearson test maximizes the correct probability of one hypothesis (i.e. the detection probability) while keeping the false alarm probability under a certain level. Although it is possible to generalize the Neyman–Pearson test to the multihypothesis case, this technique is not widely used in practice [26]. In this section, we propose the use of the sequential probability ratio test (SPRT) for QAM classification for binary as well as multihypothesis cases. With the sequential test, we can keep performing the test with more observations until a certain performance or stopping criterion is achieved.

3.1. SPRT for binary hypothesis

A binary sequential test [20] can be stated as follows. By assuming that the observed sequence of random variables $X_1, X_2, X_3, \ldots$ are i.i.d., we would like to determine which one of the two hypothesized distributions the observed sequence comes from. In mathematics, this test can be written as

$$\mathrm{H}_1\colon\ X_k \sim P_1, \quad k = 1, 2, \ldots,$$

$$\mathrm{H}_2\colon\ X_k \sim P_2, \quad k = 1, 2, \ldots,$$

where the notation '$\sim$' denotes 'obtained from a certain distribution', and P_1 and P_2 represent the hypothesized distributions. A sequential test is characterized by a set of stopping rules and a set of decision rules. The stopping rule at time n tells us whether we should stop the experiment with observations $X_1, \ldots, X_n$ or continue the experiment for one more additional observation X_{n+1}. The decision rule performs the hypothesis test based on all available data. Note that FSST of size n is in fact a special case of a sequential test, namely, it always stops at n.

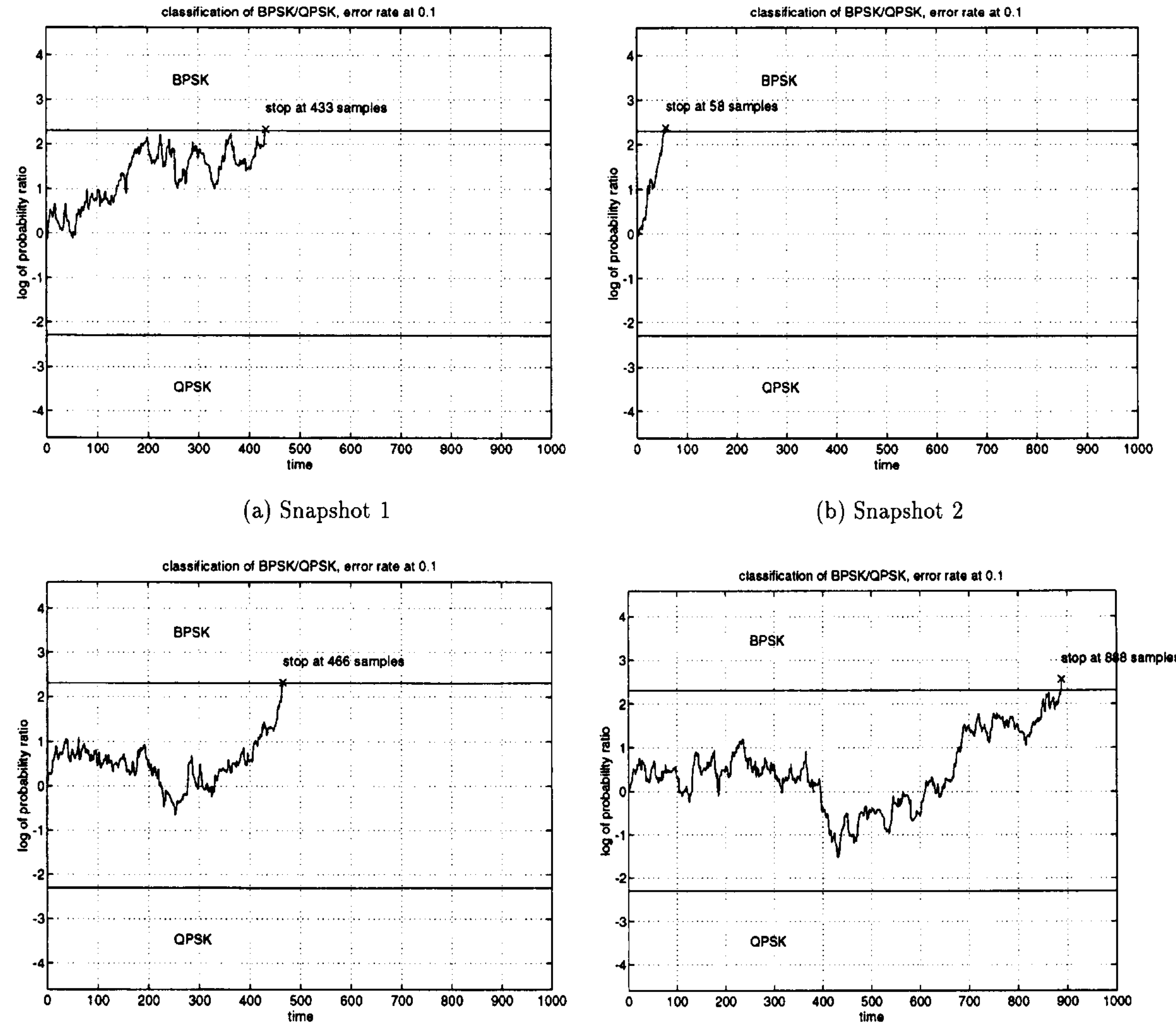
The sequential probability ratio test (SPRT) provides a very effective sequential test. With SPRT, we compute the likelihood ratio

$$\lambda(X_1, X_2, \ldots, X_n) = \prod_{k=1}^{n} \frac{p_2(X_k)}{p_1(X_k)}$$

based on samples $\{X_1, X_2, \ldots, X_n\}$, where p_1 and p_2 are probability densities associated with distributions P_1 and P_2, respectively. Then, this ratio is compared with some threshold values a and b, where $0 \leqslant a < b < \infty$. The test continues until the ratio $\lambda(X_1, X_2, \ldots, X_n)$ falls outside (a, b), and the decision rule is

$$\text{if} \quad \lambda(X_1, X_2, \ldots, X_n) < a \ \rightarrow\ \mathrm{H}_1,$$

$$\text{if} \quad \lambda(X_1, X_2, \ldots, X_n) > b \ \rightarrow\ \mathrm{H}_2.$$



(a) Snapshot 1

(b) Snapshot 2

(c) Snapshot 3

(d) Snapshot 4

Fig. 2. Four snapshots of SPRT for BPSK/QPSK classification with symbol SNR = −7 dB.

To explain SPRT better, we consider an example of classifying BPSK/QPSK and show four snapshots in Fig. 2 conditioned on that the source signals are actually BPSK modulated. The x- and y-coordinates of the figure represent the time and the log-likelihood ratio value, respectively, and the two parallel lines in each snapshot represent the decision thresholds of SPRT. Note that the decision thresholds are $\log a$ and $\log b$ because we compute the log-likelihood ratio. The four sample curves provides different variations of the log-likelihood ratio along the time axis. SPRT will not stop until the log-likelihood ratio hits one of the two decision boundaries. If the upper (or lower) boundary is hit first, we decide on BPSK (or QPSK). Time required for decision making depends on the data statistics as well as decision boundaries. If input data are clearly in favor of one of the two hypotheses or decision boundaries are close to zero, it requires a smaller number of samples (or, equivalently, shorter time delay) in decision making. On the contrary, if input data are more uncertain or decision boundaries are away from zero, more time is needed to reach a decision.

3.1.1. Optimality of SPRT

SPRT with likelihood ratio λ and interval (a,b) is usually denoted by SPRT(a,b,λ). The optimality of SPRT can be characterized by the well-known Wald–Wolfowitz theorem [27]. This theorem says that on the average SPRT requires the minimum amount of samples to achieve a given level of performance (i.e. the error probability). Thus, the average sample size of SPRT is not greater than that of FSST with the same performance.

Furthermore, we can control the decision interval (a,b) to adjust the error probabilities of SPRT. Wald showed that as long as the error decision probabilities are reasonably small, we can choose

$$a=\varepsilon_2/(1-\varepsilon_1)\approx\varepsilon_2 \quad \text{and} \quad b=(1-\varepsilon_2)/\varepsilon_1\approx 1/\varepsilon_1,$$

where ε_1 (or ε_2) denotes the error decision probability given that H_1 (or H_2) is true. The values a and b are called the Wald boundaries.

3.1.2. Average stopping time

The average stopping times for H_1 and H_2 can be approximated as [20]

$$E\{N|\mathrm{H}_1\}\approx\frac{1}{\mu_1}\left[(1-\varepsilon_1)\log\frac{\varepsilon_2}{1-\varepsilon_1}+\varepsilon_1\log\frac{1-\varepsilon_2}{\varepsilon_1}\right],$$

$$E\{N|\mathrm{H}_2\}\approx\frac{1}{\mu_2}\left[\varepsilon_2\log\frac{\varepsilon_2}{1-\varepsilon_1}+(1-\varepsilon_2)\log\frac{1-\varepsilon_2}{\varepsilon_1}\right],$$

where

$$\mu_j=E\{\log\lambda(X)|\mathrm{H}_j\} \quad \text{and} \quad \mu_j\neq 0, \quad \text{for } j=1,2. \tag{6}$$

The above approximations can be further simplified to be

$$E\{N|\mathrm{H}_1\}\approx\frac{\log\varepsilon_2}{\mu_1} \quad \text{and} \quad E\{N|\mathrm{H}_2\}\approx-\frac{\log\varepsilon_1}{\mu_2} \tag{7}$$

for reasonably small ε_1 and ε_2.

3.1.3. Efficiency

Efficiency of SPRT relative to FSST is defined as the ratio of their average sample sizes with respect to the same error level ε_j [14], i.e.

$$R_j=\frac{n(\varepsilon_j)}{E\{N|\mathrm{H}_j\}} \quad \text{for } j=1,2, \tag{8}$$

where $n(\varepsilon_j)$ is the number of samples required by FSST to achieve the error level ε_j when the hypothesis H_j is true. Let us use S_n to denote the log-likelihood ratio of i.i.d. distributed sequence $X_1,X_2,\ldots,X_n$.

$$S_n=\sum_{i=1}^{n}\log\frac{p_2(X_i)}{p_1(X_i)}=\sum_{i=1}^{n}\log\lambda(X_i).$$

The error probability given that H_j is true can be written as $\varepsilon_j=P(S_n\geqslant 0|\mathrm{H}_j)$. Assume $n\gg 1$, we can apply the central limiting theorem to S_n so that

$$\varepsilon_j\approx Q\left(\frac{-\sqrt{n}\mu_j}{\sigma_j}\right),$$

where $Q(x)=\frac{1}{\sqrt{2\pi}}\int_x^\infty \mathrm{e}^{-y^2/2}\mathrm{d}y$, μ_j is given by Eq. (6), and $\sigma_j^2=\mathrm{Var}\{\log\lambda(X)|\mathrm{H}_j\}$.

We may further simplify $Q(x)$ as [23]

$$Q(x)\approx\frac{1}{\sqrt{2\pi}x}\mathrm{e}^{-x^2/2} \quad \text{for } x\gg 0,$$

so that the log of error probability can be approximated by

$$\log\varepsilon_j\approx-n\frac{\mu_j^2}{2\sigma_j^2} \quad \text{for } \varepsilon_j\ll 1.$$

Therefore, the sample size required to achieve the error level of ε_j via FSST is

$$n(\varepsilon_j)\approx-\frac{2\sigma_j^2}{\mu_j^2}\log\varepsilon_j. \tag{9}$$

By plugging (9) and (7) into (8), we obtain the efficiency of SPRT as

$$R_j=-\frac{2\sigma_j^2}{\mu_j} \quad \text{for } j=1,2.$$

Thus, efficiency can be approximated by the mean and variance of $\log\lambda(X)$.

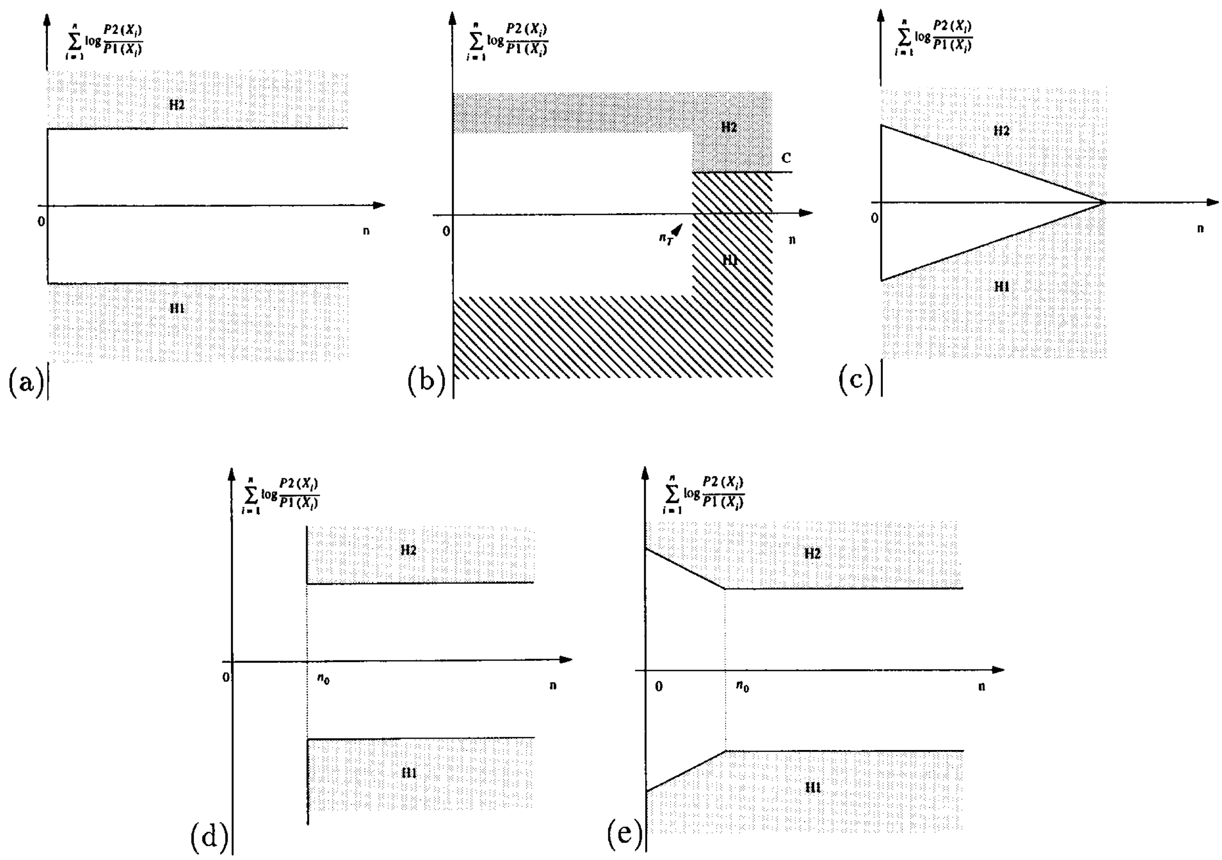


Fig. 3. Examples of decision boundaries: (a) Wald's decision boundary, (b) decision boundary of truncated SPRT, (c) converging decision boundary, (d) Read's decision boundary and (e) Baruah–Bhattacharjee's decision boundary.

3.1.4. Decision boundaries

The standard SPRT by Wald compares the likelihood ratio with two parallel boundary lines so that one of the hypotheses is accepted whenever the likelihood ratio reaches one of the boundaries. This condition is illustrated in Fig. 3(a). Although Wald's SPRT is optimal in the sense that it requires the smallest amount of samples for decision making with a given individual error probability, there have been criticisms and modifications of Wald's SPRT [9, 25]. For example, although it has been proved that the sample size is finite with probability one, the size can be too long to be practical. In practice, we may be forced to stop even though the stopping criterion has not been met yet, and the resulting performance is not as good as Wald's SPRT. This is known as *truncated SPRT* as depicted in Fig. 3(b), where we show an example of SPRT truncated at n_T samples and C denotes the threshold value for decision making after n_T samples are observed. Another example proposed in [2] is the converging boundaries as shown in Fig. 3(c), where decision boundaries are dynamically changing with time. This kind of decision strategy makes it easier to reach a conclusion for longer observation time by relaxing the requirement of making correct decisions. Figs. 3(d) and (e) are methods proposed by Read [21] and Baruah and Bhattacharjee [6], respectively. Read suggested to apply SPRT only after n_0 samples. Decision boundaries in Fig. 3(e) are in fact a combination of boundaries in Figs. 3(c) and (d). In our experiments given in Section 5, we follow Read's work as shown in Fig. 3(d).

Tantaratana [25] reviewed several decision boundaries to reduce the average sample size for a non-perfect statistical model [25]. The general approach is to combine the fixed-sample-size and the sequential tests so that the sequential test is invoked after a fixed amount of data have been collected. The purpose of this mixture design is to maintain the performance at least close to FSST when the parameters are mismatched but still keep average sample size as small as Wald's SPRT when parameters are correctly tuned.

3.2. *Multihypothesis test*

As mentioned before, one of the major drawback of FSST is that even though it provides the solution to the lowest *average* error probability, there is no assurance in minimizing the *individual* error probability. SPRT was derived for the binary hypothesis case in the previous subsection, and the relationship between individual error probabilities and Wald's boundaries provides important insights into the multihypothesis problem. Several multihypothesis sequential tests have been proposed in the past. Most of them divide the test into a couple of binary tests. We may classify them into [13] (1) the positive approach, which considers simultaneously all possible binary hypothesis tests and (2) the negative approach, which eliminates the unlikely hypotheses along the process until there is only one left. Another way to classify different approaches is to see how the alternative hypothesis is constructed by using binary hypothesis tests. There are three approaches to represent the test statistics of alternative hypothesis: (1) maximum-likelihood, (2) geometric mean and (3) algebraic mean. These approaches are explained below.

3.2.1. *Positive approach*

The Armitage test [3, 13] generalizes the binary to M-ary hypothesis test by constructing pair-wise SPRT for all hypotheses. For example, if there are m hypotheses $\mathrm{H}_1, \mathrm{H}_2, \ldots, \mathrm{H}_m$ and their corresponding probability density functions are $p_1, p_2, \ldots, p_m$, respectively. This approach stops the test at n samples and accepts H_i if

$$\prod_{k=1}^{n} \frac{p_i(X_k)}{p_j(X_k)} > b_{i,j}, \quad \text{for all } j \neq i, \tag{10}$$

where $X_1, X_2, \ldots, X_n$ are i.i.d. samples. The decision error probability ε_i given that hypothesis H_i is true can be controlled by adjusting $b_{i,j}$ and it is bounded by

$$\varepsilon_i = \sum_{j \neq i} \varepsilon_{i,j} \leqslant \sum_{j \neq i} \frac{1}{b_{i,j}},$$

where $\varepsilon_{i,j}$ is the error probability of claiming H_j given that hypothesis H_i is true.

3.2.2. *Negative approach*

The negative approach, denoted by m-SPRT [13, 17], is based on the principle that one can reject H_i if it is unlikely to be the answer compared with the current best candidate H_{j^*}. We stop the test and decide H_i if H_i is the only hypothesis left. Let H_{j^*} denote the hypothesis that gives the maximum-likelihood values of $X_1, X_2, \ldots, X_n$. m-SPRT constructs binary SPRT $(a_i, \infty, p_i/p_{j^*})$ for $i \neq j^*$ such that we reject H_i if

$$\prod_{k=1}^{n} \frac{p_i(X_k)}{p_{j^*}(X_k)} < a_i. \tag{11}$$

The error probability ε_i of rejecting H_i can also be specified by $\{a_1, a_2, \ldots, a_m\}$ such that

$$\varepsilon_i \leqslant \sum_{j \neq i} a_j.$$

Comparing this approach with the previous one, we see that m-SPRT has substantial saving on computation, since the number of hypotheses decreases once some hypothesis has been rejected. On the other hand, the number of hypotheses for the positive approach remains the same all the time. For this reason, we choose m-SPRT in our experiments.

4. Reference phase invariant classifiers

The a posteriori probabilities of received QAM waveforms given in previous sections are parameterized by $\boldsymbol{p}$, i.e. a vector of communication parameters such as the carrier frequency, carrier phase, etc. In practice, $\boldsymbol{p}$ is unknown and has to be estimated from received data. For the QAM classification problem, research effort has focused on deriving a reference-phase-free test statistics [15, 19]. The study of how the reference-phase effects the test statistics is valuable because an incorrect carrier frequency estimate will result in error in the phase term. This section will concentrate on how a reference-phase invariant test statistic is derived, and how it can be modified to be less sensitive to the frequency estimate error. From the perspective of statistics, a test with unwanted parameters is known as a test with nuisance parameters which is denoted by θ in the following.

4.1. *Nuisance-parameter-free test*

There are three approaches reported to solve this problem [4]. One approach is to replace the unwanted parameter θ by its maximum-likelihood estimate

$$\hat{\theta} = \arg\max_{\theta} \; p_i(x_1, x_2, \ldots, x_k; \theta) \tag{12}$$

and

$$\mathrm{H}_i\colon \; p_i(x_1, x_2, \ldots, x_k; \hat{\theta}), \quad \text{for } i = 1, 2, \ldots, m, \tag{13}$$

where $p_i(x_1, x_2, \ldots, x_k; \theta)$ is the probability density function parameterized by a nuisance parameter θ if H_i is true.

Wald suggested to use the average probability density function by introducing a weighting function to average out the unwanted parameter, i.e.

$$\mathrm{H}_i\colon \; \bar{p}_i(x) = \int_{\theta} p_i(x; \theta) w_i(\theta) \, \mathrm{d}\theta, \quad \text{for } i = 1, 2, \ldots, m,$$

where

$$\int_{\theta} w_i(\theta) \, \mathrm{d}\theta = 1.$$

Usually, $w_i(\theta)$ is chosen to be equally weighted if there is no prior knowledge of the unwanted parameter θ. For example, we may assign $w(\theta_c) = 1/2\pi$, where θ_c is the unknown reference phase of the received signal. In particular, if the average of reference phase is taken by one symbol period, the phase information is lost and the resulting probability density represents the amplitude distribution. Therefore, the averaging process has to be taken for a several-symbol period in order to keep the phase information, i.e.

$$\mathrm{H}_i\colon \; \bar{p}_i(x_1, x_2, \ldots, x_k) = \int_{\theta} p_i(x_1, x_2, \ldots, x_k; \theta) w_i(\theta) \, \mathrm{d}\theta. \tag{14}$$

The third approach is to transform the sample sequence $x_1, x_2, \ldots$ to another sequence $y_1, y_2, \ldots$ so that the resulting likelihood function of the new sequence is not parameterized by the nuisance parameter. For example, the phase difference classifier for MPSK signals discussed in [16] uses the transformation

$$\Delta\theta_i = \theta_i - \theta_{i-1} \pmod{2\pi}$$

to obtain the phase difference sequence $\Delta\theta_i$, where θ_i is the phase of the received signal sampled at the ith symbol period.

4.2. *Practical classifiers based on windowed data*

In practice, the carrier frequency might be unknown and has to be estimated from the received signals. The frequency offset due to the estimation error causes the rotation of signal constellation so that the constant reference-phase assumption can be violated. Depending on the degree of the frequency estimate error, we can still assume a constant reference phase for a period of time, e.g. k-symbol period. The immunity to frequency estimate error is determined by the degree that the constant reference assumption is violated in classifier design. The proposed classifiers segment the received waveform into several windows. Data within each window are used to calculate the ML phase estimate by using (12) and/or the likelihood for each hypothesis by (13) or (14). The final test statistic is the product of those test statistics calculated for each set of windowed data. A smaller window size gives a lower correct classification rate since the test statistic contains more reference-phase uncertainty. However, it gives a slower degradation of the correct rate in the presence of larger frequency estimate error. The selection of window size is a tradeoff between robustness and performance. In practice, an appropriate window size can be selected based on the knowledge of the receiving environment.

The new algorithms based on the windowed data and the decision/stopping rules discussed in Section 3.2 are summarized as follows.

Algorithm 1. Positive approach.

1. Initialize the corresponding log of likelihood L_i for hypothesis H_i, i.e. $L_i = 0$, $\forall i$.

2. Collect k samples $x_1, x_2, \ldots, x_k$, where k is the given window size.
3. Calculate the likelihood of hypothesis H_i for the k samples $x_1, x_2, \ldots, x_k$:
$p_i^*(x_1, x_2, \ldots, x_k)$, $\forall i$, by (13) or (14).
4. Update the log of likelihood L_i by $L_i + \log\, p_i^*(x_1, x_2, \ldots, x_k)$, $\forall i$.
5. Stop the test and accept H_i if $L_i - L_j > \log b_{i,j}$, $\forall j \neq i$, where $b_{i,j}$ is the decision boundary described in (10). Otherwise, go to Step 2 and collect another k samples.

Algorithm 2. Negative approach.
1. Initialize the corresponding log of likelihood L_i for hypothesis H_i, i.e. $L_i = 0$, $\forall i$. Initialize the candidate set to include all hypotheses.
2. Collect k samples $x_1, x_2, \ldots, x_k$, where k is the given window size.
3. Calculate the likelihood of hypothesis H_i for k samples $x_1, x_2, \ldots, x_k$:
$p_i^*(x_1, x_2, \ldots, x_k)$, $\forall i$, by (13) or (14).
4. Update the log of likelihood L_i by $L_i + \log\, p_i^*(x_1, x_2, \ldots, x_k)$, $\forall i$.
5. Update the candidate set by removing H_i from the set if $L_i - L_{j^*} < \log a_i$, where $L_{j^*} = \max_{\forall i} L_i$ and a_i is the rejection decision boundary in (11) for H_i.
6. Stop the test and decide H_{i^*} if H_{i^*} is the only hypothesis left in the candidate set. Otherwise, go to Step 2 and collect another k samples.

The methods apply the *generalized maximum-likelihood* approach or the *maximum average-likelihood* approach depending on Eqs. (13) or (14) used in Step 3. These algorithms have three advantages: (1) a sequential test is feasible; (2) the window size is adjustable to improve the performance for various channel conditions; and (3) a multihypothesis test is feasible.

5. Experimental results

Example 1. 8-PSK/16-PSK classification.

We show in Fig. 4 the required average sample size (ASN) of SPRT compared with FSST (MAP) to classify 8-PSK/16-PSK for symbol SNR ranging from 8 to 17 dB. Their constellations are shown in Figs. 1(c) and (d). The desired performance is to achieve 99% individual correct rate when the input modulation schemes are either 8-PSK or 16-PSK for SPRT and 99% average correct rate for FSST. We see that the efficiency of SPRT is about 3 dB for all SNR values. This means that we only need about one-half of the samples on the average to make decision to reach the same correct level with SPRT rather than FSST.

Example 2. 8-PSK/V.29 (7200 bps)/Star 8-QAM classification.

The test set is composed of three 8-QAM constellations: 8-PSK, V.29 (7200 bps) and Star 8-QAM. The corresponding constellation diagrams are given in Figs. 1(c), (d) and (f). Two thousand sets of data were simulated for every constellation. The average and individual correct rates for using 100 symbols for classification is shown in Fig. 5. We can see from the figure that the individual error rates are not evenly distributed. The low SNR correct rate is very low for V.29 but high for 8-PSK in comparison with the average correct rate. The reason why V.29 is difficult to identify can be explained by the fact that V.29 has two amplitude levels, which is the same as Star 8-QAM, and 8 phase angles, which is the same as 8-PSK. As a result, V.29 has a greater chance to be misclassified. This experiment represents a typical challenge to multihypothesis FSST. This problem could be resolved by adjusting the decision threshold of test statistics so that errors can be more evenly distributed.

To demonstrate the performance of SPRT, we generate 1000 sequences of a maximum length of 50 000 symbols for each of the three modulation schemes. By enforcing the maximum number of samples used to make decisions to be finite, we basically adopt truncated SPRT. Furthermore, the following three cases using the negative approach are examined:
- Test Case 1: symbol SNR 4 dB, known reference phase, $a_P = a_V = a_S = 0.01$;
- Test Case 2: symbol SNR 4 dB, known reference phase, $a_P = a_V = a_S = 0.05$;
- Test Case 3: symbol SNR 0 dB, known reference phase, $a_P = a_V = a_S = 0.01$.

Note that the subscripts P, V and H denote the three modulation schemes 8-PSK, V.29 (7200 bps)

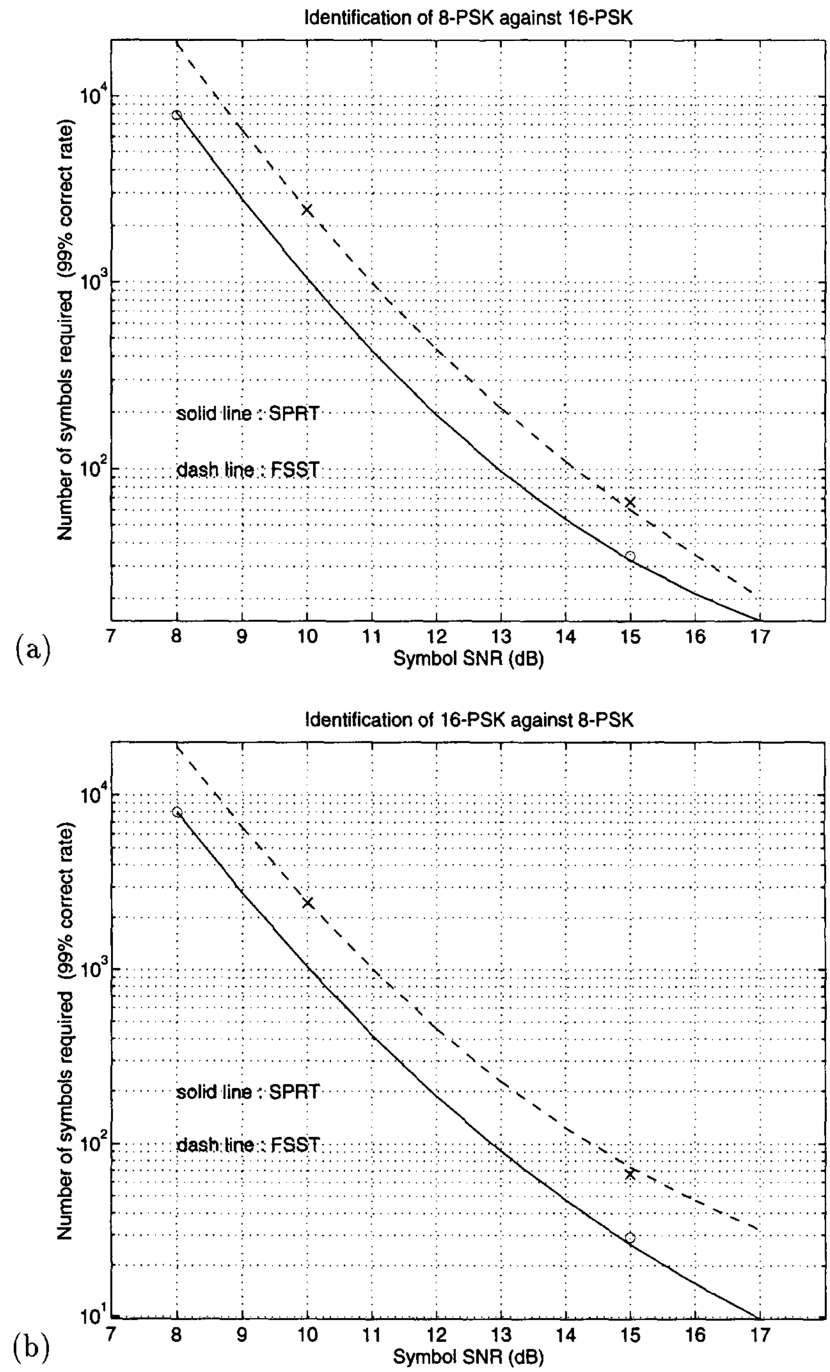


Fig. 4. The number of symbols required to achieve error level of 0.01 is plotted as a function of symbol SNR in 8-PSK/16-PSK classification with the input is (a) 8-PSK and (b) 16-PSK modulated.

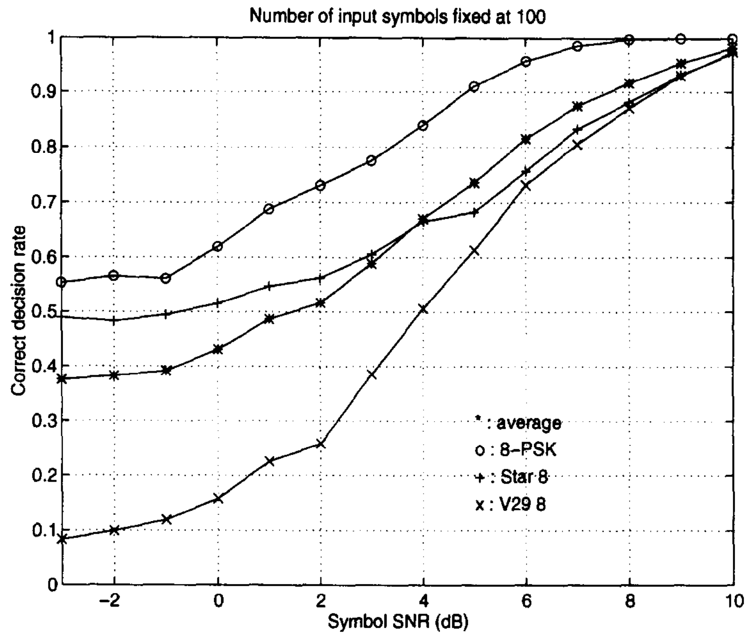


Fig. 5. Correct classification rates in classifying 8-PSK, V.29 (7200 bps) and Star 8-QAM using FSST.

and Star 8-QAM, respectively. Also, the choice of boundary decision parameter $a_P = a$ implies a correct classification rate close to $1 - 2a$ when the input signal is 8-PSK modulated. Thus, without truncation, the correct rates for Test Cases 1 and 3 should be about 98% while the correct rate for Test Case 2 is about 90%.

Results for Test Case 1 are summarized in Table 1, where RR and AST represent the rejection rate and the average stopping time, respectively. For example, the last two columns in Table 1(a) tells that, with 1000 Star 8-QAM trial sequences, it takes, on the average, 120 samples (symbols) to reject H_P (8-PSK) and the rejection rate is 99.8%, and takes on the average 1448 samples to reject H_V (V.29) and the rejection rate is 98.6%. Finally, only 1.6% of the trials make mistakes by rejecting the underlying H_S while the average stopping time is 961 samples. Note that to make the final decision among the three hypotheses, we have to reject two hypotheses. The average sample size required to make the final decision is listed at the bottom row for each input signal type. It is clear that we need six times more samples to recognize Star 8-QAM and V.29 than 8-PSK. This is consistent with our previous experience in FSST that 8-PSK is the most recognizable constellation among the three test modulations. The predicted average reject time by using Wald's approximations is shown in Table 1(b) for comparison. The predictions are very close. Table 1(c) shows the decision results of using truncated SPRT. The 'correct rate' column shows the individual average correct classification rate. We observe correct rates around 98% via design. The column P + V denotes that only Hypothesis S is rejected after 50 000 samples and P + V + S indicates that no hypothesis has been rejected. It turns out that none of these cases exist in the current experiment.

Test Case 2 is used to show how the values of a_P, a_V and a_S control the desired performance. By changing their values from 0.01 to 0.05, we see from Table 2 that the correct rate is reduced from 98% to 90%. However, the average sample size is also reduced by a factor of 60–75%. For comparison, we also performed FSST which is comparable with Test Case 2 by fixing the sample size to be 900 symbols. The average correct rate is 91.6% and the individual correct rates for 8-PSK, V.29 and Star 8-QAM are

Table 1
(a) Rejection rates (RR), average rejection time (ART) and average sample sizes (ASN)

	Input signals					
	8-PSK		V.29 7200 bps		Star 8-QAM	
Hypothesis	RR (%)	ART	RR (%)	ART	RR (%)	ART
8-PSK	0.6	199	99.2	207	99.8	120
V.29 7200 bps	99.4	229	1.4	1066	98.6	1448
Star 8-QAM	100.0	122	99.4	1435	1.6	961
ASN	231		1465		1445	

(b) predicted average rejection time (ART)

	Input signals		
	8-PSK	V.29 7200 bps	Star 8-QAM
Hypothesis	Predicted ART	Predicted ART	Predicted ART
8-PSK	NA	224	118
V.29 7200 bps	231	NA	1484
Star 8-QAM	123	1493	NA

(c) decision results and correct rates (CR) for Test Case 1

		Decisions						
Input signals	CR	P	V	S	P + V	P + S	V + S	P + V + S
8-PSK (P)	99.4	99.4	0.6	0.0	0.0	0.0	0.0	0.0
V.29 7200 bps (V)	98.6	0.8	98.6	0.6	0.0	0.0	0.0	0.0
Star 8-QAM (S)	98.4	0.2	1.4	98.4	0.0	0.0	0.0	0.0

99.9%, 88.0% and 86.8%, respectively, as shown in the last column in Table 2(c). The average correct rate of FSST is slightly lower than that of SPRT (92.9%). The average ASN for FSST and SPRT are 900 and 626, respectively. We see that, on the average, SPRT needs only about $\frac{2}{3}$ of symbols required by FSST with almost identical average correct rates in this experiment.

Results for Test Case 3 are given in Table 3. The average stopping time conditioned on a particular hypothesis and the average sample size to make decision increase significantly due to the lower symbol SNR value. Furthermore, there exists more than one non-rejected hypotheses after 50 000 samples. For example, there are 4.6% of the V.29 simulation sequences which cannot be distinguished from Star 8-QAM. In computing the final correct rate, we are forced to make decision by comparing their probabilities based on 50 000 samples. It turns out the final correct rate is only slightly less the sum

Table 2
(a) Rejection rates (RR), average rejection time (ART) and average sample sizes (ASN)

	Input signals					
	8-PSK		V.29 7200 bps		Star 8-QAM	
Hypothesis	RR (%)	ART	RR (%)	ART	RR (%)	ART
8-PSK	5.6	114	96.8	131	99.7	81
V.29 7200 bps	94.4	148	8.0	554	92.5	916
Star 8-QAM	100.0	88	95.2	789	7.8	619
ASN	157		820		901	

(b) predicted average rejection time (ART)

	Input signals		
	8-PSK	V.29 7200 bps	Star 8-QAM
Hypothesis	Predicted ART	Predicted ART	Predicted ART
8-PSK	NA	146	80
V.29 7200 bps	151	NA	965
Star 8-QAM	77	971	NA

(c) decision results and correct rates (CR) for Test Case 2

	SPRT								FSST
Input signals	CR	P	V	S	P + V	P + S	V + S	P + V + S	CR
8-PSK (P)	94.4	94.4	5.6	0.0	0.0	0.0	0.0	0.0	99.9
V.29 7200 bps (V)	92.0	3.2	92.0	4.8	0.0	0.0	0.0	0.0	88.0
Star 8-QAM (S)	92.2	0.3	7.5	92.2	0.0	0.0	0.0	0.0	86.8

of the actual correct rate and the probability of the ambiguity case. For example, as indicated in the row for V.29 of Table 3(c), the final correct rate is 97.5% which is approximately equal to the sum of 93.5% (the correct rate) and 4.6% (ambiguity between V + S). Thus, a direct probability comparison based on 50 000 samples is in favor of the right decision. The same statement applies to Star 8-QAM test sequences. This is consistent with our intuition.

Example 3. Reference phase invariant SPRT.

This experiment includes 11 QAM modulation schemes: BPSK (2P), QPSK (4P), 8-PSK (8P), 16-PSK (16P), V.29 7200 bps (8V), V.29 9600 bps (16V), 16-QAM (16Q), 32-QAM (32Q), 64-QAM (64Q), 128-QAM (128Q) and 256-QAM (256Q), where names in the parentheses are their simplified notations. One thousand symbol sequences are generated for every modulation scheme at 10 dB symbol SNR. Results of using SPRT to achieve 99% correct

Table 3
(a) Rejection rates (RR), average rejection time (ART) and average sample sizes (ASN)

	Input signals					
	8-PSK		V.29 7200 bps		Star 8-QAM	
Hypothesis	RR (%)	ART	RR (%)	ART	RR (%)	ART
8-PSK	1.4	2714	99.1	2785	100.0	1545
V.29 7200 bps	98.6	2952	1.9	11 311	94.0	19 541
Star 8-QAM	100.0	1736	94.4	18 393	2.3	7496
ASN	2996		20 218		20 429	

(b) predicted average rejection time (ART)

	Input signals		
	8-PSK	V.29 7200 bps	Star 8-QAM
Hypothesis	Predicted ART	Predicted ART	Predicted ART
8-PSK	NA	2979	1609
V.29 7200 bps	3068	NA	21 477
Star 8-QAM	1675	21 650	NA

(c) decision results and correct rates (CR) for Test Case 3

		Decisions						
Input signals	CR	P	V	S	P + V	P + S	V + S	P + V + S
8-PSK (P)	98.6	98.6	1.4	0.0	0.0	0.0	0.0	0.0
V.29 7200 bps (V)	97.5	0.9	93.5	1.0	0.0	0.0	4.6	0.0
Star 8-QAM (S)	97.5	0.0	2.3	94.0	0.0	0.0	3.7	0.0

rejection rate with the maximum average-likelihood approach over a 100 symbol window are given in Table 4. Read's decision boundaries are used by applying SPRT after receiving 100 symbols. Therefore, the minimum stopping time is 100 symbol periods. The rejection rate (RR) for every hypothesis and modulation source, the average rejection time (ART) by simulation, and the average correct rate (ACR) and the decision confusion matrix are shown in Table 4(a)–(c), respectively.

6. Conclusions

Several QAM classification algorithms based on the likelihood statistics were examined in this research. In particular, we proposed a new QAM classification algorithm based on the sequential probability ratio test (SPRT) and demonstrated that SPRT has several advantages over the classical fixed sample size test (FSST). FSST guarantees the minimum total error probability solution by choosing the hypothesis which

Table 4
(a) Rejection rates (RR)

Source	Hypotheses RR (%) 2P	4P	8P	16P	8V	16V	16Q	32Q	64Q	128Q	256Q
2P	0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
4P	100.0	0.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
8P	100.0	100.0	1.4	98.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0
16P	100.0	100.0	99.2	0.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0
8V	100.0	100.0	100.0	100.0	0.4	100.0	99.6	100.0	100.0	100.0	100.0
16V	100.0	100.0	100.0	100.0	100.0	1.0	100.0	100.0	100.0	100.0	99.0
16Q	100.0	100.0	100.0	100.0	100.0	99.0	4.0	99.0	98.0	100.0	100.0
32Q	100.0	100.0	100.0	100.0	100.0	100.0	100.0	1.4	100.0	99.4	99.2
64Q	100.0	100.0	100.0	100.0	100.0	100.0	99.0	100.0	5.0	98.0	94.0
128Q	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.0	100.0	2.0	99.0
256Q	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	98.0	98.8	3.2

(b) average reject time (ART)

Source	ASN	Hypotheses ART (simulation) 2P	4P	8P	16P	8V	16V	16Q	32Q	64Q	128Q	256Q
2P	100	NA	100	100	100	100	100	100	100	100	100	100
4P	101	100	NA	100	100	101	100	100	100	100	100	100
8P	6222	100	100	6375	6218	100	101	109	113	105	110	105
16P	6863	100	101	6852	8900	100	102	107	112	104	109	104
8V	146	100	100	100	100	100	118	144	105	135	106	132
16V	717	100	100	100	100	110	320	367	375	653	403	582
32Q	9799	100	100	106	106	101	420	416	11 850	632	9651	568
64Q	34 385	100	100	103	103	119	556	1417	615	15 540	540	31 403
128Q	9312	100	100	107	106	103	450	372	9298	633	2550	663
256Q	33 086	100	100	104	104	114	576	976	534	32 970	547	25 267

(c) decision results by using reference phase invariant SPRT

Source	ACR	Decisions (%) 2P	4P	8P	16P	8V	16V	16Q	32Q	64Q	128Q	256Q
2P	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4P	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8P	98.6	0.0	0.0	98.6	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
16P	99.2	0.0	0.0	0.8	99.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8V	99.6	0.0	0.0	0.0	0.0	99.6	0.0	0.4	0.0	0.0	0.0	0.0
16V	99.0	0.0	0.0	0.0	0.0	0.0	99.0	0.0	0.0	0.0	0.0	1.0
16Q	96.0	0.0	0.0	0.0	0.0	0.0	1.0	96.0	1.0	2.0	0.0	0.0
32Q	98.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.6	0.0	0.6	0.8
64Q	95.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	95.0	2.0	2.0
128Q	98.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	98.0	1.0
256Q	96.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	1.2	96.8

gives the maximum a posteriori probability (MAP). However, it does not have the control of the individual error probability given that one of the hypotheses is true. In contrast, SPRT provides an effective way to control the individual error probability. The number of samples required to make decision represents the delay to establish communication links. SPRT guarantees the minimum average delay with a certain level of error probability. The merit of the proposed method was supported by extensive experimental results.

Practical algorithms based on windowed data and the sequential test were presented to classify signals with an unknown reference phase. The selection of the window size is a tradeoff between the robustness and the performance in the presence of the frequency offset. In practice, an appropriate window size can be selected based on the knowledge of the receiving environment. However, it would be interesting to investigate an adaptive algorithm that can automatically adjust the window size according to the channel condition. A more detailed analysis is expected to be performed in future research.

References

[1] J. Aisbett, Automatic modulation recognition using time domain parameters, Signal Processing 13 (1987) 323–328.

[2] T.W. Anderson, A modification of the sequential probability ratio test to reduce the sample size, Ann. Math. Statist. 31 (1960) 165–197.

[3] P. Armitage, Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis, J. Roy. Statist. Soc. B (1950) 137–144.

[4] P. Armitage, Sequential methods, in: D.V. Hinkley, N. Reid, E.J. Snell (Eds.), Statistical Theory and Modeling, Chapman and Hall, London, 1991, pp. 129–151.

[5] K. Assaleh, K. Farrell, R.J. Mammone, A New Method of Modulation Classification for Digitally Modulated Signals, MILCOM, 1992, 0712–0716.

[6] H.K. Baruah, G.P. Bhattacharjee, A modification of the sequential probability ratio test for testing a normal mean, Australian J. Statist. 22 (1980) 178–187.

[7] B.F. Beidas, C.L. Weber, Higher-order correlation-based approach to modulation classification of digitally frequency-modulated signals, IEEE J. Selected Areas Comm. (January 1995) 89–101.

[8] Y.T. Chan, L.G. Gadbois, Identification of the modulation type of a signal, Signal Processing 16 (1989) 149–154.

[9] Y.T. Chien, K.S. Fu, A modified sequential recognition machine using time-varying stopping boundaries, IEEE Trans. Inform. Theory 12 (1966) 206–214.

[10] L. Vergara Domínguez, J.M. Páez Borralo, J. Portillo García and B. Ruiz Mezcua, A general approach to the automatic classification of radiocommunication signals, Signal Processing 22 (1991) 239–250.

[11] N. Ghani, R. Lamontagne, Neural networks applied to the classification of spectral features for automatic modulation recognition, MILCOM (1993) 111–115.

[12] B.K. Ghosh, Sequential Tests of Statistical Hypotheses, Addison-Wesley, Reading, MA, 1970.

[13] B.K. Ghosh, P.K. Sen (Eds.), Handbook of Sequential Analysis, Marcel Dekker, New York, 1991.

[14] Z. Govindarajulu, The Sequential Statistical Analysis of Hypothesis Testing, Point and Interval Estimation and Decision Theory, American Science Press, New York, 1981.

[15] C.-Y. Hwang, A. Polydoros, Likelihood methods for MPSK modulation classification, IEEE Trans. Comm. 43 (4) (1995) 1493–1504.

[16] F.F. Liedtke, Computer simulation of an automatic classification procedure for digitally modulated communication signals with unknown parameters, Signal Processing 6 (1984) 311–323.

[17] G. Lordon, 2-SPRT's and the modified Kiefer–Weiss problem of minimizing an expected sample size, Ann. Statist. (1976) 281–291.

[18] R. Mammone, R. Rothaker, C. Podilchuk, S. Davidovici, D. Schilling, Estimation of carrier frequency, modulation type and bit rate of an unknown modulated signal, IEEE Internat. Conf. Communications, 1987, pp. 1006–1012.

[19] A. Polydoros, K. Kim, On the detection and classification of quadrature digital modulations in broad-band noise, IEEE Trans. Comm. 38 (8) (1990) 1199–1211.

[20] H.V. Poor, An Introduction to Signal Detection and Estimation, 2nd ed., Springer, Berlin, 1994.

[21] C.B. Read, The partial sequential probability ratio test, J. Amer. Statist. Assoc. 66 (1971) 646–650.

[22] J. Reichert, Automatic classification of communication signals using higher order statistics, ICASSP V, 1992, pp. 221–224.

[23] M.K. Simon, S.M. Hinedi, W.C. Lindsey, Digital Communication Techniques, Signal Design and Detection, Prentice-Hall, Englewood Cliffs, NJ, 1994.

[24] S.S. Solima, S.-Z. Hsue, Signal classification using statistical moments, IEEE Trans. Comm. 40 (1992) 908–916.

[25] S. Tantaratana, Sequential detection of a positive signal, in: I.F. Blake, H.V. Poor (Eds.), Communications and Networks, Springer, Berlin, 1986, pp. 157–196.

[26] H.L.V. Trees, Detection, Estimation, and Modulation Theory, Part I, Wiley, New York, 1971.

[27] A. Wald, Sequential tests of statistical hypotheses, Ann. Math. Statist. (1945) 117–186.

[28] Y. Yang, S.S. Soliman, Optimum classifier for M-ary PSK signals, MILCOM (1991) 1693–1697.