14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

DIGITAL MODULATION CLASSIFICATION IN FLAT-FADING CHANNELS

Anchalee Puengnim, Nathalie Thomas and Jean-Yves Tourneret

IRIT/ENSEEIHT/TéSA, 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France
Anchalee.Puengnim@tesa.prd.fr, Nathalie.Thomas@enseeiht.fr, Jean-Yves.Tourneret@enseeiht.fr

ABSTRACT

This paper addresses the problem of classifying digital modulations in a Rayleigh fading environment. The first step of the proposed classifier consists of estimating the parameters unknown by the receiver, i.e., the fading amplitude, phase offset, and residual carrier frequency. These unknown parameters appearing in the class conditional densities are then replaced by their estimates, resulting in a so-called plug-in classifier. The performance of this classifier is compared to another classification strategy recently proposed to solve the modulation classification problem in a fading environment.

1. INTRODUCTION

Digital modulation classification consists of identifying the type of a modulated signal corrupted by noise. It is required in many communication applications including cooperative and non-cooperative scenarios [1]. The most popular modulation classifier (often referred to as *optimal classifier*) is probably the Bayes classifier which minimizes the average probability of error (or an appropriate average cost function). However, the Bayes classifier may be difficult to implement due to its high computational complexity. This is particularly true for the classification of digital modulations, because averaging over the data symbols leads to an exponential computational complexity, when there are too many parameters unknown at the receiver. Also, the Bayes classifier is not robust to model mismatch. To overcome the difficulties inherent to the Bayesian strategy, several suboptimal likelihood based classifiers have been proposed in the signal processing and communication literature (see for instance [1, 2, 3]). The main idea of these classifiers is to avoid the costly integration required to derive the posterior distribution of the unknown parameters. This integration can be avoided by estimating the unknown parameters and using the generalized likelihood ratio test [4], by approximating the average likelihood ratio test [2], or by using hybrid solutions [5]. An alternative to likelihood based classifiers is to extract interesting features from the observations and use the features for classification. In this case, the key point is to find the"appropriate" set of features depending on the considered communication system. Many features have been proposed in the literature including statistical moments [6] or higher-order statistics [1].

A considerable number of researches has been carried out on modulation classification mitigating the effect of the additive white Gaussian noise (AWGN), the phase offset, and the residual carrier frequency. However, in practice, the transmitted signal may propagate through various additional impairment environments including fading. The problem of classifying communication signals in presence of fading has received less attention in the literature. The Bayes classifier was studied in [7] for BPSK and QPSK modulations. An hybrid likelihood-based solution was studied in [3] for QAM modulations. The proposed methodology consists of estimating the unknown parameters by the method of moments and plugging these estimates in the likelihoods. The main contribution of this paper is to extend the classification rule studied in [8] for modulations subjected to Rayleigh fading. The proposed strategy is similar to the one developed in [3] except the unknown parameters (residual carrier frequency, phase offset, and fading amplitude) are estimated by using the minimum mean square error estimator. The numerical problems related to this estimator are circumvented by using Markov chain Monte Carlo (MCMC) methods. Note that the main novelty of the proposed classification rule with respect to [8] is that the fading amplitude is estimated (contrary to [8]).

This paper is organized as follows. Section 2 presents the signal model used for modulation classification. Section 3 recalls the well-known maximum likelihood classifier, which minimizes the probability of error, for known residual carrier frequency, phase offset, and fading amplitude. The problem of estimating the parameters unknown to the receiver is addressed in Section 4. A plug-in classifier combining parameter estimation and maximum likelihood classification is also proposed. Simulation results and conclusions are reported in Sections 5 and 6.

2. SIGNAL MODEL AND ASSUMPTIONS

This work considers a synchronous transmission scheme over a Rayleigh fading channel. This kind of transmission yields residual carrier frequency and phase offsets due to imperfect coherent downconversion. We assume here that there is no residual channel effects and that the amplitude factor is random due to fading as in [5, 7]. However, this study could be extended to more general models including a residual channel and timing errors (as in [9]). After preprocessing, the baseband complex envelope of the received signal sampled at one sample per symbol at the output of a matched filter can be written as in [1]:

$$x_k = \alpha e^{j(\pi \frac{k}{N_s} f_r + \phi)} s_k + n_k, \quad k = 1, 2, ..., N_s, \qquad (1)$$

where

- N_s is the number of symbols in the observation interval,
- s_k is an i.i.d. symbol sequence drawn from one of c constellations denoted $\{\omega_1, \omega_2, ..., \omega_c\}$, where ω_j is a set of M_j complex numbers $\{S_1, S_2, ..., S_{M_j}\}$,
- ϕ is a phase offset (resulting from fading phase and synchronization errors),
- $f_r = 2N_s(f_c - \hat{f}_c) \in (-1/2, 1/2]$ is a normalized residual carrier frequency also called frequency offset (f_c is the

carrier frequency and $\hat{f}_c$ is the frequency of the local oscillator). Note that these notations imply that f_r is the constellation rotation whose maximum value is $\pi/2$ for $k = N_s$),

- α is the unknown real amplitude factor,
- n_k is an independent and identically distributed (i.i.d.) complex Gaussian noise sequence which has zero-mean and variance σ_n^2 (the real and imaginary components of n_k are independent and identically distributed).[1]

When the signal is transmitted through a slow flat fading channel, the attenuation factor α can be regarded as a random variable whose probability density function (pdf) is Rayleigh

$$p(\alpha) = \frac{\alpha}{\sigma_\alpha^2} \exp\left(-\frac{\alpha^2}{2\sigma_\alpha^2}\right) \mathbb{I}_{\mathbb{R}^+}(\alpha), \tag{2}$$

where $\mathbb{I}_{\mathbb{R}^+}(.)$ is the indicator function on $\mathbb{R}^+$ (i.e. $\mathbb{I}_{\mathbb{R}^+}(\alpha) = 1$ if $\alpha > 0$ and 0 else).

3. ML CLASSIFIER (KNOWN PARAMETERS)

Bayes theory provides a minimum error-rate classifier by finding the maximum a posteriori probabilities $P(\omega_j|x)$, for $j = 1, 2, ..., c$. If all modulations are equally-likely, the optimal Bayes classifier reduces to the Maximum Likelihood (ML) classifier. The ML classifier selects the modulation of the samples $x = (x_1, x_2, ..., x_{N_s})$ as the one that maximizes the pdf $p(x|\omega_j)$. Such problem was studied in [10] in the ideal situation where α, f_r, ϕ and σ_n^2 are known a priori. The ML classifier can be defined as follows:

$$\text{Assign } x \text{ to } \omega_i \text{ if } l(x|\omega_i) \geq l(x|\omega_j), \ \forall j = 1, ..., c,$$

where

$$l(x|\omega_j) = \sum_{k=1}^{N_s} \ln\left[\frac{1}{M_j}\sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_n^2} \parallel x_k - S_i \parallel^2\right)\right] \tag{3}$$

is obtained after dropping constants in the log-likelihood of the observed signal. It is important to note that knowing all parameters α, f_r, ϕ and σ_n^2 is unrealistic in most communication systems. However, this assumption allows to obtain a reference to which suboptimal classifiers can be compared. More precisely, this ideal classifier provides an upper bound (for instance, in terms of probability of correct classification) of the expected performance for a digital modulation classifier.

4. THE PLUG-IN MCMC CLASSIFIER

This section studies a plug-in classifier for classifying digital modulations subjected to Rayleigh fading. We assume that the received signal amplitude varies from one observation interval to another and is unknown to the receiver. This assumption is realistic in a slow fading context and has been used in [7]. In this case, the classifier has to mitigate the amplitude changes to yield good classification performance. One solution to this problem is to assign some prior distribution to fading amplitude and phase, then integrate out these parameters from the likelihood. However, this strategy yields classification rules with exponential implementation complexity [3]. An alternative is to estimate the unknown parameters and then replace the unknown parameters by their estimates in the likelihood. This strategy, sometimes referred to as *plug-in rule*, has shown good classification properties in fading environment [3]. This section studies a *plug-in rule* which estimates the phase offset, residual carrier frequency and fading parameters for modulation classification purposes. Note that

4.1. Plug-in rule

Denote as $\theta = (f_r, \phi, \alpha)$ the unknown parameter vector. The *plug-in rule* is defined as follows:

$$\text{assign } x \text{ to } \omega_i \text{ if } l(x|\hat{\theta}_i) \geq l(x|\hat{\theta}_j), \forall j = 1, ..., c, \tag{4}$$

where $l(x|\hat{\theta}_j)$ is the logarithm of the likelihood associated to class ω_j (whose constellation consists of M_j symbols $S_1, S_2, ..., S_{M_j}$)

$$l(x|\hat{\theta}_j) = \sum_{k=1}^{N_s} \ln\left[\frac{1}{M_j}\sum_{i=1}^{M_j} \exp\left(-\frac{1}{\sigma_n^2} \parallel x_k - y_i \parallel^2\right)\right], \tag{5}$$

and

$$y_i = \hat{\alpha} S_i e^{j(\pi \frac{k}{N_s} \hat{f}_r + \hat{\phi})}.$$

The plug-in rule can be used as soon as estimates of the unknown parameter vector θ can be obtained conditionally upon each class ω_j. This is the purpose of the next subsection.

4.2. Parameter estimation

Estimating the parameter vector θ can be made by using the method of moments as in [3]. However, Bayesian estimators are often preferred because of their asymptotic properties. A Bayesian estimation technique was studied in [8] to estimate the unknown phase offset, carrier frequency and residual channel in absence of fading. This section shows how the method can be extended to signals subjected to fading. More precisely, the unknown parameter vector $\theta = (f_r, \phi, \alpha)$ is estimated conditionally to each possible class ω_j according to the Minimum Mean Square Error (MMSE) principle (which minimizes the standard quadratic cost function $E[(\hat{\theta} - \theta)^2|\omega_j]$)

$$\hat{\theta}_{\text{MMSE}} = E[\theta|x, \omega_j]. \tag{6}$$

The subscript ω_j will be removed in the rest of this paper, for brevity. Obviously, a closed-form expression for the MMSE estimator of θ cannot be obtained. However, the MMSE estimate (which is the mean of the *a posteriori* density) can be approximated as follows

$$\hat{\theta}_{\text{MMSE}} = \int \theta p(\theta|x) d\theta \simeq \frac{1}{N}\sum_{i=1}^{N} \theta^i, \tag{7}$$

where $\theta^i, i = 1, ..., N$ are samples drawn from $\theta^i \sim p(\theta|x)$. This result can be used to approximate the MMSE estimator $\hat{\theta}_{\text{MMSE}}$, as soon as it is possible to generate samples θ^i distributed according to $p(\theta|x)$. This paper proposes to generate θ_i by using the Metropolis-Hastings (MH) algorithm. The MH algorithm consists of drawing samples distributed according to $p(\theta|x)$ by running an ergodic Markov chain whose stationary distribution is the target distribution $p(\theta|x)$. The reader is invited to consult [11] for more details. The Markov chain state space and current state are denoted by Ω and $\theta^n = (f_r^n, \phi^n, \alpha^n) \in \Omega$, respectively. At each

[1]The parameter σ_n^2 is assumed to be known without loss of generality. Indeed, an estimate is usually available in practice, as explained in [1].

iteration, a candidate z is drawn according to an instrumental distribution $q(z|\theta^n)$. This candidate is accepted with the following probability:

$$\alpha(\theta^n, z) = \min\left\{1, \frac{p(z|x)q(\theta^n|z)}{p(\theta^n|x)q(z|\theta^n)}\right\}. \quad (8)$$

4.3. Instrumental Distribution

A fundamental property of the MH algorithm is that any instrumental distribution $q(z|\theta^n)$ can be chosen, provided that the support of $p(\cdot|x)$ is contained in the support of $q(z|\theta^n)$ [11]. This paper proposes to draw z from a local perturbation of the previous sample, i.e., $z = \theta^n + \epsilon$, leading to the well-known random-walk MH algorithm. In this case, the instrumental distribution is of the form $q(z|\theta^n) = g(z - \theta^n)$. Interestingly, the choice of a symmetric distribution for g leads to an acceptance probability which is independent on q.

Instead of updating the whole of θ *en bloc*, it is often more convenient and computationally efficient to divide θ into k blocks and to update each block one-at-a-time. This procedure has been suggested by many authors (see [11] for more details) and has been shown to improve the mixing property of the sampler. Here we propose to update θ one component at-a-time. Such strategy, indeed, exhibits good performance in classification of digital modulations, as shown in Section 5.

4.4. Reducing the computational complexity

The acceptance probability (8) depends on the pdfs $p(z|x)$ and $p(\theta^n|x)$ whose computation requires to evaluate summations of logarithm functions. This operation can be easily and efficiently conducted on MATLAB. However, in practical applications where a Digital Signal Processor (DSP) has to be used, the evaluation of a logarithm function is too expensive. Instead, by denoting $a_M = \max_{i=1}^{p} a_i$, the following approximation can be used:

$$\ln\left(\sum_{i=1}^{p} e^{a_i}\right) = a_M + \ln\left(1 + \sum_{i\neq M} e^{a_i - a_M}\right),$$

$$\simeq a_M = \max_{i=1}^{p} a_i.$$

By applying this result to $a_i = -\frac{1}{\sigma_n^2} \parallel x_k - y_i \parallel^2$, the following result can be obtained

$$l(x|\omega_j) \propto \sum_{k=1}^{N_s}\left[-\ln M_j + \ln\left(\sum_{i=1}^{M_j} e^{a_i}\right)\right],$$

$$\simeq -N_s \ln M_j - \frac{1}{\sigma_n^2}\sum_{k=1}^{N_s}\left(\max_{i=1}^{M_j} \parallel x_k - y_i \parallel^2\right). \quad (9)$$

This last expression reduces the computational cost required to evaluate the likelihood. The corresponding loss of performance is not critical in most simulations that have been conducted. This point will be illustrated in Section 5.3.3.

5. SIMULATION RESULTS

Many simulations have been carried out to evaluate the performance of the plug-in classifier. This section focuses on a four-class problem $\Omega_4 = \{BPSK, QPSK, 8PSK, 16QAM\}$ which has already been considered in the literature [1]. All constellations have been normalized (unit energy) yielding the following signal-to-noise ratio (SNR) in decibels

$$SNR = 10\log_{10}(1/\sigma_n^2).$$

5.1. Fading channel

Flat fading mobile radio channels are usually characterized by the following frequency response:

$$S(f) = \frac{1}{2\pi f_d}\left[1 - \left(\frac{f}{f_d}\right)^2\right]^{-1/2} \mathbb{I}_{[-f_d, f_d]}(f). \quad (10)$$

The output of this channel can be generated by filtering a complex white Gaussian sequence with a low-pass Butterworth filter. The cutoff frequency of this filter is the product of the symbol duration T by the Doppler shift f_d due to vehicle motion. It is possible to generate slow or fast fading channels, depending on the value of f_dT. As an example, Fig. 1 shows the output of the Butterworth filter for two different values of f_d and $T = 1$. These figures clearly show that a large (resp. small) value of f_dT induces fast (low) fading amplitude variations.

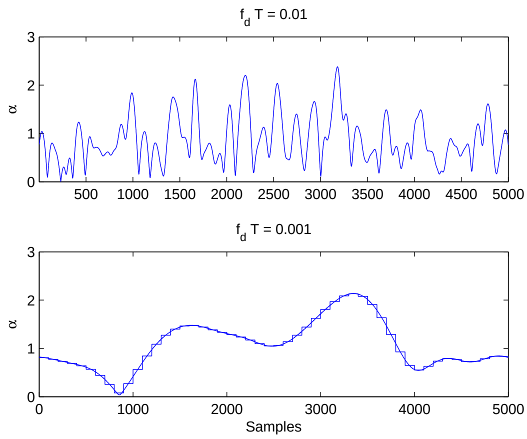


Fig. 1. Fading amplitude versus time for $f_dT = 0.01$ (Top) and $f_dT = 0.001$ (Bottom).

The simulations performed in this paper have been obtained for $f_dT = 0.001$. In this case, the fading amplitude α can be assumed approximately constant for each group of $N_s = 100$ consecutive symbols s_k. However, to consider the variations of the fading amplitude as a function of time, the values of parameter α for different Monte Carlo runs have been obtained by averaging the envelope of the Butterworth filter output over 100 consecutive samples. This point is illustrated on the bottom figure 1 which compares the real fading amplitude with its piecewise constant approximated value.

5.2. Parameter estimation

This section illustrates the performance of the MCMC-based MMSE estimator summarized in 4.2. The unknown parameter vector θ has been estimated on each burst of 100 symbols by running a Markov chain with 1000 samples including 500 burn-in samples (i.e., the first 500 samples generated by the MH algorithm have not been used for the estimation). The simulation has been conducted for a BPSK constellation with a signal to noise ratio SNR $= 5dB$. Moreover, the residual carrier frequency is constant ($f_r = 0$ without loss of generality), the random phase ϕ is uniformly distributed on the interval $[-\pi/4, \pi/4]$ and the fading amplitude is distributed according to a Rayleigh distribution (see 5.1 for more details). The actual values of the unknown parameters (continuous lines) and the corresponding estimates (circles) are depicted on Fig. 2. These results clearly show the accuracy of the proposed estimation methodology.

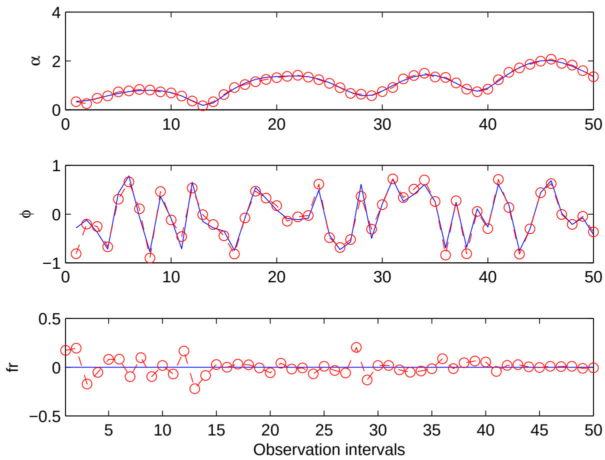


Fig. 2. Unknown parameters and their estimates for fading amplitude (top), phase offset (middle), and residual carrier frequency (bottom).

5.3. Classification

The performance of a classifier can be measured by the average probability of correct classification P_{cc} defined by

$$P_{cc} = \frac{1}{M}\sum_{i=1}^{M} P\left[\text{assigning } x \text{ to } \omega_i | x \in \omega_i\right],$$

where M is the number of classes (here $M = 4$ since $\Omega_4 = \{BPSK, QPSK, 8PSK, 16QAM\}$). All simulations presented in this section have been obtained from 1000 trials belonging to each class ω_i (i.e. a total of 4000 trials). The number of symbols in each observation interval is $N_s = 100$.

5.3.1. *Performance versus SNR*

The first simulation results depicted on Fig. 3 compare the average probability of correct classification for different classifiers as a function of SNR:

- the circle curve corresponds to the ML classifier (labeled Ref) which assumes the parameter vector θ is known,
- the star curve is obtained for the MCMC plug-in classifier (labeled MCMC),
- the diamond curve stands for the classifier derived in [3] (labeled MOM since the unknown parameter vector θ is estimated by the method of moments).

The simulation scenario is similar to the example of Section 4.2 (uniform phase offset, $f_r = 0$ and Rayleigh fading amplitude). Note again that the ML classifier cannot be implemented in practical applications since it assumes that the parameter vector θ is perfectly known. Thus, it provides an upper bound of classification performance. Fig. 3 shows that the MCMC plug-in classifier outperforms the MOM classifier specially at high SNRs. The figure also shows that the average probability of correct classification for the MCMC plug-in classifier approaches the optimal one provided by the ML classifier for high SNRs.

Fig. 4 shows the probability of correct classification of the MCMC plug-in classifier for each candidate modulation (BPSK, QPSK, 8PSK, and 16QAM). This figure indicates that modulations with large numbers of constellation points (8PSK and 16QAM) are more difficult to classify than modulations with small numbers of points (BPSK, QPSK) for the same SNR.

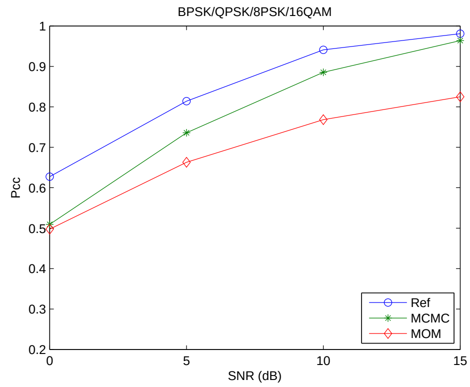


Fig. 3. Average probability of correct classification versus SNR for Ω in a slow flat fading scenario.

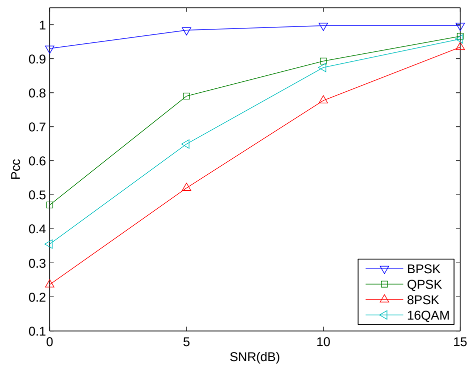
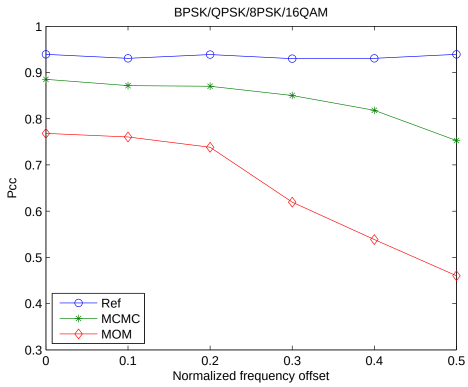


Fig. 4. Probability of correct classification versus SNR for BPSK, QPSK, 8PSK and 16QAM signals in a slow flat fading scenario.

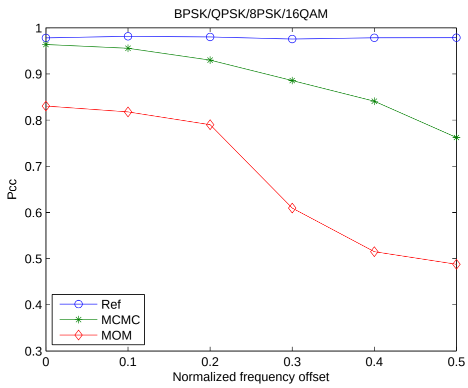
5.3.2. *Performance versus f_r*

Figure 5(b) and 5(a) show the effect of frequency offset (due to inaccuracies of the local oscillators) on classification performance for SNR $= 10$ and 15 dB. When the frequency offset is less than

0.2, the MCMC classifier still performs reasonably well. The classification performance drops very slightly at the frequency offset $f_r = 0.3$ and tends to degrade much further. However, it is important to note that the MCMC-based classifier is more robust to frequency offset than the MOM classifier particularly for $f_r > 0.2$.



(a) SNR $= 10$ dB



(b) SNR $= 15$ dB

Fig. 5. Average classification performance versus f_r in a slow flat fading scenario for different SNRs.

5.3.3. *Approximated Classification Rule*

The last simulation results illustrate the performance of the approximate MCMC classifier which uses (9) instead of (5). Fig. 6 compares the estimated posterior distributions of the residual carrier frequency f_r obtained by using the exact (dotted line) and approximate (continuous line) MCMC samplers. The number of burn-in iterations for this example is 500 and the posteriors have been estimated by using the 2500 last Markov chain samples. The two distributions are clearly in good agreement, showing that the approximate MCMC sampler can be used if the computational cost of the algorithm is an important issue.

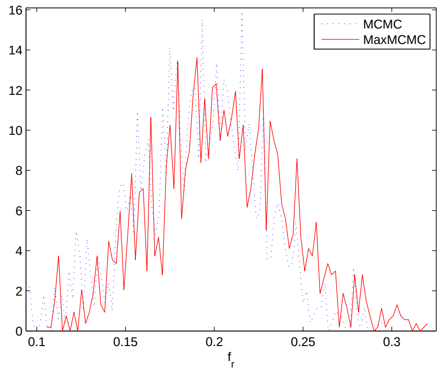


Fig. 6. Estimated posteriors for f_r obtained for exact (dotted line) and approximated (continuous line) MCMC samplers ($f_r = 0.2$).

6. CONCLUSIONS

This paper studied the important problem of digital modulation classification in a fading environment. The proposed classifier estimated the parameters unknown to the receiver (phase offset, residual carrier frequency and fading amplitude). The estimates were then plugged into the class-conditional densities, resulting in a so-called plug-in classifier. The proposed classifier showed good performance. Reducing the computational complexity of the proposed strategy is an important problem which is currently under investigation.

7. REFERENCES

[1] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, 2000.

[2] K. M. Chugg, C. S. Long, and A. Polydoros, "Combined likelihood power estimation and multiple hypothesis modulation classification," *Proc. ASILOMAR*, vol. 2, pp. 1137–1141, 1996.

[3] A. Abdi, O. A. Dobre, R. Chauchy, Y. Bar-Ness, and W. Su, "Modulation classification in fading channels using antenna arrays," *Proc. IEEE MILCOM*, vol. 1, pp. 211–217, 2004.

[4] N. Lay and A. Polydoros, "Modulation classification of signals in unknown ISI environments," *Proc. IEEE MILCOM*, vol. 1, pp. 170–174, Nov. 1995.

[5] L. Hong and K. C. Ho, "An antenna array likelihood modulation classifier for BPSK and QPSK signals," *Proc. IEEE MILCOM*, vol. 1, pp. 647–651, Oct. 2002.

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

[7] L. Hong and K. C. Ho, "Classification of BPSK and QPSK signals with unknown signal level using the Bayes technique," *Int. Conf. Circuits and Systems*, vol. 4, pp. 1–4, May 2003.

[8] S. Lesage, J. Tourneret, and P. M. Djuric, "Classification of digital modulation by MCMC sampling," *Proc. IEEE ICASSP*, pp. 2553–2555, 2001.

[9] J. Venalainen, L. Terho, and V. Koivunen, "Modulation classification in fading multipath channel," in *Proc. 36th Asilomar Conf. Signals, Syst., Computers*, (Pacific Grove, CA), pp. 1890–1894, Nov. 2002.

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

[11] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, "Introducing Markov chain Monte Carlo," in *Markov Chain Monte Carlo in Practice*, pp. 1–19, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds.. London, U.K.: Chapman & Hall, 1996.