Near ML Modulation Classification

Dongwoon Bai dongwoon.bai@post.harvard.edu
Jungwon Lee jungwon@stanfordalumni.org
Sungsoo Kim sungsoo04.kim@samsung.com
Inyup Kang inyup.kang@samsung.com

Mobile Solutions Lab, Samsung US R&D Center, San Diego, CA, USA

***Abstract*—This paper deals with the problem of classification of digital modulation. In particular, we develop and propose a practical modulation classification scheme based on the likelihood of observations. While ML classification is well known and shows the optimal performance, its computational complexity prevents it from being easily implemented in hardware. On the contrary, our proposed scheme has low computational complexity and near optimal classification performance. Moreover, this scheme is designed to perform in fast fading channels. It is shown that our proposed classifier takes advantage of the channel variation without loosing near optimality.**

I. Introduction

This paper deals with the problem of classification of digital modulation whose constellation consists of a finite number of points. A classifier needs to make an intelligent selection among a set of candidate modulations based on a series of noisy observations. While this problem has been traditionally addressed for military applications, a number of commercial applications also draw interest recently such as cognitive radio, software defined radio, and interference identification.

In [1], the benefit of joint detection has been investigated assuming that interference modulation is known. If not, for mitigation, interference modulation needs to be identified using a modulation classifier beforehand. For Long Term Evolution (LTE) systems in particular, multi-user multiple-input multiple-output (MU-MIMO) schemes can be deployed [2]. Because the precoding matrix may not be perfectly orthogonal due to imperfect channel state information at the transmitter, MU-MIMO transmission is inherently accompanied by multi-user interference. In [3], this challenge has been addressed by designing an interference-aware receiver capable of jointly detecting desired signal and multi-user interference. However, the biggest hurdle for this approach is that the modulation index of multi-user interference is not provided to the co-scheduled receiver in existing LTE systems. Keeping this setup in mind, we propose a practical yet near optimal modulation classification scheme in this paper, which can be further extended for identification of LTE multi-user interference.

Modulation classification has been investigated in two directions, i.e., likelihood-based approaches [4]–[9] and feature-based approaches as surveyed in [10]. Feature-based methods are usually designed to be robust to model mismatch or parameter estimation error and to perform without the knowledge of some parameters. If all parameters can be estimated reliably, they do not generally perform better than likelihood-based classifiers and may not be easily modified for fast fading channels.

On the other hand, it is well known that under the assumption of a uniform prior, a decision based on the method of maximum likelihood (ML) minimizes the average probability of error for classification of a finite number of candidates [11]. Although the method of ML is optimal, the following two issues of ML remain unsolved for modulation classification. First, the complexity of likelihood computation is intense for modulation classification, and thus it may not be suitable for real time applications. Second, a number of parameters may not be known to the receiver (e.g., timing, phase and frequency offsets, the channel gain, the noise variance) and the average likelihood can be used instead. This makes the likelihood computation even harder since the likelihood function needs to be averaged over unknown parameters. Another approach is to estimate unknown parameters and perform likelihood computation sequentially [9] or jointly.

Assuming all the parameters can be estimated reliably, ML modulation classification has been proposed in [4], [5]. The subsequent efforts to reduce the computational complexity of ML modulation classification have been followed. In [6], approximate ML classification is proposed, whose computational complexity is much reduced from that of ML. However, its performance is inferior to ML classification at low and moderate signal-to-noise ratio (SNR). The use of three-dimensional look-up tables (LUTs) is proposed to reduce the complexity of likelihood computation in [7], [8] but their LUT sizes are significantly large.

In this paper, we also focus on reducing the computational complexity of likelihood-based classification, i.e., resolving the first issue. Unlike those efforts in [6]–[8], we propose to compute the most discriminating portion of likelihood online and substitute the rest with the expected value, which is precomputed offline and whose LUT size is small, e.g., the minimum suggested size of LUT in [8] is $15,000$ times larger than our proposed LUT. We obtain the upper bound and prove the convergence of the residual error of this likelihood bias compensation showing that it is marginal and diminishing. Moreover, the proposed scheme has low computational complexity and near optimal classification performance for a wide range of SNR. Our classifier is also designed to perform in fast fading channels and take advantage of the channel variation without losing near optimality. If the prior probability of each modulation is known, our classification method can easily utilize this prior since it is based on the likelihood.

This paper is organized as follows. In Section II, we present our system model. In Section III, we focus on modulation

978-1-4673-1881-5/12/$31.00 ©2012 IEEE

classification in the additive white Gaussian noise (AWGN) channel, where optimal ML classification is studied and near ML classification scheme is proposed. In Section IV, we investigate how these schemes should be modified for fast fading channels. We provide simulation results in Section V. Our conclusions are drawn in Section VI.

II. The System Model

The baseband received complex signal is given by

$$y_k = h_k x_k + z_k, \quad k = 1, \cdots, K \tag{1}$$

where k is the sample index, K is the number of samples, h_k is the complex channel gain, x_k is the complex transmitted symbol, and z_k is the zero-mean circularly symmetric independent and identically distributed (i.i.d.) complex Gaussian noise with the noise variance $\sigma^2 = \mathbb{E}[|z_k|^2]$. Let C_n be one unknown candidate modulation, whose cardinality, the number of complex constellation points is given by $|C_n|$. However, C_n does not change over a block of K samples, e.g., $x_k \in C_n$ for all $k = 1, \cdots, K$. The set of candidate modulations is given by $\Gamma = \{C_n | n = 0, \cdots, N-1\}$. The prior probability of each modulation is denoted by $q_n \triangleq \Pr\{C_n\}$ subject to $\sum_{n=0}^{N-1} q_n = 1$. While we mostly focus on the case of a uniform prior $q_n = 1/N$, the extension using a nonuniform prior will be discussed. The classifier selects the most probable modulation index $\hat{n}$ (i.e., $C_{\hat{n}} \in \Gamma$) based on K noisy observations of $y_1, \cdots, y_K$. Note that the transmitted symbol x_k does not need to be estimated individually.

While the proposed classification method can be applied for any digital modulation classification, we show the classification performance when Γ consists of quadrature amplitude modulation (QAM) with different modulation orders focusing on the case of LTE systems. We include the modulation of NS (no signaling) $C_0 = \{0\} \in \Gamma$ as a special type of QAM. This covers the case when the receiver attempts to blindly decode control or data messages without knowing their presence. Except C_0, if the transmitted symbol exists, we assume that it is normalized such that $\frac{1}{|C_n|}\sum_{s \in C_n} |s|^2 = 1$ for C_n of $n = 1, \cdots, N-1$ and each constellation point is equally probable. Thus, the SNR is given by $\mathrm{E}[|h_k|^2]/\sigma^2$. Although classification schemes in this paper can be applied any general set of Γ, the simulation results are shown for QAM classification of NS (C_0), 4QAM (C_1), 16QAM (C_2), and 64QAM (C_3), i.e., $N = 4$. In the next section, we explain classification methods assuming the AWGN channel.

III. Modulation Classification in the AWGN Channel

In this section, without loss of generality, we assume $h_k = 1$ and the SNR is reflected in σ, i.e., SNR$= 1/\sigma^2$. It is assumed that SNR is known to the receiver or can be reliably estimated.

A. ML Classification

The ML modulation classifier has been studied in [4], [5]. ML classification is known to be an optimal scheme for decision problems (e.g., modulation classification) meaning that it minimizes the average probability of decision error with a uniform prior [11]. The ML classification of modulation can be summarized as follows. The average log-likelihood of C_n based on K observation $y_1, \cdots, y_K$ is given by

$$l_n \triangleq \frac{1}{K}\sum_{k=1}^{K} \ln \sum_{s \in C_n} \exp\left(-\frac{1}{\sigma^2}|y_k - s|^2\right) - \ln \pi\sigma^2 |C_n|, \tag{2}$$

where the natural logarithm $\ln(\cdot)$ is assumed without loss of generality. Since the average log-likelihood computation requires the knowledge of SNR (equivalently, the noise variance), l_n is a function of the noise variance σ^2 in (2). Any classification schemes that do not explore the knowledge of channel statistics (e.g., the noise variance) can only be strictly suboptimum. Note that the only issue preventing the simple computation of the average log-likelihood is the summation over C_n inside of the logarithm in (2). Assuming a uniform prior q_n, the ML classifier makes the following decision

$$\hat{n} = \arg\max l_n. \tag{3}$$

However, under a nonuniform prior q_n, the maximum a posteriori (MAP) decision rule becomes

$$\hat{n} = \arg\max \left(l_n + \frac{1}{K}\ln q_n\right). \tag{4}$$

Note that the decision rules of all other modulation classification methods in this paper can be easily extended to MAP decision rules similarly.

B. Approximate ML Classification

While the ML classifier minimizes the average probability of decision error for equally probable modulations, the difficulty of ML classification lies on the complexity of log-likelihood computation. In this paper, our effort is dedicated to find a suboptimal but near optimal scheme with reduced computational complexity. To reduce the complexity of computing the summation of the exponential function of the distance from various constellation points to the received signal, it has been suggested in [7], [8] to build a three-dimensional (two dimensions for real and imaginary parts of y, and one dimension for σ) LUT of the function

$$f_n(y, \sigma) = \ln\left[\sum_{s \in C_n} \exp\left(-\frac{1}{\sigma^2}|y - s|^2\right)\right] \tag{5}$$

in (2) for each hypothetical modulation C_n. Without hardware design constraint, any complicated function can be implemented by LUT if the function is smooth enough and the input range is limited. However, in practice, this implementation for handheld devices is infeasible if the size of LUT is huge. In particular, the LUT size for (5) is huge, e.g., 4.6–18.4 Mbits per modulation is suggested in [8].

On the other hand, we can approximate the average log-likelihood using only the distance from the closest constellation point to the received signal y. Let us define

$$s_{\min} \triangleq \arg\min_{s \in C_n} |y - s|. \tag{6}$$

Note that $s_{\min}$ is a function of y and C_n which can be omitted when they are obvious from the context. The average log-likelihood in (2) can be approximated as

$$\begin{aligned} l_n &= \frac{1}{K}\sum_{k=1}^{K}\left(-\frac{1}{\sigma^2}|y_k - s_{\min}|^2\right) - \ln \pi\sigma^2 |C_n| \\ &\quad \underbrace{+\frac{1}{K}\sum_{k=1}^{K}\ln\left(1+\frac{\sum_{s\in C_n/\{s_{\min}\}}\exp\left(-\frac{1}{\sigma^2}|y_k-s|^2\right)}{\exp\left(-\frac{1}{\sigma^2}|y_k-s_{\min}|^2\right)}\right)}_{\triangleq b_n} \quad (7) \\ &\approx \frac{1}{K}\sum_{k=1}^{K}\left(-\frac{1}{\sigma^2}|y_k - s_{\min}|^2\right) - \ln \pi\sigma^2 |C_n| \triangleq \hat{l}_n. \quad (8) \end{aligned}$$

It is assumed that the exponential of the smallest distance $\exp\left(-\frac{1}{\sigma^2}|y_k - s_{\min}|^2\right)$ is relatively larger than the sum of remaining terms $\sum_{s\in C_n/\{s_{\min}\}}\exp\left(-\frac{1}{\sigma^2}|y_k-s|^2\right)$ and thus their ratio becomes close to zero, which makes the last term b_n in (7) close to zero. This is not a bad assumption to make at high SNR. Note that b_n is equal to zero for modulation of one constellation point, (e.g., NS) since $C_n/\{s_{\min}\}$ becomes an empty set. However, b_n is a positive number in general. Since the approximation $\hat{l}_n$ substantially underestimates the average log-likelihood l_n, it is significantly biased at low and intermediate SNR while the use of $\hat{l}_n$ can yield good classification performance at high SNR.[1] This approximation is studied in [6], where it is also suggested (but not explicitly demonstrated) to use the second largest exponential term inside the logarithm at the first line of (7) and ignore the remaining bias. For this, we need to compute the Jacobian logarithm based on a LUT anyway [12], which may not eliminate bias completely at low and some intermediate SNR. To remove bias this way, we need to keep computing additional terms recursively using approximation until the bias is negligible, which is computationally intense. However, if only $\hat{l}_n$ is used, the complexity is reduced significantly, which only requires the closest constellation point selection (implemented using a simple circuit of hard slicing) and the computation of the squared Euclidean distance.

Our approach for this problem is different as follows. First, let us define each element of the last term b_n in (7) as

$$\begin{aligned} &\beta_n(y,\sigma) \\ &\quad \triangleq \ln\left(1+\frac{\sum_{s\in C_n/\{s_{\min}\}}\exp\left(-\frac{1}{\sigma^2}|y-s|^2\right)}{\exp\left(-\frac{1}{\sigma^2}|y-s_{\min}|^2\right)}\right), \quad (9) \end{aligned}$$

which is a nonnegative function of observation y, the noise variance σ^2, and the index of hypothetical modulation n. Note that b_n is a sample average of $\beta_n(y,\sigma) \geq 0$. Let C_m denote input (true) modulation of K observations $y_1, \cdots, y_K$. The distribution of y_k depends on C_m and σ^2, and so does the distribution of b_n. This sample average b_n can be considered as an implicit function of C_m and σ, and contributes to the negative bias in the average log-likelihood approximation in (8). By the central limit theorem, the sample average b_n converges to the ensemble average as

$$\lim_{K\to\infty} b_n = \mathbb{E}_{Y(C_m)}\left[\beta_n(Y(C_m),\sigma)\right] \triangleq B_n(C_m,\sigma), \quad (10)$$

where $B_n(C_m,\sigma)$ denotes the ensemble average of $\beta_n(Y(C_m),\sigma)$ and Y denotes a random variable following the distribution of y_k in (1), which is a function of the input modulation C_m. Thus, instead of computing each realization of the sample average b_n online, we can substitute this with the *precomputed ensemble average* $B_n(C_m,\sigma)$, whose value can be obtained by carrying out two-dimensional numerical integration offline. In other words, our approximate ML modulation classifier computes only essential discriminating portion of the likelihood online and substitutes the rest with the precomputed ensemble average. This is a major difference from the approach of [7], [8] where they perform most of computation using huge LUTs of three dimensions and the approach of [6] whose performance is much inferior than that of ML due to its crude approximation of the likelihood.

The challenge here is that we do not know the input modulation C_m of $B_n(C_m,\sigma)$. However, it is observed that $B_n(C_m,\sigma)$ not only changes marginally for different C_m but also takes the minimum value when $m = n$ for a wide range of SNR except very low SNR in the case of four class QAM classification for NS, 4QAM, 16QAM, and 64QAM. We propose the simple bias compensation using $B_n(C_n,\sigma)$ in substitute for $B_n(C_m,\sigma)$. The corresponding decision rule of approximate ML classification given by

$$\hat{n} = \arg\max(\hat{l}_n + B_n(C_n,\sigma)) \quad (11)$$

works reasonably well while the amount of bias compensation can be further optimized for slight improvement in classification performance using the Gaussian approximation of log-likelihood in [4], [5]. Two terms of (11), B_n and $\hat{l}_n$, depend on the noise variance σ^2 (or SNR equivalently), which is known to the receiver. Note that the entry of this LUT needs to be computed for each instantaneous SNR and hypothetical modulation C_n. The size of LUT is three times of the number of quantized SNR levels in the case of four class QAM classification for NS, 4QAM, 16QAM, and 64QAM because NS does not have any bias. Thus, the LUT size is quite small, e.g., only 1.2 Kbits with 8 bit quantization. The exemplary LUT is given in Table I, which will be used later for simulation. If the SNR of the channel does not match with the SNR entry of the LUT, $B_n(C_n,\sigma)$ can be interpolated.

IV. Modulation Classification in Fading Channels

In this section, we assume flat fading in (1), where the channel gain h_k may not stay constant. We investigate to extend modulation classification for the AWGN channel to fast fading channels. Note that for MU-MIMO transmission in LTE, the channel of multi-user interference can be estimated [3]. Thus, it is assumed that the channel gain h_k is either known to the receiver or estimated with respect to the noise standard deviation. However, we do not put any assumption

[1] In this paper, bias means the bias of average log-likelihood approximation, $\mathbb{E}[\hat{l}_n - l_n]$.

TABLE I
EXEMPLARY LUT OF $B_n(C_n,\sigma)$

SNR (dB)	$B_n(C_n,\sigma)$		
	4QAM	16QAM	64QAM
-20	1.2360	2.5747	3.9362
-19	1.2188	2.5525	3.9114
-18	1.1998	2.5285	3.8849
-17	1.1788	2.5017	3.8554
$\vdots$	$\vdots$	$\vdots$	$\vdots$
30	0	0	0.0000

on statistical properties of the channel gain h_k in developing modulation classification techniques while simulation results are produced under the specific channel model. For the notational convenience, let us embed the instantaneous channel gain h_k inside of σ defining the instantaneous noise variance $\sigma_k^2 \triangleq \sigma^2/|h_k|^2$.

A. ML Classification

The ML classifier for the flat fading channel can be easily extended from the ML classifier for the AWGN channel. The average log-likelihood becomes

$$l_n = \frac{1}{K}\sum_{k=1}^{K} \ln \sum_{s\in C_n} \exp\left(-\frac{1}{\sigma_k^2}|\hat{x}_k - s|^2\right) - \ln \pi\bar{\sigma}^2|C_n|, \quad (12)$$

where $\hat{x}_k \triangleq y_k/h_k$ and $\bar{\sigma} = \exp\left(\frac{1}{K}\sum_{k=1}^{K}\ln\sigma_k\right)$. However, $\bar{\sigma}$ does not need to be computed because it is a common term regardless of modulation. Again, the ML classifier makes the following decision

$$\hat{n} = \arg\max l_n. \quad (13)$$

B. Approximate ML Classification

For approximate ML classification, the average log-likelihood in (12) can be expressed similarly as

$$\begin{aligned} l_n \;=\; & \frac{1}{K}\sum_{k=1}^{K}\left(-\frac{1}{\sigma_k^2}|\hat{x}_k - s_{\min}|^2\right) - \ln\pi\bar{\sigma}^2|C_n| \\ & + \underbrace{\frac{1}{K}\sum_{k=1}^{K}\beta_n(\hat{x}_k,\sigma_k)}_{\triangleq \bar{b}_n} \end{aligned} \quad (14)$$

The expected value of $\beta_n(\hat{x}_k,\sigma_k)$ is given by

$$\mathbb{E}_{\hat{X}(C_m)}[\beta_n(\hat{X}(C_m),\sigma_k)] = B_n(C_m,\sigma_k), \quad (15)$$

where $\hat{X}$ is a random variable following the distribution of $\hat{x}_k$. The distribution of $\bar{b}_n$ is a function of C_m and $\sigma_1,\cdots,\sigma_K$. While $\bar{b}_n$ can be substituted by its expectation $\frac{1}{K}\sum_{k=1}^{K}B_n(C_m,\sigma_k)$, the convergence to the expectation cannot be easily obtained from the conventional central limit theorem as in the AWGN case because $\bar{b}_n$ is the sample average of independent but *nonidentical* distribution. Nevertheless, we have the following theorem showing the same holds for fading channels.

Theorem 1: As $K\to\infty$, $\bar{b}_n$ converges as

$$\bar{b}_n - \frac{1}{K}\sum_{k=1}^{K}B_n(C_m,\sigma_k) \to 0, \quad (16)$$

with probability 1. Moreover, as $K\to\infty$, $\bar{b}_n$ also converges in mean square as

$$\begin{aligned} \mathbb{E}\left[\left|\bar{b}_n - \frac{1}{K}\sum_{k=1}^{K}B_n(C_m,\sigma_k)\right|^2\right] \;&\le\; \frac{1}{K}(\ln|C_n|)^2 \\ &\to\; 0, \end{aligned} \quad (17)$$

where the speed of convergence is bounded by $\frac{1}{K}(\ln|C_n|)^2$. □

Proof: The proof using the strong law of large numbers [13, p.250] will be given in the full paper [14]. ■

Note that this upper bound universally holds regardless of the input modulation C_m, statistical properties of the effective channel σ_k, and the class of the digital modulation scheme of candidate C_n (not necessarily QAM). As in the AWGN channel, since we do not know C_m, we substitute C_m with C_n. Moreover, the same LUT of $B_n(C_n,\sigma)$ built for the AWGN channel such as Table I can be used universally regardless of the channel type. The only difference for fading channels is that the entry of the LUT, $B_n(C_n,\sigma)$ needs to be averaged to obtain $\frac{1}{K}\sum_{k=1}^{K}B_n(C_n,\sigma_k)$

Finally, simulation results at the end of this paper will show that the decision rule of approximate ML classification with the bias compensation given by

$$\hat{n} = \arg\max\left(\hat{l}_n + \frac{1}{K}\sum_{k=1}^{K}B_n(C_n,\sigma_k)\right) \quad (18)$$

works reasonably well.

V. SIMULATION RESULTS

We run simulation of input modulation classification of NS, 4QAM, 16QAM, and 64QAM. The sample size $K = 104$ for the AWGN channel and the fading channel is chosen based on the possible number of samples in one resource block in LTE systems because modulation classification of multi-user interference needs to be performed per resource block [3]. For simulation, we first build a LUT of $B_n(C_n,\sigma)$ offline for approximate ML evaluated at each instantaneous SNR from -20 dB to 30 dB with 1 dB scale as in Table I. Thus, the total number of entries for LUT is 153, which only requires 1224 bits when each entry is represented by 8 bits. This LUT size (1.2 Kbits) is only $1/15,000$ times of the minimum suggested size of LUT (18,400 Kbits) in [8]. We use independently generated h_k following the zero-mean circularly symmetric complex Gaussian distribution to model an i.i.d. Rayleigh fast and flat fading channel. In the fading channel, the approximate ML classifier takes linearly interpolated values of $B_n(C_n,\sigma)$ from the LUT. For extrapolation, the bias is not compensated if the instantaneous SNR becomes larger than the largest entry of SNR (30 dB) in LUT since it is negligible. If the instantaneous

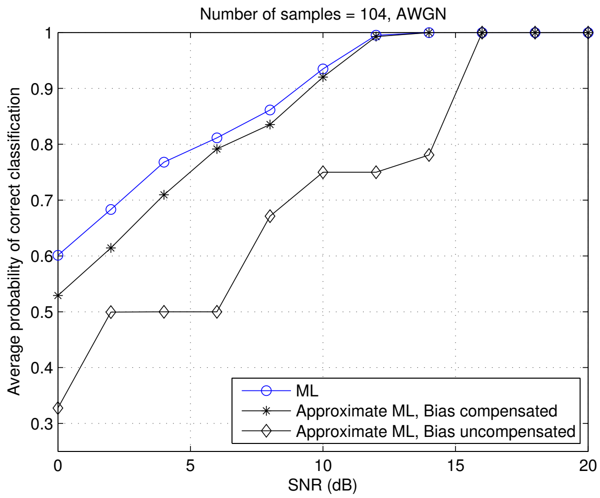

Fig. 1. Average probability of correct decision for $K = 104$ in the AWGN channel.

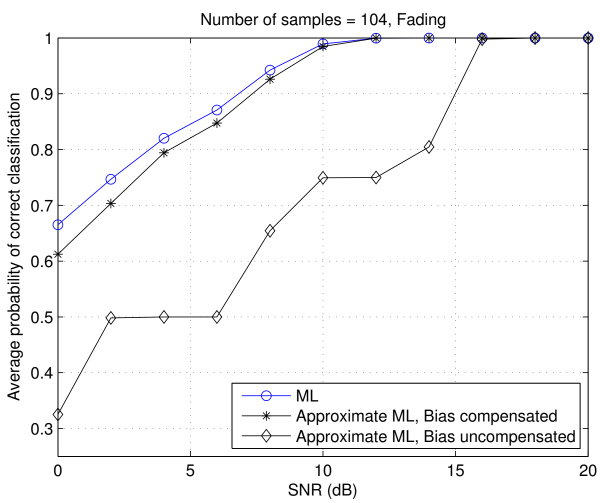

Fig. 2. Average probability of correct decision $K = 104$ in the fading channel.

SNR drops below the smallest entry of SNR (-20 dB) in LUT, the corresponding sample is just ignored because the use of this sample may not add any useful information for the classification purpose.

Fig. 1 and 2 show the average probability of correct decision in the AWGN channel and the fading channel, respectively. The probability is averaged over four input modulations. It can be confirmed that approximate ML classification with bias compensation shows the performance close to that of ML classification for a wide range of SNR while approximate ML classification without bias compensation (i.e., following the decision rule: $\hat{n} = \arg\max \hat{l}_n$ as in [6]) works only at high SNR. It is clear that bias compensation is essential for approximate ML classification. Moreover, near ML performance for the fading channel means that the bias compensation using a LUT does not degrade the approximate ML classification significantly if the interpolation and extrapolation are performed properly. Surprisingly, ML classification and approximate ML classification with bias compensation show better classification performance in the fading channel than in the AWGN channel confirming that both methods takes advantage of the channel variation.

VI. Conclusion

To address the problem of digital modulation classification, we developed and proposed a practical likelihood-based scheme. Our proposed scheme works not only for the AWGN channel but also for all flat fading channels. While ML classification is well known to show the optimal classification performance meaning that it minimizes the average probability of classification error, its computational complexity prevents it from being easily implemented in hardware. Our proposed scheme, approximate ML classification with bias compensation has low computational complexity with the use of the small precomputed LUT. Moreover, simulation results showed that the proposed method has near optimal classification performance and takes advantage of the channel variation in the fast fading channel without loosing near optimality.

References

[1] J. Lee, D. Toumpakaris, and W. Yu, "Interference mitigation via joint detection," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 6, pp. 1172 –1184, Jun. 2011.

[2] D. Bai, C. Park, J. Lee, H. Nguyen, J. Singh, A. Gupta, Z. Pi, T. Kim, C. Lim, M. Kim, and I. Kang, "LTE-Advanced modem design: challenges and perspectives," *IEEE Commun. Mag.*, Feb. 2012.

[3] J. H. Bae, S. Kim, J. Lee, and I. Kang, "Advanced downlink MU-MIMO receiver for 3GPP LTE-A," *accepted for presnetation at IEEE ICC Workshop on Telecommunications: From Research to Standards*, Jun. 2012.

[4] W. Wei and J. Mendel, "A new maximum-likelihood method for modulation classification," in *Asilomar Conference on Signals, Systems and Computers*, vol. 2, Oct.-Nov. 1995, pp. 1132–1136.

[5] ——, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 189–193, Feb. 2000.

[6] T. Yucek and H. Arslan, "A novel sub-optimum maximum-likelihood modulation classification algorithm for adaptive OFDM systems," in *IEEE Wireless Communications and Networking Conference*, vol. 2, Mar. 2004, pp. 739–744.

[7] W. Su, J. Xu, and M. Zhou, "Real-time modulation classification based on maximum likelihood," *IEEE Commun. Lett.*, vol. 12, no. 11, pp. 801–803, Nov. 2008.

[8] J. Xu, W. Su, and M. Zhou, "Software-defined radio equipped with rapid modulation recognition," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1659–1667, May 2010.

[9] W. Headley and C. da Silva, "Asynchronous classification of digital amplitude-phase modulated signals in flat-fading channels," *IEEE Trans. Commun.*, vol. 59, no. 1, pp. 7–12, Jan. 2011.

[10] O. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *Communications, IET*, vol. 1, no. 2, pp. 137–156, Apr. 2007.

[11] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, Inc., 2001.

[12] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[13] P. Billingsley, *Probability and Measure*, 1st ed. John Wiley & Sons, Inc., 1979.

[14] D. Bai, J. Lee, J. Bae, S. Kim, and I. Kang, "Near ML modulation classification and its implication on interference indentification," *in preparation*.