

Automatic Modulation Classification Using Interacting Multiple Model - Kalman Filter for Channel Estimation

Ahmed O. Abdul Salam, *Senior Member, IEEE*, Ray E. Sheriff, Yim-Fun Hu, *Senior Member, IEEE*, Saleh R. Al-Araji, *Senior Member, IEEE*, Kahtan Mezher, *Senior Member, IEEE*

Abstract—A rigorous model for automatic modulation classification (AMC) in cognitive radio (CR) systems is proposed in this paper. This is achieved by exploiting the Kalman filter (KF) integrated with an adaptive interacting multiple model (IMM) for resilient estimation of the channel state information (CSI). A novel approach is proposed, in adding up the square-root singular values (SRSV) of the decomposed channel using the singular value decompositions (SVD) algorithm. This new scheme, termed Frobenius eigenmode transmission (FET), is chiefly intended to maintain the total power of all individual effective eigenmodes, as opposed to keeping only the dominant one. The analysis is applied over multiple-input multiple-output (MIMO) antennas in combination with a Rayleigh fading channel using a quasi likelihood ratio test (QLRT) algorithm for AMC. The expectation-maximization (EM) is employed for recursive computation of the underlying estimation and classification algorithms. Novel simulations demonstrate the advantages of the combined IMM-KF structure when compared to the perfectly known channel and maximum likelihood estimate (MLE), in terms of achieving the targeted optimal performance with the desirable benefit of less computational complexity loads.

Index Terms—Automatic modulation classification, Kalman filter, interacting multiple model, channel estimation.

I. INTRODUCTION

THE desire to embed intelligent features, such as cognitive radio (CR), in emerging communication systems has recently received much attention. This is mainly attributed to the mechanisms that CRs possess to adapt to their changing surrounds [1]. The prime intention of CRs is to promote the awareness of efficient and sustainable spectrum utilization [17]. Unlicensed secondary users (SUs) featuring the intelligent strategies of CR systems are able to opportunistically share the unoccupied spectrum bands with licensed primary users (PUs) in order to achieve better spectral efficiency. This, of course, should occur without harming the

activities of PUs, and SUs should vacate particular spectrum bands upon PUs' request without any untoward delay or disturbance. Specifically, CR systems not only need to perform spectrum sensing (SS) in the interested frequency bands but also need to interrogate the sensed signals to obtain additional information, such as the modulation scheme, in order to design improved dynamic spectrum access (DSA) strategies [18, 19].

As an intermediate stage between signal detection and demodulation, automatic modulation classification (AMC), sometimes referred to as automatic modulation recognition (AMR), plays a significant role in such adaptive aspects, whether of civilian or military nature. AMC mainly aspires to identify the modulation type of a noisy signal from a possible set of known patterns [2, 3]. Such a task should be flawlessly achieved irrespective of transmission constraints, such as large variations of signal-to-noise ratio (SNR) and severe fading. The consistency between AMC and DSA, in addition to adaptive coded modulation (ACM) techniques in general wireless communication systems, and particularly in CRs, is regularly reasserted. This is happening while most developers and operators are investigating emerging mobility scenarios, such as fifth-generation (5G) applications and spectrum opportunities under inconsistent channels [1, 17-19].

Two main categories of AMC techniques are prominent in survey literature [4, 5]: likelihood-based (LB); and feature-based (FB). The LB is considered optimal in the Bayesian sense if the channel state information (CSI) is known to the receiver, and also to the transmitter in the case of cooperative environment. The FB is built on a specific signal feature and is alleged to be less costly. In either method, if the work environment lacks coordination between the transmitter and the receiver, which is the case considered in this paper, then the whole process is conducted blindly. Generally, blind AMC systems are chiefly concerned with the automatic recognition of the modulation scheme of the received signals with limited or no prior knowledge, and hence are foreseen to radically improve the performance of the CR systems.

A number of recent studies in the context of LB and FB categories can be identified in the literature and a few samples of studies are considered here [20-26]. An LB algorithm for automatically identifying different quadrature amplitude modulation (QAM) and phase-shift keying (PSK) modulation schemes was given in [20]. The proposed algorithm maximizes a log-likelihood function (LLF) based on the known probability distribution associated with the phase or amplitude of the received signals for the most suitable modulation candidates. Their approach is blind as it does not

Manuscript received XX XX, 2018; revised XX XX 2018, and accepted XX XX, 2018. First published XX XX, 2018; current version published XX XX, 2018. The associate editor coordinating the review of this paper and approving it for publication was Dr. XXXX XXXX.

Ahmed O. Abdul Salam and Yim-Fun Hu are with the Faculty of Engineering and Informatics, University of Bradford, Bradford, BD7 1DP, UK (e-mail: aoa_salam@yahoo.com, y.f.hu@bradford.ac.uk).

Ray E. Sheriff is with the School of Engineering, University of Bolton, Bolton, BL3 5AB, UK (e-mail: R.Sheriff2@bolton.ac.uk).

Saleh R. Al-Araji was formerly, and Kahtan Mezher is, with the College of Engineering, Khalifa University, P.O. Box 127788, Abu Dhabi, UAE (e-mail: saleh_alaraji@yahoo.com, kahtan.mezher@ku.ac.ae).

Copyright (c) 2019 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

need prior knowledge of the carrier frequency or the baud rate. An LB algorithm was also used in the AMC system for constellation identification of orthogonal frequency division multiplexing (OFDM) signal with index modulation, as an indicator of active subcarriers, in [21]. Two scenarios were proposed, known CSI and unknown CSI, with the latter conducted blindly by default. A rather different hybrid AMC approach that constitutes a combination of both the LB and FB schemes was proposed in [22], motivated by the claimed fact that the moments of the received signal are easy to compute and can provide a simple way for signal detection. The FB scheme in [22] was based on the high-order moments (HOMs) and high-order cumulants (HOCs). A useful complexity analysis and simulation comparisons with other state-of-the-art machine learning classifiers, such as genetic-programing based K-nearest neighbor (GP-KNN), linear support vector machine (LSVM) and Kolmogorov-Smirnov (KS) were provided [22].

Some notable studies on FB are selected for discussion [23-26] in the following, but several others also exist. The graph, constructed from the Fourier transform of the second and fourth powers of the received signal, as a discrimination feature was exploited for classification in [23]. This FB classifier was claimed to have no dependence on the CSI or SNR estimation, nor did it need timing or frequency offset corrections. An end-to-end convolution neural network (CNN) based AMC (CNN-AMC), which automatically extracts features from the long symbol-rate observation sequence along with the estimated SNR was proposed in [24]. Such a classifier was thought of as a competent track to the FB, and the complexity of which alongside the conventional LB was presented. Another blind AMC method, based on the combination of elementary and cyclic cumulants, was proposed in [25]. The elementary cumulant was used to decide whether the constellations are from real, circular, or rectangular class, which was referred to as macro classifier. While, on the other hand, the cyclic cumulant was used to classify modulation within a subclass, which was referred to as micro classifier. A modulation-constrained clustering with unknown channel matrix and noise variance for multiple-input multiple-output (MIMO) systems was addressed in [26]. The AMC task was altered into a number of clustering problems, by projecting the sliced received signals into clusters, one for each modulation pattern. The final outcome was then decided by using the ML criterion applied to each intrinsic cluster.

Concluding from the above, the demand for accurate channel modelling and estimation has been noticeably emphasized. The finite state Markov chain (FSMC) has recently gained much attention in the modelling of wireless channels [1, 6, 7, 11]. Such channels are broadly categorized as a type of jump Markov linear systems (JMLSs), whose parameters evolve according to the realization of an FSMC model. A recent study, considering an LB classifier and featuring the Baum-Welch (BW) procedure for forward-backward computation of fading variable using the hidden Markov model (HMM) can be identified in [27]. From another perspective, the interacting multiple model (IMM) has been

commonly adopted to cope with the dynamic changes of various aerospace, navigation and signal processing systems [6, 7]. The IMM algorithm is widely involved in determining various processes, the underlying parameters of which evolve in the premises of FSMC and HMM approximations.

The Kalman filter (KF), on the other hand, is generally considered as one of the best optimum estimation algorithms that can be applied to a variety of vibrant systems [6, 10]. In this paper, we propose to employ the robust IMM geared to a bank of KFs for the purpose of CSI estimate adaptation and, thereby, render the AMC performance more efficient in CR systems. That is in the sense that several individual KFs can be constructed to mimic different fading effects, which are anticipated to evolve similarly to the FSMC or the HMM processes. The IMM algorithm then statistically combines the KFs' outputs as per the weighing factor assigned to each KF branch based on how accurately each copes with the presumed dynamic range of fading variants.

To the best of our knowledge, this paper, for the first time, proposes integrating IMM with the KF structures to adaptively track and follow the CSI and fading attributes. Amending the combined IMM-KF structure for adaptive CSI estimation and aiming at the consolidation of AMC systems is proposed in this paper, which has not been addressed elsewhere in the literature. Simulations are conducted with reference to the classical maximal likelihood (ML) channel estimators [2, 3]. The results validate the performance and efficiency of the integrated IMM-KF structure in the AMC paradigm.

A. Contributions

The demand for reliable AMC strategies in the design of CR systems has become increasingly pressing with the advent of new communication technologies. With this in mind, this paper aims to design a consistent and seamless AMC model with acceptable computational budget. The most notable contributions are as follows:

- Employing the SVD, equipped with the MRC, to perform the MIMO eigenchannel decomposition for space reduction and to render tractable CSI estimation.
- Developing a new approach of adding up the effective eigenmode transmission parameters instead of relying on individual ones and hence maintaining the power of all.
- Featuring a novel AMC approach based on the combined IMM-KF structure for resilient CSI estimation to enhance the LB classifier performance.
- Applying the IMM-KF output to the EM algorithm and performing the iterative computation of overlapped cycles concerned with the CSI estimation and signal detection.
- Assessing the underlying complexity and computational requirements of the AMC based on IMM-KF, which advantageously reveals that it has less overheads compared to the classical ML for CSI estimation.

B. Mathematical Notations

In this work: boldface uppercase letters denote matrices; non-boldface uppercase letters denote scalars; boldface

lowercase letters denote vectors; and non-boldface lowercase letters denote scalar variables. The operators $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^\dagger$ and $\|\cdot\|^2$ denote conjugate, transpose, Hermitian (conjugate transpose) and Frobenius norm, respectively. The operator \forall means for all elements that belong to, or \in , a certain space. The minimization is $\min(\cdot)$, maximization is $\max(\cdot)$, the summation is $\text{sum}(\cdot)$, while $\text{diag}[\dots]$ represents the diagonal entries of a matrix. $\mathbb{E}(\cdot)$ is the statistical expectation, $\Pr(\cdot)$ is the probability, $\ln(\cdot)$ is the natural logarithmic, and $\exp(\cdot)$ is the exponential function.

C. Outline

The remainder of this paper is organized as follows. Section II introduces the system modelling; Section III presents the space reduction techniques; Section IV discusses the basics of modulation classifications; Section V discusses the channel modeling and KF estimation; Section VI addresses the FSMC channel modeling and IMM adaptation, Section VII presents the joint channel estimation and data detection; Section VIII is concerned with the expectation-maximization; Section IX elaborates on the complexity analysis, and Section X is devoted to the numerical results and is followed by the concluding remarks.

II. SYSTEM MODELING

Consider a linearly modulated signal transmitted through a time-variant fading channel of Rayleigh characteristics using a $N_t \times N_r$ MIMO system. The received baseband signal in this case is denoted by [2-5]

$$y_i(k) = \sum_{j=1}^{N_t} [h_{i,j}(k)x_{j,m}(k) + n_i(k)], \quad \forall k \in K \quad (1)$$

where $y_i(k)$ is the K -length complex-valued sequence received on antenna i , $h_{i,j}$ is the channel parameters, $x_{j,m}(k)$ is the transmitted signal on antenna j and associated with the constellation set mapping, \mathcal{C}_m as the m^{th} modulation and $m = 1, \dots, M$, which can be of any *a priori* known type, and $n_i(k)$ is the additive white Gaussian noise (AWGN) of zero mean and variance $\sigma_{n_i}^2$. The symbols $\{x_{j,m}(k)\}_{m=1}^M$ are assumed uniform and independent over \mathcal{C}_m space with zero mean and variance one. Revising (1) in matrix form yields

$$\mathbf{Y}(k) = \mathbf{H}(k)\mathbf{X}(k) + \mathbf{N}(k), \quad \forall k \in K \quad (2)$$

where \mathbf{H} is the $N_t \times N_r$ channel matrix and the coefficients of which are assumed to be statistically independent Gaussian distributed complex circular random variables of zero mean and unknown variance, which is to be estimated. The AWGN vector \mathbf{N} is given by $\mathcal{CN} \sim (0, \sigma_n^2 \mathbf{I})$ and \mathbf{I} is an identity matrix.

III. SPACE REDUCTION

Modeling simplification using space reduction techniques is highly desirable, especially for online streaming applications such as the AMC systems. Aiming at an effective and straightforward CSI estimate, two powerful yet essentially equivalent simplification strategies can be applied. The first

strategy is called eigenmode transmission (ET), while the second is beamforming (BF); both are prevalent in the literature [8, 9]. While the singular value decomposition (SVD) can be applied to both strategies, BF is more related to maximum-ratio combining (MRC) at the transmitter, often called (MRT). Both strategies, however, separate the MIMO channel into $R = \min\{N_t, N_r\}$ virtually separated parallel channels [12-14] with a gain of $g_r = \sqrt{\lambda_r}$, which is the square root of the r^{th} eigenvalue, λ_r of $\mathbf{H}\mathbf{H}^\dagger$ or $\mathbf{H}^\dagger\mathbf{H}$.

The $N_r \times N_t$ channel matrix \mathbf{H} , with rank $R = \min(N_t, N_r)$ has an SVD representation given by $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$, where \mathbf{U} and \mathbf{V} are $N_r \times R$ left and $N_t \times R$ right matrices to satisfy $\mathbf{U}\mathbf{U}^\dagger = \mathbf{V}\mathbf{V}^\dagger = \mathbf{I}_R$, $\mathbf{\Sigma} = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_R]$ with $\sigma_i \geq 0$ and $\sigma_i \geq \sigma_{i+1}$, where σ_i is the i^{th} singular-value of the channel. The columns of \mathbf{V} and \mathbf{U} are also known as the left input and right output singular vectors, respectively. Now, $\mathbf{H}\mathbf{H}^\dagger$ is a $N_r \times N_r$ semi-definite Hermitian matrix. Let the eigen decomposition of $\mathbf{H}\mathbf{H}^\dagger$ be $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\dagger$ where \mathbf{Q} is an $N_r \times N_r$ matrix satisfying $\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\dagger = \mathbf{I}_{N_r}$ and $\mathbf{\Lambda} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_R]$ with random variable eigenvalues $\lambda_i \geq 0$, sorted in a descending order $\lambda_i \geq \lambda_{i+1}$ and have Chi-square distribution and $2N_t N_r$ degrees of freedom [14]. Then, $\lambda_i = \sigma_i^2$ for $i \in \{1, 2, \dots, R\}$ and $\lambda_i = 0$ for $i \in \{R+1, R+2, \dots, N_r\}$.

After applying any of the transmission strategies and with the support of spatial-temporal matched filtering (STMF) to maximize the energy of displaced signals, (1) and (2) can be reduced to the following expression of total received signal

$$y(k) = \sum_{r=1}^R [\sqrt{\lambda_r} x_{r,m}(k) + n_r(k)], \quad \forall k \in K, \forall m \in M \quad (3)$$

where the noise statistical characteristics remain intact after all mathematical manipulations and the received signal elements are hence decoupled and can be processed separately, since they are independent and identically distributed (*iid*) statistical processes. Further simplification can also be proposed here by having an effective channel parameter expressed by summing the diagonal entries $\tilde{h} = \text{sum}.\text{diag}[\sqrt{\lambda_r}]$, *i.e.*, summing up the square-root of singular values (SRSV), to alter (3) as follows

$$y(k) = \tilde{h}(k)x_m(k) + n(k), \quad \forall k \in K, \forall m \in M \quad (4)$$

where the above also assumes the signal is identical on all channel paths and the noise is uncorrelated and not adding up.

It is worth noting that the BF refers to signals being transmitted using the principal eigenmode of a channel [12, 13]. This is exactly what is meant to be for the ET strategy as a MIMO channel is decoupled into separate and non-interfering ETs. The ET and BF are mainly motivated owing to the treatment of different eigenmodes as scalar channels and, thereby, maintain lesser complexity, which is credited to the SVD involvement. For both strategies to perform optimally, the provision of full CSI knowledge is a necessity; otherwise, blind or partial CSI knowledge may yield suboptimal performance, which is also acceptable. In either way, the employment of ET or BF can push for higher peaks

in the adaptive transmission rates and channel capacities and full diversity can be also reached under certain conditions [12, 13]. As this paper is favorably inclined towards the ET strategy, a blind CSI estimation on the receiver side is considered without any back return to the transmitter. This is the normal situation in vast non-cooperative wireless communication systems; otherwise there will be no need for the AMC if both the transmitter and the receiver are originally cooperative.

Other interesting subspace reduction forms aligned with the task of signal identification can also be found in the literature. For example, the quadrature left decomposition (QLD) and the quadrature right decomposition (QRD) techniques were discussed in [15, 16]. Both studies suggested further space reduction by using the WRD, or WLD, approach, where W, R, and L stand for matrices' names used in the analysis therein, which can be attended by swapping particular layers in the punctured structures. The complexity overheads were evaluated and major savings were claimed while achieving near optimum classification performances.

Inspired by the above, the substantiation of the new paradigm proposed in this paper becomes quite apparent. The proposed approach here departs from other roadmaps suggested elsewhere in the sense that it is not concerned with the factorization of the channel covariance matrix. The core idea of the proposed paradigm is mainly centered on the eigenvalues of the decomposed channel and, instead of being treated individually, they are summed to generate a new single parameter. This parameter symbolizes the variant effects of a block-fading channel that is readily available to streamline the IMM-KF estimation algorithm. Such an approach is believed to be unique, with no evidence that it has been experimented elsewhere previously.

The new concept of total effective SRSV, which is mathematically equivalent to the Frobenius norm, is dubbed Frobenius ET (FET) in this paper. The simplified diagram of the proposed eigen-based adaptive IMM-KF for CSI estimation is portrayed in Fig. 1 below.

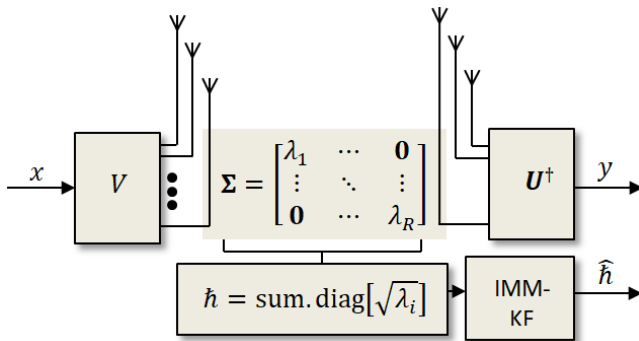


Fig. 1. Simplified diagram of the proposed adaptive CSI estimator.

IV. MODULATION CLASSIFICATION

This section is centered on the LB analysis, which is intrinsic to the premise of Bayes system. The eventual ML classifier is realized by conducting a multiple hypothesis test (MHT), whereby \mathcal{H}_m is arbitrarily assigned to the m^{th}

modulation type out of M potentials. The MHT is a composite methodical problem with strong links to the joint and conditional probability density functions (PDFs) of unknown parameters [2-5]. The PDF below is for the system in (4)

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\Theta}) = \prod_{k=1}^K \sum_{m=1}^M \frac{1}{M} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{|y_k - \hat{h}_k x_{m,k}|^2}{2\sigma_n^2}\right) \quad (5)$$

where \mathbf{y} is the vector for all received samples $\forall k \in K, \boldsymbol{\Theta} = [\hat{h} \ \sigma_n]^T$ is the vector of unknown parameters and the time index is changed for simplicity to down-script. Optimum or semi-optimum solutions to the composite MHT task can be attempted by appraising the unknown parameters as random variables (RVs). Four different ML application scenarios can consequently be distinguished, namely: the average likelihood ratio test (ALRT); generalized likelihood ratio test (GLRT); hybrid likelihood ratio test (HLRT); and quasi likelihood ratio test (QLRT) [2-5]. Probing into the ALRT reveals the likelihood of \mathbf{y} under \mathcal{H}_m is attained by averaging the PDF over unknown signal constellations and nuisance parameters

$$\Lambda_m^{\text{ALRT}}(\mathbf{y}) = \mathbb{E}_x \mathbb{E}_{\boldsymbol{\Theta}} [p(\mathbf{y}|\mathbf{x}, \boldsymbol{\Theta})] \quad (6)$$

where $\mathbb{E}_x[\cdot]$ and $\mathbb{E}_{\boldsymbol{\Theta}}[\cdot]$ are the expectations operators. The above ALRT algorithm assumes the statistical characteristics of unknown RVs are either exactly defined or fitted to the best, which is a vastly tedious task in reality. It is therefore practically not feasible in most CR applications, especially under large constellation size, due to its highly nonlinear complexity and computational burdens. The ALRT merely signifies the upper-bound (UB) limit for other algorithms.

From another perspective, the GLRT algorithm is taken as being equivalent to ALRT, providing that the ML estimate (MLE) of nuisance parameters is used and, therefore, the performance of which is unavoidably inferior. Applying the GLRT algorithm under nested constellations is also considered awkward [4]. While on the other hand, the averaging is conducted over the unknown symbols only in the HLRT and QLRT algorithms, and eventually the suitable estimate $\hat{\boldsymbol{\Theta}}$ replaces $\boldsymbol{\Theta}$ as follows

$$\Lambda_m^{(\text{Q})\text{HLRT}}(\mathbf{y}) = \mathbb{E}_x [p(\mathbf{y}, \hat{\boldsymbol{\Theta}}|\mathbf{x})] \quad (7)$$

The MLE for the sought parameters is at the heart of HLRT, while the QLRT may use other estimation tools. The non-data aided (NDA) provision suggests that both algorithms can be realized blindly or semi-blindly.

Recall that the natural logarithm is monotonically increasing and Λ_m is positive-definite, the log-likelihood ratio (LLR) defined as $\mathcal{L}_m = \mathcal{L}(\mathbf{y}|\mathcal{H}_m) = \ln(\Lambda_m)$ is proved useful to simplify the MC decision making. The decision criterion for binary HT (BHT) problem is hence constructed as the ratio between two observations \mathcal{H}_m and $\mathcal{H}_{m'}$ with prior probabilities P_m and $P_{m'}$, respectively, as given below

$$\ln \frac{\mathcal{L}(\mathbf{y}|\mathcal{H}_m)}{\mathcal{L}(\mathbf{y}|\mathcal{H}_{m'})} = \mathcal{L}_m - \mathcal{L}_{m'} \stackrel{> \mathcal{H}_m}{< \mathcal{H}_{m'}} \ln \frac{P_m}{P_{m'}} \quad (8)$$

The rightmost term in (8) is the threshold; its value equals 0 for equally likely modulation types, i.e., $P_m = P_{m'}$. This

BHT can be uniformly generalized overall applicable M hypotheses space to yield the following detection rule

$$\hat{c}_m = \arg_{m \in \{1, 2, \dots, M\}} \max(\mathcal{L}_m) \quad (9)$$

The main task of MC is to maximize the LLR functions comprised in MHTs and decide the best modulation candidate with as fewer errors as possible. The classifier performance can be quantified in terms of the individual probability of correct classification P_{CC} , which is the probability under hypothesis \mathcal{H}_m given that modulation m is the accurate modulation

$$P_{CC} = \Pr(\text{correct}|\mathcal{H}_m) = \Pr[\mathcal{L}_m > \mathcal{L}_{m'}], \quad \forall m \neq m' \quad (10)$$

The overall average probability of correct classification is not a favorable performance indicator since the individual probabilities cannot be assessed separately [3].

V. CHANNEL MODELING AND KF ESTIMATION

Efficient CSI estimation is crucial to the performance of higher-order modulations; KFs are very attractive in such domains [6, 10]. An accurate approximation for the channel modeling needs is to be investigated first before embarking on the development of KF algorithms. The following 1st-order autoregressive (Gauss-Markov) AR(1) model can embrace most of the channel dynamics and lead to effective tracking

$$\hat{h}(k+1) = \hat{h}(k) + w(k) \quad (11)$$

where $w(\cdot)$ is the state AWGN denoted by $\mathcal{CN} \sim (0, \sigma_w^2)$. The Jakes-Clarke (JC) channel model is commonly adopted as one of the best statistical models for inferring the random fading effects [6, 10, 11]. It can be easily approximated by using the AR(1) model as described in (11) above, and in this case \hat{h} represents a zero-mean complex random process with autocorrelation function $\mathbb{E}\{\hat{h}(k)\hat{h}^*(k-m)\} = \sigma_h^2 J_0(2\pi f_D \tau / f_s)$, where $J_0(\cdot)$ is the zero-order Bessel function, f_D is the maximum Doppler frequency, τ is the time delay, and f_s is the sampling frequency [6, 10, 11].

The channel state and observation models given in (11) and (4), respectively, are essential to construct the recursive scalar KF algorithms. The definitions of the following terms are due [7, 8]; $\hat{h}_{k|k-1}$ and $\hat{h}_{k|k}$ for the state prediction and estimate at step k , and $P_{k|k-1} \triangleq \mathbb{E}\{(\hat{h}_{k|k} - \hat{h}_{k|k-1})^2\}$ and $P_{k|k} \triangleq \mathbb{E}\{(\hat{h}_{k|k} - \hat{h}_{k|k})^2\}$ for the prediction and estimation errors' covariances, respectively. The predictions for the system model described in (11) can be obtained as below

$$\begin{cases} \hat{h}_{k|k-1} = \hat{h}_{k-1|k-1} \\ P_{k|k-1} = P_{k-1|k-1} + \sigma_w^2 \end{cases} \quad (12)$$

and the state estimate can be recursively updated as follows

$$\begin{cases} \hat{h}_{k|k} = \hat{h}_{k|k-1} + G_k z_k \\ G_k \triangleq P_{k|k-1} \hat{x}_{l,k-1}^T S_k^{-1} \\ P_{k|k} = (1 - G_k) P_{k|k-1} \end{cases} \quad (13)$$

where G_k is the KF gain at step k , the residual sequence $z_k = (y_k - \hat{x}_{l,k-1} \hat{h}_{k|k-1})$ is of covariance

$S_k = \hat{x}_{l,k-1} P_{k-1|k-1} \hat{x}_{l,k-1}^T + \sigma_w^2$, and $\hat{x}_{l,k-1}$ is *a priori* classified symbol whereas superscript T denotes transpose.

VI. FSMC AND IMM ADAPTATION

The information-theoretic approach of popular 1st-order FSMC is a good approximation of random channel variations and agrees with the JC model. A single transition probability in the FSMC model $\pi_{\hat{h}_1 \rightarrow \hat{h}_2, k}$ represents the prior probability of a channel transiting from state \hat{h}_i to state \hat{h}_j at a particular step k as given below [1]

$$\pi_k^{i,j} = \pi_{\hat{h}_i \rightarrow \hat{h}_j, k} = \Pr\{\hat{h}_{j,k} | \hat{h}_{i,k-1}\} \quad (14)$$

The current channel is hence only associated with the previous state, and statistically independent of all past and future states, i.e., $\pi_k^{i,j} = 0$ for $|i-j| > 1$. The transition probability π_k matrix can be formed with entries as denoted in (14) for self and immediate neighboring cells accordingly.

Building on the above, a scalar KF can be arbitrated for a span of discrete channel states and hence facilitate the IMM algorithms appropriately. For a \mathcal{J} -bank of KFs, these algorithms comprise the following core processing steps [7].

1) Interact: Action conditional and predicted probabilities

$$\mu_{k-1|k-1}^{i|j} = \frac{\pi_k^{i,j} \mu_{k-1|k-1}^i}{\mu_{k|k-1}^j}; \quad \mu_{k-1|k-1}^i = \sum_{j=1}^{\mathcal{J}} \pi_k^{i,j} \mu_{k-1|k-1}^j \quad (15)$$

And then mix the state estimates and covariance

$$\begin{aligned} \bar{h}_{k-1|k-1}^i &= \sum_{j=1}^{\mathcal{J}} \mu_{k-1|k-1}^{i|j} \hat{h}_{k-1|k-1}^j \\ \bar{P}_{k-1|k-1}^i &= \sum_{j=1}^{\mathcal{J}} \mu_{k-1|k-1}^{i|j} \left[P_{k-1|k-1}^j \right. \\ &\quad \left. + (\hat{h}_{k-1|k-1}^i - \hat{h}_{k-1|k-1}^j)^2 \right] \end{aligned} \quad (16)$$

2) Filter: Individual KF performs its usual filtering tasks.

3) Update: Update each KF probability as per residual error

$$\begin{aligned} \Delta_k^i &= \frac{\exp\left(-\frac{1}{2}(z_k^i)^2 (S_k^i)^{-1}\right)}{\sqrt{|2\pi S_k^i|}} \\ \mu_{k|k}^i &= \frac{\mu_{k|k-1}^i \Delta_k^i}{\sum_{j=1}^{\mathcal{J}} \mu_{k|k-1}^j \Delta_k^j} \end{aligned} \quad (17)$$

4) Combine: Weight the state estimate and its covariance

$$\begin{aligned} \hat{h}_{k|k} &= \sum_{i=1}^{\mathcal{J}} \mu_{k|k}^i \hat{h}_{k|k}^i \\ P_{k|k} &= \sum_{i=1}^{\mathcal{J}} \mu_{k|k}^i \left[P_{k|k}^i + (\hat{h}_{k|k} - \hat{h}_{k|k}^i)^2 \right] \end{aligned} \quad (18)$$

The estimated channel and noise attributes are eventually enabled in (5) and the ensuing MC algorithms. This can be explicitly put in the form of Euclidean distance metric, which is measured between the observed data and all ideal symbols in each candidate modulation, and selecting the shortest one

This apparently turns out to be a typical integer least-squares (LS) or minimum variance (MV) detection problem.

VII. JOINT CHANNEL ESTIMATION AND DATA DETECTION

Capitalizing on the above, it has become apparent for the requirement to promote a joint link between the AMC and CSI estimation strategies. In the strict sense of an AMC framework, the awareness of such joint channel estimation and data detection, dubbed here (JCEDD), is well established. The unknown CSI parameters that are accurately and seamlessly estimated would decisively contribute towards the successful operation of the AMC systems. Given such estimations are mostly encountered in uncertain environments, the AMC empowered by CSI estimate should be sufficiently intelligent to operate under adverse and complex situations, while the structure needs to be less bulky in order to afford real-time processing as practically feasible. Therefore, this section aims to shed more light into the JCEDD approach along with supplementary options, which have been proved to be useful for the AMC and CSI estimation integrated performance.

Research on the JCEDD approach in the framework of blind AMC applications is extensive, and only a few notable studies are mentioned herein [10, 28-34]. The interest in JCEDD techniques is not new and growth is expected to continue, not only with AMC applications, but also with other pinnacle functions of CR systems like the SS and ACM.

An overview of the JCEDD blind techniques can be obtained from [28]. Generally, blind techniques can be branched into two main streams, either deterministic or statistical, depending on how the input signal is modelled [28]. If the modelling assumes the input is a random variable with predefined statistics, the corresponding blind technique is to be statistical. On the other hand, if the modelling does not have any *a priori* description of the input random variable, or the statistics of which are available but are not exploited, and does not have a statistical description, the corresponding blind technique algorithm is called deterministic.

The JCEDD can be applied to any kind of unknown parameters of interest, which evolve around the core parameter represented by the alphabet symbols. Moreover, the JCEDD algorithm can be approached via iteratively exchanging information between the CSI estimator and the signal detector, noting that the detection and classification words can be used interchangeably in the literature and also in this work. The information exchange process is decomposed into two optimization loops, the outer and the inner. The outer loop is a global optimization algorithm, which searches for an optimal CSI estimate, while the inner loop is an ML algorithm that identifies the transmitted symbols [28-33]. These loops are interweaved, and sometimes called the upper level and the lower level, in tandem.

In [28], the repeated weighted boosting search algorithm was used for CSI estimate at the outer loop, which searches the MIMO channel space by evolving a population of MIMO channel matrices. While at the inner loop, the optimized hierarchy reduced search algorithm aided detector, which is an

$$\min \|\mathbf{y} - \mathbf{h}\mathbf{x}_m\|^2 \underset{> \mathbf{h}}{< \mathbf{h}_{m'}} \min \|\mathbf{y} - \mathbf{h}\mathbf{x}_{m'}\|^2, \forall m, m' \in M \quad (19)$$

advanced extension of the complex sphere decoder, was used [28]. The signal identification by classifying both the modulation type and the STBC scheme in MIMO systems was considered as a joint classification problem in [32]. The JCEDD was also extended to MIMO-OFDM systems in the framework of sparse Bayesian learning techniques in [10]. It was also applied to the unknown co-scheduled user's modulation constellation size as part of the MU-MIMO detector in [33]. The per-survivor processing of best optimization path using the VA algorithm as an effective trellis search engine in the context of JCEDD was elaborated in [29, 30]. The KF algorithm for CSI estimation was employed in [10, 34], whereas the GA procedure was implemented in [30].

On the side of design tools necessary for the JCEDD algorithm, the first step usually begins by tailoring a minimization cost function, which should be suitably assigned to reflect on all the parameters of concern. Considering the MIMO configuration, this can be done by applying the MLE on the transmitted symbols \mathbf{X}_i and the channel matrix \mathbf{H} and then maximizing the joint conditional PDF over \mathbf{X} and \mathbf{H} together. Alternatively, this can be achieved by minimizing the following joint cost function [28-34]

$$\mathcal{F}(\tilde{\mathbf{X}}, \tilde{\mathbf{H}}) = \sum_{k=1}^K \|\mathbf{Y} - \mathbf{H}\mathbf{X}\|^2 \quad (20)$$

where all the constants have been ignored as they will cancel each other while calculating LRTs. Namely, the joint ML for channel and data estimation is obtained as below [28-34]

$$(\hat{\mathbf{X}}, \hat{\mathbf{H}}) = \arg\{\min_{\tilde{\mathbf{X}}, \tilde{\mathbf{H}}} [\mathcal{F}(\tilde{\mathbf{X}}, \tilde{\mathbf{H}})]\} \quad (21)$$

As can be seen from (21), the search for the optimal joint ML result is over the entire discrete space of all the possible transmitted alphabets and the MIMO channel matrix, which is computationally prohibitive in reality. Therefore, the complexity of such an optimization process needs to be reduced to tractable levels. This can be done if each single search cycle is decomposed into two iterative loops: firstly, over all the possible data symbols; and then over all the channel matrix entries. This can be formed as below [30, 31]

$$(\hat{\mathbf{X}}, \hat{\mathbf{H}}) = \arg\{\min_{\tilde{\mathbf{H}}} \{\min_{\tilde{\mathbf{X}}} [\mathcal{F}(\tilde{\mathbf{X}}, \tilde{\mathbf{H}})]\}\} \quad (22)$$

Such interlaced loops are repeated over the complete space of all data and channel entries in effect, and until a breakdown of the most suitable per-survivor paths are obtained with minimal errors as possible [29, 30]. The efficient implementation of the above iteration tool is explored in the next section.

VIII. EXPECTATION-MAXIMIZATION

The statistical optimization expression, whether described in (21) or equivalently in (22), is in general difficult because it is nonconvex in the premise of an ML function. The expectation-maximization (EM) algorithm can be applied to

transform the complicated optimization to a sequence of quadratic optimizations [10, 28, 34]. Not only can the EM be found in non-cooperative work conditions, but it can also be very useful under cooperative circumstances [35]. The main intention of developing the EM algorithm is to render the MLE of the unknown CSI parameters more tractable.

The EM algorithm comprises two steps, namely: the expectation step (E-step); and the maximization step (M-step). In the E-step, the expectation of the LLF of the complete data given the observed data is evaluated. While in the M-step, new estimates of the unknowns are obtained by maximizing the expectation computed in the E-step. The EM algorithm possesses the property of monotonically increasing at each step to produce the expected maximum of LFs when it converges.

Keeping the above in mind, the EM algorithm can now be explicitly sought. Like the case of IMM-KF, let the only unknown scalar parameter be $\Theta = \hat{h}$, then the MLE of which is denoted as follows [28, 34, 35]

$$\begin{aligned} \hat{h}_{ML} &= \underset{\hat{h}}{\operatorname{argmax}} \{ \ln(p(\mathbf{y}|\hat{h})) \} \\ &= \underset{\hat{h}}{\operatorname{argmax}} \left\{ \ln \left[\sum_x \ln(p(\mathbf{y}|x, \hat{h})p(x)) \right] \right\} \end{aligned} \quad (23)$$

From (23), it is apparent that there is no closed-form solution for \hat{h}_{ML} since $p(\mathbf{y}|\hat{h})$ is of a mixed Gaussian distribution nature. A common circumvent to this issue is the expectation-maximization (EM) algorithm to iteratively compute the MLE by using two data sets. The first data set is incomplete and comprises the observations \mathbf{y} , while the second is complete and represented by $\mathbf{z} = [\mathbf{y}^T \mathbf{x}^T]^T$, which is the observations and unknown or missing modulation symbols.

The E-step is thereby attended as below [10, 35]

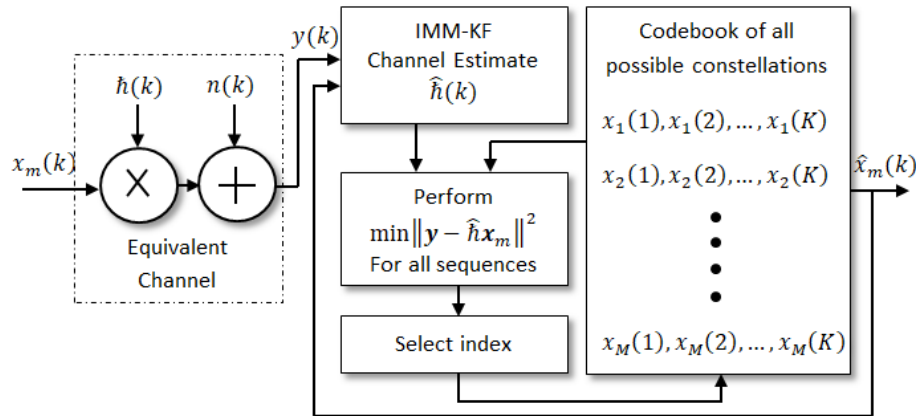


Fig. 2. Simplified AMC system diagram based on IMM-KF for channel estimate.

IX. COMPLEXITY ANALYSIS

The computational complexity of the proposed IMM-KF for AMC is assessed and compared against the classical MLE approach in this section. In addition to the optimum performance of IMM-KF algorithm, which will be demonstrated in the following simulation exercises, this section is intended to emphasize the main feature of IMM-KF

$$\begin{aligned} Q(x, \hat{h}_{ML}^{i-1}) &= \mathbb{E}_{\mathbf{z}} \left\{ \ln \left(p(\mathbf{z}|\hat{h}, \hat{h}_{ML}^{i-1}) \right) \right\} \\ &= \mathbb{E}_{\mathbf{x}} \left\{ p(\mathbf{x}|\mathbf{y}, \hat{h}_{ML}^{i-1}) \ln \left(p(\mathbf{z}|\hat{h}_{ML}^{i-1}) \right) \right\} \end{aligned} \quad (24)$$

where $p(\mathbf{x}|\mathbf{y}, \hat{h}_{ML}^{i-1})$ is the *a posteriori* probability of the modulation symbol vector \mathbf{x} conditioned on \mathbf{y} and \hat{h}_{ML}^{i-1} , and i here denotes the i^{th} iteration of estimation and classification.

The parameter \hat{h}_{ML}^i maximizes Q in the M-step as follows [10, 35]

$$\hat{h}_{ML}^i = \underset{\hat{h}}{\operatorname{argmax}} \{ Q(x, \hat{h}_{ML}^{i-1}) \} \quad (25)$$

The M-step (25) signifies the most surviving path essential for the MLE algorithm, while it is directly realized by the IMM-KF. The E-step (24) forms the basis to maximize (6) and (7) or minimize (19) and relates to both the IMM-KF and MLE. Fig. 2 shows a block diagram of the proposed AMC system based on the IMM-KF adaptive channel estimation.

As can be seen from Fig. 2, there is one loop involved in the search algorithm performed over all incoming data samples and constellation candidates. Upon discovering a constellation match, the next cycle of channel estimate is thereby directly calculated using the encoded sequence. On the other hand, the MLE algorithm constitutes another loop in addition to the above mentioned one, which is related to the search of best channel estimate as per optimized convex function. Evidently, such interlacing loops of exhaustive search for constellation index and channel estimate could be permissible for small data alphabet size and small data record only. As the data extents get larger, the IMM-KF admirably promises for better convergence as its reckoning load is half of that of the MLE method, as will be seen in the next section.

algorithm of retaining less computational demands compared to the MLE. Such complexity reduction is of significant importance especially in real-time tactical scenarios where fast processing and decisions on the incoming signals need to be attended seamlessly. Few studies on the computational analysis of AMC systems can be found in the literature, and most of which attempt the MLE in the general context of optimum LB algorithms under MIMO conditions. The

computational analysis of IMM-KF featuring space reduction for the purpose of AMC implementation is illustrated for the first time in this paper.

Without loss of generality, the complexity of any ACM scheme is proportional with respect to the numbers of code sequence and alphabets. Since the code sequence is assumed intact here, the complexity will hence be governed by the alphabets order only. Further assumptions also need to be considered while facilitating the computational approximations. Firstly, the computations are performed over one data sample; secondly, the eigenvectors are uncorrelated and without inter-modal interference; and thirdly, the channel statistics do not vary.

The computational complexity usually involves the overall operation of mathematical addition, subtraction, multiplication and division procedures. The computational complexity loads of the essential operations $\mathcal{O}(\cdot)$ for both the IMM-KF and MLE in comparison with other schemes [5, 16] can be easily deduced as shown in Table I below.

TABLE I

COMPUTATIONAL COMPLEXITY FOR COMPLETE EM CYCLE

Item	Load	Comments
IMM-KF	$\mathcal{O}(M)$	FET space reduction
MLE	$\mathcal{O}(2M)$	FET space reduction
[5]	$\mathcal{O}(2N_t N_r M^{N_t})$	MIMO without space reduction
[16]	$\mathcal{O}(2M^{N_t})$	MIMO with WRD/WLD space reduction

Recall that the subspace is reduced by combining the MRC and SVD algorithms, the classifiers' complexity can be easily assessed as it is dropped from $\mathcal{O}(N_t N_r M^{N_t})$ [5] and $\mathcal{O}(M^{N_t})$ [16] to $\mathcal{O}(M)$ for each step in the EM algorithm. The optimal MLE requires $\mathcal{O}(M)$ for each iteration in (24) and (25), while the proposed IMM-KF needs one iteration in (24) only and its estimate can be readily inserted into (25) without extra burdens. Such a complexity reduction result remains the same whether dealing with single or multiple antennas, as per the guidance provided in this paper. The figures of $\mathcal{O}(N_t N_r M^{N_t})$ and $\mathcal{O}(M^{N_t})$ in reality can grow overly unfeasible for larger constellations and antennas' aggregates, while the IMM-KF and MLE overheads are fixed and thereby converge easily. The IMM-KF even has half the computational requirements compared to the MLE and, hence, can be advocated for better economical JCEDD enactment in the real-time AMC systems. The superiority of the proposed scalar IMM-KF algorithm using the FET approach for space reduction is tangible; otherwise, the vector computation of KF [10], and hence the resulting IMM, would be tediously high.

X. SIMULATION RESULTS AND DISCUSSION

The performance corroboration of the proposed adaptive IMM-KF estimator is presented using simulation exercises in this section. The innovative paradigm of FET subspace reduction is incorporated in the signal model of 3×3 MIMO

configuration. The signaling is evoked by a BPSK sequence of 129 samples length and unity power sent in the direction of a Rayleigh channel. The channel is assumed with plausible flat fading over certain block samples. An AWGN is also incurred in these channels and the SNR is 10 dB. In each simulation scenario, the 129 samples are divided into three block segments; the length of each is 43 samples, to manifest fixed fading. The segmented channel fading magnitudes along with their corresponding SRSVs are given in Table II. The three segments of the first channel exhibit severe, medium and mild fading behaviors, respectively, while the second channel is assumed to have the opposite behavior. These arbitrary fading attributes are plotted in Fig. 3 and Fig. 4, respectively.

TABLE II
SCENARIO OF (3×3) MIMO CHANNEL PARAMETERS.

Item	MIMO
H blocks	[0.9 0.9 j 0.9; 0.9 j 0.9 0.9; j 0.9, 0.9, 0.9] [0.5 0.5 j 0.5; 0.5 j 0.5 0.5; j 0.5 0.5 0.5] [0.1 0.1 j 0.1; 0.1 j 0.1 0.1; j 0.1 0.1 0.1]
$\sqrt{\lambda}$ SRSV	[2 1.27 1.27] [1.11 0.7 0.7] [0.22 0.14 0.14]
FET	4.54 2.51 0.5

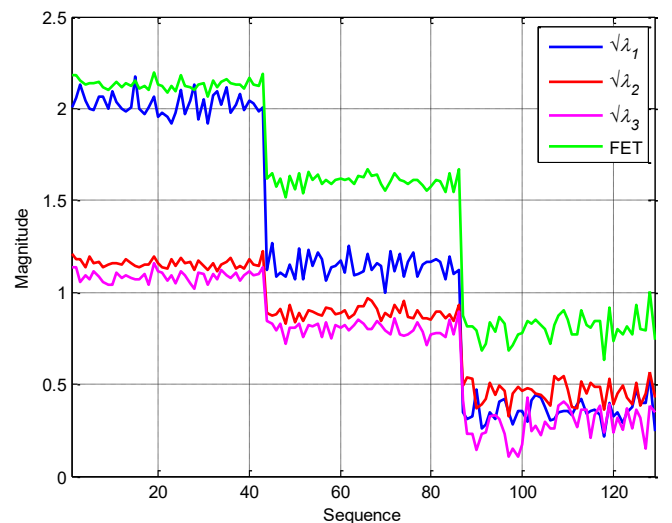


Fig. 3. SRSVs of block-fading Rayleigh channel in AWGN.

From Fig. 3, it is quite evident that each single path of decomposed channels carries important components of the overall transmission power and hence it is better to be summed rather than being treated individually or merely use the dominant maximum eigenmode for transmission only, which is the case of ET. This is the main concept underlining the meaningful postulation of the new paradigm of FET proposed here. Such an approach consequently keeps as much channel power as possible and maintains an efficient track of its varying conditions, instead of being aborted and vanished to nowhere. The power components of different channel paths have a direct impact on the transmitted signal power and hence it is more advisable to be all consolidated and exploited.

Let us now assess the adaptive CSI estimation of IMM-KF and compare it to a single KF and to the other two conventional estimators, namely; the least-square estimator (LSE) and the minimum mean-square estimator (MMSE). These classical estimators are equivalent to the MLE and maximum a-posteriori estimator (MAPE) under uncorrelated Gaussian noise impairments. The initialization of KF and IMM-KF are assigned values as shown in Table III, while the other estimators do not need such an initialization process. As shown in the table, a 3-KF bank is employed to facilitate the IMM algorithm with initial probabilities of 0.1, 0.8 and 0.1 for each individual filter, while the transitional probabilities of 0.8, 0.1 and 0.1 represent the local, neighbor and remote FSMC states, respectively. Each KF has a single σ_w^2 value, such as 0.01, 0.1 and 1, to correspond to mild, moderate and severe channel conditions, respectively.

TABLE III
INITIALIZATION OF KF AND IMM-KF ALGORITHMS.

Item	Parameters Initialisation	Notes
KF	$\hat{h}_{0 0} > 0, P_{1 0} \geq 1$ $\sigma_n^2 = 0.01, \sigma_w^2 = 0.1$	Same for each KF in the IMM except process noise
IMM-KF	$\sigma_{w1}^2 = 0.1, \sigma_{w2}^2 = 1, \sigma_{w3}^2 = 5$ $\mu_{0 0}^i = [0.8 \ 0.1 \ 0.1], \forall i \in [1, 2, 3]$ $\pi_0^{i,j} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}, \forall i, j \in [1, 2, 3]$	Individual process strengths Modal probabilities Transitional probabilities

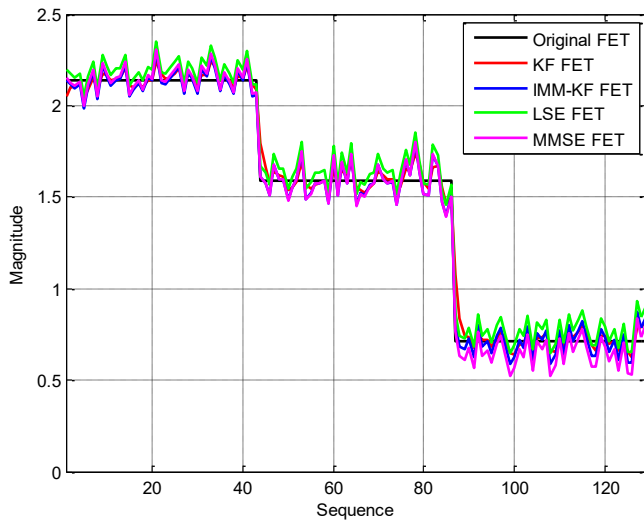


Fig. 4. Estimation of SRSVs over block-fading Rayleigh channel in AWGN.

The estimators are separately allocated through the given channels and the simulation outcomes are compiled in Fig. 4 for the (3×3) MIMO transmission scenario. The figure shows that all estimators maintain a good track of the channels' SRSV variations; yet some perform better than others. The KF and IMM-KF clearly outperform the LSE and MMSE and the IMM-KF is preferred, especially during the channels' deep fading parts. The IMM-KF is also superior through the channels' transitions between states, which is inherited from the FSMC's natural behavior. The main driver behind the IMM-KF's robust adaptation performance is

centered on the attractive switching competence of the IMM structure itself. If the channel has mild fading, the KF with the lowest state noise power $\sigma_{w1}^2 = 0.1$ is assigned the highest weight in the output of probability combining. As the fade moderates, the state noise power $\sigma_{w2}^2 = 1$ is allocated the next highest weight than other KFs in the IMM bank. The deep fading effect is treated with the highest state noise power $\sigma_{w3}^2 = 5$ and the output of combining probability receives the highest weight accordingly. The first encounter is similar to the situation of having a lone KF with action virtually reminiscent of a low-pass filter (LPF). This LPF reduces the channels' harsh fluctuations and smooths out the output as revealed in the two plots.

Moreover, the transitional probabilities of 0.8 and 0.1 represent the percentage of time a state spends for about 80% in its current positions and makes a transition to adjacent cells, whether forward or backward, with 10% likelihood occurrence of each. However, the modal and transitional probabilities change with respect to the innovation behavior of each KF branch as the time elapses. The perspectives of such IMM-KF switching probabilities to cope with the various jumping fades are clearly distinguished in Fig. 5.

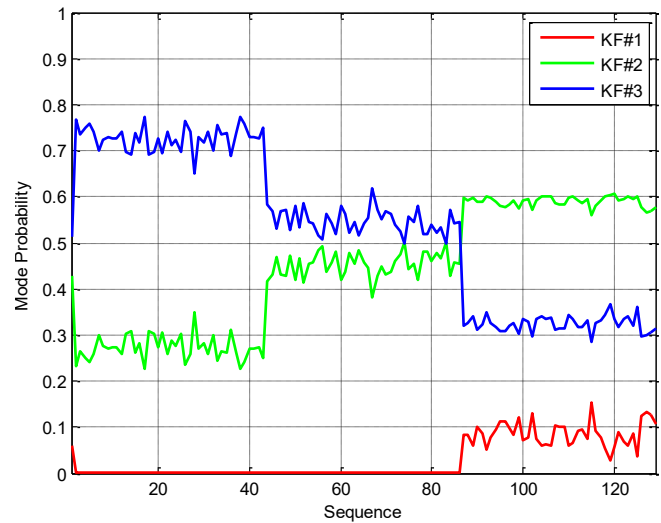


Fig. 5. Modal probabilities of 3-bank KFs over block-fading channel.

The remaining simulation exercises are conducted to assess the AMC performance using the proposed IMM-KF and compared to that of MLE and perfectly known channel (KC). The same initialization settings are used again for the IMM-KF structure, but now rather different combinations of antennas are used. A challenging 3-tuple set of nesting constellations {BPSK, 8-PSK, 16-PSK} is generated using the celebrated procedure. The results of 10^3 Monte Carlo runs over an uncorrelated slow-fading Rayleigh channel and 1×1 and 2×2 antennas are shown in Figs. 6 and 7, respectively. The results illustrate the probability of correct classification P_{CC} calculated for the IMM-KF, MLE and KC, where the latter serves as an upper performance bound. Table IV shows their correct classification rates at 0 dB.

TABLE IV
CORRECT CLASSIFICATION RATES AT 0 dB.

	1 × 1 antennas			2 × 2 antennas		
	KC	MLE	IMM-KF	KC	MLE	IMM-KF
BPSK	0.99	0.99	0.99	1	1	1
8-PSK	0.85	0.85	0.85	0.98	0.98	0.98
16-PSK	0.57	0.57	0.57	0.75	0.75	0.75

The same simulation scenarios are repeated using another 3-tuple set of different constellations, namely {QPSK, 16-QAM, 64-QAM}, which is tailored to realistic common systems. The ensuing P_{CC} trends are depicted in Figs. 8 and 9, for 1 × 1 and 2 × 2 antennas, respectively, and the values of which calculated at 0 dB are given in Table V below.

TABLE V
CORRECT CLASSIFICATION RATES AT 0 dB.

	1 × 1 antennas			2 × 2 antennas		
	KC	MLE	IMM-KF	KC	MLE	IMM-KF
QPSK	0.75	0.75	0.75	0.99	0.99	0.99
16-QAM	0.27	0.27	0.27	0.67	0.67	0.67
64-QAM	0.04	0.04	0.04	0.30	0.30	0.30

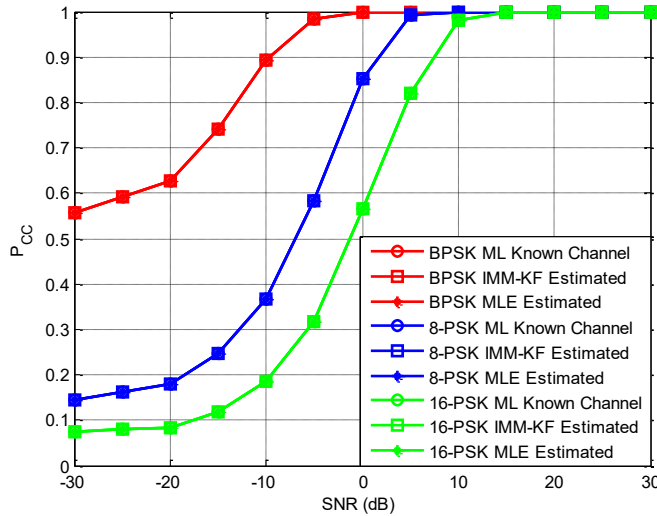


Fig. 6. P_{CC} using 1 × 1 antennas.

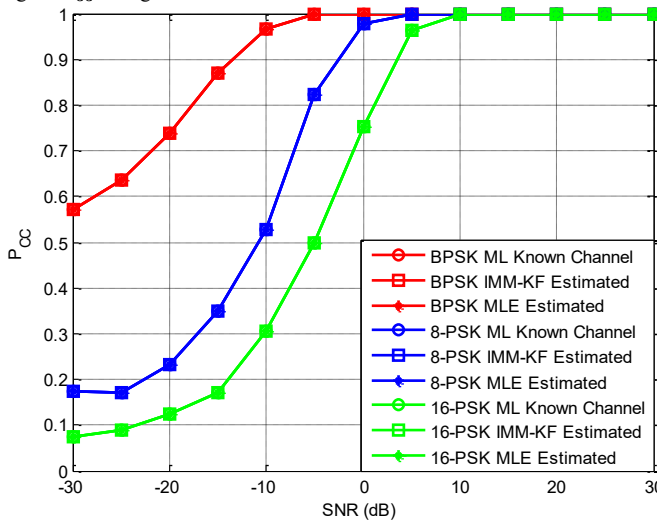


Fig. 7. P_{CC} using 2 × 2 antennas.

The results show that the proposed IMM-KF performs similarly to the optimal MLE and both coincide with the KC

irrespective of antennas employed. The increased SNR yields for a gradual dominance of the signal covariance matrix and to spread its eigenvalues, which enhances the signal detection accordingly. All methods are also effective in combating the nesting effect to a certain limit, whereas the trends of correct classification are clearly shifting down for larger constellations' sizes. A classifier to attain correct decisions hence becomes more difficult with the severity of lesser spaces between constellations' points. Such a demand is very evident in the cases of 16-PSK, 16-QAM and 64-QAM as the spaces between adjacent constellation points becomes more crowded and the detection hence becomes very tasking. For the same M -ary, the QAM suffers from magnitude and phase variants that are even harder to identify.

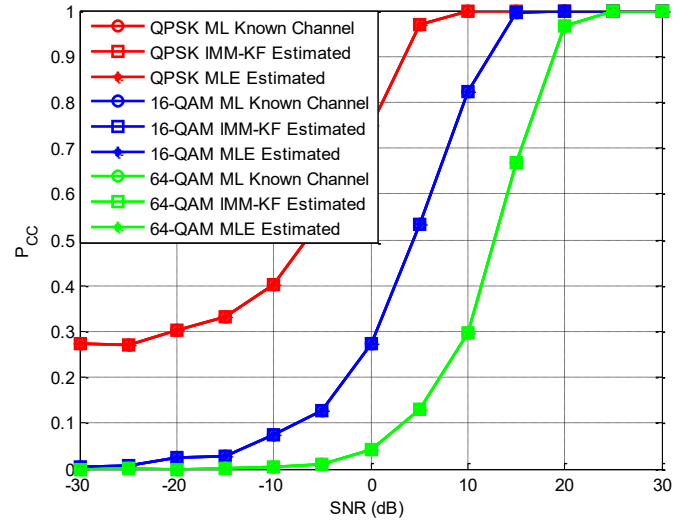


Fig. 8. P_{CC} using 1 × 1 antennas.

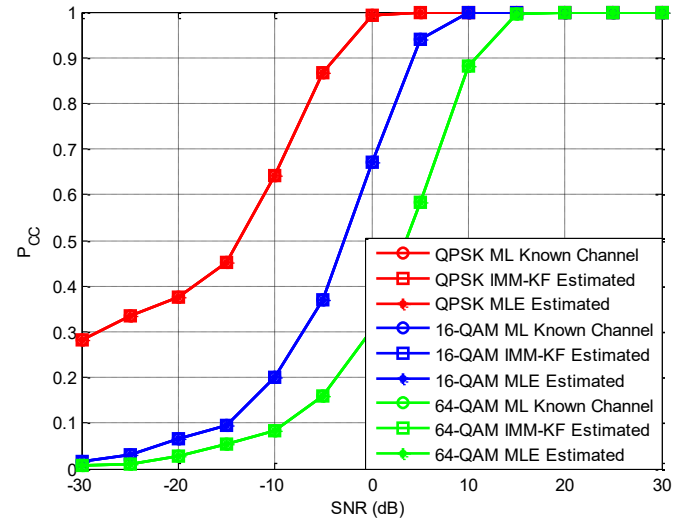


Fig. 9. P_{CC} using 2 × 2 antennas.

To show more appreciation for the superiority of complexity saving accredited to the proposed FET for space reduction in comparison with other methods, the next exercise is aimed to accentuate this major advantage. Due to the massive number of resulting operations while increasing the antennas and alphabets spaces, only a limited number of these parameters are considered in the simulation. The trends

depicted in Fig. 10, which are based on the entries of Table I, demonstrate the computational figures using constellation sizes of 2, 4, and 8 and for few transmit antennas 1, 2, and 3, respectively. From Fig. 10 it is clear as the numbers of constellation size and transmit antennas increase, the computational demands increase drastically. For a practical MLE-AMC system employing 64-ary alphabet and 4-MIMO antennas, the computational cost is well anticipated to be in the range of 268.43×10^6 operations without space reduction and 16.77×10^6 operations with WRD or WLD space reduction. Such computations increase intensely for massive MIMO and large constellations and become more unbearable in reality, and hence the proposed FET approach provides an attractive alternative choice to produce radical computational savings. One last note concerning Fig. 10 is that the IMM-KF and MLE schemes employing FET perform similarly regardless of the number of antennas. Their computations remain very minimal compared to other channel decompositions aiming at space reduction, and hence one more exercise would be beneficial to reveal the full situation as will be exposed below.

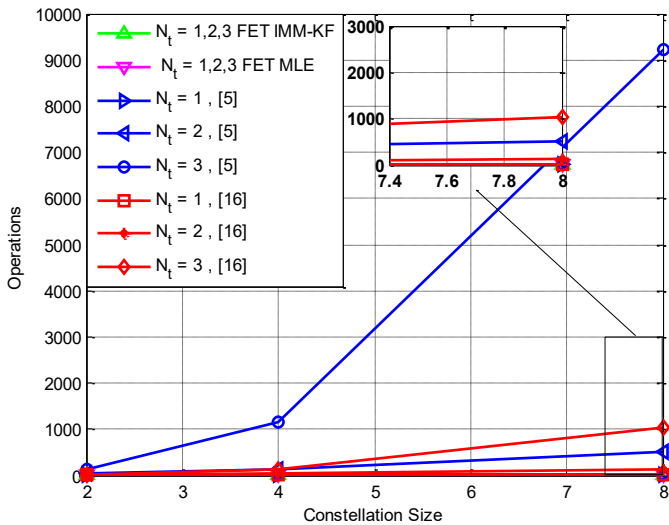


Fig. 10. Number of operations versus constellation size for different antennas.

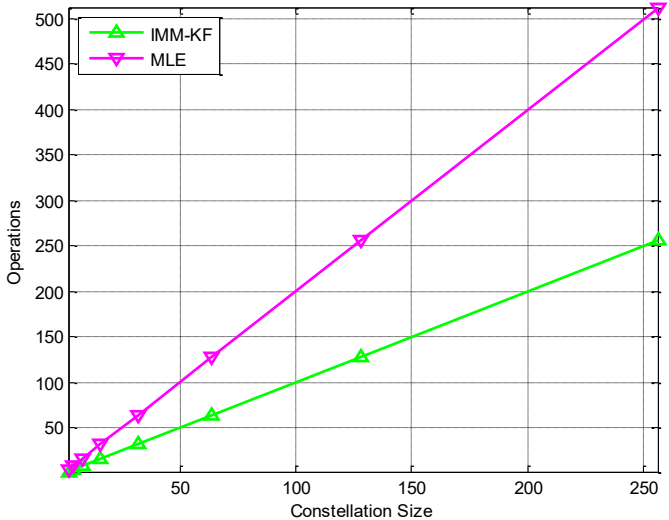


Fig. 11. Number of operations in IMM-KF and MLE, regardless of antennas.

Now, in order to make a distinct differentiation between the IMM-KF and MLE operations using the FET procedure, another exercise is performed and the results of which are depicted in Fig. 11. Based on the entries of Table I, the computational loads for both IMM-KF and MLE featuring FET have a linear relationship with respect to the constellation size. The effectiveness of applying the IMM-KF is very apparent compared to the MLE as it retains only half of the computations required to implement the JCEDD for one complete EM cycle. Reducing the dimensionality by facilitating the FET approach and getting further gain in lowering the computational cost is chiefly accredited to the IMM-KF. Such breakthrough constitutes a corner stone in the design of many mission-critical and tactical systems in reality.

XI. CONCLUSIONS

An efficient AMC paradigm featuring the IMM-KF mixture to elicit a robust and reliable CSI estimate is proposed in this paper. The adaptive IMM-KF estimation is based on decomposing the CSI parameters into a small number of parallel components, by applying the well-known SVD methodology, and then adding up their dominant ETs, which is newly called Frobenius ET (FET) in this paper. This accordingly will maintain the total power of all ETs instead of the waste resulting from relying on individual ETs only. The estimate is applied over Rayleigh MIMO channel and directly admitted to the EM recursion to invoke the Q(HLRT) for AMC core framework. Further to the viable performance, the overall IMM-KF computation is significantly less demanding compared to the optimal MLE especially with large constellation sizes. The IMM-KF only needs one search loop to find the best constellation candidate and is hence more economical than the MLE, which requires two extensive loops to fulfil the EM complete cycle.

REFERENCES

- [1] B. Li, S. Li, J. Hou, J. Fu, C. Zhao and A. Nallanathan, "A Bayesian Approach for Adaptively Modulated Signals Recognition in Next-Generation Communications", *IEEE Trans. on Sig. Pro.*, Vol. 63, No. 16, Aug. 2015, p. 4359-4372.
- [2] F. Hameed, O. A. Dobre and D. C. Popescu, "On the Likelihood-Based Approach to Modulation Classification", *IEEE Trans. on Wireless Com.*, Vol. 8, No. 12, Dec. 2009, p. 5884-5892.
- [3] J. L. Xu, W. Su and M. C. Zhou, "Likelihood-Ratio Approaches to Automatic Modulation Classification", *IEEE Trans. on Sys., Man, and Cyb.-Part C: Apps. and Revs.*, Vol. 41, No. 4, July 2011, p. 455-469.
- [4] O. A. Dobre, A. Abdi, Y. Bar-Ness and W. Su, "Survey of Automatic Modulation Classification Techniques: Classical Approaches and New Trends", *IET Com.*, Vol. 1, No. 2, 2007, p. 137-156.
- [5] Y. A. Eldemerdash, O. A. Dobre and M. Oner, "Signal Identification for Multiple-Antenna Wireless Systems: Achievements and Challenges", *IEEE Com. Survs. and Tuts.*, Vol. 18, No. 3, 2016, p. 1524-1551.
- [6] C. Kominakis, C. Fragouli, A. H. Sayed and R. D. Wesel, "Multi-Input Multi-Output Fading Channel Tracking and Equalization Using Kalman Estimation", *IEEE Trans. on Sig. Pro.*, Vol. 50, No. 5, May 2002, p. 1065-1076.
- [7] T. Sathyan and T. Kirubarajan, "Markov-Jump-System-Based Secure Chaotic Communication", *IEEE Trans. on Cir. and Sys. -I*, Vol. 53, No. 7, Jul. 2006, p. 1597-1609.

- [8] M. R. McKay, A. J. Grant and I. B. Collings, "Performance Analysis of MIMO-MRC in Double-Correlated Rayleigh Environments", *IEEE Trans. on Com.*, Vol. 55, No. 3, Mar. 2007, p. 497-507.
- [9] L. G. Ordóñez, D. P. Palomar and J. R. Fonollosa, "Ordered Eigenvalues of a General Class of Hermitian Random Matrices and Performance Analysis of MIMO Systems", *Proc. ICC*, 18-23 May 2008, Beijing, p. 3846-3852.
- [10] R. Prasad, C. R. Murthy and B. D. Rao, "Joint Channel Estimation and Data Detection in MIMO-OFDM Systems: A Sparse Bayesian Learning Approach", *IEEE Trans. on Sig. Pro.*, Vol. 63, No. 20, Oct. 2015, p. 5369-5382.
- [11] P. Sadeghi, R. A. Kennedy, P. B. Rapajic and R. Shams, "Finite-State Markov Modeling of Fading Channels: A Survey of Principles and Applications", *IEEE Sig. Pro. Mag.*, Sep. 2008, p. 57-80.
- [12] H. Zhang, S. Wei, G. Ananthaswamy and D. L. Goeckel, "Adaptive Signaling Based on Statistical Characterizations of Outdated Feedback in Wireless Communications", *Proc. of the IEEE*, Vol. 95, No. 12, Dec. 2007, p. 2337-2353.
- [13] S. H. Ting, K. Sakaguchi and K. Araki, "A Robust and Low Complexity Adaptive Algorithm for MIMO Eigenmode Transmission System with Experimental Validation", *IEEE Trans. on Wireless Com.*, Vol. 5, No. 7, Jul. 2006, p. 1775-1784.
- [14] G. B. Giannakis, Z. Liu, X. Ma and S. Zhou, "Space-Time Coding for Broadband Wireless Communications", John Wiley & Sons, 2007.
- [15] M. M. Mansour, "A Near-ML MIMO Subspace Detection Algorithm", *IEEE Sig. Pro. Let.*, Vol. 22, No. 4, Apr. 2015, p. 408-412.
- [16] H. Sarieddeen, M. M. Mansour and A. Chehab, "Modulation Classification via Subspace Detection in MIMO Systems", *IEEE Com. Let.*, Vol. 21, No. 1, Jan. 2017, p. 64-67.
- [17] Y.-C. Liang, K.-C. Chen, G. Y. Li, and P. Mahonen, "Cognitive Radio Networking and Communications: An Overview," *IEEE Trans. Veh. Tech.*, vol. 60, no. 7, p. 3386-3407, Sep. 2011.
- [18] J. J. Popoola and R. van Olst, "A Novel Modulation-Sensing Method", *IEEE Veh. Tech. Mag.*, Sep. 2011, p. 60-69.
- [19] F. K. Jondral, "Cognitive Radio: A Communications Engineering View", *IEEE Com. Mag.*, Aug. 2007, p. 28-33.
- [20] D. Zhu, V. J. Mathews and D. H. Detienne, "A Likelihood-Based Algorithm for Blind Identification of QAM and PSK Signals", *IEEE Trans. on Wireless Com.*, Vol. 17, No. 5, May 2018, p. 3417-3430.
- [21] J. Zheng and Y. Lv, "Likelihood-Based Automatic Modulation Classification in OFDM With Index Modulation", *IEEE Trans. on Veh. Tech.*, Vol. 67, No. 9, Sep. 2018, p. 8192-8204.
- [22] M. Abu-Romoh, A. Aboutaleb and Z. Rezki, "Automatic Modulation Classification Using Moments and Likelihood Maximization", *IEEE Com. Let.*, Vol. 22, No. 5, May 2018, p. 938-941.
- [23] Y. A. Eldemerdash, O. A. Dobre, O. Ureten and T. Yensen, "A Robust Modulation Classification Method for PSK Signals Using Random Graphs", *IEEE Trans. on Ins. and Mea.*, Vol. 68, No. 2, Feb. 2019, p. 642-644.
- [24] F. Meng, P. Chen, L. Wu and X. Wang, "Automatic Modulation Classification: A Deep Learning Enabled Approach", *IEEE Trans. on Veh. Tech.*, Vol. 67, No. 11, Nov. 2018, p. 10760-10772.
- [25] J. Tian, Y. Pei, Y.-D. Huang and Y.-C. Liang, "Modulation-Constrained Clustering Approach to Blind Modulation Classification for MIMO Systems", *IEEE Trans. on Cog. Com. and Net.*, Vol. 4, No. 4, Dec. 2018, p. 894-907.
- [26] S. Majhi, R. Gupta, W. Xiang and S. Glisic, "Hierarchical Hypothesis and Feature-Based Blind Modulation Classification for Linearly Modulated Signals", *IEEE Trans. on Veh. Tech.*, Vol. 66, No. 12, Dec. 2017, p. 11057-11069.
- [27] J. Zhang, F. Wang, Z. Zhong and S. Wang, "Continuous Phase Modulation Classification via Baum-Welch Algorithm", *IEEE Com. Let.*, Vol. 22, No. 7, Jul. 2018, p. 1390-1393.
- [28] L. Tong and S. Perreau, "Multichannel Blind Identification: From Subspace to Maximum Likelihood Methods", *Proc. of the IEEE*, Vol. 86, No.10, 1998, p.1951-1968.
- [29] N. Seshadri, "Joint Data and Channel Estimation and Using Blind Trellis Search Techniques", *IEEE Trans. on Com.*, Vol. 42, No. 2/3/4, Feb/Mar/Apr 1994, p. 1000-1011.
- [30] S. Chen and Y. Wu, "Maximum Likelihood Joint Channel and Data Estimation Using Genetic Algorithms", *IEEE Trans. on Sig. Proc.*, Vol. 46, No. 5, May 1998, p. 1469- 1473.
- [31] M. Abuthinien, S. Chen and L. Hanzo, "Semi-Blind Joint Maximum Likelihood Channel Estimation and Data Detection for MIMO Systems", *IEEE Sig. Pro. Let.*, Vol. 15, 2008, p. 202-205.
- [32] O. Bayer and M. Oner, "Joint Space Time Block Code and Modulation Classification for MIMO Systems", *IEEE Wireless Com. Let.*, Vol. 6, No. 1, Feb. 2017, p. 62-65.
- [33] A. Gomaa, L. M. A. Jalloul, M. M. Mansour, K. Gomadam, and D. Tujkovic, "Max-Log-MAP Optimal MU-MIMO Receiver for Joint Data Detection and Interferer Modulation Classification", *IEEE Com. Let.*, Vol. 20, No. 7, Jul. 2016, p. 1389-1392.
- [34] C. Cozzo and B. L. Hughes, "An Adaptive Receiver for Space-Time Trellis Codes Based on Per-Survivor Processing", *IEEE Trans. on Com.* Vol. 50, No. 8, Aug. 2002, p. 1213-1216.
- [35] J. Zhang, D. Cabric, F. Wang and Z. Zhong, "Cooperative Modulation Classification for Multipath Fading Channels via Expectation-Maximization", *IEEE Trans. on Wireless Com.*, Vol. 16, No. 10, Oct. 2017, p. 6698-6711.